PETER GÄRDENFORS

COGNITIVE SCIENCE: FROM COMPUTERS TO ANT HILLS AS MODELS OF HUMAN THOUGHT

1. Before cognitive science

In this introductory chapter some of the main themes of the development of cognitive science will be presented. The roots of cognitive science go as far back as those of philosophy. One way of defining cognitive science is to say that it is just *naturalized philosophy*. Much of contemporary thinking about the mind derives from René Descartes' distinction between the body and the soul. They were constituted of two different substances and it was only humans that had a soul and were capable of thinking. According to him, other animals were mere automata.

Descartes was a *rationalist*: our minds could gain knowledge about the world by rational thinking. This epistemological position was challenged by the *empiricists*, notably John Locke and David Hume. They claimed that the only reliable source of knowledge is sensory experience. Such experiences result in *ideas*, and thinking consists of connecting ideas in various ways.

Immanuel Kant strove to synthesize the rationalist and the empiricist positions. Our minds always deal with our inner experiences and not with the external world. He introduced a distinction between the thing in itself (*das Ding an sich*) and the thing perceived by us (*das Ding an uns*). Kant then formulated a set of *categories of thought*, without which we cannot organize our phenomenal world. For example, we must interpret what happens in the world in terms of cause and effect.

The favourite method among philosophers of gaining insights into the nature of the mind was *introspection*. This method was also used by psychologists at the end of the 19th and the beginning of the 20th century. In particular,

this was the methodology used by Wilhelm Wundt and other German psychologists. By looking inward and reporting inner experiences it was hoped that the structure of the conscious mind would be unveiled.

However, the inherent subjectivity of introspection led to severe methodological problems. These problems set the stage for a scientific revolution in psychology. In 1913, John Watson published an article with the title "Psychology as the behaviourist views it" which has been seen as a *behaviourist* manifesto. The central methodological tenet of behaviourism is that only objectively verifiable observations should be allowed as data. As a consequence, scientists should prudently eschew all topics related to mental processes, mental events, and states of mind. Observable behaviour consists of *stimuli* and *responses*. The brain was treated as a black box. According to Watson, the goal of psychology was to formulate lawful connections between such stimuli and responses.

Behaviourism had a dramatic effect on psychology, particularly in the United States. As a consequence, animal psychology became a fashionable topic. Laboratories were filled with rats running in mazes and pigeons pecking at coloured chips. An enormous amount of data concerning *conditioning* of behaviour was collected. There was also a behaviourist influence in linguistics: the connection between a word and the objects it referred to was seen as a special case of conditioning.

Analytical philosophy, as it was developed in the early 20th century, contained ideas that reinforced the behaviourist movement within psychology. In the 1920s, the so-called Vienna circle formulated a philosophical programme which had as its primary aim to eliminate as much as possible of metaphysical speculations. Scientific reasoning should be founded on an *observational* basis. The observational data were obtained from experiments. From these data knowledge could only be expanded by using logically valid inferences. Under the headings of *logical empiricism* or *logical positivism*, this methodological programme has had an enormous influence on most sciences.

The ideal of thinking for the logical empiricists was logic and mathematics, preferably in the form of *axiomatic systems*. In the hands of people like Giuseppe Peano, Gottlob Frege, and Bertrand Russell, arithmetic and logic had been turned into strictly formalized theories at the beginning of the 20th century. The axiomatic ideal was transferred to other sciences with less success. A background assumption was that all scientific knowledge could be formulated in some form of *language*.

2. The dawn of computers

As a part of the axiomatic endeavour, logicians and mathematicians investigated the limits of what can be computed on the basis of axioms. In particular, the focus was put on what is called *recursive functions*. The logician Alonzo Church is famous for his thesis from 1936 that everything that can be computed can be computed with the aid of recursive functions.

At the same time, Alan Turing proposed an abstract machine, later called the *Turing machine*. The machine has two main parts: an infinite tape divided into cells, the contents of which can be read and then overwritten; and a movable head that reads what is in a cell on the tape. The head acts according to a finite set of instructions, which, depending on what is read and the current state of the head, determines what to write on the cell (if anything) and then whether to move one step left or right on the tape. It is Turing's astonishing achievement that he proved that such a simple machine can calculate all recursive functions. If Church's thesis is correct, this means that a Turing machine is able to compute everything that can be computed.

The Turing machine is an abstract machine – there are no infinite tapes in the world. Nevertheless, the very fact that all mathematical computation and logical reasoning had now been shown to be mechanically processable inspired researchers to construct real machines that could perform such tasks. One important technological invention was the so-called logical circuits that were constructed by systems of electric tubes. The Turing machine inspired John von Neumann to propose a general architecture for a real computer based on logic circuits. The machine had a central processor which read information from external memory devices, transformed the input according to the instructions of the program of the machine, and then stored it again in the external memory or presented it on some output device as the result of the calculation. The basic structure was thus similar to that of the Turing machine.

In contrast to earlier mechanical calculators, the computer *stored* its own instructions in the memory coded as binary digits. These instructions could be modified by the programmer, but also by the program itself while it was operating. The first machines developed according to von Neumann's general architecture appeared in the early 1940s.

Suddenly there was a machine that seemed to be able to think. A natural question was then to what extent computers think like humans. In 1943, McCulloch

and Pitts published an article that became very influential. They interpreted the firings of the neurons in the brain as sequences of zeros and ones, by analogy with the binary digits of the computers. The neuron was seen as a logic circuit that combined information from other neurons according to some logical operator and then transmitted the results of the calculation to other neurons.

The upshot was that the entire brain was seen as a huge computer. In this way, the metaphor that became the foundation for cognitive science was born. Since the von Neumann architecture for computers was at the time the only one available, it was assumed that the brain too had essentially the same general structure.

The development of the first computers occurred at the same time as the concept of *information* as an abstract quantity was developed. With the advent of various technical devices for the transmission of signals, such as telegraphs and telephones, questions of efficiency and reliability in signal transmission were addressed. A breakthrough came with the mathematical theory of information presented by Claude Shannon. He found a way of measuring the amount of information that was transferred through a channel, independently of which code was used for the transmission. In essence, Shannon's theory says that the more improbable a message is statistically, the greater is its informational content (Shannon, Weaver, 1948). This theory had immediate applications in the world of zeros and ones that constituted the processes within computers. It is from Shannon's theory that we have the notions of bits, bytes, and baud that are standard measures for present-day information technology products.

Turing saw the potentials of computers very early. In a classical paper from 1950, he foresaw a lot of the developments of computer programs that were to come later. In that paper, he also proposes the test that nowadays is called the *Turing test*. To test whether a computer program succeeds in a cognitive task, such as playing chess or conversing in ordinary language, let an external observer communicate with the program via a terminal. If the observer cannot distinguish the performance of the program from that of a human being, the program is said to have passed the Turing test.

3. 1956: Cognitive science is born

There are good reasons for saying that cognitive science was born in 1956. That year a number of events in various disciplines marked the beginning of a new era. A conference where the concept of *Artificial Intelligence* (AI) was used

for the first time was held at Dartmouth College. At that conference, Alan Newell and Herbert Simon demonstrated the first computer program that could construct logical proofs from a given set of premises. This event has been interpreted as the first example of a machine that performed a cognitive task.

Then in linguistics, later the same year, Noam Chomsky presented his new view of *transformational grammar* that was to be published in his book *Syntactic Structures* in 1957. This book caused a revolution in linguistics and Chomsky's views on language are still dominant in large parts of the academic world. A central argument is that any natural language would require a Turing machine to process its grammar. Again we see a correspondence between a human cognitive capacity, this time judgements of grammaticality, and the power of Turing machines. No wonder that Turing machines were seen as what was needed to understand thinking.

Also in 1956, the psychologist George Miller published an article with the title "The magical number seven, plus or minus two: Some limits on our capacity for processing information" that has become a classic within cognitive science. Miller argued that there are clear limits to our cognitive capacities: we can actively process only about seven units of information. This article directly applies Shannon's information theory to human thinking. It also explicitly talks about cognitive processes, something which had been considered to be very bad manners in the wards of the behaviourists that were sterile of anything but stimuli and responses. However, with the advent of computers and information theory, Miller now had a *mechanism* that could be put in the black box of the brain: computers have a limited processing memory and so do humans.

Another key event in psychology in 1956 was the publication of the book *A Study of Thinking*, written by Jerome Bruner, Jacqueline Goodnow, and George Austin, who had studied how people group examples into categories. They reported a series of experiments where the subjects' task was to determine which of a set of cards with different geometrical forms belong to a particular category. The category was set by the experimenter, for example the category of cards with two circles on them. The subjects were presented one card at a time and asked whether the card belonged to the category. The subject was then told whether the answer was correct or not. Bruner and his colleagues found that when the concepts were formed as conjunctions of elementary concepts like "cards with red circles", the subjects learned the category quite efficiently; while if the category was generated by a disjunctive concept like "cards with circles *or* a red object" or negated concepts like "cards that do *not* have two circles," the subjects had severe

problems in identifying the correct category. Note that Bruner, Goodnow, and Austin focused on *logical* combinations of primitive concepts, again following the underlying tradition that human thinking is based on logical rules.

4. The rise and fall of artificial intelligence

Newell and Simon's program was soon to be followed by a wealth of more sophisticated logical theorem-proving programs. There was great faith in these programs: in line with the methodology of the logical positivists, it was believed that once we have found the fundamental axioms for a particular domain of knowledge we could then use computers instead of human brains to calculate all the consequences of the axioms.

But thinking is not logic alone. Newell and Simon soon started on a more ambitious project called the General Problem Solver that, in principle, should be able to solve any well-formulated problem. The General Problem Solver worked by means-end analysis: a problem is described by specifying an initial state and a desired goal state and the program attempts to reduce the gap between the start and the goal states. However, work on the program was soon abandoned since the methods devised by Newell and Simon turned out not to be as general as they had originally envisaged.

The first robot programs, like for example STRIPS developed at Stanford Research Institute, also followed the symbolic tradition by representing all the knowledge of the robot by formulas in a language that was similar to predicate logic. The axioms and rules of the program described the results of various actions together with the preconditions for the actions. Typical tasks for the robots were to pick up blocks in different rooms and stack them in a chosen room. However, in order to plan for such a task, the program needed to know all the consequences of the actions taken by the robot. For instance, if the robot went through the door of a room, the robot must be able to conclude that the blocks that were in the room did not move or ceased to exist as a result of the robot entering the room. It turned out that giving a complete description of the robot's world and the consequences of its actions resulted in a combinatorial explosion of the number of axioms required. This has been called the *frame problem* in robotics.

The optimism of AI researchers and their high-flying promises concerning the capabilities of computer programs were met with several forms of criticism. Already in 1960, Yehoshua Bar-Hillel wrote a report on the fundamental prob-

lems of using computers to perform automatic translations from one language to another. And in 1967, Joseph Weizenbaum constructed a seductive program called ELIZA that could converse in natural language with its user. ELIZA was built to simulate a Rogerian psychotherapist. The program scans the sentences written by the user for words like "I", "mother", "love" and when such a word is found, the program has a limited number of preset responses (where the values of certain variables are given by the input of the user). The program does very little calculation and understands absolutely nothing of its input. Nevertheless, it is successful enough to delude an unsuspecting user for some time until its responses become too stereotyped.

Weizenbaum's main purpose in writing ELIZA was to show how easy it was to fool a user that a program has an understanding of a dialogue. We are just too willing to ascribe intelligence to something that responds appropriately in a few cases – our human-centred thinking extends easily to computers. Weizenbaum was horrified because some professional psychiatrists suggested ELIZA as a potential therapeutic tool that might be used in practice by people with problems.

In spite of the critics, AI lived on in, more or less, its classical shape during the 1970s. Among the more dominant later research themes were the so-called *expert systems* that have been developed in various areas. Such systems consist of a large number of symbolic rules (that have normally been extracted from human experts) together with a computerized inference engine that applies the rules recursively to input data and ends up with some form of solution to a given problem.

The most well-known expert system is perhaps MYCIN, which offers advice on infectious diseases (it even suggests a prescription of appropriate antibiotics). MYCIN was exposed to the Turing test in the sense that human doctors were asked to suggest diagnoses on the basis of the same input data, from laboratory tests, that was given to the program. Independent evaluators then decided whether the doctors or MYCIN had done the best job. Under these conditions, MYCIN passed the Turing test, but it can be objected that if the doctors had been given the opportunity to see and examine the patients, they would (hopefully) have outperformed the expert system.

However, expert systems never reached the adroitness of human experts and they were almost never given the opportunity to have the decisive word in real cases. A fundamental problem is that such systems may incorporate an extensive amount of knowledge, but they hardly have any knowledge about the *validity* of their knowledge. Without such meta-knowledge, a system cannot make valid *judgements* that form the basis of sound decisions.

5. Mind: the gap

A unique aspect of our cognitive processes is that we experience at least part of them as being *conscious*. The problem of what consciousness is has occupied philosophers for centuries, and there is a plethora of theories of the mind.

Cartesian dualism, which treats the body and the mind as separate substances, has lost much of its influence during the 20th century. Most current theories of the mind are *materialistic* in the sense that only physical substances are supposed to exist. But this position epitomizes the question of how conscious experiences can be a result of material processes. There seems to be an unbridgeable gap between our physicalistic theories and our phenomenal experiences.

A theory of the mind that has been popular since the 1950s is the *identity theory*, which claims that conscious processes are identical with material processes in the brain. As a consequence, the phenomenal is in principle *reducible* to the physical. It should be noted that according the identity theory it is only processes in the brain that can become parts of conscious experiences.

However, the new vogue of cognitive theories based on the analogy between the brain and the computer soon attracted the philosophers. In 1960, Hilary Putnam published an article with the title "Minds and machines" where he argued that it is not the matter of a brain or a computer that determines whether it has a mind or not, but only what *function* that brain or computer performs. And since the function of a computer was described by its program, the function of the brain was, by analogy, also identified with a program. This stance within the philosophy of mind has become known as *functionalism*.

The central philosophical tenet of the AI approach to represent cognitive processes is that mental representation and processing is essentially *symbol manipulation*. The symbols can be concatenated to form expressions in a *language of thought* – sometimes called Mentalese. The different symbolic expressions in a mental state of a person are connected only via their logical relations. The symbols are manipulated exclusively on the basis of their form – their *meaning* is not part of the process.

The material basis for these processes is irrelevant to the description of their results – the same mental state can be realized in a brain as well as in a computer. Thus, the paradigm of AI clearly presupposes the functionalist philosophy of mind. In brief, the mind is thought to be a computing device, which generates symbolic expressions as inputs from sensory channels, performs logical opera-

tions on these sentences, and then transforms them into linguistic or non-linguistic output behaviours.

However, functionalism leaves unanswered the question of what makes certain cognitive processes conscious or what gives them *content*. As an argument against the strongest form of AI that claims that all human cognition can be replaced by computer programs, John Searle (1980) presents his "Chinese room" scenario. This example assumes that a person who understands English but no Chinese is locked into a room together with a large set of instructions written in English. The person is then given a page of Chinese text that contains a number of questions. By meticulously following the instructions with respect to the symbols that occur in the Chinese questions, he is able to compose a new page in Chinese that comprises answers to the questions.

According to functionalism (and in compliance with the Turing test) the person in the room who is following the instructions would have the same capacity as a Chinese-speaking person. Hence functionalism would hold that the person *together with* the equipment in the room understands Chinese. But this is potently absurd, claims Searle. For analogous reasons, according to Searle, a computer lacks *intentionality* and can therefore not understand the meaning of sentences in a language. Searle's argument has spawned a heated debate about the limits of functionalism and what it would mean to *understand* something.

6. First heresy against high-church computationalism: thinking is not only by symbols

6.1. Artificial neuron networks

For many years, the symbolic approach to cognition was totally dominant. But as a result of the various forms of criticism which led to a greater awareness of the limitations of the "symbol crunching" of standard AI programs, the ground was prepared for other views of the fundamental mechanisms of thinking. We find the first signs of heresy against what has been called "high-church computationalism".

For empiricist philosophers like Locke and Hume, thinking consists basically in the forming of *associations* between "perceptions of the mind." The basic

idea is that events that are similar become connected in the mind. Activation of one idea activates others to which it is linked: when thinking, reasoning, or daydreaming, one thought reminds us of others.

During the last few decades, associationism has been revived with the aid of a new model of cognition – *connectionism*. Connectionist systems, also called *artificial neuron networks*, consist of large numbers of simple but highly interconnected units ("neurons"). The units process information in *parallel* in contrast to most symbolic models where the processing is serial. There is no central control unit for the network, but all neurons act as individual processors. Hence connectionist systems are examples of *parallel distributed processes* (Rumelhart, McClelland, 1986).

Each unit in an artificial neuron network receives activity, both excitatory and inhibitory, as input; and transmits activity to other units according to some function of the inputs. The behaviour of the network as a whole is determined by the initial state of activation and the connections between the units. The inputs to the network also gradually change the strengths of the connections between units according to some *learning rule*. The units have no memory in themselves, but earlier inputs are represented indirectly via the changes in strengths they have caused. According to connectionism, cognitive processes should not be represented by symbol manipulation, but by the *dynamics* of the patterns of activities in the networks. Since artificial neuron networks exploit a massive number of neurons working in parallel, the basic functioning of the network need not be interrupted if some of the neurons are malfunctioning. Hence, connectionist models do not suffer from the computational brittleness of the symbolic models and they are also much less sensitive to noise in the input.

Some connectionist systems aim at modelling neuronal processes in human or animal brains. However, most systems are constructed as general models of cognition without any ambition to map directly to what is going on in the brain. Such connectionist systems have become popular among psychologists and cognitive scientists since they seem to be excellent *simulation tools* for testing associationist theories.

Artificial neuron networks have been developed for many different kinds of cognitive tasks, including vision, language processing, concept formation, inference, and motor control. Among the applications, one finds several that traditionally were thought to be typical symbol processing tasks like pattern matching and syntactic parsing. Perhaps the most important applications, however, are models of various forms of *learning*.

Connectionist systems brought a radically new perspective on cognitive processes: cognition is *distributed* in the system. In contrast, a von Neumann computer is controlled by a central processor. In favour of this architecture it has been argued that if the brain is a computer, it must have a central processor – where would you otherwise find the "I" of the brain? But the analogy does not hold water – there is no area of the brain that serves as a pilot for the other parts: there is no one in charge. The neuronal processes are distributed all over the brain, they occur in parallel and they are to a certain extent independent of each other. Nevertheless, the brain functions in a goal-directed manner. From the connectionist perspective, the brain is best seen as a *self-organizing system*. Rather than working with a computer-like program, the organization and learning that occur in the brain should be seen as an *evolutionary* process (Edelman, 1987).

On this view, the brain can be seen as an *ant hill*. The individual neurons are the ants who perform their routine jobs untiringly, but rather unintelligently, and who receive signals from other neurons via their dendrite antennas. From the interactions of a large number of simple neurons a complex well-adapted system like an ant hill *emerges* in the brain. In other words, cognition is seen as a holistic phenomenon in a complex system of distributed parallel processes.

Along with the development of symbolic and connectionist programming techniques, there has been a rapid development in the *neurosciences*. More and more has been uncovered concerning the neural substrates of different kinds of cognitive processes. As the argument by McCulloch and Pitts shows, it was thought at an early stage that the brain would function along the same principles as a standard computer. But one of major sources of influence for connectionism was the more and more conspicuous conclusion that neurons in the brain are not logic circuits, but operate in a distributed and massively parallel fashion and according to totally different principles than those of computers. For example, Hubel and Wiesel's (1968) work on the signal-detecting functioning of the neurons in the visual cortex were among the path-breakers for the new view of the mechanisms of the brain. It is seen as one of the strongest assets of connectionism that the mechanisms of artificial neuron networks are much closer to the functioning of the brain.

Another talented researcher who combined thorough knowledge about the brain with a computational perspective was David Marr. His book *Vision* from 1982 is a milestone in the development of cognitive neuroscience. He worked out connectionist algorithms for various stages of the visual processing from

the moment the cells on the retina react, until a holistic 3D model of the visual scene is constructed in the brain. Even though some of his algorithms have been questioned by later developments, his methodology has led to a much deeper understanding of the visual processes during the last two decades.

6.2. Non-symbolic theories of concept formation

Both the symbolic and the connectionist approaches to cognition have their advantages and disadvantages. They are often presented as competing paradigms, but since they attack cognitive problems on different levels, they should rather be seen as complementary methodologies.

However, there are aspects of cognitive phenomena for which neither symbolic representation nor connectionism seems to offer appropriate modelling tools. In particular it appears that mechanisms of *concept acquisition*, which is paramount for the understanding of many cognitive phenomena, cannot be given a satisfactory treatment in any of these representational forms. Concept learning is closely tied to the notion of *similarity*, which has also turned out to be problematic for the symbolic and associationist approaches.

To handle concept formation, among other things, a third form of representing information that is based on using *geometrical* or *topological* structures, rather than symbols or connections between neurons, has been advocated. This way of representing information is called the *conceptual* form. The geometrical and topological structures generate mental *spaces* that represent various domains. By exploiting distances in such spaces, judgements of similarity can be modelled in a natural way.

In the classical Aristotelian theory of concepts that was embraced by AI and early cognitive science (for example, in the work of Bruner, Goodnow, and Austin presented above) a concept is defined via a set of *necessary and sufficient properties*. According to this criterion, all instances of a classical concept have equal status. The conditions characterizing a concept were formulated in linguistic form, preferably in some symbolic form.

However, psychologists like Eleanor Rosch showed that in the majority of cases, concepts show *graded membership*. These results led to dissatisfaction with the classical theory. As an alternative, *prototype theory* was proposed in the mid-1970s. The main idea of this theory is that within a category of objects, such

as those instantiating a concept, certain members are judged to be more *representative* of the category than others. For example robins are judged to be more representative of the category "bird" than are ravens, penguins, and emus; and desk chairs are more typical instances of the category "chair" than rocking chairs, deck chairs, and beanbag chairs. The most representative members of a category are called *prototypical* members. The prototype theory of concepts fits much better with the conceptual form of representing information than with symbolic representations.

6.3. Thinking in images

When we think or speak about our own thoughts, we often refer to inner scenes or pictures that we form in our fantasies or in our dreams. However, from the standpoint of behaviourism, these phenomena were unspeakables, beyond the realm of the sober scientific study of stimuli and responses. This scornful attitude towards mental images was continued in the early years of AI. Thinking was seen as symbol crunching and images were not the right kind of building blocks for computer programs.

However, in the early 1970s psychologist began studying various phenomena connected with *mental imagery*. Roger Shepard (Shepard, Metzler, 1971) and his colleagues performed an experiment that has become classical. They showed subjects pictures representing pairs of 3D block figures that were rotated in relation to each other and asked the subjects to respond as quickly as possible whether the two figures were the same or whether they were mirror images of one another. The surprising finding was that the time it took the subject to answer was linearly correlated with the number of degrees the second object had been rotated in relation to the first. A plausible interpretation of these results is that the subjects generate mental images of the block figures and *rotate* them in their minds.

Stephen Kosslyn (1980) and his colleagues have documented similar results concerning people's abilities to imagine maps. In a typical experiment, subjects are shown maps of a fictional island with some marked locations: a tree, a house, a bay, etc. The maps are removed and the subjects are then asked to focus mentally on one location on the map and then move their attention to a second location. The finding was that the time it takes to mentally scan from one location

to the other is again a linear function of the distance between the two positions on the map. The interpretation is that the subjects are *scanning a mental map*, in the same manner as they would scan a physically presented map.

Another strand of mental imagery has been developed within *cognitive semantics*. In the Chomskian theory of linguistics, syntax is what counts and semantic and pragmatic phenomena are treated like Cinderellas. In contrast, within cognitive semantics, as developed by Ron Langacker (1987) and George Lakoff (1987) among others, the cognitive representation of the meaning of linguistic expressions is put into focus. Their key notion for representing linguistic meanings is that of an *image schema*. A common assumption is that such schemas constitute the representational form that is common to perception, memory, and semantic meaning. The theory of image schemas also builds on the prototype theory for concepts. Again, this semantic theory replaces the uninterpreted symbols of high-church computationalism with image-like representations that have an inherent meaning. In particular, our frequent use of more or less conventional *metaphors* in everyday language can be analysed in an illuminating way using image schemas.

7. Second heresy: cognition is not only in the brain

7.1. The embodied brain

The brain is not made for calculating – its primary duty it to control the body. For this reason it does not function in solitude, but is largely dependent on the body it is employed by. In contrast, when the brain was seen as a computer, it was more or less compulsory to view it as an isolated entity. However, there is little hope that such a scenario would ever work. As a consequence, there has recently been a marked increase in studies of the *embodied* brain.

For example, the eye is not merely seen as an input device to the brain and the hand as enacting the will of the brain, but the eye-hand-brain is seen as a *coordinated* system. For many tasks, it turns out that we think faster with our hands than with our brains. A simple example is the computer game Tetris where you are supposed to quickly turn, with the aid of the keys on the keyboard, geometric objects that come falling over a computer screen in order to fit them with the pattern at the bottom of the screen. When a new object appears, one can mentally

rotate it to determine how it should be turned before actually touching the keyboard. However, expert players turn the object faster with the aid of the keyboard than they turn an image of the object in their brains. This is an example of what has been called *interactive thinking*. The upshot is that a human who is manipulating representations in the head is not the same cognitive system as a human interacting directly with the represented objects.

Also within linguistics, the role of the body has attracted attention. One central tenet within cognitive semantics is that the meanings of many basic words are embodied, in the sense that they relate directly to bodily experiences. George Lakoff and Mark Johnson show in their book *Metaphors We Live By* (1980) that a surprising variety of words, for instance prepositions, derive their complex meaning from a basic embodied meaning that is then extended by metaphorical mappings to a number of other domains.

7.2. Situated cognition

There is one movement within cognitive science, known as *situated cognition*, which departs even further from the traditional stance. The central idea is that in order to function efficiently the brain does not only need the body but also the surrounding world. In other words, it is *being there* that is our primary function as cognitive agents (Clark, 1997). Cognition is not imprisoned in the brain but emerges in the interaction between the brain, the body and the world. Instead of representing the world in an inner model the agent in most cases uses the world as its own model. For example, in vision, an agent uses rapid movements of the eyes to extract what is needed from a visual scene, rather than building a detailed 3D model of the world in its head.

In many cases it is impossible to draw a line between our senses and the world. The captain of a submarine "sees" with the periscope and a blind person "touches" with her stick, not with the hand. In the same way we "think" with road signs, calendars, and pocket calculators. There is no sharp line between what goes on inside the head and what happens in the world. The mind leaks out into the world.

By arranging the world in a smart way we can afford to be stupid. We have constructed various kinds of artefacts that help us solve cognitive tasks. In this way the world functions as *scaffolding* for the mind (Clark, 1997). For example, we have developed a number of memory aids: we "remember" with the aid of

books, DVDs, hard-disks, etc. In this way, memory is placed in the world. For another practical example, the work of an architect or a designer is heavily dependent on making different kinds of sketches: the sketching is an indispensable component of the cognitive process (Gedenryd, 1998).

The emphasis on situated cognition is coupled with a new view of the basic nature of the cognitive structures of humans. Instead of identifying the brain with a computer, the *evolutionary origin* of our thinking is put into focus. The key idea is that we have our cognitive capacities because they have been useful for survival and reproduction in the past. From this perspective, it becomes natural to compare our form of cognition with that of different kinds of animals. During the last decade, *comparative cognition* has grown considerably as a research area. The methodology of this branch is different from that of traditional cognitive psychology. Instead of studying subjects in laboratories under highly constrained conditions, evolutionary psychology focuses on data that are *ecologically valid* in the sense that they tell us something about how humans and animals act in natural problem-solving situations.

7.3. The pragmatic turn of linguistics

The role of culture and society in cognition was marginalized in early cognitive science. These were problem areas that were to be addressed when an understanding of individual cognition had been achieved. However, when the focus of cognitive theories shifted away from symbolic representations, semantic and pragmatic research reappeared on the agenda. Pragmatics consists of the rules for linguistic actions; semantics is conventionalized pragmatics; and finally, syntax adds grammatical markers to help disambiguate when the context does not suffice to do so. This tradition connects with several other research areas such as anthropology, psychology, and situated cognition,

This shift of the linguistic programme can also be seen in the type of data that researchers are considering. In the Chomskian research programme, *single sentences* presented out of context are typically judged for their grammaticality. The judgements are often of an introspective nature when the researcher is a native speaker of the language studied. In contrast, within the pragmatic programme, recordings of actual *conversations* are recorded or video-taped. For the purpose of analysis, they are normally transcribed by various methods. The conversational analyses treat language as part of a more general interactive cognitive setting.

7.4. Robotics

The problem of constructing a robot is a good test of progress in cognitive science. A robot needs perception, memory, knowledge, learning, planning, and communicative abilities, that is, exactly those capacities that cognitive science aims at understanding. Current industrial robots have very little of these abilities – they can perform a narrow range of tasks in a specially prepared environment.

In contrast, nature has, with the stamina of evolution, solved cognitive problems by various methods. Most animals are *independent* individuals, often extremely flexible. The simplest animals are classified as *reactive systems*. This means that they have no foresight, but react to stimuli as they turn up in the environment. So, given nature's solutions, why can we not construct machines with the capacity of a cockroach?

The current trend in robotics is to start from reactive systems and then add higher cognitive modules that amplify or modify the basic reactive systems. This methodology is based on what Rodney Brooks calls the *subsumption architecture*. One common feature of such robots is that that they *learn by doing*: linguistic or other symbolic inputs play a minor role in their acquisition of new knowledge. One factor that was forgotten in classical AI is that animals have a *motivation* for their behaviour. From the perspective of evolution, the utmost goals are survival and reproduction. In robotics, the motivation is set by the constructor.

8. The future of cognitive science

The goal of contemporary cognitive science is not primarily to build a thinking machine, but to increase our understanding of cognitive processes. This can be done by various methods, including traditional psychological experiments, observations of authentic cognitive processes in practical action, or by simulating cognition in robots or programs. Unlike the early days of AI when it was believed that one single methodology, that of symbolic representation, could solve all cognitive problems, the current trend is to work with several forms of representations and data.

Furthermore, the studies tend to be closely connected to findings in neuroscience and in other biological sciences. New techniques of *brain imaging* will

continue to increase our understanding of the multifarious processes going on in the brain. Other techniques, such as *eye-tracking*, will yield rich data for analysing our cognitive interaction with the world and with the artefacts in it.

As regards the practical applications of cognitive science, a main area is the construction of *interfaces* to information technological products. The aim is that IT products should be as adapted to the demands of human cognition as possible. In other words, it should be the goal of information technology to build scaffolding tools that enhance human capacities. To give some examples of already existing aids, pocket calculators help us perform rapid and accurate calculations that were previously done laboriously with pen and paper or even just in the head. And word processors relieve us from the strain of retyping a manuscript.

Donald Norman (1988) started a tradition in *cognitive design* with his classical book *The Design of Everyday Things*. He showed by a wealth of provocative examples that technical constructors very often neglect the demands and limitations of human cognition. The user-friendliness of computer programs, mobile phones, remote TV controls, etc., has increased, but there is still an immense potential to apply the findings of cognitive science in order to create products that better support our ways of thinking and remembering.

Another area where cognitive science ought to a have a great impact in the future is *education*. There is a strong trend to equip schools at all levels with more and more computers. Unfortunately, most efforts are spent on the technical and financial aspects and very little on the question of *how* the computers should be used in schools. A number of so-called educational computer programs have been developed. With few exceptions, however, these programs are of a drill-and-exercise character. In particular, various kinds of *simulation* programs may be supportive for the learning process. For example, when teaching physics, a program that simulates the movements of falling bodies and displays the effects on the screen, allowing the student to interactively change the gravitational forces and other variables, will give a better grasp of the meaning of the physical equations than many hours of calculation by hand. Another promising area is the use of *virtual agents* in virtual environments.

For the development of truly educational computer programs, collaboration with cognitive scientists will be mandatory. Those who design the programs must have a profound knowledge of how human learning and memory works, of how we situate our cognition in the world and of how we communicate. Helping educationalists answer these questions will be one of the greatest challenges for cognitive science in the future.

As a last example of the future trends of cognitive science, I believe that research on the processing of sensory information will be useful in the development of *tools for the handicapped*. The deaf and blind each lack a sensory channel. Through studies of multimodal communication, these sensory deficits can hopefully be aided. If we achieve better programs for speech recognition for example, deafness can be partly compensated for.

In conclusion, we can expect that in the future, cognitive science will supply man with new tools, electronic or not, that will be better suited to our cognitive needs and that may increase the quality of our lives. In many areas, it is not technology that sets the limits, but rather our lack of understanding of how human cognition works.

References

Bruner, J.S., Goodnow, J.J. and Austin, G.A. (1956). *A Study of Thinking*. New York: Wiley.

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.

Clark, A. (1997). *Being there*. Cambridge, MA: MIT Press.

Edelman, G.M. (1987). *Neural Darwinism.* New York, NY: Basic Books.

Gedenryd, H. (1998). *How Designers Work*. Lund: Lund University Cognitive Studies 75.

Hubel, D.H., Wiesel, T.N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology* 195: 215–243.

Kosslyn, S. (1980). *Image and Mind*. Cambridge, MA: Harvard University Press.

Lakoff, G. (1987). *Women, Fire, and Dangerous Things*. Chicago, IL: University of Chicago Press.

Lakoff, G., Johnson, M. (1980). *Metaphors We Live By*. Chicago, IL: University of Chicago Press.

Langacker, R.W. (1987). *Foundations of Cognitive Grammar*, vol. 1. Stanford, CA: Stanford University Press.

Marr, D. (1982). *Vision*. San Francisco, CA: Freeman.

McCulloch, W.S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysic*s 5: 115–133.

Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 63: 81–97.

Norman, D.A. (1988). *The Design of Everyday Things*. New York, NY: Basic Books.

Putnam, H. (1960). Minds and machines. In: S. Hook (ed.), *Dimensions of Mind*. New York, NY: New York University Press.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General* 104: 192–233.

Rumelhart, D.E., McClelland, J.L. (1986). *Parallel Distributed Processing*, vol. 1–2. Cambridge, MA: MIT Press.

Searle, J.R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences* 3: 417–457.

Shannon, C.E., Weaver, W. (1948). *The Mathematical Theory of Communication*. Chicago, IL: The University of Illinois Press.

Shepard, R.N., Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science* 171 (3972): 701–703.

Turing, A. (1950). Computing machinery and intelligence. *Mind* 59: 433–460.

Watson, J.B. (1913). Psychology as the behaviourist views it. *Psychological Review* 20: 158–177.

Weizenbaum, J. (1967). *Computer Power and Human Reason*. New York, NY: Freeman and Company.