2 Artificial Intelligence: The Landscape

In the winter term of the 2003–2004 academic year, I (Rolf) gave a series of lectures on modern artificial intelligence that was broadcast from the University of Tokyo over the entire globe, to Beijing (China), Jidda (Saudi Arabia), Warsaw (Poland), Munich (Germany), and Zurich (Switzerland). This global virtual lecture hall was connected via video conferencing technology, enabling the full participation of the students from all the sites; they could ask questions, and could also show video clips or presentations from their laptops. The main topic of this series was the impact of embodiment on a theory of intelligence, or in other words how intelligence and body are related to one another. Every week, the last half hour of these global lectures was devoted to the presentation of the latest research in the field of artificial intelligence, mostly from Japanese researchers. Most of these top-notch researchers presented robots that locomote: robots that move like snakes, or two-legged robots that walk like humans, or that can stand up from a lying-down position. This observation raises the question of what this walking and locomotion business has to do with intelligence; with thinking. Why do research on how robots, animals, and people move if you are interested in understanding intelligence? One of the goals of this book is to try and answer this rather puzzling question. We hope that as we go along it will become clear that the question is very sensible, that the relations between moving and thinking are in fact quite straightforward, and that intelligence cannot be understood if we do not understand basic movement—a point that we have already argued in the previous chapter.

But before we embark on this endeavor, in order get a better feel for the research area that we are talking about, we would like to outline the landscape of artificial intelligence: that is, the structure of this scientific discipline, the kind of research performed, and how the various disciplines relate to one another.

The first thing to note is that there is a clear distinction between a traditional or classical approach, also called the symbol-processing approach, and a modern, embodied one, a distinction that will be explained in more detail just below (see figure 2.1). It is interesting to observe that when you type "embodied artificial intelligence" into a search engine such as Google, you do not find many books or articles with this term in their title. And, as closer analysis shows, the results from the search do not reflect in any way what researchers in this field actually investigate these days. Now what does this mean for the field? This is one of the questions this chapter tries to answer.

After outlining the successes and problems of the classical approach we will turn to what we have called "the embodied turn," i.e., the new paradigm for artificial intelligence research. We will discuss how the role played by neuroscience in this endeavor has changed over time, and then look at how the field of classical AI split into many disciplines. This will be followed by an overview of the disciplines most relevant to embodied intelligence, such as biorobotics, developmental robotics (including humanoid robotics), ubiquitous computing and interfacing technology, artificial life and multiagent systems, and evolutionary robotics.

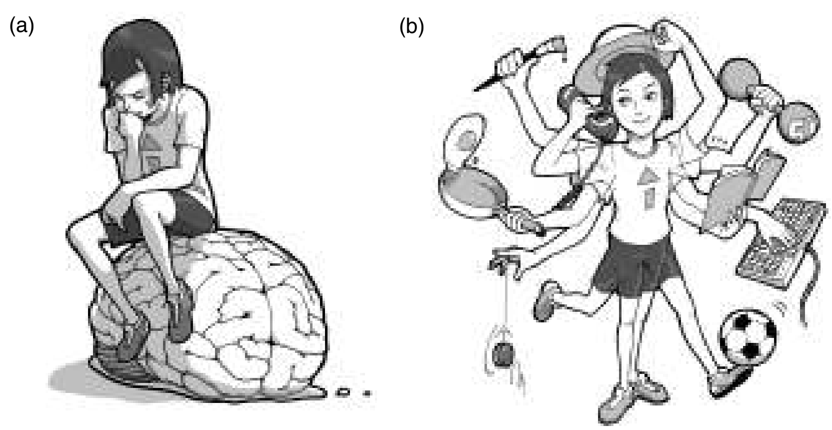

Figure 2.1
Two ways of approaching intelligence. (*a*) The classical approach. The focus is on the brain and central processing. (*b*) The modern approach. The focus is on the interaction with the environment. Cognition is emergent from the system-environment interaction, as we will argue throughout the book.

2.1 Successes of the Classical Approach

The term *embodied intelligence* was introduced in the mid-1980s in the field of artificial intelligence as a reaction against the classical approach, which views intelligence as merely a matter of abstract symbol processing. What matters in the classical approach is the algorithm or the program—the software, if you like—and not the hardware (the body or brain) on which it runs. Abstract functioning that is independent of the specifics of a particular hardware is an extremely powerful idea and constitutes one of the main reasons why computing has conquered the world, so to speak: all that matters are the programs that run on your computer; the hardware is irrelevant. This line of thinking goes back to the famous Dartmouth conference, held in 1956 in the small town of Hanover, New Hampshire, when "artificial intelligence" was officially launched as a new research discipline (for a very short history of artificial intelligence, see focus box 2.1). The American philosopher John Haugeland of the University of Chicago, author of the well-known book *Artificial Intelligence: The Very Idea*, an excellent philosophical treatise on traditional or classical artificial intelligence, coined the term GOFAI—"Good Old-Fashioned Artificial Intelligence"—to designate this approach (Haugeland, 1985).

In the classical perspective of artificial intelligence the human being was placed at center stage, with human intelligence as the main focus. As a consequence, the favorite areas of investigation were natural language, knowledge representation and reasoning, proving mathematical theorems, playing formal games like checkers or chess, and expert problem solving. This last area became extremely popular in the 1980s. Expert systems, as these models were called, were intended to replace human experts, or at least take over parts of their tasks, in areas like medical and technical diagnosis, configuration of complex computer systems, commercial loan assessment, and portfolio management. These systems epitomize the classical approach of viewing humans as symbol processing systems, i.e., as systems that manipulate symbols as computer programs do. This so-called information-processing approach strongly influenced researchers not only in artificial intelligence but also in psychology and the cognitive neurosciences. And now it seems that scientists as well as people in general see human intelligence as information processing: "What else could it be?" is the standard defense of this view. Computer scientists and psychologists teamed up to develop information-processing models of human problem-solving behavior, in particular expert systems.

Focus Box 2.1
The History of AI

Some authors (Brighton, 2004) consider the history of AI to begin around 3000 BC, apparently in Luxor, where a papyrus has been found that reports medical knowledge in expert system form: "If patient has this symptom, THEN he has this injury with this prognosis IF this treatment is applied." But usually it is agreed that the field really began with the famous Dartmouth conference in 1956 where, among others, the "fathers of AI," Marvin Minsky, John McCarthy, Allen Newell, Herbert Simon, and Claude Shannon convened to proceed on "the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it." (Dartmouth Artificial Intelligence Project Proposal, McCarthy et al., Aug. 31, 1955). The discussions revolved around the question of how or whether human thinking and processes could take place in a computer. Addressing this question required, and still requires, knowledge from many different disciplines. Finally, a common language had been found with which researchers from different disciplines could talk to each other and formulate their theories; this was the language of information processing and abstract symbol manipulation. The field started to take off and spread across the United States. Natural-language programs, programs for proving mathematical theorems, for manipulating formulas, for solving abstract problems, for playing formal games like checkers and chess, for planning, and for solving real-world problems—the expert systems—emerged and multiplied everywhere: the field was booming.

Expert systems were developed specifically for medical diagnosis, analysis and repair of malfunctioning devices, commercial loan assessment in banks, configuration of complex computer systems, and portfolio management, to name but a few. The idea was to model a human expert, such as an experienced physician, using sets of rules such as "IF the patient has red spots on skin, and patient has high fever, and . . . THEN the infection is most likely caused by . . ." (note the similarity to the Egyptian system). Herbert Simon, in 1965, predicted that by 1985 machines would be capable of doing any mental work a man can do. However, toward the end of the 1980s most companies that had started developing expert systems went bankrupt, and the goal of building systems capable of autonomously solving problems—and thereby replacing human expertise—was largely abandoned. It had become clear that conceptualizing human experts as symbol-processing machines was inappropriate and did not lead anywhere. Practitioners changed the focus from autonomous problem solving to supporting human intelligence.

Besides the field of expert systems, there were serious setbacks and disappointments in the areas of computer vision and speech processing. Human-level performance in perception—recognizing objects at various distances, orientations, lighting conditions, and partial occlusions—has not even remotely been achieved in artificial systems. Similarly, in spite of huge investments in speech systems, their capacity, accuracy, and therefore their practical utility has remained below expectations. Vision and speech are particularly challenging because they are natural phenomena that rely heavily on the real world. Trying to model human visual perception and language through (typically computationally intensive) algorithms did not seem to work either.

Luckily for many researchers in these fields, a new discipline arose in the early 1980s, connectionism, which tries to model phenomena in cognitive science with neural networks. Neural networks are computational models that are inspired by biological brains, and therefore many of them inherit the brain's intrinsic ability for adaptation, generalization, and learning. Because they are based on pattern processing rather than symbol manipulation, researchers were hoping that neural networks would be better able to describe natural mental phenomena, after expert systems and related algorithms had failed to do so. In fact connectionism was not exactly a new discipline: neural networks had been around since the 1940s, when

Focus Box 2.1
(continued)

they were first suggested as models of biological neural networks (e.g., McCulloch and Pitts, 1943). Their reappearance in the 1980s as computational devices was more like a renaissance. However, although there was definite progress, because most of these models were just algorithms like all the others, they did not end up solving the big problems of mastering the interaction with the real world either. Despite the progress there were no real breakthroughs in the use of neural networks for capturing expert knowledge, for building speech systems, or for perception of the environment. The recognition of this fact was another frustration for artificial intelligence researchers.

After these setbacks, the field was in dire need of a real paradigm shift. In the mid-1980s Rodney Brooks suggested that all of this focus on logic, problem solving, and reasoning was based on our own introspection—how we tend to see ourselves and our own mental processes—and that the way artificial intelligence was proceeding was misguided. Instead, he proposed, essentially, that we should forget about symbol processing, internal representation, and high-level cognition, and focus on the interaction with the real world: "intelligence requires a body" was the slogan of the new paradigm of embodied intelligence. With this change in orientation, the nature of the research questions also started to shift: the community got interested in locomotion, manipulation, and, in general, how an agent can act successfully in a changing world.

As a consequence, many researchers around the world started working with robots. However, even working on robots did not automatically solve the problems: the performance of most robots on real-world tasks—walking, running, perception, and object manipulation—remained unsatisfactory. So, there was still something missing. The reason for this, we strongly suspect, was that the robots were often used in the classical way: researchers programmed the robots directly to do their tasks. This often led to computationally expensive solutions that not only produced unnatural behavior, but were also too slow to achieve, for example, running behavior. Thus, the concept of embodiment not only implies that the agent must have a body—obviously robots do have bodies—it also means that one should follow a particular style of thinking when building robots or generally intelligent agents; one should design with a particular theoretical attitude in mind, as we will elaborate in this book. Although we are convinced of the potential of this approach, only time will tell whether it results in greater success than the previous ones.

In the 1980s there was a lot of hype surrounding expert systems and many companies started to develop them—alas, many soon went bankrupt after this way of conceptualizing human expertise and human intelligence in general turned out to be flawed, as discussed in the next section (see also Clancey, 1997; Pfeifer and Scheier, 1999; and Winograd and Flores, 1986).

By the mid-1980s, the classical approach had grown into a large discipline with many facets and with fuzzy boundaries, but despite some of its flaws, it can now claim many successes. Whenever you switch on your laptop computer you are starting up many algorithms that have their origin in artificial intelligence. If you use a search engine on the internet you are, for example, making use of clever machine-learning algorithms.[1]

If you use a text-processing system, it in turn uses algorithms, which try to infer your intentions from the context of what you have done earlier, and will often volunteer advice. Natural-language interfaces, computer games, and controls for appliances, home electronics, elevators, cars, and trains abound with AI algorithms. More recently, data-mining systems have been developed that heavily rely on machine-learning techniques, and chess programs have been designed that can beat just about any human on Earth, which is a considerable achievement indeed! The development of these kinds of systems, although they have their origin in artificial intelligence, has now become indistinguishable from applied informatics in general: they have become an integral part of today's computer technology.

2.2 Problems of the Classical Approach

However, the original intention of artificial intelligence was not only to develop clever algorithms, but also to understand natural forms of intelligence, which requires a direct interaction with the real world. It is now generally agreed that the classical approach has failed to deepen our understanding of many intelligent processes. How do we make sense of an everyday scene or recognize a face in a crowd, for example? How do we manipulate objects, especially flexible and soft objects and materials like clothes, string, and paper? How do we walk, run, ride a bicycle, and dance? What is common sense all about, and how are we able to understand and produce everyday natural language? Needless to say, trying to answer these questions requires us to consider not just the brain, but how the body and brain of an intelligent agent interact with the real world.

Classical approaches to computer vision (which is one form of artificial perception), for example, have been successful in factory environments where the lighting conditions are constant, the geometry of the situation is precisely known (i.e., the camera is always in the same place, the objects always appear on the conveyor belt in the same position, the types of possible objects are known and can therefore be modeled), and there is always ample energy supply. However, when these conditions do not hold, such systems fail miserably, and in the real world, stable and benign conditions are never assured: the distance from an object to your eyes changes constantly, one of the many consequences of moving around; lighting conditions and orientation are always changing; objects are often entirely or partially blocked from view; objects themselves move; and they appear against very different and changing backgrounds.

Vision systems with capacities similar to human vision, which can deal quickly with such conditions, are far from being realized artificially.

Animals and humans—including simple animals like insects—are enormously skilled at manipulating objects. Ants, for example, are known for their great ability to carry large, bulky objects such as leaves, and they do so by cooperating with other ants. Watch a dog chew on a bone by controlling it with its paws, mouth, and tongue: unbelievable! Although there are specialized machines that can out perform humans on virtually any given manipulation task—like driving a screw, picking up objects for packaging in production lines, lifting heavy objects on construction sites, or making very precise movements in minimally invasive surgical operations—the general-purpose manipulation abilities of natural systems are still unparalleled.

Locomotion is another case in point. Animals and humans move with an astonishing flexibility and elegance. Watching insects fly in complex patterns and with enormous precision is simply mind-boggling, especially since we know how small their brains really are: a million times smaller than the human brain. Watching a cheetah running at great speed is an esthetic pleasure. Monkeys move through the rain forest by climbing, swinging, walking, and running with uncanny talent. Humans can walk with a bag in one hand, an arm around a friend, up and down stairs, while looking around and smoking a cigarette, or they can walk in arbitrarily silly ways, as demonstrated by John Cleese in the famous Monty Python sketch "The Ministry of Silly Walks"; no robots can even come close to any of these feats of agility yet. And building a running robot is still considered one of the great challenges in robotics.

Although there has been a considerable amount of work on robots since the early days of classical artificial intelligence, starting in the 1960s, the performance of these robots has not been very impressive in terms of orientation ability, speed, and capacity to manipulate objects. One of the important reasons for this is that in the classical view, the ability to figure out where you are is based on detailed inner models or representations of the outside world—which implies that these representations either have to be programmed into the robots (which is done, for example, in industrial robotics) or the robots have to learn them as they interact with their environment; and they have to be continuously updated in order to remain consistent with the real world. The more complex these models are, the more effort is needed to acquire the relevant data to maintain them. Take a map of a city as an example of a model of part of the real world. The more detail the map contains, the harder it will be to keep it in tune with reality. If construction sites, temporary roadblocks, or current traffic density are taken into account, the entire map has to be updated almost continuously. If the map is intended for car drivers, as are those used in car navigation systems, information about traffic density and diversions is extremely useful, but keeping it up to date requires considerable resources. For other purposes, such as a geography class, a coarse map is more than sufficient and requires very little updating.

An issue which has attracted a lot of attention is that of common sense, because it is fundamental to mastering our everyday lives and is also crucial for understanding natural language. In the classical approach, common sense has been viewed as "propositional": the building blocks of common-sense knowledge are considered to be statements—propositions—such as "cars cannot become pregnant," "objects (normally) do not fly," "people have biological needs (they get hungry and thirsty)," "viruses cause infections," "diseases should be avoided," "if you drop a glass it will normally break," etc. Building systems that incorporate this type of common-sense knowledge has been the goal of many classical natural-language and problem-solving systems like CYC (see Guha and Lenat, 1990, for a report on the first five years of the project). The letters CYC stand for encyclopedia, which indicates what the researchers in this project were after, namely this kind of encyclopedic, or propositional, knowledge. The Stanford computer scientist and artificial intelligence pioneer Doug Lenat started this controversial project in 1984, and in 1991 he predicted that by the mid-1990s his software would be able to obtain new knowledge by simply reading text rather than being programmed by humans (Wood, 2002); this is one of the many predictions in AI that have not materialized. Surprisingly, some researchers continue to believe that a large collection of propositions—logic-based statements—together with a set of rules of inference is, in essence, all that is needed to represent common-sense knowledge: in 2004 DARPA, the Defense Advanced Research Project Agency, the American military's research arm, awarded two American researchers a $400,000 research contract to try and build a machine that could learn only by reading text. One of the problems with the CYC project, and with all succeeding projects with a similar aim was (and still is) that common sense cannot be captured by a set of rules, but requires interaction with the real world.

For example, we all have an intuitive understanding of the word *drinking*. If you now freely associate to *drinking*, what comes to mind might be: thirsty, liquid, beer, hot sunshine, the feeling of liquid in your mouth, on the lips, on your tongue when you are drinking, how it runs down your throat and how it feels in your stomach, and the experience of relief after drinking when you have been really thirsty, the experience of seeing a cold drink being served in a seaside bar on a hot summer day, the frustration at the stain on your new suit as wine is spilled over it, the sensation of wetness as water is poured over your pants, etc. It is this kind of common sense that forms the basis of everyday language communication, and it is firmly grounded in our own specific embodiment; in our experience of interacting with objects in the real world. And to our knowledge, there are currently no artificial systems capable of dealing with this kind of knowledge in a flexible and adaptive way, because it is not propositional and thus hard to formalize in a symbolic system.

Speech systems are another offshoot of classical AI. Natural language—which is different from formal languages like mathematics or computer programs—is one of the most striking abilities of an intelligent being, and the quest to understand and build systems capable of natural language has a long history in artificial intelligence. Initially, efforts were mostly geared toward processing written language. Later on, speech captured the interest of many researchers, but expectations and false predictions about the speed of development of such systems have abounded. Consequently there have been many disappointments, and the reputation of the field of artificial intelligence has suffered as a result. While in restricted applications speech systems are helpful, especially where single-word commands are sufficient as in some mobile phone applications, speech systems that can handle complete sentences or continuous streams of speech, in a robust way and in noisy environments, have not yet appeared on the market.

Speech-to-text systems—also called "phonetic typewriters"—have to be tuned to the speaker's voice, and typically a lot of post-editing needs to be done on the text produced by the software, i.e., the text usually contains many errors and needs to be corrected. This may be one of the reasons why speech systems have not really taken off, even though the idea of not having to type anymore—of producing text rapidly by simply talking into a microphone—is highly appealing. But although some of the systems may function to some degree and have turned out to be quite useful, there are still no general-purpose natural-language systems whose performance even remotely resembles that of humans in everyday conversation. (It is also interesting to note that major companies dealing in speech systems have gone bankrupt in recent years. The most famous example is L&H, Lernout and Houspie Speech Products in Belgium, which marketed speech-to-text systems as one of the three major players in the field worldwide, the others being Dragon Systems and IBM. The bankruptcy is officially due to illicit financial transactions and incorrect sales figures, but we would speculate that while there certainly have been financial and legal problems, the current immaturity of the underlying speech technology probably made matters worse.)

Another way of looking at the successes and failures of classical artificial intelligence is that it has been successful at those tasks that humans normally consider difficult—playing chess, applying rules of logic, proving mathematical theorems, or solving abstract problems—whereas actions we experience as very natural and effortless, such as seeing, hearing, speaking, riding a bicycle, walking, drinking from a glass, assembling a car from a Lego kit, talking, getting dressed, putting on makeup, or brushing our teeth—all skills requiring common sense—have proved notoriously hard. The successes in achieving these latter skills in artificial systems have been very limited, to say the least; the algorithmic approach has simply not helped much in understanding intelligence (see also Pfeifer and Scheier, 1999).

2.3 The Embodied Turn

These failures, largely due to the lack of rich interaction between system and environment, have led some researchers to pursue a different avenue; that of embodiment. With this change of orientation, the nature of the research questions also began to change. Rodney Brooks, director of the MIT Computer Science and Artificial Intelligence Laboratory, a laboratory of about a thousand researchers, was one of the first promoters of embodied intelligence. Brooks argued in a series of provocative papers entitled "Intelligence Without Representation" and "Intelligence Without Reason" that intelligence always requires a body and that we should forget about complex internal representations and models of the outside world; that we should not focus on sophisticated reasoning processes but rather capitalize on the system-environment interaction (Brooks, 1991a). "The world is its own best model" was one of his slogans at the time. Why build sophisticated models of the world when you can simply look at it? In the second half of the 1980s he started studying insect-like locomotion, and building, for example, the famous six-legged walking robot "Ghengis."

Why did he choose insects as his object of investigation? Brooks made a case that because it took evolution so much longer to move from inorganic matter to insects than it took to get from insects to humans, we should start by studying insects. Once we understand insect-level intelligence—thus Brooks's argument—it will be much easier and faster to understand and build human-level intelligence because achieving insect-level intelligence from scratch should be a much harder problem than moving from insect-level intelligence to human-level intelligence. To gain some perspective on this claim, consider this greatly abridged history of evolution on Earth. Single-cell entities arose out of the primordial soup roughly 3.5 billion years ago. A billion years passed before photosynthetic plants appeared. After almost another billion and a half years—around 550 million years ago—the first fish and vertebrates came into being, and 100 million years later insects emerged. Let us quote directly from Brooks's argument:

> Then things started moving fast. Reptiles arrived 370 million years ago, followed by dinosaurs at 330 and mammals at 250 million years ago. The first primates appeared 120 million years ago and the immediate predecessors to the great apes a mere 18 million years ago. Man arrived in roughly his present form 2.5 million years ago. He invented agriculture a mere 19,000 years ago, writing less than 5,000 years ago and "expert" knowledge only over the last few hundred years. (Brooks, 1990, p. 5)

Because of this interest in insects, walking and locomotion in general became important research topics. This, of course, represents a fundamental change from studying chess, theorem proving, and abstract problem solving, and it is not so obvious what the two areas have to do with one another (an issue we will elaborate on later). Other topics that people started investigating include orientation behavior: finding one's way in only partially known and changing environments, which includes searching for "food" (symbolized by certain kinds of objects such as small cylinders); bringing the food back to the "nest," a behavior also called homing; or generally exploring an environment. A lot of effort has also been invested in the study of very elementary behaviors such as wall following, moving toward a light source, and obstacle avoidance. It is interesting to note that researchers in the field started using vocabulary like "search for food," "homing," "going back to the nest," etc., suggesting that the robots developed in fact have animal-like properties. Attributing lifelike properties to inanimate objects has a long history in artificial intelligence, where researchers since the very beginnings have ascribed humanlike properties to their computers or computer programs, calling them intelligent or clever, claiming that they understand when replying to questions, and so on. Attribution of lifelike properties to artifacts seems to be a characteristic intrinsic to humans, or, as David

McFarland, Oxford University behavior scientist and inventor of the field of "animal robotics," put it: "Anthropomorphization, the incurable disease." But then, anthropomorphization has been around for centuries: think how many talking animals or objects there are in fairy tales or Disney movies. McFarland's point was that we have to be careful with the attributes we ascribe to animals, computers, or robots when we observe their behavior, for instance when we say that the animal "wants" to eat or that the robot "sees" a person. How do we know the animal "wants" something, and what do we really mean by this? But more about that later.

Now, the perspective of embodiment requires working with real-world physical systems, such as robots. Although computers and robots are often mentioned in one phrase, suggesting that they are roughly the same, they are in fact quite different: the input to computers consists of keystrokes or mouse clicks, and because keystrokes are discrete, the user has to prepare whatever he or she wants to enter into the computer for further processing in terms of the limited number of keys on the keyboard. By contrast, biological agents—animals and humans—have complex sensors that provide a lot of continuously changing stimulation and thus, potentially rich information about the real world. But the real world does not come with labels: we have to try to make sense of this sensory stimulation on our own, whereas in the case of the computer this job has to be taken over by the user. Thus, truly autonomous robots, those that are largely independent of human control, have to be situated, i.e., they have to be able to learn about the environment through their own sensory systems, something computers simply cannot do. Also, computers are neat and clean, and almost anybody can understand, use, and program them, and they lend themselves well to performing simulations. But building robots requires engineering expertise which is typically not present in computer science laboratories; it is messy, you have to get your hands dirty, which is something that, in the age of information technology, many people strongly dislike.

Generally speaking, the interaction of an embodied system with the real world is always "messy" and ill defined, and there are many issues one has to deal with, such as deciding on the kinds of environments in which the robot has to function (e.g., office environments, factories, outdoors in the city streets, in rough terrain, in homes, under water, in the air, in outer space), the kinds of sensors to use (cameras, microphones, infraread, ultrasound, touch), the actuators (hands, arms, legs, wings, fins, wheels, or perhaps hooks or magnets), the energy supply (a notoriously hard problem), and the materials from which the robot should be constructed. To make matters worse, the physics of the agent-environment interaction must also be considered. This includes accounting for the forces, torques, and friction that the robot will experience: the environment changes rapidly and is predictable only to a very limited extent, and the information about the world is always very limited. Most of these considerations are normally not associated with the notion of intelligence. The design principles for intelligent systems that will be introduced in part II of this book try to capture all of the design considerations that must be taken into account for embodied systems in the real world.

So, the nature of the field of artificial intelligence changed dramatically when embodiment entered the picture. While in the traditional approach the relation to psychology—in particular, cognitive psychology—had been very prominent, the interest, at least in the early days of the embodied intelligence approach, shifted more toward nonhuman biological systems such as insects, snakes, or rats. Also, at this point, the meaning of the term *artificial intelligence* started to change, or rather started to adopt two meanings: the first implies GOFAI, the traditional algorithmic approach, while the other more generally designates a paradigm in which the goals are to understand biological systems while at the same time exploiting that knowledge to build artificial systems. As a result the modern, embodied approach started to move out of computer science laboratories and into robotics, engineering, and biology labs.

2.4 The Role of Neuroscience

It is also of interest to look at the role of neuroscience in the context of the shift to an embodied approach. In the 1970s and early 1980s, as researchers in artificial intelligence started to recognize the problems of the traditional symbol-processing approach, they began to search for alternatives. Artificial neural networks seemed to provide the solution. Although they had been around since the 1950s, neural networks only started to really take off in the 1980s, just when artificial intelligence was in a deep crisis and desperately looking for a way out. Loosely speaking, artificial neural networks, or simply neural networks, are models that implement "brain-style computation," as some researchers call it. Neural networks are collections of abstract models of neurons that are connected to many other neurons to form large networks that function in a massively parallel fashion. Although inspiration was drawn from the brain, neural networks relate to brain activity only at a very abstract level and neglect many essential properties of biological neurons and brains. Despite these abstractions, the algorithms based on these simple networks demonstrate impressive performance and can achieve, for example, difficult classification and pattern-recognition tasks like deciding from an X-ray image whether some tissue contains a cancerous tumor or not, or distinguishing bags containing plastic explosives from innocuous ones at airports. In chapter 5 we will provide a more detailed account of neural networks (see also focus box 5.1).

In the field of cognitive psychology, artificial neural networks became very popular for modeling a variety of phenomena such as categorization (making distinctions between different types of objects) and perception in general, but also language acquisition (how children learn to master language) and memory. An exciting new discipline called connectionist psychology emerged as a result (e.g., Ellis and Humphreys, 1999). Using neural network models of this kind was definitely a step in the right direction, as they have highly desirable properties. For example, like natural brains, they are massively parallel; they can learn, i.e., they improve their behavior over time; they are noise and fault tolerant, i.e., they still function if the inputs are distorted and if some of the artificial neurons cease to operate; and they can generalize, meaning they continue to work in situations that have never been encountered by the network before, as long as those situations are similar to what they have already learned. The main problem with the approach, however, was that the networks were mostly disembodied, which means that they were trained on data prepared by the designer; the networks did not collect their own data in the environment using a body. With some exceptions, real-time response was not required, because the models were not connected to the outside world. In particular, they were not used in robots.

In the embodied approach, by contrast, the connection to the outside world is crucial. As artificial intelligence researchers realized that because natural neural systems are so skillful at controlling their host body's interaction with the real world, they might benefit by paying more attention to biological detail, interest in neuroscience was renewed and strengthened.[2] The kinds of networks suitable for these sorts of interactions are different from the connectionist ones used in psychology because they have to deal with real physical bodies and have to act in real time. As a result, the artificial neural networks developed for these purposes paid closer attention to biological properties, and researchers in artificial intelligence started cooperating much more closely with neurobiologists. Around the same time, a new breed of neuroscientist started to appear, the so-called computational neuroscientists, and university departments with names such as Computational Neuroscience or Neuro-Informatics emerged almost overnight. Rather than performing experiments with real brains, however, they developed detailed models either of individual neurons or of specialized collections of neurons in the brain such as the cerebellum, which plays a key role in motor control, or the hippocampus, an area thought to be involved in memory functions, as well as a host of models about aspects of the visual system. These are but a few examples; the literature in the field is awesomely vast. And some researchers in computational neuroscience became interested in issues similar to the ones artificial intelligence researchers had started tackling, e.g., locomotion, categorization, and sensory-motor coordination. Most would not consider themselves to be doing research in artificial intelligence, even though their research topics strongly overlap; for the most part computational neuroscience has not (yet!) taken a strong interest in embodiment. Finally, along a different but related line of development, engineers have started cooperating with neuroscientists to connect electronic and electromechanical devices directly to neural tissue (as we will see when we discuss cyborgs in chapter 8).

2.5 Diversification

So, in terms of research disciplines participating in the AI adventure, in the classical approach it was computer science (of course), psychology to a greater degree, and neuroscience to a lesser degree. A very close cooperation with linguistics and computational linguistics became popular due to the seminal—but somewhat misleading—work on grammatical structures pioneered by the outspoken linguist and political activist Noam Chomsky of MIT; and finally there was a very close connection with philosophy. This last connection specifically involved the field of philosophy of mind, which is an attempt to unravel the mysteries of the human psyche, of thinking, intelligence, emotion, and consciousness. At least in some areas of philosophy, there was a lot of optimism about the potential contributions of the computer metaphor toward a scientific understanding of the mind, as shown in the enthusiastic book by the British philosopher and AI researcher Aaron Sloman, *The Computer Revolution in Philosophy* (Sloman, 1978). Alas, this hope has not yet been fulfilled.

In the embodied approach, the picture altered considerably. Computer science and philosophy are still part of the game as before, but now also engineering, robotics, biology, biomechanics (the discipline studying how humans and animals move), material science, and neuroscience have come into play, whereas psychology and linguistics have—at least temporarily—if not disappeared, at least lost their status as core disciplines. So we see somewhat of a shift of interest from high-level processes (as studied in psychology and linguistics) to more low-level sensory-motor processes. Recently psychology, especially developmental psychology, has reentered the game in the context of developmental robotics, where the grand goal is to mimic in robots the processes by which babies develop into capable adults.

Although, as mentioned above, a certain amount of robotics work was done in the initial years of artificial intelligence, as exemplified by the research on the world-famous robot "Shakey" at Stanford Research Institute in Palo Alto, California, robotics at the time played only a marginal role (Shakey earned its name by its hesitant, jerky way of moving). Moreover, even though Shakey was indeed a physical robot acting in the real world, the focus was very much on its internal processing; on the kinds of computations it would have to do to navigate and orient in the real world. In this sense, although Shakey had a body, it was very much computational, and therefore in line with the classical paradigm. Because of this, it could only operate in simple and judiciously designed static environments. But, as always, it is easy to criticize with hindsight, and this in no way diminishes the value of Shakey's contribution to the development of artificial intelligence. Just recently it was elected to the Robot Hall of Fame of the Carnegie-Mellon Foundation, where historically significant robots are on display. Other "laureates" include HAL 9000 from Stanley Kubrick's movie *2001: A Space Odyssey*, the Mars Sojourner, Honda's Asimo, C3PO from *Star Wars*, and Astroboy. (Astroboy—called Tetsuwan or "Iron Arm" Atom in Japan—the hero from an extremely successful comic strip of the 1950s in Japan, has inspired many researchers and visionaries in Japan who, today, build robots in the most highly respected institutions. Astroboy is very much the spiritual father of the contemporary intelligent robotics movement in Japan.)

As the participating disciplines have changed, the terms for describing the research area have also shifted: researchers using the embodied approach no longer refer to themselves as doing artificial intelligence but rather robotics, engineering of adaptive systems, artificial life, adaptive locomotion, or bio-inspired systems. But more than that, not only have researchers in artificial intelligence moved into neighboring disciplines, scientists who have their origins in these other fields have started to play an important role in the study of intelligence. Computational neuroscience is a case in point, although researchers in that field typically do not perceive themselves as part of artificial intelligence. Thus, on the one hand the field of artificial intelligence has significantly expanded, while on the other hand its boundaries have become even fuzzier than they were before.

So we now have a partial answer to the question of why we do not get a representative sample of the research being done in modern artificial intelligence when we type "embodied artificial intelligence" into a search engine. Because the communities started to split, researchers in embodied intelligence started going to other kinds of conferences that were not purely artificial intelligence–based, as the names of these conferences indicate: "Intelligent Autonomous Systems," "Simulation of Adaptive Behavior—From Animals to Animats," "International Conference on Intelligent Robotics and Systems," "Adaptive Motion in Animals and Machines," "Artificial Life Conference," "Evolutionary Robotics," the "International Joint Conference on Neural Networks" (among many other neural network conferences), the "Genetic and Evolutionary Computation Conference" (there are several other conferences dedicated to artificial evolution, a topic we will explore in chapter 6), or the various IEEE conferences (International Society of Electrical and Electronics Engineering), and so on. In the early 1990s, when I (Rolf) tried to convince people at AI conferences that embodiment is not only interesting but essential for intelligence, and that unless we understand embodiment we will never crack the conundrum of high-level intelligence, I mostly got negative reactions, and no real discussion took place. So, I and many colleagues turned to these other conferences, where people were more receptive to the ideas of embodiment. More recently, perhaps because of the stagnation in the field of classical AI in terms of tackling the big problems about the nature of intelligence, there has been a growing interest in the issue of embodiment. Most AI conferences have started hosting workshops and special tracks on issues related to embodiment. But by and large the communities of classical artificial intelligence and of the embodied approach to intelligence are still separate, and will probably remain so for a while.

2.6 Biorobotics

This diversification has resulted in a number of interesting developments. One, as already mentioned, is the move away from human toward more animal-like intelligence, which was originally triggered because the efforts to achieve human-level intelligence had not met with success. Others include the appearance of the fields of biorobotics, developmental robotics, ubiquitous computing, artificial life, interface technology, and multiagent systems. We will look into all these different areas briefly throughout the course of this book.

Let us start with biorobotics. Biorobotics is a branch of robotics dedicated to building robots that mimic the behaviors of specific biological organisms. A good illustration is the work done by the mathematician and engineer Dimitri Lambrinos while he was working at the Artificial Intelligence Laboratory at the University of Zurich. He started to cooperate with the world leader of ant navigation research, Ruediger Wehner, also of the University of Zurich. Jointly, the two laboratories built a series of robots, the Sahabot series (the name stands for Sahara robot). The Sahabots mimic the long- and short-term navigation behaviors of the desert ant *Cataglyphis*, an extraordinary animal that lives in a salt pan, a very flat sandy ecological niche, in southern Tunisia. One of the challenges was to provide a proof of existence for the navigational mechanisms that biologists proposed to explain how this animal gets around. In other words, the goal was to demonstrate that these mechanisms could, in principle, on a robot, reproduce the orientation behavior of the desert ants. Note that this does not imply that the processes underlying the ant's behavior are indeed the same or similar to the one used on the robot.

One such mechanism, and a very simple one at that, is the so-called snapshot model, which was originally postulated by the British insect biologist Tom Collett of Sussex University (Cartright and Collett, 1983), who has worked with Wehner for many years. According to Collett, the snapshot model is used by the ant (and other insects) for precise short-range navigation to find the nest as it returns from a food-searching trip (also known as foraging in biology). This model posits that as the ant leaves the nest, which is essentially just a hole in the ground, it takes a snapshot, a photographic picture of the horizon as seen from the position of the nest, which is then stored in the ant's brain (ants, unlike humans, have almost omnidirectional vision, i.e., they see not only in the front, but all around them). The ant then goes out on a foraging trip, traveling sometimes up to 200 meters away from the nest, and returns to the vicinity of the nest using a second navigation system, which is based on an estimate of the distance from the nest and on polarized sunlight. The polarized sunlight provides the ant with direction information and can be used as a kind of compass. This system is especially suited for long-term navigation, but because long-term navigation systems always accumulate error, the ant has to use the short-term navigation system—the snapshot method—to find the exact location of the nest. From the long-term navigation system the ant gets a signal that it is near the nest and that another system must take over. The snapshot method then guides it to the nest entrance. This model, which has been verified in literally hundreds of experiments with real ants (e.g., Wehner et al., 1996), has also been tested on robots in the very environment in which the ants live, in the Sahara desert, with impressive success. While this does not imply that the model used on the robot is the one actually employed by the ants, it does show that such a mechanism could work in principle. Lambrinos, together with his colleague Ralf Moeller, developed another navigation model, the so-called average landmark vector model (Lambrinos et al., 2000), which is even simpler than the snapshot model. Both of these navigation models can be used to make predictions of the animals' behavior in certain situations that can be tested on the robots and with real ants.

Note that in this navigation system the agents—the ant and the robot—do not need a map of the environment in order to navigate successfully. In other words, it does not need a model of the real world in order to behave successfully, even though the ant cannot see the nest from a distance! This is in contrast to the standard assumption that detailed environmental information, like a map, is necessary for this kind of navigation. The only "model" of the world consists of the estimate of distance and direction to the nest for the long-term system, and the snapshot for the short-term system.

Just to illustrate the richness of the field, here is a selection of other successful biorobotics projects: the insect-like flying robots (Miki and Shimoyama, 1999) and the silkworm moth robots with pheromone sensors (Kuwana et al., 1999) developed by the futurist engineer Isao Shimoyama of the University of Tokyo; the fantastically realistic snake robots developed by the renowned roboticist Shigeo Hirose of the Tokyo Institute of Technology (Hirose, 1993); Barbara Webb's work at the University of Edinburgh in Scotland on the phonotactic behavior of crickets, i.e., how males are attracted by and move toward the sound of females undeterred by the complexity, ruggedness, and noisiness of their environment (Webb, 1996); the Robot Tuna developed at the MIT Ocean Engineering Lab by Michael Triantafyllou (e.g., Triantafyllou and Triantafyllou, 1995); Joseph Ayer's projects on lobster and lamprey robots (Ayers 2004) at Northeastern University in Boston; Auke Ijspeert's work on the simulated robot salamander at the Swiss Federal Institute of Technology in Lausanne, Switzerland (Ijspeert, 2001); the "artificial mouse" developed at the University of Zurich to investigate the role of whiskers in rodent behavior (e.g., Fend et al., 2003); and Frank Kirchner's research on robotic scorpions (Klaassen et al., 2002). There are many additional examples of biorobots which have all been very productive and have significantly contributed to our understanding of locomotion and orientation behavior (for a collection of pertinent papers see, for example, Webb and Consi, 2001, or the proceedings of the Adaptive Motion in Animals and Machines Conference, e.g., Kimura et al., 2006). The list could be continued almost indefinitely. In the meantime, locomotion and orientation have become important research topics in artificial intelligence.

2.7 Developmental Robotics

The research in biorobotics is still gaining momentum and multiplying throughout research laboratories worldwide. Toward the mid-1990s, however, Brooks, who had been one of the initiators of the biorobotics movement, argued that we had now achieved "insect-level intelligence" with robots and we should move ahead toward new frontiers. But what does it mean to say that we have achieved insect-level intelligence? Ghengis, Attila, and Hannibal, three of Brooks's six-legged robots, have achieved impressive walking performance in terms of obstacle avoidance and walking over uneven ground. However, insects can do many more things. For example they can manipulate objects with their legs and mouth, they can orient in sophisticated ways in different kinds of environments (even in the desert!), they can build complex housing, they have highly organized social structures, they reproduce and they care for their offspring. Many of these abilities, for example reproduction or complex social organizations, are far from being realized in robotic systems. So, before we have achieved true insect-level intelligence, there is still much research to be done.

But it is true that even though insects are fascinating, human-level intelligence is even much more exciting; so it is understandable that after a number of years of research on insect-level intelligence, Brooks and many others wanted to do more interesting things. This seemed a good time to tackle something more challenging: the human. Thus we are back to the goals of traditional artificial intelligence, but now we can tackle them with the experience of biorobotics. Throughout the book we will give many examples of how the insights gained have changed our thinking about intelligence. While in Japan humanoid robots had been a research topic for many years already, these activities were not directly related to artificial intelligence. This seems to be the reason why Brooks's move into humanoids had a strong impact on the research community, although it was initially, and still is, met with considerable skepticism: many researchers believe human intelligence is still way out of reach. Nevertheless, in the early 1990s Brooks started the "Cog" project for the development of a humanoid robot with the goal of eventually reaching high-level cognition (Brooks and Stein, 1994).

The term *humanoid robot* is used for robots that typically have two arms and legs, a torso and a movable head with a vision system, and sometimes additional sensory modalities such as audio and touch. They are called humanoid because there is a superficial visual resemblance to humans. Because of their anthropomorphic shape, people have a strong tendency to project humanlike properties onto these robots. But, careful: remember David McFarland's reference to anthropomorphization as an incurable disease. Some science-fiction movies can also be misleading by suggesting humanlike properties in their robots: Hollywood robots typically have a very high level of intelligence. Some are mean and want to enslave mankind, reflecting a fear that, given the current state of the art in robotics, is entirely unjustified. (Of course, we don't have to wait for superintelligent killer robots to be enslaved by machines—we are already almost entirely dependent on our cars, computers, and mobile phones, and we do many things just to please the machines, not because we want to. A case in point was the Y2K problem, the year 2000 problem, where companies and governments all over the world invested billions of dollars in order to cope with the issue. We were forced to do so by our computers: it was definitely not an act of free will. The only question is whether we attribute evil intentions to the computers; but this is a philosophical question—a matter of argument—not an empirical one. An empirical question is one for which experiments can be devised to support or falsify a hypothesis, and for this question—Are machines evil?—that is not possible.)

But back to the Cog project. "Cog" is a pun, alluding both to its cognitive abilities and to the cogs of a cogwheel, insinuating that cognition or intelligence is really based on many simple cogs—processes—that function together. Inspired by this project, many researchers were attracted by the idea of moving toward human-level intelligence, which had been the target of artificial intelligence all along, both classical and embodied. Around this time the field of developmental robotics emerged. Its pertinent conferences come under many labels: "Emergence and Development of Embodied Cognition," "Epigenetic Robotics," "Developmental Robotics," "Development of Embodied Cognition," "Humanoids," etc. This was, of course, a happy change of direction for those who might have been disappointed by the turn the field had taken—insects simply are not as sexy as humans! And human intelligence happens to be the most fascinating type of intelligence that we know of. But once again, this strand of conferences is separate from the traditional ones in artificial intelligence, and although the terms *embodiment* and *emergence* might appear in the pertinent publications, "embodied artificial intelligence" most often does not.

In the meantime, developmental robotics has grown into a considerable research and engineering community in its own right. Many people in the field started developing humanoid robots, and in Japan, for example, the research in this area is really exploding. In 1998 the powerful Ministry of Economy, Trade, and Industry (METI) in Japan launched a large five-year program for building humanoid robots: the HRP, or Humanoid Robotics Program. The program was directed by the grand old man of Japanese robotics, Hirochika Inoue, at the time professor of engineering at the University of Tokyo, who has been a pioneer in robotics since 1965. The HRP had the long-term goal of developing a partner for humans, especially for the elderly, that could take over many of their household chores, thus providing independence and autonomy for as long as possible. This endeavor unites researchers from mechanical and electronics engineering, robotics, artificial intelligence, developmental psychology, and developmental neuroscience, and most of them would probably not object to being classified as working in artificial intelligence. But not only in Japan has the field gained momentum: Europeans have also warmed to the topic, and the EU is sponsoring a number of large projects in the field, such as the RobotCub (Robotics Open Platform for Cognition, Understanding, and Behavior) (not to be confused with the better-known Robocup competitions, in which robot teams play soccer against each other), and Cogniron, the Cognitive Robot Companion. We will discuss in more detail the research issues being tackled in this exciting field when we embark on the challenge of building high-level intelligence from the bottom up (chapter 5), and when we look at robotic technology in everyday life (chapter 11).

It is perhaps worth mentioning that—fortunately—not everybody has moved into humanoid or developmental robotics because there are a vast number of fascinating research issues to be tackled in animal behavior, and biorobotics seems to be a highly productive way of doing so.

2.8 Ubiquitous Computing and Interfacing Technology

Another line of development that must not be overlooked is that of ubiquitous computing and interfacing technology. We will discuss ubiquitous computing in detail in chapter 8. Here we only discuss what is needed to map out the research landscape of artificial intelligence.

Like artificial intelligence, computer science in general has undergone dramatic change: the "core" areas of computer science—software engineering, algorithm development, operating systems, and the virtual machine—are topics that we by now understand relatively well, so people have begun switching their focus to other, more challenging areas, such as the largely unexplored territory of how computers can interact with the real world beyond the typical keyboard-and-mouse setup. The very primitive interaction of computers with humans and, by extension, with the outside world in general, has for many years been one of the greatly bemoaned facts of computer technology. There is a great deal of activity in the human-computer interaction research community, aimed at improving this situation. One way toward more sophistication in the interaction with the environment is, of course, to put sensors and more interesting input-output devices into the computer such as microphones, cameras, and touch sensors. But the interaction of computers with humans is not the only focus of interest. Rather than having computers as "boxes" or devices separate from the rest of the world, it would be nice if the computing technology were integrated with the world around us so that humans could smoothly interact with it and no longer have to push keys on a keyboard as in the old days. Computers should disappear; they should become "invisible."

The original idea was, as a first step, simply to put sensors everywhere: into rooms, cars, furniture, clothes and so on and so forth. We are already surrounded by systems working around the clock, doing work for us without our being aware of it: this would just be a further step in that direction.

More recently, ubiquitous computing researchers have also begun exploring actuation: ways in which systems can not only sense, but also influence and act upon their environments. The simplest example, the thermostat, has been around for a very long time: based on a temperature measurement, the furnace is turned on or off. Another very well-known example is the garage door that opens automatically when it senses the right car entering the driveway. It is one of the fundamental discoveries in (embodied) artificial intelligence that the close coupling between sensory and motor systems is essential for intelligent behavior (see chapter 4). This insight is starting to make its way into the ubiquitous computing community.

Even though user interfaces have always been an important topic in computer science, the main problem has been the low bandwidth of communication, so to speak: normally only a mouse and keyboard are used to get information into a computer. As we have already pointed out, a lot of effort has been directed toward making speech an easy input method for computers, but these efforts, for various reasons, have not been extremely successful (yet). Just recently more interesting and rich interfaces have been developed, such as the use of pressure sensors to provide information about the user's level of aggression, and to some extent vision, using cameras that watch the user and try to collect information about gaze direction (where is the user looking?) and emotional state. There is also work on smell, but that, although very promising, has not yet advanced significantly. Whether we actually want a computer that can smell us, especially after a 14-hour nonstop programming session, is another issue altogether. The study of wearables—computers that are actually a part of our clothing—is related to ubiquitous computing, and also raises fascinating ideas about the future of human-computer interaction. What is interesting about all of these "movements"—human-machine interfaces, wearables and ubiquitous computing—is that now virtually all computer science departments are venturing into the real world. They are not doing robotics per se, but many have started hiring engineers and are establishing workshops where they can build hardware, because now real-world devices need to be constructed. So far as we can tell, there has been little theoretical development yet, but there is a lot of creative experimentation going on. We feel that the set of design principles that we have worked out for embodied systems, and which we will describe in detail through chapters 4, 5, 6, and 7, will be extremely useful in designing such systems. We will return to the topic of ubiquitous computing in chapter 8.

In conclusion, it seems that a highly innovative and dynamic part of computer science has moved from disembodied algorithms to embodied real-world computing, or rather real-world interaction, just as artificial intelligence has. Researchers in ubiquitous computing and interfacing technology are—directly or indirectly—making important contributions to artificial intelligence. Conversely, advances in artificial intelligence—from a perspective of embodiment—and robotics, specifically in sensing and actuation technology, will contribute significantly to ubiquitous computing and thus to modern computer science in general.

2.9 Artificial Life and Multiagent Systems

Another interesting development has its origins in the field of artificial life, also called ALife for short. The classical perspective of artificial intelligence had a strong focus on the individual, just like psychology does, and as we have seen, psychology was the major discipline with which artificial intelligence researchers cooperated at the time. ALife has strong roots in biology rather than psychology, and focuses on the emergence of behavior in large populations of agents. In other words artificial life research is interested in multiagent systems. We have to be a bit careful with the term multiagent systems: in ALife research, the term *complex dynamical systems* is usually preferred, because it also includes physical inorganic systems, where the individual agents or components, such as molecules or sand grains, only have limited agent characteristics. An agent is assumed to have certain elementary sensory-motor abilities, so that it can perceive aspects of the environment and, depending on this information and its own state, perform certain behaviors. Molecules, rocks, or other "dead" physical objects do not have this ability.

One early success of this field of study was the realization that complex global behavior can emerge from simple rules and local interactions (e.g., Langton, 1995). Cellular automata are the typical representatives of this approach, where the "agents" are individual cells of a grid. The next state of each cell is determined by the cell's own state and the state of its neighbors. John Conway's "game of life" (Gardner, 1970) is probably the best-known example of cellular automata behavior: the cells on a two-dimensional grid have two states, "on" or "off" ("alive" or "dead"), and are controlled by four rules: If a live cell has less than two neighbors, then it dies (loneliness); if a live cell has more than three neighbors, then it dies (overcrowding); if a dead cell has three live neighbors, then it comes to life (reproduction); otherwise, a cell stays as it is. The fascination of the game of life is the enormous variety of fun and sophisticated spatiotemporal[3] patterns that emerge from these very simple rules. People have given many of them names, such as oscillators, blinkers, flip-flops, gliders, glider cannons, and so on; dozens of live demonstrations of this game can be found on the Internet.

What counts in typical artificial life systems is the entire population of agents, not the individual. In the case of cellular automata, the individual "agents" are the cells on the grid, but these individuals are only of interest in the context of many cells. Work on self-organization in insect societies, for example by Jean-Louis Deneubourg of the Université Libre de Bruxelles (at the Center for Nonlinear Phenomena and Complex Systems), who studies social insects, also capitalizes on a population perspective and has attracted many researchers: "ant algorithms" (Dorigo et al., 2002) and "swarm intelligence" (Bonabeau et al., 1999) are among their coinages (see also Dorigo and Stützle, 2004). Deneubourg and Dorigo were both inspired by the intellectual atmosphere created by the physicist Ilya Prigogine, who was awarded the Nobel Prize in 1977 for his work on dissipative structures. His thinking on self-organization and complex systems has influenced many researchers in artificial life. Prigogine, who had been living in Brussels for many years as the director of the famous Solvay Institutes for Physics and Chemistry, had become known outside the physics community for, among others, the book with the provocative title *Order out of Chaos* (Prigogine and Stenger, 1984).

Self-organization is indeed one of the concepts that continually pops up in modern artificial intelligence (see for example Camazine et al., 2001), and we will encounter it throughout this book. By self-organization we mean that some structure or pattern—for example, patterns on butterfly wings, stripes on the fur of a zebra, or a particular social organization in insect societies—comes about as a result of the local interaction of many components, rather than by external direction, manipulation, or global, centralized control. Self-organization is an extremely powerful concept but hard to grasp intuitively because we always try to understand the phenomena around us in terms of control. However, once we grasp the idea, it becomes very natural and then it seems hard to understand how we could have done without it before, as we will see in chapters 6 and 7.

A beautiful example of how self-organization can lead to highly sophisticated behavior is the formation of ant trails. Certain species of ants are able to find the nearest food source among several sources present in the vicinity of their nest, so the ants are somehow solving a complex optimization problem. Deneubourg and Goss (1989) asked the question of whether this ability is due to the intelligence of the individual ants or due to their social interaction. Attributing this capacity to the individual ants would imply that the ants compare the distances to the various food sources and based on this knowledge choose the nearest source. This in turn would require ample calculations and considerable exploration and knowledge of the environment on the part of the individual ants. But there is a much simpler solution. Ants mark their paths with pheromones—chemicals with a strong scent—as they leave the nest to search for food and when they come back from this journey. The ants follow the pheromones, and at the crossings where several paths intersect they choose the most heavily marked one with a certain probability. Ants return sooner from nearer food sources and as a consequence shorter paths are marked more intensively than those leading to sources farther away. Because shorter paths are more heavily marked, they will attract more ants which will accelerate the speed at which the shorter paths are marked. This kind of process is an example of a positive feedback loop, and is often called an autocatalytic or self-reinforcing process. Thus, we have a very simple explanation of how ants find their way to the nearest food source in terms of self-organization rather than the cognitive power of the individual.

Modular robotics, a research area that has drawn a lot of inspiration from ALife research, also relates to multiagent systems. In this case the individual agents are robotic modules capable of assembling into robots with different morphologies (see, for example, the volume by Hara and Pfeifer, 2003, for illustrations of modular robotic systems). One of the goals of this research is to design systems capable of self-repair, a property that all living systems have to some extent: a minor bruise or a cut will automatically heal without any external intervention. Self-assembly and self-reconfiguration are fascinating topics that will become increasingly important as systems have to operate over extended periods of time in remote, hostile environments, like the deep sea or other planets. The seminal work by the futurist engineer Satoshi Murata of the Tokyo Institute of Technology and his coworkers (Murata et al., 2004) demonstrates how self-reconfiguration can be achieved not only in simulation but with real robotic systems (see figure 7.1 in chapter 7). It should be mentioned, however, that to date self-repair and self-reconfiguration is tightly controlled by a centralized algorithm, rather than emerging from local interactions. But more about this in chapter 7.

Evolutionary systems are another example of so-called population thinking, where the adaptivity of entire populations is studied rather than the adaptivity of individuals. We will discuss the impact of evolutionary thinking in chapter 6. Because of its close relation to biology, economics has also taken inspiration from evolutionary thinking and created the discipline of agent-based economics (e.g., Epstein and Axtell, 1996).

Often, evolutionary algorithms and ant algorithms are used not as biological models, but rather as powerful optimization techniques: several large industrial companies now make use of evolutionary and ant-based algorithms for design and optimization (for an overview of the use of ant algorithms in industry, see Dorigo and Stützle, 2004).

Interestingly, the term *multiagent systems* has quickly been adopted by researchers in classical artificial intelligence, but their use of multiagent systems is somewhat different. Rather than looking for emergence, as is common in the field of ALife, they usually employ multiagent systems to achieve particular tasks, for example search tasks on the Internet (e.g., Ferber, 1999). Often in this line of research the individual agents are endowed with centralized control similar to that employed in the classical approach. So in many cases the multiagent approach in artificial intelligence does not in fact study emergence.

In robotics also there has also been a growing interest in multiagent systems. The recent surge of interest in robot soccer clearly demonstrates this point. This movement, known as RoboCup, is passionately promoted by the Japanese researcher and robot enthusiast Hiroaki Kitano and his colleagues (Kitano et al., 1997), and interest in the project is not limited to the scientific community but has spread to the population at large. During the RoboCup world championship in 2002 in the Fukuoka Dome, a stadium in the southwestern city of Fukuoka on the island of Kyushu in Japan, there were more than 100,000 passionate, emotional spectators, just like at a real soccer championship! One of the problems in multi-agent robotics has been that often only a few robots are available for study—making copies of real-world robots is so much harder than making copies of software—so that no truly interesting emergent phenomena have been observed. In robot soccer, winning the game, rather than emergence, is the goal. Recently, RoboCup teams have achieved impressive performance: the games are beginning to look like real soccer where the individual players are not only extremely fast but cooperate with each other to score a goal.

One of the important research problems so far has been the achievement of higher levels of intelligence in the simulations created by the multi-agent community. Typically, as in the work of the ethologist turned ALife researcher Charlotte Hemelrijk, who studies groups of virtual primate-like agents, hierarchies among the agents and separate subgroups emerge on their own, or migration patterns materialize based only on agent-agent interaction, without the need for preprogrammed "desires" to form social hierarchies or to migrate. Thinking, reasoning, and language have typically not been topics of interest in the ALife field. An exception is perhaps the work by the artificial intelligence researcher and linguist Luc Steels, who, in his "Talking Heads" experiment (not to be confused with the rock band of the same name), attempts to investigate high-level cognition—natural language—from a population perspective (Steels, 2001, 2003). In an ingenious set of experiments he and his students demonstrated how, for example, a common vocabulary emerges through the interaction of the agents with their environment and with each other. There is also some preliminary work on the emergence of syntax. In this research, much insight has been gained into how communication systems establish themselves—how they self-organize—and how something like grammar could emerge without being predefined in the individual agents. Although this approach is fascinating and highly promising, the jury is still out on whether it will indeed lead to something resembling natural language.

Because of the fundamental differences in goals, the distributed agents community that has its origin in the artificial life community, and the one that developed out of artificial intelligence and robotics, have so far remained largely separate. Generally speaking, the artificial life community has more of a focus on populations, distributed systems with local interactions, self-organization, and complex dynamics and somewhat less on embodied systems, but researchers in this field are definitely contributing to (embodied) artificial intelligence—again, whether they realize it or not.

2.10 Evolutionary Robotics

One of the principal research topics within ALife is trying to understand how life originated on Earth, and for all we know, evolution played the key role in this process. Thus it comes as no surprise that much of the research within ALife is devoted to evolution: this includes trying to understand natural evolution and designing creatures using artificial evolution. Since the 1960s when artificial evolution was invented, so to speak (see chapter 6), there have been many intriguing developments that have led to insights into the general nature of evolution and have yielded fascinating technological results. For example, using automated evolutionary design methods, devices have been produced that at times surpass the performance of those designed by humans, such as electronic circuits (e.g., Koza et al., 2004) or antennas (e.g., Lohn et al., 2004). For our purposes, because of our interest in embodiment, the area known as evolutionary robotics is especially relevant. Methods from artificial evolution can be used to design various aspects of robots. Traditionally, in evolutionary robotics only the controller—the brain—of the robot was evolved. But more recently, with the advent of more sophisticated concepts such as models of genetic regulatory networks, entire robots—including their body and neural systems—have been evolved. The Japanese-Canadian evolutionary robotics enthusiast and entrepreneur Takashi Gomi was one of the first to recognize the importance of this field beyond its scientific interest, and he attempted to incorporate evolutionary methods not only into robotics but into business. He organized a highly successful conference series on evolutionary robotics at the Canadian embassy in Tokyo. Since then, the field has become very popular not only in Japan but throughout the world and a considerable research community has been established. Understanding how embodied systems emerge from an evolutionary process is an important contribution to artificial intelligence. But once again, few evolutionary roboticists consider what they are doing to be artificial intelligence. We will explore evolutionary robotics more deeply in chapter 6.

2.11 Summary

In summary, we can see that the landscape of artificial intelligence has changed significantly in recent years: while originally the field was clearly a computational discipline dominated by computer science, cognitive psychology, linguistics, and philosophy, it has now turned into a more multidisciplinary field requiring the cooperation and talents of researchers in many other fields such as biology, neuroscience, engineering (electronic and mechanical), robotics, biomechanics, material sciences, and dynamical systems. And this exciting new transdisciplinary community, which is very different from the traditional AI community, has been called "embodied artificial intelligence" or "embodied cognitive science." But since this is the modern view in artificial intelligence, we will no longer employ the term *embodied* artificial intelligence: what we have described in this chapter *is* what the discipline has become; it is not merely a subset of the "real" or overarching field of artificial intelligence: embodied artificial intelligence is now artificial intelligence.

Although for some time psychology and linguistics have not been at center stage, with the rise of developmental robotics there has been renewed interest in these disciplines. The ultimate quest to understand and build systems capable of high-level thinking and natural language, and ultimately consciousness, has remained unchanged. What has changed is the path—the methodology—to get there. Although the emergence of ideas of embodiment can be found throughout the history of philosophy, the recent developments in artificial intelligence that enable not only the analysis but also the construction of embodied systems are supplying ample new intellectual material for philosophers.

In spite of the multifaceted nature of artificial intelligence, there is a unifying principle: the synthetic methodology that we will describe in detail in the next chapter. Briefly, the synthetic methodology states that by actually building physical agents—real robots—we can learn a lot about the nature of intelligence. Moreover, and this is crucial for such a diverse field, physical agents, by bringing together results from all the different areas described in this chapter, have a highly integrative function. In addition, they allow for concrete testing of ideas in an objective way: a robot either works or it does not; there is no glossing over details. Moreover, robots serve as excellent platforms for transdisciplinary research and communication. By building systems using the synthetic methodology, we not only produce fun and—at least sometimes—useful artifacts, but we can acquire a deeper understanding of natural forms of intelligence. Again, the impact of applying an embodied perspective is astonishing: the insights are surprising and change the way we view ourselves and the world around us in very fundamental ways. This is what our book is all about.