I The Study of Intelligence—Foundations and Issues

This book's central goal is to allow the reader to acquire a deeper understanding of intelligence. A number of consequences follow from this goal. First, we have to define what we mean by ''intelligence.'' Second a ''deeper understanding'' implies that our current understanding is insufficient and needs to be improved. Thus, we need to ferret out in what respect it is not satisfactory, which in turn requires analyzing our current view, its underlying assumptions, and its ramifications. Part I is devoted to the elaboration of these points.

Although we all have a good idea of what we mean by ''intelligence,'' there is no general agreement on a particular definition. Moreover, questions like ''Are animals intelligent?'' ''Can animals think?'' ''Can computers (or robots) be intelligent?'' ''How can we measure intelligence?'' ''Is intelligence inherited or can it be acquired?'' and ''To what extent are emotions involved in intelligence?'' provoke a great deal of disagreement. Chapter 1 conveys a flavor of all the aspects and the variety of ideas involved by looking at definitions, commonsense notions, and ways of testing intelligence. Just to illustrate the topic's complexity and controversial nature, the chapter presents the IQ test, ''emotional intelligence,'' and the nature-nurture debate, as well as a test for machine intelligence, the Turing test. Once the parameters of the field ''intelligence'' have been delineated, the chapter discusses the various ways intelligence can be and has been investigated. Finally, it introduces the main methodology to be used in this book, the synthetic methodology; in particular, the use of so-called autonomous agents to investigate intelligence is considered.

Once we are clear about what we mean by ''intelligence'' and how it can be investigated, we are in a position to analyze the different theoretical positions. In cognitive science, empirical and theoretical research on intelligence has been dominated by the computer metaphor: intelligence as information processing, as the manipulation of abstract symbols—the essence of the cognitivistic paradigm. The cognitivistic paradigm—elaborated in chapter 2—is intuitively highly appealing and has attracted many of the leading researchers over the last half century or so. As it has turned out, however, the paradigm has a number of undesirable implications that cannot be resolved within the framework it sets up. Very broadly speaking, they all concern the fact that humans, animals, and robots have to interact with the real world, whereas the computer metaphor has focused on abstract virtual or computational worlds and has neglected their relationship to the real world. Chapter 3 discusses the problems and issues this neglect of the real world entails. One very prominent problem, the symbol grounding problem, concerns how the symbols used in a model acquire meaning, that is, how they relate to an organism's experience. Although many solutions to these problems have been suggested, radically different approaches are required if we are to come to grips with them, and this is the crux of the entire book: elaborating these alternative approaches.

1 The Study of Intelligence

Intelligence has always been a controversial topic. Science fiction stories involving intelligent robots abound. Superintelligent machines have, for a long time, been the stuff of nightmares. Computers and, even more so, robots have inspired people's fantasies. Because of the enormous developments in digital electronics and microtechnology in recent years, true artificial intelligence seems to be drawing near. So it is not really surprising that discussions concerning artificial intelligence are often highly emotional. But nightmares and science fiction do not entirely explain the issue's emotional charge. Intelligence was an emotional topic long before computers started to spread. Just think of IQ tests. There has been a long and heated debate about what IQ tests actually measure: Is it really intelligence, or something else? And what about the recent hype about "emotional intelligence"? Is emotional intelligence, rather than IQ, the real intelligence? Another question often asked: Are ants intelligent? Or ant colonies? Are rats intelligent? Maybe not, but they are certainly *more* intelligent that ants. And humans are more intelligent than rats—at least in many respects. Most adults can speak and write and many can play chess—activities no animal can perform. But among humans, talking or playing chess (at least at a basic level) is not considered something exceptional. I, Rolf Pfeifer, know how to play chess, but nobody who has seen my performance in a game would attribute extraordinary intelligence to me. However, if a one-year-old child did exactly the same thing, we would think that the kid was superintelligent. If a dog did it, we would think the dog was a genius. So what we consider intelligent depends also on our expectations. But not only that: Assume you are playing chess against a computer. If you win, you can be happy. But even if you lose, you might still argue that you were playing intelligently, whereas the computer was only testing many alternatives in a completely unintelligent way, as figure 1.1 illustrates.

Well, you might have been able to make that argument, at least, until the May 11, 1997. On that date the world was focusing on a particular room on the 35th floor of 787 Seventh Avenue in New

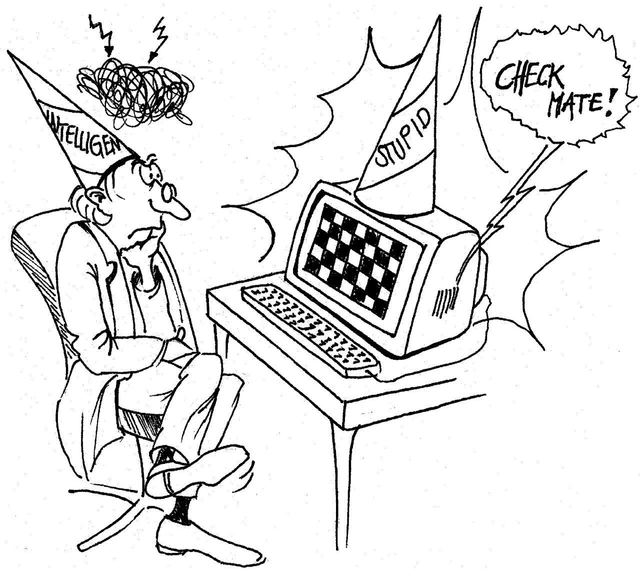


Figure 1.1 A human playing chess against a computer. Although the human is losing, he still feels he is intelligent, whereas he considers the computer to be stupid. Even after the historic victory of IBM's Deep Blue over world champion Garry Kasparov in 1997, the reaction of the human is still justified. Deep Blue's success is due largely to processing speed.

York City, where, for the first time in history, a chess program won an entire match against the reigning world champion. The hapless champion was Garry Kasparov, the chess program, Deep Blue, developed by a research team at IBM. Kasparov won the first of six games and lost the second. The next three games were draws. At this point, both Kasparov and Deep Blue had 2.5 points, with just one game to go. As we all know, Kasparov lost the final one. What does that mean? Is the person's reaction in figure 1.1 still justified? Or is it indeed the case that now computers have achieved human level intelligence? Deep Blue's victory is certainly a milestone in the history of artificial intelligence. After all, chess was considered the hallmark of intelligence in the old days of artificial intelligence. But we hope to demonstrate in this book that the person in figure 1.1 can relax: Nothing has changed fundamentally. The decisive factor in Deep Blue's victory was the speed of the computer. So this victory was a logical development, to be expected sooner or later. More is required, however, before we can speak of intelligence.

This book is about intelligence. So we should somehow be able to tell what we mean by the term. This is not an easy task, as we have already begun to see. There is very little agreement on what does and does not constitute intelligence. For the most part, the discussion of what intelligence is and isn't seems to concern what people find interesting and what they don't. Some find it interesting that termites can construct enormous buildings and that birds can fly in flocks with marvelous shapes. Others are amazed that humans can speak and recognize a particular face in a large crowd. Still others wonder about dogs catching Frisbees. Almost everybody is impressed with Einstein's achievements in general relativity. And most are still fascinated by grand masters playing a game of chess. To do justice to this variety, we start with a *tour d'horizon* of what many people have said about the phenomenon of intelligence. As we do so, we have to be aware that intelligence is a descriptive term: It describes certain properties of individuals or groups of individuals. Descriptive terms are largely arbitrary, and it is therefore unlikely that descriptive definitions of complex ideas can satisfy everybody. Nevertheless, all definitions of intelligence have a common denominator related to novelty and adaptivity. This forms the starting point of our investigation.

An exact characterization of intelligence is not all that important to understanding it. What does matter is that we work on the relevant issues. Rather than arguing whether a particular behavior should be called intelligent or not—a point that is always debatable—we try to provide answers to the following question: Given some behavior—say of a human, an elephant, an ant, or a robot—that we find interesting in some ways, how does the behavior come about? If we can give good answers to this question for a broad range of behaviors, we can say that we have gained an understanding of the principles underlying intelligence. This is precisely what we are after in this book. Of course, we have to define exactly what we mean by "good answers": Our entire conception depends on this. We do just that, in detail, throughout the book. Thus we are suggesting that we replace the original question of defining intelligence with the more profitable one of how a particular behavior in which we are interested comes about.

Before we start, let us introduce a few terms. By *cognitive science* we mean the interdisciplinary investigation of intelligence, or more generally, the mind. We are mostly interested in that part of cognitive science that applies a synthetic methodology, that is, the methodology of "understanding by building." Cognitive science is also concerned with exploring general principles of intelligence, not only those related to the human mind. It has a large overlap with *artificial intelligence* (*AI*). The difference between the two fields is that cognitive science has closer ties to empirical sciences like psychology, biology, and neurobiology, whereas AI is more closely associated with computer science, algorithms, and logic. But many researchers in AI would consider themselves cognitive scientists. We sometimes use the terms *classical AI* to distinguish the traditional approach from the more recent one described in this book, which we call *embodied cognitive science*. When talking about intelligence, we often do not want to make any distinction among humans, animals, and artificial creatures like robots or simulated organisms. In these cases we normally use the term *agent*.

1.1 Characterizing Intelligence

We start our *tour d'horizon* with a few definitions of intelligence, move on to commonsense notions, then discuss intelligence testing, a particular way of characterizing intelligence. We then turn to a very special kind of intelligence test, the Turing test, and a famous thought experiment, the Chinese Room. From this cursory review, we then define our starting point.

Definitions

As we said, it is hard to define intelligence, and not much agreement has been achieved. The introductory comments on intelligence in the Penguin *Dictionary of Psychology* reflect this lack of consensus: ''Few concepts in psychology have received more devoted attention and few have resisted clarification so thoroughly'' (Reber 1995, p. 379). Nevertheless, definitions can provide a source of intuition, so let's examine some. In 1921, the *Journal of Educational Psychology* (Vol. 12, pp. 123–147, 195–216) asked fourteen leading experts in the field at the time to provide their definitions of intelligence. As one might expect, the journal got 14 different answers back. Some of the responses received can be summarized as follows: The ability to carry on abstract thinking (L. M. Terman); Having learned or ability to learn to adjust oneself to the environment (S. S. Colvin); The ability to adapt oneself adequately to relatively new situations in life (R. Pintner); A biological mechanism by which the effects of a complexity of stimuli are brought together and given a somewhat unified effect in behavior (J. Peterson); The capacity to acquire capacity (H. Woodrow); The capacity to learn or to profit by experience (W. F. Dearborn). Although the definitions are different, they all make certain points that we find important. Note the very different levels involved. Terman talks about the ability for abstract thinking. By contrast, Peterson refers to biological mechanisms. A crucial point: Some mention the environment, some don't. In many investigations of intelligence, the environment was largely neglected.

The quotations above represented the opinion of experts. Let us now look at what people in general think about intelligence, at commonsense notions of intelligence. You may be surprised at some inclusions in this list.

Commonsense Notions

It is important to understand commonsense notions of intelligence, first because they are a great source of inspiration, and second because, ultimately, the scientific study of intelligence must relate to them: It must provide a better understanding of precisely these concepts. Commonsense notions often specify certain capabilities typical of intelligent beings. They include, among others, thinking and problem solving; the competence to speak, read, and write; intuition and creativity; learning and memory; emotions; surviving in a complex world; and consciousness. They also include the distinction of degrees of intelligence.

A GRADUATED PROPERTY

The first thing to note is that people clearly distinguish levels of intelligence. Albert Einstein (figure 1.2) was certainly extremely intelligent. If you want to go to college, you have to be intelligent. The word is often used in this sense, namely as a synonym for ''very intelligent,'' ''more intelligent than others.'' When we say a person is intelligent, we normally mean that the person has an above average level of intelligence.

Obviously, some people are more intelligent than others. Humans are more intelligent than animals, and among animals, dolphins and apes are considered more intelligent than cows or ants. We have a tendency to order living beings as being more or less intelligent—intelligence is not a characteristic that is either present or not, rather one that is present in degrees. But it is also

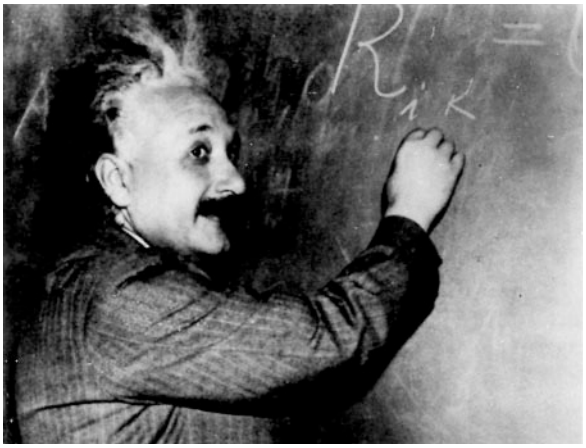


Figure 1.2 Portrait of an intelligent person. There is universal agreement that Einstein, an enormously creative thinker, was highly intelligent.

clear that ordering intelligence on a linear scale is not possible. Some students are good at writing essays, others can do math, still others play music, and a fourth group might excel at camping out in the wild: How should we compare their intelligence? It is not obvious how such a comparison can be made in a sensible and profitable way.

THINKING AND PROBLEM SOLVING

The ability to think is often mentioned as an essential characteristic of intelligence. Thinking, in its commonsense meaning, includes problem solving and logical reasoning but also less structured forms of mental activity such as those we use in our everyday lives, when doing household chores or planning a weekend trip. Most people would probably agree with the ordering of the degrees of intelligence of animals mentioned in the previous section. This implies that animals also have a certain level of intelligence. But do animals think? The capacity to think is a characteristic of an intelligent being in commonsense belief. Well, maybe some animals do think, and others don't. We have no way of really knowing. To find out, however, we could conduct an experiment. For example, we could give a horse an arithmetic problem in some form, as is sometimes seen on TV shows, and if it comes up with the right answer, say by knocking on the ground the correct number of times, we might say that it has been thinking. The fact that these demonstrations have been shown to be tricks is beside the point here. What matters is that we never *know* whether another agent is thinking or not: We can only speculate about it.

Problem solving is closely related to thinking. Typical problem solving tasks are finding a bug in a computer program, diagnosing the disease of a patient, finding a solution to a high school physics problem, designing an experiment with animals to test a hypothesis, or compiling a portfolio for a particular customer.

In its everyday meaning, the term ''thinking'' is often associated with conscious thought. This is compatible with Terman's view of intelligence as abstract thinking. It is also what the philosopher René Descartes had in mind when writing his famous statement ''Cogito ergo sum.'' Abstract thinking is perceived as especially hard by most, and individuals with this ability often command respect and admiration. This ability to think in abstractions is the first one mentioned, almost universally, by most people when asked to define intelligence. Upon further reflection, they come up with all sorts of additional conceptions. Let's look at some of them.

LEARNING AND MEMORY

Good students are usually perceived as the ones that learn easily. We also say that they have a good memory. They study the words once, for example, and they know them and they do not forget them. Many people view learning as the core property of intelligence. That learning per se does not make people intelligent, but the capacity to learn, is also a popular view. So learning to learn appears to be the key point.

Memory is considered equally important, in popular conceptions of intelligence, as capacity to learn. However, rote learning, merely memorizing facts out of context, is generally judged a pointless activity, basically a waste of time, requiring no intelligence. Memory for useful knowledge is what counts. A doctor with extensive experience who can remember all his patients and their diseases and can apply this knowledge to treat new cases is considered intelligent. Transfer of knowledge is the point, not merely storing it.

LANGUAGE

The capacity to communicate in natural language, as we know it from humans, is often considered to be the hallmark of intelligence. Clearly, natural language requires a high level of intelligence. The ability to talk to one another, to read and write, is one of humans' distinguishing features. No animal species has abilities even remotely resembling human natural language. Those who speak multiple languages are often regarded as particularly intelligent. Their ability is a combination of good learning, memory capacity, and talent for languages.

INTUITION AND CREATIVITY

Einstein was creative; so were Beethoven and Picasso. They also had a lot of intuition. Leaders and managers have intuition, too. In fact, all (or most) people do. Intuition is often taken to mean arriving at conclusions without a train of logical thought that can be traced to its origins. Likewise, creativity is a highly complex notion that includes not only the individual but the society as whole. It cannot be defined for an individual in isolation but must be discussed with respect a particular society's value criteria. Many regard creativity as the highest form of human intelligence.

Both intuition and creativity seem in some ways to go beyond thinking. Thinking can be executed in a "cold" manner, independent of emotion, whereas intuition and creativity require the engaging of emotions. Creating something new also has a somewhat mysterious flavor: How does the new thing come about? Can creativity be learned?

CONSCIOUSNESS

Consciousness is often seen as an essential ingredient of intelligence. Like creativity and language, it is a property that we can attribute with certainty only to humans. And like creativity, there is also something mysterious about consciousness: It is hard to grasp, but considered essential for many other abilities. Thinking, language, and creativity are understood as requiring consciousness. Creativity, for example, is seen as the result of a combination between conscious thought and unconscious processes. Because of its subjectivity, consciousness is an elusive concept; it is hard to know what it really is all about. Academic psychology has deliberately tried to avoid dealing with consciousness at all, arguing "that the role of consciousness in mental life is very small, almost frighteningly so. The aspects of mental life that require consciousness have turned out to be a relatively minor fraction of the business of the brain" (Bridgeman 1990, cited in Rosenfield 1992). Although Bridgeman may indeed be right, consciousness is nevertheless seen as important for the study of intelligence by many people.

EMOTIONS

Humans have emotions. Like consciousness, they are something we consider essential for humans. Moreover, most people think that higher mammals, in particular apes and dolphins, but also dogs and cats, have emotions. Whether emotions should be considered an essential feature of intelligent beings, however, is debatable. Recently, so-called emotional intelligence, introduced by Peter Salovey and John Mayer (1990), has been the subject of much discussion. Emotional intelligence refers to the ability to recognize emotions in others, using emotions to support thinking and actions, understanding emotions, and regulating emotions. The general idea is that if you recognize your own emotions, you are better able to perceive the emotions in others and to react appropriately in social situations (Goleman 1995). Apparently this ability can be improved through appropriate practice. Pertinent seminars are already being marketed worldwide.

It is generally agreed that the degree of sophistication of emotions depends to a large degree on intelligence. Humans can be jealous; they can be ashamed or feel guilty. We would normally not attribute such emotions to ants. We would also not, for example, ascribe guilt to a lion that has just killed a deer, whereas we would certainly attribute guilt as a likely emotion for a human who has killed another human.

SURVIVING IN A COMPLEX WORLD

Animals (and humans, for that matter) can survive in highly complex environments, and they sometimes display astounding behaviors. Termites build fantastic towers, and bees dance and communicate, in sophisticated ways, the location of food sources. Other animals use tools in skilled ways. Certain vultures hurl a stone at an ostrich egg to break it, Galapagos woodpecker finches probe for insects in the bark of trees by holding a cactus spine in their beaks, and chimpanzees use twigs to probe for termites. Primates exhibit sophisticated social behavior. We cannot help attributing some kind of intelligence to these creatures and those that engage in similarly sophisticated survival behaviors.

PERCEPTUAL AND MOTOR ABILITIES

Most people don't consider perceptual and motor abilities essential for intelligence. Presumably, seeing the things around us seems so natural and works so automatically that we are not aware of the complexities involved. By contrast, science considers understanding perception one of the most important research issues. Recognizing complex objects in our environment, making out a face in a crowd are amazing abilities to a scientist trying to explain them. Medical doctors, experienced diagnosticians, can sometimes find out what's wrong with a person simply by looking at him. Such perceptual competences are sometimes seen as intelligence. Motor abilities, on the other hand, especially basic ones like walking, are usually thought to require no intelligence. As the complexity of the motor task increases, however, it becomes less and less clear to what extent intelligence is required. Assembling a complex electronic device requires high sensory-motor skills, but does it call for intelligence?

This discussion of commonsense notions of intelligence is, of course, neither complete nor empirically sound. The aim was to provide a sense of the variety of abilities and components involved in what we, scientists and laypeople, think of as intelligence. As we have seen, intelligence is multifaceted and not restricted to one characteristic, like abstract thinking. We have also seen that, in addition to humans, animals often exhibit impressive levels of intelligence. Moreover, there seems to be agreement that intelligence is a gradual rather than an absolute characteristic, though it is not obvious how it should be measured. This is the task of intelligence testing.

Intelligence Testing

Numerous tests for assessing intelligence have been developed. A case in point are IQ tests. The general idea of an IQ test is to measure a capacity that is not dependent on particular knowledge but is, in a sense, a ''general intelligence capacity'' or ''factor g,'' as it is sometimes called.

The original IQ test was invented in 1905 by French psychologist Alfred Binet, essentially to find out whether children with certain learning deficiencies would be better off in a special school. German psychologist William Stern in 1912 turned the test into a general intelligence test for children, and David Wechsler in 1939 developed it into one for adults. He proposed a Gaussian distribution of test results: two thirds should be between 85 and 115 (100 being the mean), and only 2.3 percent above 130 and below 70. Figure 1.3 depicts a typical item on a modern IQ test.

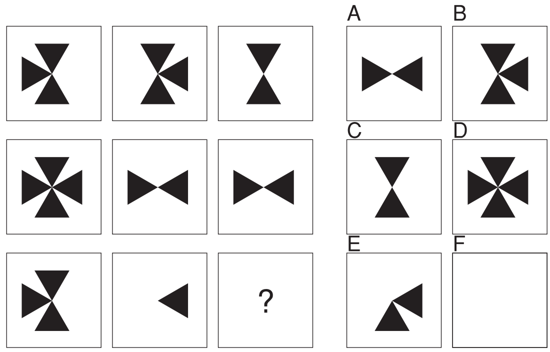


Figure 1.3 Typical problem from an IQ test. One item from the panel on the right (A through F) has to be chosen for the field with the question mark.

In 1904, English psychologist Charles Spearman, in a paper entitled "'General Intelligence' Objectively Determined and Measured" (Spearman 1904), used factor analysis, a method he invented, to support his claim that factor g indeed exists. Spearman based his argument on the finding that there are positive correlations between the different test items on an IQ test. According to Spearman, these results suggest that an underlying factor is responsible for the correlations. Although some psychologists still regard factor g as the most fundamental measure of intelligence, others postulate multiple intelligences, a view supported by recent evidence that more than seventy different abilities can be distinguished by currently available tests (Carroll 1993).

We can conclude that it is problematic to reduce a highly complex phenomenon like intelligence to a single number. This is also the essential point in Howard Gardner's theory of multiple intelligences, or multiple competences. According to Gardner, there is not one intelligence or factor g but multiple ones: linguistic intelligence, musical intelligence, logical-mathematical intelligence, spatial intelligence, bodily-kinesthetic intelligence, and personal intelligences (for perceiving your own and other people's moods, motives, and intentions). Gardner's list of intelligences suggests that there is no simple mapping of intelligence onto one dimension, one number (Gardner 1985). He also argues that some of these competences cannot be measured using standard tests, hence the German translation of Gardner's book has the title *Abschied vom IQ*, which means "Farewell, IQ."

But before we dismiss IQ entirely, let us recall one of the definitions of intelligence provided by the experts in 1921, namely, the

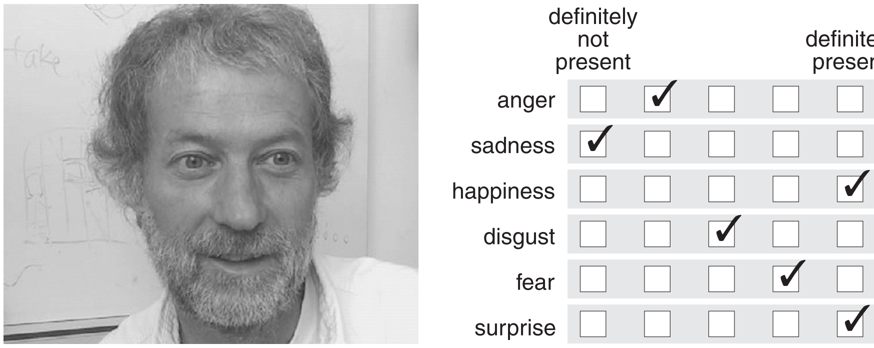


Figure 1.4 Example of a problem from an EQ test. EQ tests typically consist of four parts, one for identifying emotions, one for using emotions, one for understanding emotions, and one for regulating emotions. The figure shows test items from the test for understanding emotions.

ability to profit from your experience, to be successful in a particular environment. If we take as the environment an industrialized society, it seems that IQ is a good predictor of success in school and in professional life (e.g., Neisser et al. 1996). Recently, some have suggested that emotional intelligence might be equally important for a successful career (e.g., Goleman 1995). Because tests for emotional intelligence (EQ tests) on the one hand are controversial and on the other have only been around for a short period of time (at least compared to IQ tests), it is unclear how exactly they relate to IQ tests. To provide a feel for what these tests are like, we have included an item from an EQ test in figure 1.4.

Testing to measure intelligence has raised the question of whether intelligence is genetically predetermined and to what extent it is influenced by factors other than heredity. This has sparked a heated debate that keeps reemerging periodically: the nature-nurture debate.

The Nature-Nurture Debate

Generally speaking, the nature-nurture debate concerns the origins of knowledge. Those in the nature camp think that development is largely the expression of genetically predetermined factors. For example, it has been suggested that children are born with innate knowledge about basic principles of grammar (e.g., Pinker 1994), physics (Spelke 1994), or mathematics (Wynn 1992). By contrast, people in the nurture camp posit that most abilities are acquired during development and can be learned. The last violent eruption of this debate was in 1994, when Herrnstein and Murrey published their controversial book *The Bell Curve*, in which they claimed that the decisive factor in whether we will be successful in life is not our social environment, but intelligence as measured by IQ. They also maintained that IQ is largely innate, genetically predetermined. This position has, of course, far reaching consequences. For example, it suggests that some social programs are useless because the intended beneficiaries cannot be helped because of their innate limitations in intelligence, as expressed in low IQ scores. This view has a number problems. (See Gould 1996 for an excellent discussion of the main issues.) We mention only two. First, it assumes that intelligence can be captured by a single number, the IQ. Given our discussion of intelligence so far, this is clearly questionable. Second, it is not clear what is meant by the claim that intelligence is innate. Does it mean ''coded in the genes''? Genes interact with their environment at all levels, so that ''there is virtually no interesting aspect of development that is strictly 'genetic' '' (Elman et al. 1996, p. 21). Although there is a certain truth to both extremes in this debate—there are genetic factors in intelligence, and there are strong environmental components—the ''solution'' presumably lies somewhere in the middle, that is, that the origins of intelligence are to be found in the interaction between nature (genetic factors) and nurture (environmental factors). The problem then becomes determining how development actually works; that is, how precisely genetic and environmental factors interact in the developing organism. Computer simulations of how this interaction might be achieved in very simple organisms are given in chapter 8. These simulation studies lead to additional insights and new ways of thinking about the nature-nurture debate. Meanwhile the nature-nurture war continuous to be waged.

The nature-nurture debate is by no means the only controversy in the study of intelligence. Let us look at another, the intelligence of machines.

The Turing Test and the Chinese Room

So far we have dealt mostly with natural intelligence, because people normally associate intelligence with natural creatures, in particular humans. But what about machines? Can machines be intelligent? This question has led to long, emotionally loaded, and

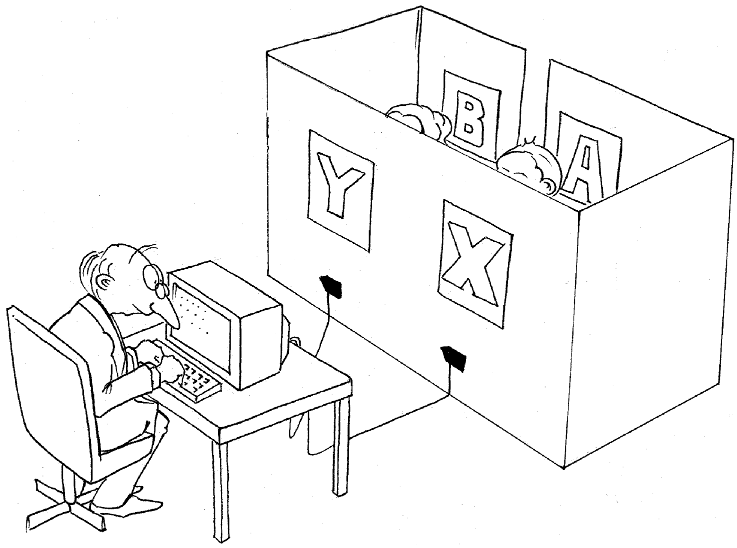


Figure 1.5 Basic setup of the Turing test. There are three participants, a man (A), a woman (B), and an interrogator (C). The interrogator is in a separate room, connected to the participants only via a computer terminal. His task is to find out who is the man and who the woman. A's goal is to confuse C whereas B tries to help C make the correct identification. The Turing test consists of replacing A by a computer: Can C then find out which is a computer and which a human?

generally nonproductive debates. Frustrated with discussions about the nature of intelligence, in which it is impossible ever to reach consensus because of the strongly subjective components involved, the brilliant English mathematician Alan Turing proposed an operationalization of the question whether machines could be intelligent at all. In 1950, in a seminal paper entitled "Computing Machinery and Intelligence (Turing 1950)" he proposed a procedure now widely known as the Turing test. The refreshing point about the Turing test is that it is an experiment, not speculation. Its results can be assessed objectively, and it does not refer to any kind of thinking or mental processes.

The Turing test consists of an imitation game. Figure 1.5 shows the basic setup. Let us quote Turing himself:

It (the imitation game) is played by three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. He knows them by labels X and Y, and at the end of the game he says either "X is A and Y is B" *or "X is B and Y is A." The interrogator is allowed to put questions to A and B thus:*

C: Will X please tell me the length of his or her hair?

Now suppose X is actually A, then A must answer. It is A's object in the game to try and cause C to make the wrong identification. His answers might therefore be:

"My hair is shingled, and the longest strands are about nine inches long." (Turing 1950; reprinted in Feigenbaum and Feldman 1963, p. 11)

In order that tones of voice may not help the interrogator, the answers should be written, or better still, typewritten. The ideal arrangement is to have two computer terminals (at the time the test was originated, teleprinters) communicating between the two rooms. Alternatively the questions and answers could be repeated by an intermediary. The object of the game for the third player (B) is to help the interrogator. The best strategy for her is probably to give truthful answers. She can add such things as "I am the woman, don't listen to him!" to her answers, but that will be of no avail, because the man can make similar remarks.

We now ask the question, "What will happen when a machine takes the part of A in this game? Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman?" These questions replace our original, "Can machines think?" (Turing 1950; reprinted in Feigenbaum and Feldman 1963, pp. 11–12)

The original Turing version of the test involves three parties (the interrogator, one person trying to help the interrogator, and one trying to confuse him), simpler versions have later been proposed in which the interrogator is interacting with a system (human or machine) and has to find out whether the system is a human or a machine.

There has been much discussion about whether the Turing test is a good test of intelligence. Many criticisms have been voiced. One often heard is that the test is constrained to measuring a particular form of natural language communication. One of the prominent critics of the Turing test, philosopher John Searle, has argued, in essence, that observing behavior is not enough, because by merely observing behavior we cannot find out whether a system really understands the questions it is given (Searle 1980). As a thought experiment, he proposed the famous Chinese Room (figure 1.6). In

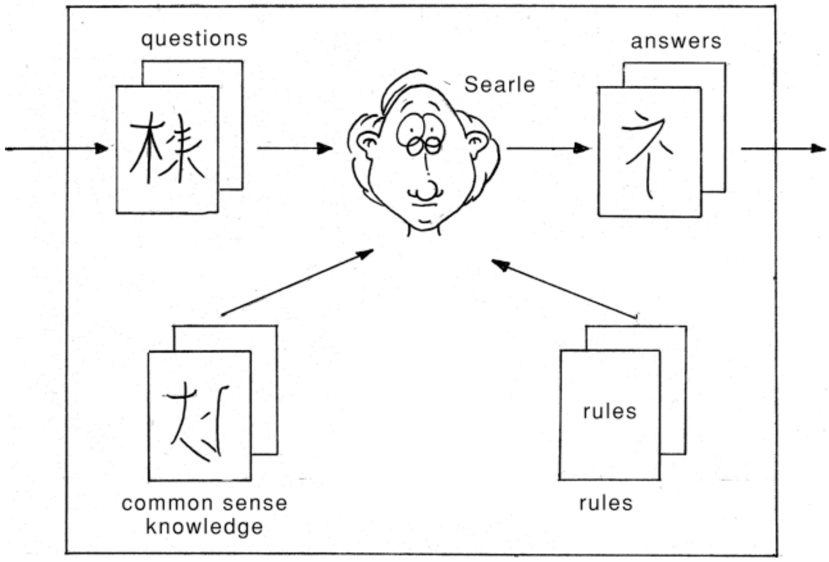


Figure 1.6 Searle's Chinese Room experiment. Using the rules and the commonsense knowledge, Searle is producing an answer to a question that is handed through a window in the room. Even though he does not understand Chinese, he can produce meaningful Chinese sentences.

his original paper, the person locked in the Chinese Room was Searle himself. The argument holds for anyone else, as long as he doesn't speak Chinese. Initially Searle is given two large batches of writing, one with Chinese characters and one written in English. The batch with the Chinese characters represents a data base of commonsense knowledge required to answer questions handed to him through the opening on the left of the room. The second batch consists of rules containing the instructions on how to ''process'' the questions, that is, they tell Searle how to produce an answer from the questions written with Chinese characters. This is done by comparing the characters of the question to the characters in the commonsense knowledge base and by choosing certain characters that will make up the answer. When this process is finished, the answer is handed through the opening on the right of the room. Note that the comparison of Chinese characters and the choice of characters that make up the answer is done entirely on the basis of their shapes, that is, on a purely formal or syntactic basis. Let us now suppose that Searle keeps playing this game for a while and gets really proficient at following the instructions for manipulating the Chinese symbols. From an external point of view, that is from the point of view of somebody outside the Chinese Room, Searle's answers to the questions are indistinguishable from those of native Chinese speakers. Nobody looking at Searle's answers can tell that he doesn't speak a word of Chinese. He has produced answers by manipulating uninterpreted formal symbols.

Searle, quite in contrast to Turing, is not willing to accept a definition (or a test) of intelligence that relies entirely on behavior. It is not sufficient for him that a system produce the same output as a human. He does not view the Turing test as a good means to judge the intelligence of a system. For true understanding, true intelligence—in his view—something else is required. Many papers have been written about the Chinese Room, and we cannot do justice to the entire discussion. Instead of going into that debate, let us, just for the fun of it, ask the following question: According to Searle, the Chinese Room does not understand Chinese. Now, how do we know Searle understands English? All we can do is say something, observe Searle's behavior and what he says in a particular situation, and if that makes sense, we attribute understanding to him. Just like the Chinese Room! But more probably, we know that Searle is human, we are humans and we understand English, so we simply assume that he also understands.

So far, in our description of the Chinese Room thought experiment, we basically followed Searle's line of reasoning. However, there is a serious problem with the argument. It suggests that there could indeed be a set of rules capable of producing the appropriate outputs based only on manipulation of meaningless characters. Remember that to Searle, the Chinese characters, the symbols, are entirely meaningless. If we interpret the rules as a computer program, then he suggests that there could be a computer program capable of producing the appropriate outputs (the answers) to the inputs (the questions), based on purely syntactic manipulation of some system of characters, the meaningless symbols. From half a century of computer linguistics research, it is well known that this does not work (e.g., Winograd and Flores 1986). At a minimum, this casts doubt on the primary assumption of the thought experiment (Clancey 1997).

To conclude this section on the Turing test, we mention one of its major limitations. If we are willing to attribute at least some level of intelligence to ants, rats, or elephants, the Turing test is obviously out as a tool for assessing intelligence. It can be applied only to systems capable of dealing with ''human'' natural language.

Whether it is a good test for human intelligence is still subject to debate (e.g., Crockett 1994; Epstein 1992).

The Common Denominator

We have now looked at various ways in which we can characterize intelligence. Our ultimate goal is to understand all of them: abstract thinking, learning and memory, natural language, medical diagnosis, surviving in the wild. But we have to start somewhere. If we look at the various characterizations from an abstract perspective, there seems to be one underlying common theme that involves ''coming up with something new.'' The ability to speak, for example, implies generating new utterances appropriate to the situation. ''Appropriate'' means that the speaker gets some benefit or value from his utterance—otherwise he wouldn't say it. We would not attribute the ability to speak, for example, to a person who always utters the same five sentences. Nor would we attribute intelligence to a factory robot that only repeats the same movements over and over again. The Turing test becomes interesting only when the interrogator asks new questions, questions that he suspects could not have been preprogrammed. When Terman talks about intelligence as the ability to carry out abstract thinking, what he really means is the ability to come up with something new, a solution to an abstract problem, a mathematical proof, an answer to a hard question, something that did not exist before. Surviving in the wild means coping with novel situations which in turn implies behaving in new ways. Or let us look at Pinter's characterization of intelligence as the ability to adapt oneself adequately to new situations in life. The term ''adapt'' often suggests something passive, conforming to existing rules. This is exactly what most people do *not* mean by intelligence. But there is another meaning to the term ''adapt'': to exploit a situation in order to benefit from it. For example, the business world has changed dramatically in recent years. Computer technology, electronic communication systems, in particular the Internet, are by now everywhere. Companies that have adapted to these changes by changing their business practices, by inventing new ways of doing business, have survived; the others have largely disappeared. Note that this innovation requires conforming to the rules of information technology. Both components, conforming and generating are always present. The key point is generation of diversity while complying with the givens. We call this the *diversity-compliance trade-off.*

And now we ask: What are the mechanisms enabling organisms to adapt to, cope with, environmental changes? As we noted, adaptation always contains two components: complying with existing rules and generating new behavior; only if both components are present do we speak of adaptivity. It then makes sense to tie intelligence to adaptive behavior. The term ''rule'' has been used in a very broad sense. It can refer to the rules of information technology, social rules, the rules of grammar, the laws of nature (e.g., physiology) and physics. This characterization of intelligence as the capacity to adapt is independent of levels. It applies to a mathematician carrying out abstract thinking, to a child talking to his parents (using natural language), and just as much to an animal escaping a predator or searching for food.

These dual meanings of adaptivity, the conservative component, and the innovative component, can be found throughout the literature on intelligence. The famous Swiss psychologist Jean Piaget coined the terms ''assimilation'' and ''accommodation'' to designate these two aspects of intelligence (e.g., Piaget 1952). In learning theory, this has been called the stability-flexibility trade-off (e.g., Carpenter and Grossberg 1988). We will encounter these concepts in various guises throughout the book.

Before concluding this section, we should remark that the study of intelligence often does not take interaction with the environment explicitly into account, even though it may be implicitly present. This aspect, which we call *embodiment*, emerges as one of the key factors in understanding intelligence, and embodied cognitive science capitalizes on it. The terms ''adaptation,'' ''behavior,'' and ''generation of behavioral diversity'' by their very nature imply the existence of a body interacting with an environment.

1.2 Studying Intelligence: The Synthetic Approach

Now that we know what we want to investigate, we have to specify how we are going to proceed. We can distinguish between analytic and synthetic approaches, as shown in figure 1.7. The analytic approach is universally applied in all empirical sciences. Typically, experiments are performed on an existing system, a human, a desert ant, or a brain region, and the results are analyzed in various

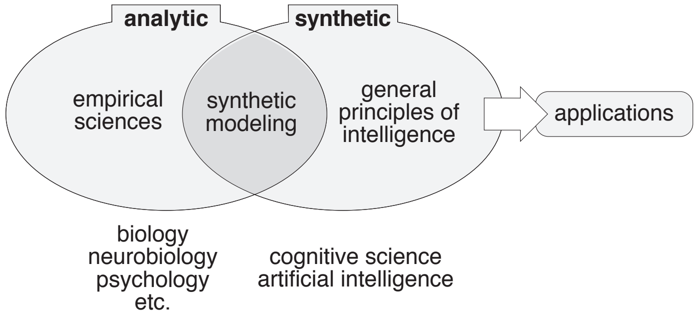


Figure 1.7 Overview of approaches to the study of intelligence. On the left, we have the empirical sciences like biology, neurobiology, and psychology that mostly follow an analytic approach. On the right, we have the synthetic ones, namely cognitive science and AI, which can either model natural agents (this is called synthetic modeling, the shaded area) or alternatively can simply explore issues in the study of intelligence without necessarily being concerned about natural systems. From this latter activity, industrial applications can be developed.

ways. Often the goal is to develop a model to predict the outcome of future experiments. By contrast, the synthetic approach works by creating an artificial system that reproduces certain aspects of a natural system. This is another important function of models. Rather than focusing on producing the correct experimental results, that is, the correct output, we can try to reproduce the internal mechanisms that have led to the particular results. In a memory experiment, we could predict, say, the number of items recalled, based on a statistical model. Alternatively we could try to model the memory processes themselves. An ethologist may want to predict where an ant path will be formed. Again, he can use statistical modeling, but he can also attempt to model the behavioral rules by which the ants interact with the environment and with each other. Such models are typically computer models that, when run, are expected to reproduce the experimental results. The focus of interest shifts from reproducing the results of an experiment, although that is still an important aspect, to understanding why the results come about. This kind of approach is called *synthetic modeling* and is extremely productive. It is at the core of the discipline central to this book, *embodied cognitive science*. Such an approach can be characterized as ''understanding by building.'' In the study of intelligence, this approach has been championed by AI and cognitive science and it is the approach that we have adopted in this book. The analytic and the synthetic approaches are complementary, however, not contradictory. In many sciences, the computational approach, an instance of the synthetic methodology, has become an integrated part, complementing the experimental method.

Synthetic Modeling: AI and Cognitive Science

Traditional AI and cognitive science proceed by developing computer models of mental functions. As a consequence, intelligence in these disciplines is closely tied to computers. Very roughly, the main idea is that intelligence—thinking—can be understood in terms of computer programs: Input is provided, the input is processed, and finally an output is generated. By analogy, the human brain is viewed in some sense as a very powerful computer. It receives inputs from the outside world through sensors (e.g., eyes, ears, skin). These inputs are processed: for example, stimulation received through the eyes is mapped onto an internal representation or model, and you recognize a cup of coffee standing in front of you. Depending on your internal state, your motivation, this percept generates the intention or plan to drink coffee: the processing phase. Finally, the action is executed: the output. In this view, called the *information processing metaphor*, the brain is seen as the "seat of intelligence," as illustrated in figure 1.8. Input-processing-output in computers corresponds to sensing-thinking-acting in intelligent agents such as humans or robots.

Like no other approach, this view of intelligence, together with the synthetic methodology, has revived the study of the mind and has provided major impulses to the field. It was more than a lucky coincidence that these types of computer models seemed almost perfectly suited to the study of the mind. It has greatly inspired many scientists, in particular psychologists and computer scientists. It has generated a lot of exciting research and applications. Moreover, this approach has strongly influenced psychology and has become known as *information processing psychology*. The focus in this perspective is mostly on thinking, reasoning, and abstract problem solving.

When researchers in AI started applying these ideas to building robots, to developing systems that interact with the real world, however, they found that it was simply not possible to build robots that would do a good job in the real world with this view of intelligence. It proved extremely difficult to get robots to do even simple things like moving around, picking up objects, and bringing them to a designated location. The problems were so serious that many

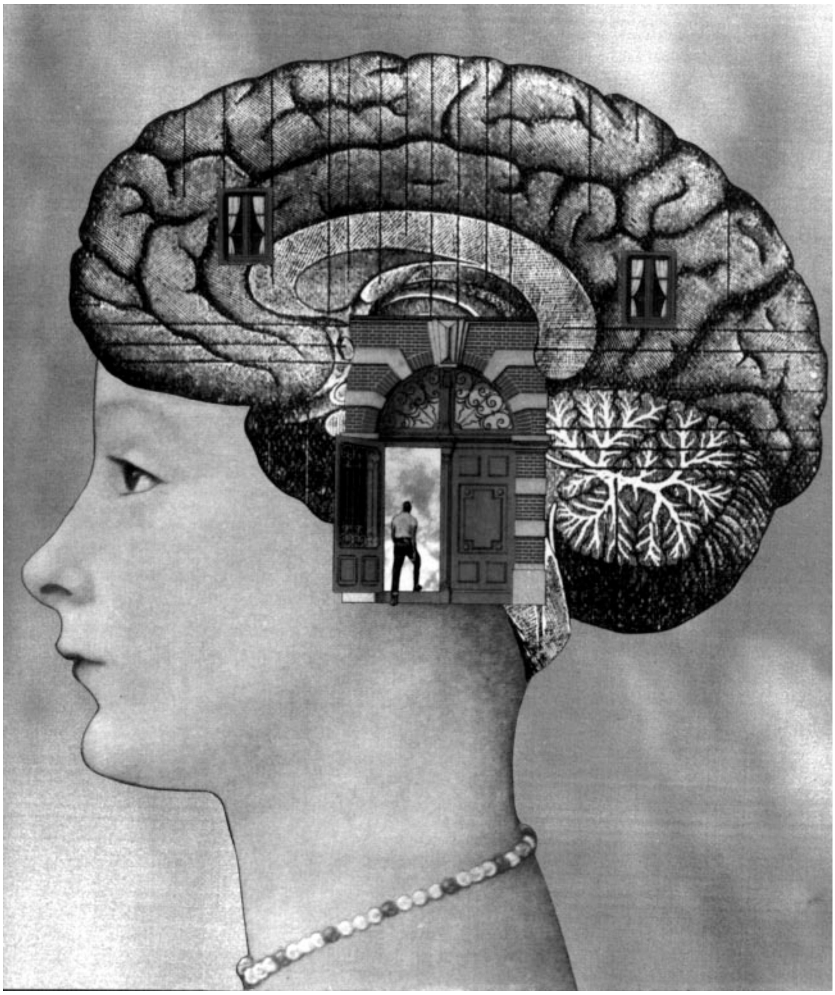


Figure 1.8 The brain as the "seat of intelligence." Sensory stimulation enters the brain, is processed (perception), and is integrated into a model of the environment (modeling). This model is used for planning and task execution, and finally a motor action is performed. This is the "input—processing—output" perspective. (From Uni Magazin 1995, reprinted with permission.)

started looking for alternatives. This resulted in the new field of embodied cognitive science.

Rodney Brooks, the director of the MIT Artificial Intelligence Laboratory and one of the founders of this new field, argued that the traditional approach to AI was fundamentally flawed. He maintained that all of AI's ideas concerning thinking, logic, and problem solving were based on assumptions that come from our own introspection, from how we see ourselves. He suggested that we drop these assumptions, do away with thinking and reasoning, and focus on the interaction with the real world. In a seminal paper in 1986, Brooks proposed the so-called subsumption architecture. He suggested that intelligent behavior could be achieved using a large number of loosely coupled processes that function predominantly in a asynchronous, parallel way. He argued that only minimal internal processing is required and that sensory signals should be mapped relatively directly to motor signals. Such an architecture leads to a tight system-environment coupling. Intelligence, in this view, emerges from the interaction of an organism with its environment, where the organism is equipped with a large number of parallel processes connected only loosely to one another. Such a conception of intelligence contrasts strongly with the information processing view. Note that in this perspective, the agent has a body, sensors, a motor system; in other words, it is embodied. Moreover, it needs to be autonomous. Let us examine this in more detail.

Autonomous Agents

In traditional AI and cognitive science, computer models have been the predominant tools. Synthetic methodology, however, can be extended to include not only simulations, but also physical systems, artificial creatures, behaving in the real world. These systems are called *autonomous agents*. The term ''autonomous'' designates independence from human control. Typically, autonomous agents have the form of mobile robots and can be used as models of biological systems, humans or animals. We now have a novel situation: The autonomous agents actually behave in the real world without the intervention of a human: They have sensors to perceive the environment, and they perform actions that change the environment. These are the key properties of agents. They are behaving systems in their own right. This is why they are also well suited to explore issues in the study of intelligence in general, not only of biological systems. We can perform experiments with our robots as we like, creating artificially intelligent systems. And because the robots physically interact with the real world, they can also be used for applications, to perform tasks that humans cannot or do not want to do themselves. Thus we can pursue three potential goals with the synthetic methodology: We can model biological systems, we can explore principles of intelligence in general, and we can develop applications.

It is highly instructive and productive to work with physical robots. Depending on what we intend to study, it may even be necessary. But often we can achieve the desired results in simulation. We can simulate the behavior and environment of an animal or robot, or we can produce creatures living in virtual worlds that are not simulations of real systems. The latter is the business of the fields of virtual reality and artificial life. The essential point is to have agents—physical or virtual, because agents interact with their environment on their own. This is why they represent the main tool of embodied cognitive science.

Figure 1.9 provides an overview of different types of agents. *Biological agents* exist in nature—we don't have to build them. Of the two categories of robotic agents, *research agents* and *industrial agents*, we will primarily focus on research agents, because of our focus on cognitive science. But we do believe, and discuss later, that industrial agents have fascinating applications: This is the business of engineering. Among computational agents, we have *simulated agents*, those that simulate an animal or a robot, and

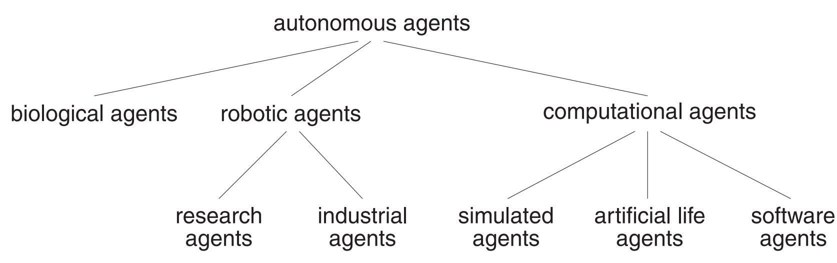


Figure 1.9 Classification of agents. The relevant category of agents for the study of intelligence are the autonomous agents. They can be subdivided into biological agents, robotic agents, and computational agents. Biological agents are naturally occurring. Robotic agents are further divided into research agents and industrial agents. Research agents are used to model natural agents, and to explore general principles of intelligence. Industrial agents are used for practical applications. Computational agents are subdivided into simulated agents (i.e., agents simulating a biological or robotic agent), artificial life agents, and software agents.

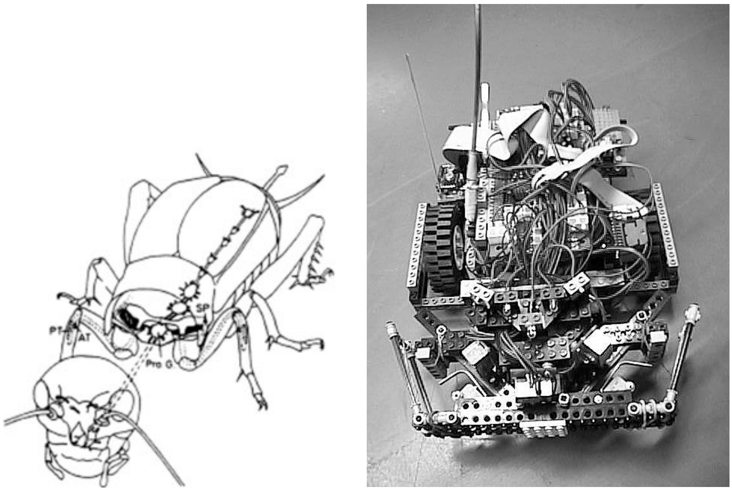
artificial life agents that do not necessarily simulate something but are creatures of their own type, digital creatures. There has been considerable hype about the last category of computational agents, called *software agents*. In essence, software agents are computer programs that perform a certain task and interact with real-world software environments and humans by issuing commands and interpreting the environment's feedback. Typical tasks are filtering electronic mail, sending routine messages such as reminders for meetings or announcements of seminars, collecting information on the Internet, scheduling meetings, performing system maintenance tasks like continuous intrusion detection, and assisting in purchasing a car or finding an apartment. Especially with the advent of the Internet, software agents have become enormously popular. They come in many varieties (e.g., Riecken 1994), and it is sometimes hard to distinguish them from other kinds of computer programs. Software agents have a great potential for application, especially in a networked society. However, a detailed treatment would be beyond the scope of this book.

Let us now look at the different ways in which autonomous agents are used.

MODELING

One application of autonomous agents in cognitive science is to model the behavior of biological agents. An example of the modeling approach is shown in figure 1.10, where an autonomous robot is used to model the phonotactic behavior of a cricket (Webb, 1993, 1994). We designate as phonotaxis those processes by which female crickets move towards a particular sound, the calling song of a potential mate. This robot model can be used to generate (biological) hypotheses about cricket behavior; these hypotheses can be tested in experiments with real (biological) crickets.

As a further example, assume that you want to replicate another idea from nature, say, an artificial retina. Once you have developed your conception of how the retina functions, you may very quickly find that your hypothesis about the functioning of the (biological) retina is flawed if you actually build and test it on a robot. Or let's take an example from neurobiology. We know that the control mechanisms of animals are based on neural structures. Biological neural systems have inspired artificial neural network models. Neural architectures, as it turns out, can be understood only in the context of the physical system in which they are embedded. Intro-



a b

Figure 1.10 Illustration of the synthetic methodology. The robot cricket (b) can be used to investigate the behavior of the real cricket (a). (Reprinted with permission.)

ducing mobile robots—that is, real, behaving systems—brings a novel perspective to modern neuroscience.

Many researchers have capitalized on this fascinating interaction between biology and autonomous agents research. Lambrinos et al. (1997) have developed a robot that navigates according to the same principles as the desert ant *Cataglyphis*. Ferrell (1994) and others have developed walking robots, applying principles known from insect walking as described by Cruse (1991). And some robots move around using control circuitry just like that of the housefly (Franceschini et al. 1992). Robot modeling has also been successfully employed in the area of psychology. An example is the humanoid robot Cog (Brooks and Stein 1993), used for development studies of how human infants learn to reach for a ball or play with toys, for example. There are also many attractive simulation studies, such as walking insects (Beer 1990), fish learning to swim in simulated water, for example, a shark preying on other fish (Terzopoulos et al. 1994), and a humanoid robot used for developmental and social interaction studies (Kuniyoshi and Nagakubo 1997).

EXPLORING GENERAL PRINCIPLES OF INTELLIGENCE

A second application of autonomous agents in cognitive science is to explore principles of intelligence. This approach draws inspiration from nature, but offers us more freedom than the modeling approach. Experiments can be conducted using any type of sensor, even sensors that do not exist in nature (like laser scanners, or radio emitters-receivers). We can use wheels, magnets, and batteries in our systems; we can exchange pieces of code, place sensors in different positions, add another lens here and there; in short, we can perform experiments. We can build systems that we have invented using artificial devices. Developing systems different from the ones we observe in nature is an extremely productive way of doing research. By doing things differently from nature, we may learn a great deal about how things might, in fact, function in nature.

One of the main motivations to employ autonomous agents is the idea of *emergence*. Autonomous agents, by definition, behave in the real world without human intervention. One of the fascinating features of autonomous agents is that they exhibit so-called *emergent* behaviors, that is, behaviors not programmed into the agents by the designer. Robots programmed only to follow a light start helping each other, or they are cleaning up though programmed only to avoid obstacles. Or a group of simulated birds are flocking, but were programmed only with local rules, that is, rules that make reference only to their immediate neighbors.

In exploring principles of intelligence, the search for emergence is an important motivation. We show many examples of this approach throughout the book. To mention but a few: The famous Braitenberg vehicles are used to explore how very simple mechanisms can lead to truly amazing emergent behaviors; the robot Polly, which used to give tours at the AI Laboratory at MIT, was used to study principles of cheap visual navigation; ''boids,'' a kind of artificial bird, were used to investigate how flocking behavior could emerge from local rules; and fantastic creatures, created by Karl Sims, living in a virtual world of simulated physics were used to explore the evolution of morphology and neural controllers.

The line between exploring principles of intelligence and modeling can be fuzzy. Often agents used for modeling purposes are modified so that they deviate from the model. On the other hand, a robot used to explore general principles might be applied for modeling purposes because it develops interesting related behavior. SMC agents (which we discuss in detail in chapter 12) were originally used to study neural architectures for sensory-motor coordination. Then developmental psychologists became interested in using them to model category learning in human babies. The insights thus gained can then be applied to develop systems that perform useful tasks in the real world.

APPLICATIONS

To date, the enormous potential for applications of autonomous agents technology has hardly been explored. Robots, for example, can be used for marking the mines on a minefield with color, for monitoring sewage systems for leakages, for cleaning up hazardous waste sites, for distributing mail, and for surveilling an industrial plant. Autonomous wheelchairs are another possible application. Computational agents hold tremendous promise for applications, especially in the areas of evolutionary robotics and artificial life. Ideas from natural evolution are employed, for example, for optimization problems and they have also been successfully applied to industrial problems. Simulated agents are used widely in the field of computer graphics and the entertainment industry.

THE DESIGN PERSPECTIVE

The synthetic methodology is closely coupled with the notion of design. Autonomous agents, whether robotic or computational, in order to be built, must be designed. Although we normally think about design as an activity for engineers, the design perspective has proven extremely fruitful in cognitive science for studying natural intelligence. Evolution can be viewed, in a sense, as a designer (e.g., Dawkins 1988), perhaps a blind one, but nevertheless an extremely effective one: Natural systems have truly impressive capabilities. What we are asking is how we would design a system that behaves in a particular way that we find interesting? We devote a great deal of effort in this book to exploring design. In fact, one of our main goals is to elaborate a set of design principles for autonomous agents that, in a sense, constitute our understanding of the nature of intelligence.

Issues to Think About

Issue 1.1: Is IQ Irrelevant for Intelligence?

We have argued, as most people these days do, that IQ is a poor measure of intelligence because it tries to reduce a complex phenomenon to a single number. On the other hand, there is evidence that IQ is a good predictor of some kinds of success. When discussing the nature-nurture debate, we concluded that intelligence originates from a highly complex interaction between genetic and environmental factors, the interaction between nature and nurture. This consideration suggests a generally valid strategy for cognitive science. Experience teaches us that studying an individual's development, rather than the individual in its current appearance, often leads to a better understanding of its behavior, because simply inspecting the organism itself offers us only little insight into the constraints, the history, the personal experiences, the interactions of the individual with the environment. As many developmental studies have shown, concepts, the ability to make distinctions, are a direct consequence of sensory-motor behavior. Thus high intellectual ability resulting in a high IQ score may well be due to a complex mix of sensory-motor abilities that in turn depend on the particular social environment. In other words, before an individual is capable of solving the problems on an IQ test, he has to master many other things including many that do not relate directly to abstract thinking but to other notions of intelligence mentioned above. The reason IQ is a good predictor for certain types of success may eventually be explained on the basis of a developmental perspective, but it remains an open research question.

Issue 1.2: The Diversity-Compliance Trade-off: The Common Denominator?

We have argued that the common denominator underlying the various notions of intelligence discussed in this chapter is the diversity-compliance trade-off, in particular, that the two core aspects of intelligent behavior are generation of diversity and compliance with rules. In other words, there is always a trade-off between generating new solutions, being flexible and innovative, and complying with the existing rules, exploiting what is already known. This characterization, we have argued, holds for a vast range of agents, from companies that face new challenges in the information age down to ants surviving in the desert. If this view is correct, then it should be applicable to the reader. We would like you, before you continue reading, to reflect for a few moments on whether you feel your own behavior can be described in this way. Think about how you usually solve a problem: Can your approach be described in terms of the diversity-compliance trade-off? Or do you think you behave according to different principles?