4 Embodied Cognitive Science: Basic Concepts

In this chapter, we introduce the concepts that we need later on when exploring the various approaches. Moreover, we need such a framework if we actually want to build agents. One important concept that we discuss is that of the complete agent. Complete agents are inspired by natural agents, animals and humans, which are—quite obviously—capable of surviving in the real world. They are ‘‘complete’’ because they incorporate everything required to perform actual behavior. (Standard computer programs, for example, are not complete because they cannot behave in the real world.) We argue that it is such complete agents that we want study and synthesize. We provide a characterization of what we mean by complete agents, and we show that if we want to model, to synthesize such agents, we must take into account some special considerations relating to the idea of emergence, that is, to the fact that behavior emerges from the agent-environment interaction. Emergence is in turn a consequence of the frame-of-reference problem, which conceptualizes the relationships among those involved in the design process, namely the designer (who is often also the observer), the natural agent (if we are doing modeling work), the agent to be designed, and the environment. One important implication of frame-of-reference considerations is that behavior cannot be reduced to an internal mechanism. This in turn necessitates a new design methodology, which is this chapter’s central topic.

We begin the chapter with a characterization of complete agents and discuss a number of basic concepts like adaptivity, autonomy, self-sufficiency, embodiment, and situatedness. We then turn to agents—both simulated and real robots—and discuss how they can be used as modeling tools. We examine the pros and cons of working with real robots and with agent simulations. We also compare this new kind of agent simulation with more traditional forms of simulation. We then outline the framework for design that focuses on emergence, including a description of the frame-of-reference problem. Finally, we discuss what we mean by a good explanation and how we can find explanations of agent behavior by running experiments.

This chapter is difficult and covers a lot of ground. This is unavoidable. At first reading, all the points may not become immediately clear. All the issues raised here, however, will be illustrated in greater detail later on. The reader may find it helpful to return to this chapter after having read through some of the subsequent chapters.

4.1 Complete Autonomous Agents

Biological agents have to perform a number of tasks: searching for food, eating and drinking, grooming, reproducing, and caring for their offspring. The term ''task'' is normally used in a design context to designate something the agent needs to get done. Typical tasks for autonomous robots, for example, are marking all the mines in a mine field with color, or mowing the lawn of a soccer field. Note that the task of mowing the lawn implies certain desired behaviors on the part of the agent. What is really meant is that the agent's task is to keep the grass short. And because the designer can't think of any other way to accomplish the job, he simply equates the task with the method, that is, with the behavior by which the task is to be achieved, namely mowing. Note that animals don't have tasks. Rather, a task is an observer-based attribution summarizing the effect of certain behaviors of the animals. In the field of embodied cognitive science, researchers often talk about tasks of animals. What they mean is either the behavior involved—collecting food—itself or the effect of the behavior, that is, the fact that if the animals behave in a particular way, the food ends up in the nest. What is important is that we observe the frame-of-reference problem: There need be no internal representation of the task within the agent. Often, the distinction is not so relevant: Both task and desired behaviors can be used to specify what an agent should do.

The ability to survive in complex environments is a given for all biological systems. Achieving this ability in artificial agents turns out to be an extremely hard problem. Complete autonomous agents are physical systems that are able to resolve these issues. For fun and for historical reasons we also call these complete autonomous systems ''Fungus Eaters.'' Let us briefly look at the story of these

''Fungus Eaters.'' They illustrate the main intuitions underlying the embodied cognitive science framework.

In 1961 the Japanese psychologist Masanao Toda[1] proposed to study ''Fungus Eaters'' as an alternative to the traditional methods of academic psychology (Toda 1982, chap. 7). Rather than performing ever more restricted and well-controlled experiments on isolated faculties (memory, language, learning, perception, emotion, etc.) and narrow tasks (memorizing lists of nonsense syllables, letter perception on degraded stimuli, etc.), we should study ''complete'' systems, though perhaps simple ones. ''Complete'' in this context means that the systems are capable of behaving autonomously in an environment without a human intermediary. Such systems have to incorporate capabilities for classification, for navigation, for object manipulation, and for deciding what to do. The integration of these competences into a system capable of behaving on its own, according to Toda's argument, will yield more insights into the nature of intelligence than looking at fragments of the complex human mind.

The ''Solitary Fungus Eater'' is a creature—in our terminology, an autonomous agent—sent to a distant planet to collect uranium ore (see figure 4.1). The more ore it collects, the more reward it will get. If feeds on a certain type of fungus that grows on this planet. The ''Fungus Eater'' has a fungus store, means of locomotion (e.g., legs or wheels), and means for decision making (a brain) and collection (e.g., arms). Any kind of activity, including thinking, requires energy, if the level of fungus in its fungus store drops to zero, the Fungus Eater dies. The Fungus Eater is also equipped with sensors, one for vision and one for detecting uranium ore (e.g., a Geiger counter).

The scenario Toda describes is interesting in a number of respects. Fungus Eaters must be autonomous: They are simply too far away to be controlled remotely. This autonomy in turn implies situatedness: Because they cannot be remote controlled, they have to view the world from their own perspective; that is, the only information the agent has available is acquired through the sensors in interaction with the environment. Fungus Eaters must be self-sufficient, because there are no humans to exchange their batteries and to repair them. They must be embodied, otherwise they would not be able to collect anything in the first place. All this implies

[1] This is our own interpretation of his paper; Toda may not agree with it.

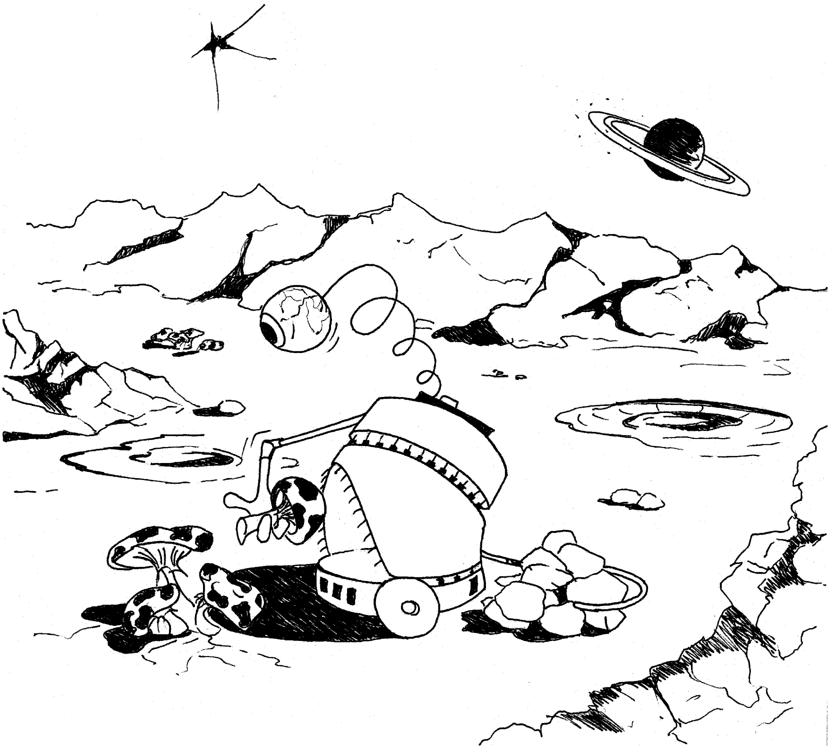

Figure 4.1 Toda's Fungus Eater, a complete autonomous agent. The robot is operating on a distant planet. Its task is to collect uranium ore. It feeds on a certain type of fungus. It is autonomous (too far away for remote control), self-sufficient (it must take care of its own energy supply which, in this case, is a particular type of fungus that grows on this planet, thus the name Fungus Eater), embodied (it exists as a physical system), and situated (its knowledge about the environment is acquired through its own sensory system). In the figure, it is in the process of devouring fungus.

that they must be adaptive, because the territory in which they have to function is largely unknown. These concepts are fundamental to embodied cognitive science, and we now discuss each in turn.

Before we do so, however, let us first examine another reason why Fungus Eaters are of particular interest for the study of intelligence, one that relates to evolutionary considerations. Nature has always produced Fungus Eaters, that is, creatures capable of surviving in the real world. There are, for example, the single-cell entities that emerged from the primordial soup 3.5 billion years ago. Only 550 million years ago, the first fish and vertebrates arrived, insects 450 million years ago. Reptiles came 370 million years ago, dinosaurs 330, and mammals 250 million years ago. Primates appeared 120 million years ago, the great apes 18 million years ago, man in its present form only 2.5 million years ago. Writing was invented less than 5,000 years ago. Based on these considerations, Brooks (1991a) argues that the really hard part for nature was to get to the level where creatures could move around and had sensory abilities. Once that was in place, things became much simpler. If we do not understand this sensory-motor basis, we have no chance of ever understanding intelligence. This is another fundamental reason why we must study Fungus Eaters, that is, complete autonomous systems.

Self-Sufficiency

MULTIPLE TASKS AND BEHAVIORS

Self-sufficiency means an agent's ability to sustain itself over extended periods of time. This implies that the agent must maintain its energy supply. A biological agent must eat and drink. Moreover, it has to eat and drink the right combination of foods. A prerequisite of eating and drinking is that the food and drink be there: Humans have to go to the grocery store or a restaurant; an animal typically has to look for food in the environment, an activity called foraging. An agent must also take care of itself; that is, it has to stay sufficiently clean, and it has to try not to get hurt. In other words, it also has to avoid predators. Moreover, it has to get enough sleep. If these conditions are fulfilled, the biological agent can engage in activities leading to reproduction. (Note that this description in terms of tasks is our description as observers. It has nothing to do with what is going on inside the animal.)

Similar considerations apply to artificial systems. A robot, for instance, has to maintain its battery level, or if it is fuel driven, it has to maintain a sufficient fuel supply. To be considered self-sufficient, the robot should be able to maintain its energy supply without external human intervention. Thus, a robot running off a power cable is not self-sufficient. A robot should also maintain a certain operating temperature. If it gets too hot or too cold, it might be damaged. Moreover, it should not bump into things, and it should avoid perils. In addition, robots are always designed for a particular task, or several tasks. They have to clean a factory floor, vacuum a carpet, mow a lawn, deliver mail in an office, collect soda cans, give tours of a university institute, and so on. Hence, agents in the real world, be they animals or robots, always have to engage in multiple behaviors. From an observer's perspective, we can say that they are able to perform multiple tasks.

TRADE-OFFS AND DEFICITS

In the real world, there are always trade-offs. If a robot is collecting soda cans or food or cleaning a park, it always expends energy. So at some point, it must replenish its energy resources; that is, it must go to the charging station and plug itself into an outlet. While doing that, it cannot collect soda cans: It must remain at the charging station until its energy supply is sufficiently high again. So there is a trade-off: Doing one thing implies not being able to do another.

Note that losing energy while collecting soda cans or mowing a lawn is a given, determined by the physics of the agent: It will happen without the agent's knowing about it. If a cleaning robot is recharging, the office space gets cluttered with soda cans or the grass keeps growing without the robot's doing anything about it: Remember, the real world has its own dynamics. If it remains at the charging station for a long time, enough soda cans might have accumulated so that it is no longer possible for the robot ever to collect all of them again. Or, to put it differently, it has incurred an irrecoverable deficit. Another way of defining self-sufficiency, then, is as follows: An agent is self-sufficient if it can avoid irrecoverable deficits. In nature, evolution has ''solved'' this problem, but robot designers must explicitly deal with it. Figure 4.2 shows a robot that has incurred an irrecoverable deficit.

CIRCADIAN CYCLES

Natural environments have circadian cycles: environmental conditions that change over one day, such as lighting conditions, temperature, or humidity. Similarly artificial environments often have cycles: day-night cycles, or cycles in the frequency of people attending a place (coffee rooms are attended more during day time than at night), and so forth. Conditions for certain types of tasks are usually better during one segment of the cycle than during another. For example, an agent equipped with vision is better off during the day, whereas one with infrared (IR) sensors is better off at night, for the following reason. IR sensors are active sensors: They send out an IR signal and measure the intensity of the reflected IR light, a process that works well in the dark. By contrast, a robot equipped only with IR sensors has trouble during the day. Daylight contains

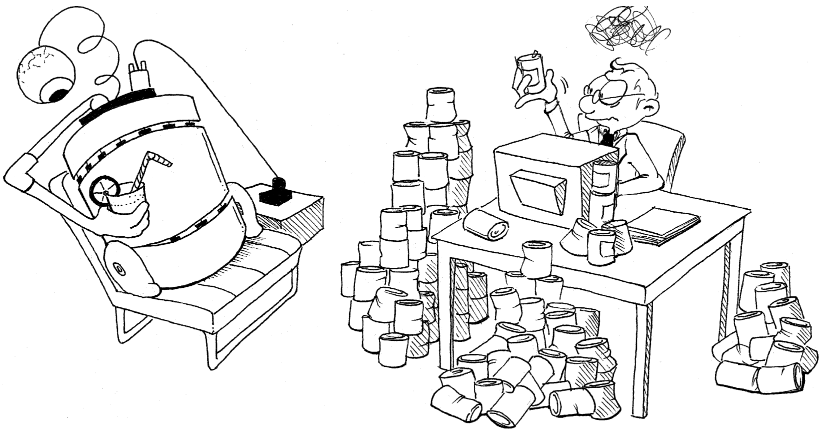

Figure 4.2 Robot incurring an irrecoverable deficit. Because the robot has been sitting at the charging station for too long, the soda cans have piled up in the meantime to a level where the robot is no longer capable of removing them all, even if it were to spend all of its "spare time," that is, all of the time it has available when not at the charging station, on can collecting. This robot is not self-sufficient.

a certain amount of IR light, which may cause interference with the reflected IR light. For the robot in figure 4.2, soda cans typically accumulate more quickly during the day. The target for a self-sufficient agent is always based on a circadian cycle: It should not incur a deficit over one cycle. If it does, then the deficit is likely to increase indefinitely, because the following day will typically bring an additional deficit. The concept of circadian cycles has not been widely used in embodied cognitive science and will not be further elaborated.

THE PROBLEM OF BEHAVIOR CONTROL

Complete systems always have several behaviors in which they must engage. Some of the behaviors will be compatible, others mutually exclusive. Because not all behaviors are compatible, a decision must be made as to which behaviors to engage in at each point in time. This is the problem of *behavior control.*

The most straightforward solution to this problem is to assume that there is an internal module or representation for each observed behavior category. For example, if we observe that a rat (or a robot) is following a wall, we might postulate that it has an internal module or a representation for wall following. Such a representation is often called an action. Because there are always multiple actions an agent has to engage in, to control behavior under this assumption, you need a mechanism for deciding which action to choose for execution at any given point in time, that is, which internal module to excute. In other words, you have to solve the *action selection problem*.

The problem with this approach to behavior control is that the assumption of a straightforward, one-to-one mapping from a specific behavior to a specific internal action does not reflect what actually occurs in natural systems. (Even the concept of an internal action represents an assumption.) To illustrate this point, let us look at an example. Assume that you are sitting in the cafeteria talking to a friend. Your friend has to attend a class and you are trying to describe his behavior. He gets up and starts moving toward the exit, avoiding chairs, tables, and people who stand around. To describe his behavior, you may want to use terms like ''avoiding a chair,'' ''going toward the exit,'' or ''going to class,'' implying that you somehow carve up your friend's behavior into distinct segments. There are two issues of which to be aware: First, the segmentation of an agent's behavior is observer-based and largely arbitrary. For example, you could also choose a more fine-grained segmentation such as ''getting up from chair,'' ''moving left leg forward,'' ''moving right leg forward,'' and so forth. Not surprisingly, segmentation of behavior is a notorious problem in psychology and ethology. For empirical purposes such a segmentation obviously has to be made, but we need then to make explicit that we are talking about purely observer-based categories. Second, it is not appropriate to conclude that for each of these behavioral segments there is an internal module.

There are mechanisms for behavior control, however, that do not require the existence of internal actions. Chapter 6 discusses an example, Braitenberg vehicles. In fact, we think that the problem of behavior control should be approached differently than described above. This follows from one of our design principles, the principle of loosely coupled, parallel processes (see chapters 10 and 11).

Autonomy and Situatedness

We have been using terms like ''autonomous agents'' and ''autonomous mobile robots.'' In this context, autonomy generally means freedom from external control. Autonomy is not an all-or-nothing issue, but a matter of degree. Complete, total autonomy does not exist; no agent is totally autonomous. It always depends to some

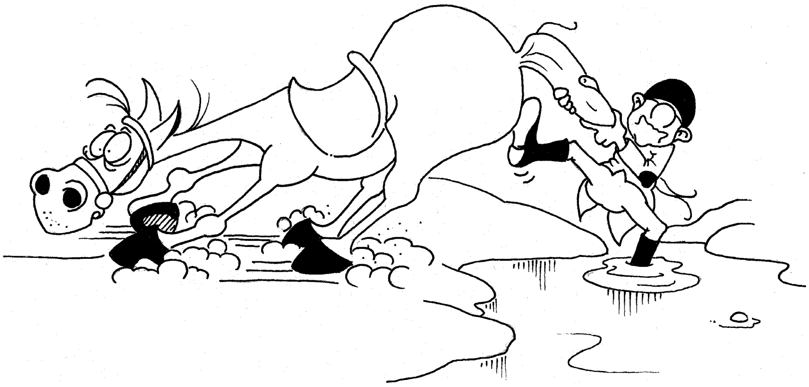

Figure 4.3 A horseback rider trying to control his horse. He is trying to force his horse to drink, not very successfully. The rider does exert some influence on the horse, and the horse is dependent on the rider for some things, but the horse is also to some degree autonomous. This is why the adage that "you can lead a horse to the water but you can't make him drink" has the ring of truth to it.

degree on external factors, factors beyond the agent's control. There are two aspects of autonomy here: dependence on the environment and dependence on other agents. Organisms depend on the environment for food, drink, oxygen, building materials, and the like. If agents are not capable of acquiring these resources on their own, they depend on other agents—they are less autonomous.

The main difference between dependence on the environment and dependence on other agents is that we do not attribute intentions to an environment, whereas an agent may want another agent to do certain things. Most parents want their children to do their homework and to perform well in school. We know, however, that parents have only a limited influence on their children: The latter have some degree of autonomy. The same holds for animals. We can get horses to do certain things we want them to do. But as the saying goes, "You can lead a horse to the water but you can't make him drink," again implying that the horse does have a certain degree of autonomy. So, in general, agents can be influenced, and they depend on others, but they are not completely controllable, as figure 4.3 illustrates.

From this discussion it becomes clear that when we use the term "autonomous agent," we mean an agent that has a certain degree of autonomy. It is not the case that an agent is either fully autonomous or not at all. From our discussion of self-sufficiency, it should be evident that self-sufficiency increases an agent's degree of autonomy, because a self-sufficient agent does not depend on another agent for its energy supply. The extent to which one agent can control another depends on the controlling agent's knowledge of the state and the internal mechanism of the agent to be controlled. The more precisely parents know what their children feel and think, the better they can influence them toward desired behaviors. One important reason that humans have only a very limited degree of controllability is that they have their own history, which is not, or is only indirectly and to a very limited extent, accessible to others.

Controllability and the capability of acquiring one's own history are correlated: The more an agent can have its own history, the less controllable it will be. The less parents know what their children do and what sorts of experiences they have, the less they know about what they feel and think. If they knew everything about them (including their reaction to all types of events)—which, of course, is impossible—they could easily make them do whatever they wanted, simply by manipulating the consequences of the children's actions according to what they knew the children's reactions would be. Because parents actually have only limited knowledge of their children's reactions, they have only limited control over them. Abstractly speaking, if the controlling agent (A) has access to the controlled agent's (B) internal state, and if he knows the laws by which the state of B can be influenced, A can control B completely, that is, A can get B into whatever state A wants B to be in. The less knowledge A has about B's internal state, the less A can control B. Thus, autonomy is not so much a property of an agent as a property of the relationship between agents (i.e., what one agent knows about the other). Stated differently, B has a certain amount of autonomy relative to A, and the amount of B's autonomy is—qualitatively speaking—inversely proportional to the amount of knowledge A has about B's internal state.

This property can be translated to robots. If a robot is equipped with a learning system, it can have its own experiences; that is, it can acquire its own knowledge over time. Note that this requires the agent to be situated. Recall the notion of situatedness from chapter 3: An agent is situated if it acquires information about its environment only through its sensors in interaction with the environment. A situated agent interacts with the world on its own, without an intervening human. It has the potential to acquire its own history if it is equipped with the appropriate learning mechanisms. Such an agent is potentially more autonomous than its preprogrammed, purely reactive counterpart. One implication of learning is that if the agent, after learning, encounters the same situation it has previously encountered, it will react differently than earlier on. Thus the more the agent has learned in the meantime, the more experiences of its own it has had, the less it will do the same as before, and thus, the less another agent will be able to control it, because its internal state will have changed, and the second agent will now have less knowledge of its internal state than it did previously. From this we can conclude that if we are interested in building autonomous agents, we must design them with learning components, because the capacity to learn increases an agent's autonomy. An agent's degree of autonomy can, in principle, be further increased by applying evolutionary methods (described in chapter 8). If he designs a robot not directly but via an additional evolutionary process, the designer has less control over how the robot will work and how it will behave in a particular situation. Applying evolutionary techniques often makes it difficult for designers—and for other agents in general—to understand why the agent is doing what it is doing; as the agent evolves and acquires its own history, it is progressively more difficult for the designers to understand (and manipulate) its behavior. Evolution makes the agent more independent of designers, and therefore evolved agents have the potential for higher levels of autonomy.

Embodiment

Autonomous agents are real physical agents; in other words, they are embodied. Because we have talked so far exclusively about biological agents (humans or animals) or about robots, it has been implicit that the agents of interest have to be embodied. Embodiment has proven to be an essential characteristic whose importance can hardly be overemphasized. A fundamental consequence of embodiment is that embodied agents must interact with their environments. To understand this interaction, we have to study, for example, how organisms acquire experience: knowledge about the environment obtained by interacting with it. This is one of the hardest problems in the study of intelligence. The vast research field of perception is devoted to elucidating the underlying mechanisms and processes.

Embodiment implies that the agent is continuously subjected to physical forces, to energy dissipation, to damage, in general to any influence in the environment. On the one hand, this complicates matters considerably. On the other, this often leads to substantial simplifications, because advantage can be taken of the physics involved. It has been demonstrated, for example, that walking robots can be built that require no electronic control: They are entirely brainless machines, their actions governed totally by the laws of physics.

The focus on embodied agents often leads to surprising insights, and throughout the book, we provide examples of such insights. We discuss embodied perspectives on learning, categorization, perception, memory, and sensory-motor processing. As the name of the field indicates, embodiment is at the core of embodied cognitive science. It is one of the central constituents in Brooks's (1991a,b) approach, which he called ''embodied intelligence.'' The idea that intelligence can emerge only from embodied agents is one of the fundamental assumptions of embodied cognitive science. (For other perspectives on embodiment see, for example, Lakoff 1987 and Varela, Thompson, and Rosch 1991).

Adaptivity

CHARACTERIZATION AND DEFINITION

Adaptivity is really a consequence of self-sufficiency. If an agent is to sustain itself over extended periods of time in a continuously changing, unpredictable environment, it must be adaptive. Remember that several of the definitions of intelligence given in chapter 1 alluded, in one way or another, to the concept of adaptivity, that is, the ability to adjust oneself to the environment. Thus, adaptivity and intelligence are directly related.

By adaptation, we mean that some structure is maintained in changing environmental conditions. Ashby (1960) used the term ''homeostasis,'' meaning that certain variables, the essential variables, remain within given limits (figure 4.4). Within those limits the organism can function and stay alive. This is called the ''viability zone'' (Meyer and Guillot 1990).

KINDS OF ADAPTATION

The term ''adaptation'' has various meanings and is used in different ways by different people. In our discussion, we follow McFarland (1991):

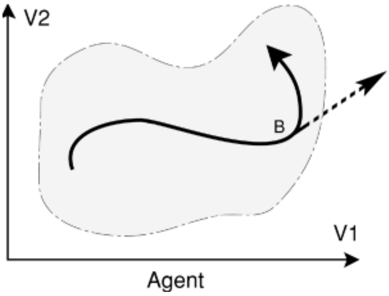

Figure 4.4 Adaptivity. The figure shows the viability zone (enclosed area) between two variables *V1* and *V2* (e.g., level of blood sugar and body fluid). Within this zone, the agent can stay alive and function. The solid arrow marks the agent's trajectory, that is, the development of the two variables over time. At point B, there is a danger that the agent might leave the viability zone (marked by the broken line) if it does not act. The agent is adaptive because it takes corrective action to prevent itself from leaving the viability zone. (Adapted from Meyer and Guillot 1991.)

Biologists usually distinguish between (1) evolutionary adaptation, which concerns the ways in which species adjust genetically to change in environmental conditions in the very long term; (2) physiological adaptation, which has to do with the physiological processes involved in the adjustment by the individual to climatic changes, changes in food quality, etc.; (3) sensory adaptation, by which the sense organs adjust to changes in the strength of the particular stimulation which they are designed to detect; and (4) adaptation by learning, which is the process by which animals are able to adjust to a wide variety of different types of environmental change.'' (p. 22)

Here are a few illustrations of the types of adaptation McFarland discusses (see also McFarland 1991):

1. *Evolutionary Adaptation:* An illustration of evolutionary adaptation is the peppered moth (*Biston betularia*). Originally these moths were light in color, which made them well camouflaged against lichen-covered, light-colored trunks of trees. In regions that became industrialized, industrial smoke darkened the tree trunks. Gradually the peppered moth population in industrial areas became predominantly composed of a dark variety, which was well camouflaged against the dark trees.
2. *Physiological Adaptation:* Many species can adapt to changes in environmental temperature: sweating, in man, is an example of adapting to heat changes.

3. *Sensory Adaptation:* If we are in a dark room and then the light is turned on, the eye adjusts to the change in a sensory stimulus, light intensity, by changing the diameter of the pupil.
4. *Adaptation by Learning:* This is a very general form of adaptation and is exploited in many ways. Animals can learn which food is most nutritious, where food can be found, which place gives the most shelter, and so forth.

Note that these different kinds of adaptations work on different timescales. Typically, sensory adaptation is the quickest, whereas evolutionary adaptation takes many generations. In this book, we focus mainly on adaptation by learning and through evolution.

Ecological Niches and Universality

DEFINITION

If we look at biological agents—animals—we find that they require a particular kind of environment for survival that is suited to satisfy their needs. Such an environment is called an animal's ''ecological niche''. Wilson (1975) defines ''ecological niche'' as follows: ''The range of each environmental variable such as temperature, humidity, and food items, within which a species can exist and reproduce'' (p. 317). It should be added to this definition that niche occupancy by a particular species usually implies competition. Different occupants of the niche compete for the same resources like food and space.

In nature, there is no such thing as a ''universal animal.'' Animals (and humans) are always ''designed'' by evolution for a particular niche. (We put the term ''designed'' between quotation marks to indicate that it is meant metaphorically: Evolution does not have a particular design goal.) Agents behave in the real world. As we pointed out, they always require certain conditions for their survival. A robot always requires some kind of energy source. It must be equipped with sensors and effectors in order to perform its task in a particular environment, or more precisely, in a particular ecological niche. To take the earlier example, if the robot has to work at night, it may be better to equip it with IR devices rather than with vision sensors. So, the idea of an ecological niche holds for robots as well (focus 4.1). It follows that there can be no universal robot, because the robot must perform in the real world, which consists of many varied environments to which a particular