1 Intelligence, Thinking, and Artificial Intelligence

"I think, therefore I am"! Thus the famous and popular quote by the seventeenth-century French philosopher René Descartes, from his *Discourse on Method*, published in 1637. What is surprising about this quote is that it implies that the reason I exist is not the body, but the mind! In Descartes's view there are two separate systems: the body and the mind. This division raises the problem of how these two systems relate to one another, an issue that is referred to as the mind-body problem (see focus box 1.1). One of the main challenges posed by the mind-body problem is the question of how a thought—something happening in the immaterial mind—can potentially influence the body. For example, I can decide in my mind to pick up a cup to drink a sip of coffee, and subsequently my arm and hand begin to move to perform the action. This is the way we like to think about ourselves: the mind controls our actions, which implies that we are in control of our behavior and therefore our lives—which is, so to speak, the "Cartesian heritage" of Western culture. The importance of the individual—individualism—and being in control are two extremely cherished values in Western societies: We, as individuals, decide about something—a goal that we want to achieve, such as becoming a doctor or catching a Frisbee—and then we make plans and go about doing it. Or when at a party, we decide that we would like to meet someone, so we start talking to that person. It all seems very natural, the way things should be. But is it really? In other words, is this an accurate way of describing how we as intelligent beings function? As you might expect, after what we have said so far, our answer is "no." While there may be some truth to this way of viewing ourselves, it is largely based on wishful thinking; on how we would like to see ourselves rather than on how things actually are (see also focus box 1.1 for more details). It turns out that instead of our ideas—our minds—controlling our actions,

Focus Box 1.1
The Mind-Body Problem

The so-called dualist position, as laid out by René Descartes in the seventeenth century, states that there are two separate systems within a human being: a mental thing, the *res cogitans*, and a physical thing, the *res extensa.* Descartes was concerned about how these two worlds—the mental and physical—talk to one another. His ideas have raised many deep issues, which together are known as the mind-body problem. This problem is probably the most famous in the philosophy of mind, and is concerned with the relation between the mental and the physical, or between mind and matter: more specifically, how can the physical processes of our bodies and brains give rise to abstract mental phenomena such as consciousness? David Chalmers, one of the leading philosophers of consciousness, is very clear about how important this question is: "Consciousness is the biggest mystery. It may be the largest outstanding obstacle in our quest for a scientific understanding of the universe." (Chalmers, 1997, p. ix). There is a vast literature on this issue, but rather than providing a systematic review—the interested reader is referred to David Chalmers's and Thomas Metzinger's popular online bibliographies on the philosophy of mind—we would like to point out just one particularly enticing issue; that of conscious will.

Most people would probably agree that mental phenomena, such as thinking and cognition, originate from brain processes. Assume for a moment that your hand is on the desk in front of you and you are about to move your finger. The neuroscientist Benjamin Libet and his colleagues, in an often-cited landmark experiment (Libet et al., 1983), asked people to move their finger spontaneously, whenever they liked. In addition, the subjects had to look at a clock with a revolving point of light, and report where the dot was on the clock when they experienced "conscious awareness of 'wanting' to perform a given self-initiated movement" (quoted in Wegner, 2002, p. 52). Moreover, he recorded brain activity, the so-called readiness potential, from electroencephalography (EEG) sensors attached to the scalp, and he measured actual finger movement using electromyography (EMG), a method for detecting muscle movement. The results were stunning: the onset of brain activity starts more than half a second before the actual finger movement and over 300 msec before the subjects become aware that they want to move a finger! In other words, the conscious will of wanting to move the finger occurs a significant interval—after the onset of the relevant brain activity. So the experience of conscious will kicks in after the brain has already started preparing for the action. In other words, the mental will to move the finger could not have been the initiating agent of the movement. This is quite contrary to what we would expect, and runs counter to the subjective experience of the individual: we "feel" that our decision to move our finger is what kicks off the proper brain processes necessary to move the finger. The surprising conclusion from this experiment—whether we like it or not—is that the initiation of the voluntary act of moving the finger seems to be caused by unconscious neural activity, not the other way around! Needless to say, this is a serious blow to the notion of free will. Or is it? Libet notes that even if the movement is indeed initiated by unconscious forces, there is still enough time to veto an act—to decide not to move the finger—once one is aware of one's intentions. Perhaps this keeps the door open to the notion of free will. As you can imagine, these findings have created a flurry of discussions in the scientific community. The issue of free will, however, is just one of the many scientific debates that are currently raging about how the mental and physical aspects of a person influence each other. And it is not just a question of how they interact: in the extreme case, many philosophers hold, we may never be able to know how the mental and the physical communicate. To use the words of the legendary German brain physiologist DuBois Reymond: ignoramus ignorabimus (we do not know, we will not know). Or will we? The deep issues raised by Descartes still await final explanation, but the progress in modern neuroscience—and artificial intelligence—provides a scientific way, rather than just a philosophical one, for dealing with them.

to a surprising extent our body determines our thoughts. And this is what we will explore in this book: how the body shapes the way we think. We are convinced that the exploration of this relationship between body and thinking will clarify the conundrum of intelligence in interesting ways; we hope that it will indeed lead to a new view of intelligence, as suggested in the title.

In this chapter we will proceed as follows. First, because it has a prominent position in the title of the book, we will briefly examine the term *thinking* and how it relates to cognition and intelligence. We will talk about why the topic of intelligence has captured the attention of philosophers, scientists, and people at large throughout the history of humankind. Then we will explain how, in this book, we as researchers attempt to tackle these issues, namely by employing the methodology of artificial intelligence. We will end the chapter by introducing the notion of embodiment from which the major contributions to "a new view of intelligence" have originated and which, we believe, holds the most promise for our future understanding of intelligence.

1.1 Thinking, Cognition, and Intelligence

So far, we have used the term *thinking* without much reflection, with the assumption that everyone has a fairly clear notion of what it is all about. But let us look a bit more closely. Intuitively—and this is the way it is defined in psychological dictionaries—thinking is associated with conscious or deliberate thought, with something high-level or abstract. The trouble with this conception is that it relies on the assumption that a process either is or is not conscious. But perhaps matters are not as clear as they might seem at first sight. Here is one possible reason why.

Do newborns think? We cannot be sure, but perhaps they don't. Or maybe it would be better to say that they think less than adults. What about after a few days? Or after a few weeks? Certainly after a few months or years, and clearly as adults, we do think. But if this is true, it raises the question at what age children actually do start thinking. Again, this is difficult to answer, but it is clear that their skills gradually improve as they grow older; perhaps then their ability to think also improves gradually over time. This way of viewing thinking—and more generally, intelligence or cognition—is referred to as a developmental approach, i.e., it posits that the ability to think develops over time. From this perspective, the question shifts from whether an agent—an animal, robot, or human—is thinking or not to how much thinking is actually going on.

In other words, we can escape the limiting view that thinking is a binary property: i.e., an agent either thinks or it does not. (Throughout this book we use the term *agent* whenever we do not want to make a distinction between humans, animals, or robots, i.e., when what we say applies to all three.) Much of what we have to say about intelligence in this book is general: it applies not only to humans, but, to a greater or lesser degree, to animals and robots as well. For example, agents have interesting properties related to intelligence that other nonagents, like cups or rocks, do not have: we will discuss this in more detail in chapter 4.

It seems obvious that the ability to think increases over time as the organism grows and matures. But even as adults, "thinking" remains a vague term that for most people implies conscious thought. However, consciousness is an equally vague concept, and again we can imagine that there is a continuum rather than an all-or-none property. We would suspect that, for example, bacteria, insects, birds, rats, dogs, chimpanzees, and humans are conscious to a greater or lesser extent, rather than being either conscious or not. Moreover, in clinical psychology there is the concept of unconscious thoughts, which are thoughts that, even though we are not consciously aware of them, influence our behavior, often in undesirable ways. Therefore, rather than trying to come up with a definition for thinking or consciousness, it is probably best to agree that we are dealing with a continuum, with a gradual phenomenon. We side with Douglas Hofstadter, who, in his clever and entertaining book *Metamagical Themas*, laments the fact that people seem to have a compulsion for "black-and-white cutoffs when it comes to mysterious phenomena such as life and consciousness." And he adds that "the onward march of science seems to force us ever more clearly into accepting intermediate levels of such properties." (Hofstadter, 1985).

Consciousness is a peculiar, fascinating, but highly elusive sort of thing. Because it is tied to subjective experience, it is hard to investigate scientifically. However, recent advances in brain imaging and neuroscience in general have yielded stunning but also puzzling results (e.g., Crick and Koch, 2003). A particularly enticing issue concerns the role of consciousness in free will, which we briefly describe in focus box 1.1. In this book we will not go into the subject of consciousness. Some people appear to believe that unless we have explained consciousness, we have understood nothing about intelligence. We hope that we can convince the reader that this is not the case and that we can acquire a deep understanding of intelligence by pursuing the idea of embodiment. But we also feel that because we discuss the issue of how cognition can emerge from a physically embodied system—and most people seem to agree that consciousness is related to cognition—we will ultimately contribute to the understanding of consciousness.

Cognition, closely related to intelligence, is another vague and general term that is often used to designate those kinds of processes of an agent that are not directly related to sensor or motor mechanisms. Examples of cognitive processes are abstract problem solving and reasoning, memory, attention, and language. Again, as we will see, if we inspect the underlying mechanisms of these phenomena we find that cognition cannot really be distinguished from other (noncognitive) kinds of sensory-motor processes. As we will argue later, even simple activities such as walking or grasping a cup have cognitive qualities, so to speak. And perception, which is obviously related directly to sensor processes, is an important subfield of cognitive psychology. Lachman et al. (1979), in their well-known book *Cognitive Psychology and Information Processing*, described the field using a computer metaphor: "[cognitive psychology is about] how people take in information, how they recode and remember it, how they make decisions, how they transform their internal knowledge states, and how they translate these states into outputs" (p. 99). Cognition is sometimes employed as a more general term than thinking because it does not necessarily imply consciousness. However, it is important to keep in mind that despite the more abstract connotations of thinking as compared to cognition, thinking is not a disembodied process: as we will see, it seems to be directly tied to sensory-motor and other bodily (i.e., physiological) processes, as is cognition.

The last term to be characterized is that of intelligence, which closely resembles thinking and cognition, but is typically used in an even more general way. There is no good definition for intelligence, but we do not feel this is a bad thing. Throughout the book we will always take care to clarify what we are talking about, but at the same time we will try not to get bogged down in debates about definitions. We will see that some of the concepts that are defined in the literature—e.g., between learning, memory, and perception—are not, from the perspective of the underlying mechanisms, clearly separable. For example, learning and memory are always involved in perception; what we perceive—for example, the sight of a friend in a bar—is determined by our memory, and of course, our memory is affected by what we perceive. As we will also see later on, these terms are used by an external observer to characterize certain behaviors, and are therefore largely arbitrary: the definitions depend more on the observer than on the observed phenomena themselves.

But back to intelligence. The entry for "intelligence" in the Penguin Dictionary of Psychology starts by stating that "Few concepts in psychology have received more devoted attention and few have resisted clarification so thoroughly" (Reber, 1995). If Reber's comment is about the definition of the term, we fully agree. However, we disagree with the idea that intelligence itself has resisted clarification. This book, we claim, clarifies many aspects of it. Before we turn toward elucidating the mystery of intelligence, though, we should introduce a bit of additional terminology.

We have been using the term *agent* to indicate that an argument holds whether we are talking about a human, animal, or robot. We do not use it in its everyday sense, referring to an insurance agent who offers us particular services, a secret agent who unearths information for a government, or a chemical agent that reacts with other substances. In this book, an agent is "anything that can be viewed as perceiving its environment through sensors and acting upon that environment through effectors," as defined by Russell and Norvig in their classical textbook on artificial intelligence (1995, p. 33). In other words, an agent differs from other kinds of objects such as a rock or a cup, which are only subject to physical forces: they cannot react on their own. Moreover, we are particularly interested in embodied agents, which are agents that have a physical body with which they can affect, and be affected by their environment. Software agents, which is a term used to designate certain types of computer programs, such as internet agents that search for information, are not embodied and will not be further considered here.

Finally, we use the term *robot* in a relatively broad sense. The original sense of the word—it derives from the Czech *robota*, meaning something like "work" or "forced labor"—implies that robots were initially meant to do work for humans. So, factory robots are the "species" that most closely conform to this idea. However, for the purpose of this book they are not of central interest; they not be further discussed because their behaviors are essentially preprogrammed and they do not tell us much about the nature of intelligence. We expect these robots to do precisely what we want them to do—they should not all of a sudden come up with some interesting, unexpected ideas or behaviors on their own. The term *robot* as used here refers to machines that have at least some agent characteristics in the sense discussed above, irrespective of whether they do useful work for humans or not. This includes humanoids, pet robots, entertainment robots, service robots, rescue robots, etc. In chapter 11 we will review and analyze different types of robots. Whether or not a particular machine or device deserves to be called a robot is largely arbitrary and cannot be precisely defined.

With all of this in mind, let us now familiarize ourselves intuitively with intelligence and explore its fascination and its mysteries.

1.2 The Mystery of Intelligence

Intelligence is obviously an important issue. Literally hundreds of books have been written about it, and here we add yet another book on the subject. Well, yes and no. Yes, this is another book about intelligence, but we feel that it is very different from its predecessors. The fact that there is an enormous literature on the topic is not really surprising. Throughout human history, philosophers, psychologists, artists, teachers, and more recently neuroscientists and artificial intelligence researchers have been wondering about it, have been fascinated by it, and have devoted much of their lives to its investigation. And many of them have written books about it. Still, there are good reasons why it makes sense to write—yet another!—book about this topic because, we believe, it presents some novel points that previously have not even been considered to be part of the field of intelligence. These novel points all relate, one way or another, to the notion of embodiment, the seemingly simple idea that intelligence requires a body. As we will see in this section, and as we hope to demonstrate throughout the book, this new perspective of embodiment has led to often surprising insights and new research issues for studying intelligence.

Intelligence is a highly sensitive topic because we tend to believe that intelligence is what distinguishes us from animals: we are so much more intelligent than them, we tell ourselves—and in many ways this is certainly the case. In our societies, Western or Eastern, an enormously high value is attached to intelligence. Our schools and universities are almost universally considered our highest cultural resource: indeed many of them look like temples built to honor the gods. Universities are monuments with strongly symbolic character. The goals of these institutions are, in one way or another, to preserve and further increase the level of intelligence in our societies. "You are very intelligent" is one of the highest compliments one can give or receive. We are constantly reminded that intelligence is good, positive, and desirable. Parents always think that their children are highly intelligent. You are allowed to say virtually anything about someone's children—you can call them lazy, cheeky, aggressive, nervous, easily distracted, shy—but never, ever, say they are not intelligent! We continue to place this premium on rational intelligence despite the recent surge of interest in emotional intelligence, which argues that rationality is limited and that we should also take emotions into account when measuring intelligence. In other words, in this view, intuition and the ability to emotionally judge a situation is considered just as important as the "cold" kind of intelligence required to pass high school exams or to achieve high scores on intelligence tests. This perspective is documented by the famous books of the well-known neuroscientist Antonio Damasio (e.g., *Descartes' Error*), and by the tests developed by the American psychologist Daniel Goleman to measure emotional intelligence (Goleman, 1997). Regardless of these developments, rational, logical intelligence is still considered to be one of the most enviable characteristics of human beings.

But there is another reason why intelligence is a sensitive issue. For many decades the question of whether intelligence is inherited or can be acquired during a person's lifetime has been hotly debated: this is the famous (or infamous) nature-nurture debate (see, for example, Ridley, 2003; Ceci and Williams, 2000, for a collection of articles on the topic). We assume that part of the reason this debate is so emotionally charged is because it is about intelligence. Other personality traits besides intelligence cause much less controversy. For example, whether a person has an honest character or high moral standards, and how these traits are acquired, is not discussed as much, although honesty and morality are still considered desirable qualities. Having a high IQ (intelligence quotient), or more generally scoring high on the many standard intelligence tests now on the market (in spite of all the current interest in emotional intelligence), is still considered one of the most desirable personality traits to have. In order to be politically correct we hesitate to attribute value to IQ scores publicly; however, privately, we suspect, most people do value them. When the two Harvard psychologists Richard Herrnstein and Charles Murray published their controversial analysis of the IQ in their famous 1994 book *The Bell Curve: Intelligence and Class Structure in American Life*, they spurred another extremely emotional debate in America and throughout the world. Among other findings, they reported, with a number of qualifications, that Asians have the highest scores on IQ tests, Caucasians are second, and black people have the lowest! It seems easy to conclude from this result that class structure is a result of intelligence, regardless of whether intelligence is inborn or acquired. The interesting scientific question in this seemingly eternal debate is not whether intelligence is inherited or acquired during the lifetime of an individual, but how evolution and development interact such that intelligence arises in an agent. This topic will be broached in chapters 5 and 6, where we discuss the relationship between development, evolution, and intelligence.

Intelligence is highly mysterious, and we all wonder what it is: How was it possible that something so sophisticated could have been produced by evolution? How does it develop as a baby grows to become an adult? How can we walk, talk, or solve a problem? And how can we, without effort, recognize a face in a crowd, or play a piece of music? Just to take one example of a process essential for intelligence, memory is a highly enigmatic phenomenon, and nobody really understands how it works. Memory performance varies greatly depending on the person's mood or physical conditions; sometimes people are really forgetful, and sometimes we are astonished by their accuracy of recall. How do we retrieve something from memory? In a computer, the stored items have addresses that can be used for this purpose. But where are the addresses in the brain? There are events that have long passed of which we have the most vivid memories, whereas others are murky and dark, at least temporarily. Then, suddenly—we have all had the experience—we remember something long forgotten. The tip-of-the-tongue phenomenon, a mostly frustrating experience, is also something that everyone has experienced: we know that we know something, but we just cannot seem to spit it out. For example, just before, I was thinking about the name of the author of *Descartes' Error*, but I just could not seem to mentally call it up. But five minutes later, it was there without effort, even though I hadn't been thinking about it any more in the meantime. How do we know that we know something if we cannot remember it? Why is it so easy to recognize the face of a casual acquaintance when he appears, but so hard to describe in his absence? And how come we firmly believe certain facts to be true which are demonstrably false? That such phenomena exist is easy to verify, but hard to explain.

But memory—and, by extension, intelligence—is not just mysterious, but incredibly valuable and necessary. Having no memory implies the inability to learn, and not being able to learn is incredibly debilitating. Hollywood has a long-standing love affair with amnesia or memory loss, because it challenges those affected in interesting ways. In the Hollywood thriller *Memento*, the protagonist loses the ability to make new memories through a blow to the head. The excitement in the movie comes from watching how he tries, through various adventures, to reconstruct what happened to him. In the comedy *50 First Dates*, the main character one day encounters Lucy, who also has lost the ability to make new memories because of an accident. The comic side of the film is that whenever the lead character talks to her, she has forgotten that she ever met him! Memory is of fundamental importance, not only to intelligence but also to our own well-being, yet many fascinating problems relating to memory remain to be answered. Like intelligence, memory is a very important but still poorly understood phenomenon. For this reason we have devoted an entire chapter (chapter 10) to memory. Moreover, we believe that the perspective of embodiment developed in this book may clarify at least some of the issues surrounding memory and, more generally, intelligence.

1.3 Defining Intelligence

So intelligence is important, sensitive, and mysterious, but what is it really? We start from the assumption that everyone has a good intuition of what intelligence is all about. It has to do with consciousness, thinking, and memory (as already mentioned), along with problem solving, intuition, creativity, language, and learning, but also perception and sensory-motor skills, the ability to predict the environment (including the actions of others), the capacity to deal with a complex world (which may result from a combination of other abilities), and performance in school and on IQ tests and the like. In general, a good definition should capture at least some of the intuitions. But given the length of the list and the vagueness of the concepts, it seems unlikely that we will ever agree on a single one.

Here are some sample definitions from an inquiry by the *Journal of Educational Psychology* in 1921, wherein leading experts of the time were asked for their suggestions. L. Terman: “the ability to carry on abstract thinking”; W. F. Dearborn: “the capacity to learn or profit by experience”; S. S. Colvin: “having learned or ability to learn to adjust oneself to the environment” (this definition is so general that it can hardly be wrong, depending on what we mean by “adjust oneself to the environment”); R. Pintner: “the ability to adapt oneself adequately to relatively new situations in life” (similar to the previous one; the question here is what an expression like “relatively new” means in a definition); V. A. C. Henmon: “the capacity for knowledge, and knowledge possessed,” and so on and so forth. We could go on for quite a while, but it is not clear what we would gain by adding more definitions to this list.

One of the reasons for the difficulty in coming up with a good definition is the breadth of the concept, as illustrated by the many intuitions it encompasses. Another is that our definitions will depend on our professional and personal background, subjective expectations, and individual interests and preferences. Gregory (1987), in *The Oxford Companion to the Mind* (p. 378), points out that biologically minded researchers tend to stress concepts such as adaptation and capacity for adjustment to the environment (e.g., Colvin and Pintner), whereas the more philosophically minded intellectual is likely to emphasize the element of abstraction (as in Terman's definition). Such a concept will always have many definitions; there is little hope that there will ever be general agreement on any particular one.

Also, trying to come up with a definition suggests that a property—in this case, intelligence—is either there or not, which is obviously not the case: Are ants intelligent? Perhaps to some degree, but an entire ant colony might be. This idea that not just a single agent, but also a whole group of agents might together be considered intelligent is known as collective intelligence, and we will look into this in some depth in chapter 7. Biologists studying ants are obviously fascinated by the richness of the behaviors they observe, but whether they, or we, would term such behavior "intelligent" is another matter. If we do, though, is the intelligence of an ant colony comparable to the intelligence of a human, or to that of a single ant? One point in our favor is that ant colonies cannot speak, while humans can. So, if we consider language to be an important part of intelligence, we might be tempted to conclude that all humans are more intelligent than ant colonies. Maybe ants, or their colonies as a whole, are not really intelligent, but what about rats or dogs? They are certainly more intelligent than ants, because they can do things that ants cannot, such as learning to navigate in a maze or catching a Frisbee while running. But humans are clearly more intelligent than rats and dogs. Perhaps dogs and cats are more intelligent than us in certain respects: again, dogs and rats cannot speak, write, or build cars, but when it comes to finding survivors at disaster sites or drugs in luggage at airports, dogs are far superior to humans, which is why they are employed for these tasks. It also seems obvious that some humans are more intelligent than others, but when we really think about it, what do we mean by this? Is it because they do some things better than others, for example they perform better at an intelligence test? Or is it because they are more successful in their careers than others? Or is it because they can do math? But then what about those who can sing or survive in the wild? So we see that the issue is very involved and multifaceted, and trying to come up with a clear-cut definition seems doomed to failure from the very start.

So, rather than trying to come up with a definition of intelligence, our suggestion for how to make progress is to look for a topic of interest (such as how dogs can run or catch a Frisbee; how rats learn so quickly to orient in a maze; how ants find their way back to the nest as they return from a trip searching for food; or how humans walk and recognize a face in a crowd) and then try to understand how this particular behavior comes about. Whether one would want to call any of these behaviors intelligent is largely a matter of taste and not really important.

In spite of all the difficulties of coming up with a concise definition, and regardless of the enormous complexities involved in the concept of intelligence, it seems that whatever we intuitively view as intelligent is always vested with two particular characteristics: compliance and diversity. In short, intelligent agents always comply with the physical and social rules of their environment, and exploit those rules to produce diverse behavior. These ideas will be discussed in detail in chapters 3 and 4. Here, just to provide some intuition, we give a brief example to illustrate the idea of diversity-compliance. All animals, humans, and robots have to comply with the fact that there is gravity and friction, and that locomotion requires energy: there is simply no way out of it. But adapting to these constraints and exploiting them in particular ways opens up the possibility of walking, running, drinking from a cup, putting dishes on a table, playing soccer, or riding a bicycle. Diversity means that the agent can perform many different types of behavior so that he—or she or it—can react appropriately to a given situation. An agent that only walks, or only plays chess, or only runs is intuitively considered less intelligent than one that can also build toy cars out of a Lego kit, pour beer into a glass, and give a lecture in front of a critical audience. Learning, which is mentioned in many definitions of intelligence, is a powerful means for increasing behavioral diversity over time. This general characterization of intelligence will be discussed in more detail in chapter 3.

Intelligence can be studied in many different ways, e.g., by performing experiments with humans as in psychology; by studying brain processes as in neuroscience; or by thinking about it in different ways, as in philosophy. In this book we will employ the method of artificial intelligence, which we consider especially productive for this purpose. So, let us briefly get acquainted with it.

1.4 Artificial Intelligence

By artificial intelligence we mean the interdisciplinary research field that has, in essence, three goals: (1) understanding biological systems (i.e., the mechanisms that bring about intelligent behavior in humans or animals); (2) the abstraction of general principles of intelligent behavior; and (3) the application of these principles to the design of useful artifacts. It is important to note that "mechanism" implies not only neural mechanisms or brain processes, but also the body of the agent and its interactions with the real world: the fact that muscles are elastic, and that the weight on one leg increases if the other one is lifted are just as much part of the mechanism of walking as are the reflexes and brain centers involved in this behavior.

In the next chapter we will give a more detailed history of the field, but here we present a very short introduction. Artificial intelligence dates back to 1956 when John McCarthy of MIT invited many leading researchers of the time to a workshop where he introduced the term *artificial intelligence*. Among the participants were Marvin Minsky, Herbert Simon, and Allan Newell, the founding fathers, so to speak, of artificial intelligence. Very roughly, they were convinced at the time that, by using the notion of computation or abstract symbol manipulation, it would soon become possible to reproduce interesting abilities normally ascribed to humans, such as playing chess, solving abstract problems, and proving mathematical theorems. What originated from this meeting, and what came to be the guiding principles until the mid-1980s, was what is now known as the classical, symbol-processing paradigm, also known as the cognitivistic paradigm. We might want to characterize this approach with the slogan "cognition as computation": what matters for intelligence in this approach is the abstract algorithm or the program, whereas the underlying hardware on which this program runs is irrelevant. An implication of this way of thinking is that not only can intelligence arise in biological systems and run on wet, biological brains, but it can also arise in artificial systems and run on computers.

The cognitivistic paradigm is still very popular among scientists. Some choose to view computer programs as *models* of actual thinking, a position called "weak AI," while others claimed and still claim that these programs *are* actually thinking—this is known as the "strong AI" stance. The weak AI position is unproblematic and generally accepted: the nature of the simulation model is clearly different from the thing it simulates. Just as in a simulation of rain the computer does not get wet; the model of thinking is different from the thinking process itself. It is the strong AI stance with which people often take issue. This is not surprising. It is unsettling for many people to believe that a computer is actually thinking, rather than just simulating the process. For more details on the history of AI and on the different positions, see, for example, McCorduck's thoughtful book *Machines Who Think* (1979), with many entertaining anecdotes; or Pfeifer and Scheier (1999); or consult focus box 2.1, which outlines the history of AI. Unlike the cognitivistic view of intelligence, which is algorithm-based, the embodied approach envisions the intelligent artifact as more than just a computer program: it has a body, and it behaves and performs tasks in the real world. It is not only a model of biological intelligence, but a form of intelligence in its own right.

As we will explain in chapter 2, the classical paradigm has had its definite successes, but it has failed to make clear the nature of intelligence, which is the main purpose of this book. Our intention here is not to give a comprehensive overview of the field—for that purpose, the interested reader is referred to the classic by Russell and Norvig (1995)—but rather to investigate recent advances that not only have fundamentally changed the field, but have led to a host of surprising insights. The most significant of these novel insights by far is the importance of embodiment.

1.5 Embodiment and Its Implications

By embodiment, we mean that intelligence always requires a body. Or, more precisely, we ascribe intelligence only to agents that are embodied, i.e., real physical systems whose behavior can be observed as they interact with the environment. Software agents, and computer programs in general, are disembodied, and many of the conclusions drawn in this book do not apply to them. As simple as the statement "intelligence requires a body" may sound, the implications are overwhelming, as we will see. There are some consequences of embodiment that are obvious, and some that are not. For example, if a system is embodied, it is subject to the laws of physics and has to somehow deal with gravity, friction, and energy supply in order to survive. While this is interesting and poses new challenges for our view of intelligence, the real importance of embodiment comes from the interaction between physical processes and what we might want to call information processes. In biological agents, this concerns the relation between physical actions and neural processing—or, to put it somewhat casually, between the body and the brain. The equivalent in a robot would be the relation between the robot's actions and its control program. Since the whole book is about precisely these issues, we will not go into any detail about this here. Instead, we would like to provide a flavor of what is to come, and for now it is all right if the reasons why embodiment is necessary for intelligence are not one hundred percent clear. Also, as a kind of preview, we merely mention the claims without substantiating them; we will do that in later chapters. Here are a few examples.

First, embodiment is an enabler for cognition or thinking: in other words, it is a prerequisite for any kind of intelligence. So, the body is not something troublesome that is simply there to carry the brain around, but it is necessary for cognition. It seems that the body is required even for functions such as mathematical thinking—something we often assume is a purely abstract, mental process—as argued by Lakoff and Núñez. Second, many tasks become much easier if embodiment is taken into account. For example, grasping objects requires much less control if stiffness and deformability of materials are used properly: just consider how the soft, deformable tissue of your fingertips makes the grasping of hard objects easier; imagine if you had to grasp a glass wearing thimbles on all your fingers! The reason the task becomes easier is that part of the neural control that would otherwise be required for grasping is in fact taken over by the morphological and material properties of the hand: if you were to grasp a glass with thimbles, you would have to be much more careful about how and where you placed your fingers. Third, if the sensors of a robot or organism are physically positioned on the body in the right places, some kind of preprocessing of the incoming sensory stimulation is performed by the very arrangement of the sensors, rather than by the neural system. That is, through the proper distribution of the sensors over the body, "good" sensory signals are delivered to the brain; it gets good "raw material" to work on. For example, grasping an object is easy because the anatomy of the human hand is such that the fingertips will tend to touch an object, rather than the backs of the fingers, and there are many more touch sensors in our fingertips than in the backs of our fingers and hands. Fourth, if the material properties of an agent's muscle-tendon system are exploited, rapid movements such as running can be achieved very easily even though the neural system would be too slow to control all the details of the movement. For example, when your foot hits the ground, the elastic stretching and recoil of the ankle is taken over by the springy material of the muscle-tendon system and need not be controlled by the neural system (this point will be elaborated in detail in chapter 4). Fifth, through an agent's physical interaction with its environment, informative and correlated sensory signals are generated in different sensory channels. This idea sounds complicated—and in fact it *is* complicated—but it lies at the heart of intelligent action, and we will explore it in great detail later on. For example, when you walk, the environment seems to flow past your eyes at the same time that the sensors in your leg muscles register the strains of moving. For example, when an agent moves, objects closer to the agent seem to move by faster than those farther away, which provides the agent with distance information. This kind of "information structuring" will be explored in later chapters. So, there exists a subtle interplay or balance between an agent's neural activity (the brain), its morphology (the body's shape and its material properties), and its interaction with the environment, and that interplay can be exploited to achieve certain tasks. Recall that the elasticity of the muscle-tendon system, or the deformability of the tissue on the fingertips, in a sense takes a load off the brain.

In addition to laying the groundwork for a new theory of intelligence using these ideas, we will attempt to dismantle the widely held assumption that the brain controls the body. This may be disconcerting for some, because it is an idea that runs very deep in our society and has a long history, as we have already pointed out. Rather than postulating that there is a hierarchical structure in which one part—the brain—controls another—the body—the new theory focuses on the interaction between these two systems. We will argue that although clearly of great importance, the brain is not the sole and central seat of intelligence; and that intelligence is instead distributed throughout the organism. We will dig even deeper and show that the notion of control itself needs to be revised. We will also make a case that brain processes cannot be understood by looking at the brain alone: in order to understand the function of the brain, we must consider embodiment; we must deal with the coupling between brain, body, and environment. It may be easier for us to think about hierarchical systems where one person or thing, e.g., the brain, is in control, rather than about distributed, flat systems where components influence each other—but that doesn't mean it's the way things really are. It is one goal of this book to demonstrate how things—especially ourselves—can be viewed differently.

We will argue—convincingly, we hope!—that the notion of intelligence as computation, which underlies the cognitivistic paradigm, is misleading, and that speculations about the future of artificial intelligence by extrapolating from Moore's law—the law that computing power doubles roughly every one to two years—are fundamentally flawed. The futurist, entrepreneur, and computer scientist Ray Kurzweil, author of *The Age of Spiritual Machines*, is a case in point. Because he assumes that intelligence is exclusively a function of computational power, he sketches a scenario where in the near future computers will outperform human brains simply because they will have as much or more number-crunching power. We hope to convince the reader that computational theories of intelligence are doomed to failure from the very outset. Also, we will show that in much of the literature on the subject there is confusion between what exists within the agent itself and what is present within the head of the person observing the agent: this is the frame-of-reference problem that we will encounter many times throughout the book.

We will also demonstrate that in spite of its limitations, artificial evolution (a class of computer algorithms modeled on biological evolution that will be described in chapter 6) is a very powerful design tool, especially for designing intelligent agents. We will in fact show that computers have automatically designed complex artifacts, and that in some cases these artifacts are superior in performance to those designed by human engineers. These results deal a heavy blow to the common belief that computers cannot be creative. But when we want to design an artifact that has to function in the real world, the designs have to be tested either in physically realistic simulations or directly in the real world, and needless to say this means that the artifact cannot be merely abstract, but must have a body.

The last implication of embodiment to be discussed here concerns the synthetic methodology, an approach that we will employ throughout the book and which we describe in detail in the next chapter. It can be characterized by the slogan "understanding by building." When studying embodiment, it is essential to build actual physical systems, which, because we are interested in intelligent systems, will most likely be robots. For example, if we are trying to understand human walking, the synthetic methodology requires that we build an actual walking robot. Of course, simulations can also be employed, but they have to replicate the actual physical processes of walking in order to tell us something about walking in general. And there is always the question of the accuracy of a simulation. Experience has shown that building a real physical system always yields the most new insights. It is easy to "cheat" with simulation: a real-world walking agent, like a human or a physical robot, has to somehow deal with bumps in the ground, while this problem can be ignored in a simulation (where each problem has to be explicitly programmed in). The synthetic methodology contrasts with the more classical analytical ways of proceeding as in biology, psychology, or neuroscience, where an animal or human is analyzed in detail by performing experiments on it. Having said that, it is interesting to note that the sciences in general have become more synthetic lately, as the brisk rise of the computational sciences demonstrates: physicists increasingly prepare experiments in simulation; surgeons prepare operations in simulation; and pharmacologists test the effects of drugs in simulation. If these simulations are to be useful, they of course have to be as accurate as possible. But even if there is a high level of simulation accuracy, it will always be necessary eventually to perform experiments in the real world.

1.6 Summary

Let us briefly summarize the main points we have made so far. We started by inspecting Descartes's famous quote, and the mind-body problem. Then we introduced the terms *thinking*, *cognition*, and *intelligence*, and showed that even though we all have a pretty clear idea of what we mean by these terms, they still are ill defined. Moreover, they are best conceived of as a continuum: intuitively, we view some behaviors as requiring more thinking than others, and some animals as being more intelligent than others. Because these are all descriptive terms, we should not spend too much time on trying to find clear-cut definitions. That being said, in normal usage, *thinking* is often associated with conscious thought, *cognition* is somewhat more general and is used for behaviors not directly coupled with sensory-motor processes, and *intelligence* is even more general and encompasses any kind of behavior—including abstract behaviors such as cognition and thinking—that is beneficial to the agent. We then highlighted a few of the reasons intelligence is so fascinating, e.g., because it is a sensitive issue in that it distinguishes us from other species, and because of the nature-nurture debate, which is about the extent to which intelligence is inherited or acquired during one's lifetime. We pointed out some intelligence-related phenomena that are hard to explain, such as perception and memory. Next, we outlined the difficulties and issues involved in actually defining intelligence, e.g., its subjective nature, the large variety of types of intelligence, and its continuous character. We then very briefly introduced the research field of artificial intelligence, which is about understanding biological systems, abstracting principles of intelligent behavior, and designing and building artificial systems. We then gave a rough idea of what embodiment is, touching on some of its far-reaching implications. We also stressed the importance of actually building physical systems.

Somewhat provocatively, we said that we will challenge the classical notion of the brain controlling the body, and we will try to show that computational theories of intelligence are doomed to failure. We will also, along the way, attempt to dismantle the myth that machines cannot be creative.

In summary, the import of assuming the embodied perspective for understanding and designing intelligent systems can hardly be overestimated. In the next chapter we will outline the conceptual landscape of artificial intelligence as it now stands: we will take a crack at clarifying the structure of this scientific discipline, describing the kinds of research that are being conducted, and explaining how the various subdisciplines relate to one another.