16 The third age of computing

Every 30 years there is a new wave of things that computers do. Around 1950 they began to model events in the world (*simulation*), and around 1980 to connect people (*communication*). Since 2010 they have begun to engage with the physical world in a non-trivial way (*embodiment*).
Butler Lampson[1]

The next revolution

The first age of computing was concerned with using computers for *simulation*. As we have seen, the first computers were built to do complex calculations. The initial motivation for building the ENIAC was to calculate artillery tables showing the angles at which guns should be fired based on the distance to the target and other conditions. After World War II, scientists used the ENIAC to explore possible designs for a hydrogen bomb. More generally, computers were used to simulate complex systems defined in terms of a mathematical model that captured the essential characteristics of the system under study. During the first thirty years of computing, from about 1950 until the early 1980s, researchers increasingly used computers for simulations of all sorts of complex systems. Computer simulations have transformed our lives, from designing cars and planes to making weather forecasts and financial models. At the same time, businesses used computers for performing the many, relatively simple calculations needed to manage inventories, payroll systems, and bank transactions. Even these very early computers could perform numerical calculations much, much faster than humans.

The second age of computing was about using computers for *communication*. The last thirty years, from the early 1980s until today, have seen computers become personal, not only for scientists and businesses but also for consumers. We now routinely use laptops, mobile phones, and tablets for a variety of activities, such as word processing, sending emails, searching the web, sharing photos, reading ebooks, and watching videos. Huge improvements in processing power together with astonishing miniaturization have come from the steady advance of computer technology predicted by Moore's law. It is these dramatic improvements in power and size during the last thirty years that have made possible the wide range of compact, portable computing devices we have available today. But this miniaturization of computing has been accompanied by an equally dramatic increase in *connectivity*, the ability to communicate with other computers.

Today's global Internet emerged, along with the increasing availability of wireless networks, from early experiments with the ARPANET, a computer network created by the U.S. Department of Defense in the late 1960s and the 1970s as a means of communication between research laboratories and universities. And by the 1990s, the World Wide Web had arrived to transform our online lives. The web made it possible for people with little computer proficiency to surf the Internet. It also enabled electronic commerce sites such as Amazon to emerge as serious competitors to bricks-and-mortar businesses. The increasing connectivity of our computing and communication devices has also led to the rise of social networking sites like Facebook and Twitter and the emergence of *crowdsourcing*, the practice of gathering services, ideas, information, or money by soliciting contributions from a large group of people online. Crowdsourcing websites include Amazon's Mechanical Turk, a service that uses humans to perform tasks that people do better than computers, such as comparing colors or translating foreign languages. Another is Wikipedia, a free encyclopedia that permits anyone to write and edit almost all its articles. Still another is Galaxy Zoo, an astronomy project that invites people to help classify large numbers of galaxies.

Butler Lampson's third age of computing is about using computers for *embodiment* – that is, using computers to interact with people in new and intelligent ways:

> The most exciting applications of computing in the next 30 years will *engage* with the physical world in a non-trivial way. Put another way, computers will become *embodied.*[2]

He asserts that the present state of computer applications, such as robotic surgery, remote-controlled drones, robotic vacuum cleaners, and cruise controls for cars, are still in their infancy. In the next few decades, Lampson predicts, medical science will develop prosthetic eyes and ears that will enable people to really see and hear; cars will drive themselves; sensors in our homes and bodies will continuously monitor our health and well-being; and we will have intelligent, robot personal assistants to help us both at work and at home.

For computer systems to achieve such engagement, Lampson believes, they will have to handle uncertainty and probability as well as they now handle facts:

> Probability is also essential, since the machine's model of the physical world is necessarily uncertain. We are just beginning to learn how to write programs that can handle uncertainty. They use the techniques of statistics, Bayesian inference and machine learning to combine models of the connections among random variables, both observable and hidden, with observed data to learn parameters of the models and then to infer hidden variables such as the location of vehicles on a road from observations such as the image data from a camera.[3]

In addition to managing uncertainty, many of the applications of embodiment will need to be much more dependable than today's computer systems. Computers driving cars or performing surgical procedures are obvious examples of applications in which reliability is critical for safety. We need methods for specifying the desired behavior of a computer program and proving that the resulting code actually fulfills these specifications. Today, such methods work only for small-scale systems, and there is still much research needed to scale these methods up to handle the large, critical safety applications of tomorrow.

In this chapter we will look at two key trends – the coming robotics revolution and the "Internet of Things" – and end the chapter with some words about consciousness and realistic neural networks.

Fig. 16.1. George Devol's original patent for the first programmable robotic arm in 1954 was the foundation for the modern robotics industry.

The rise of the robots

The word *robot* was introduced to the world by the Czech author Karel Čapek in his play *R.U.R.*, an abbreviation for Rossum's Universal Robots. The play, first performed in 1921, portrayed mass-produced artificial humans that could manufacture products much more cheaply than real people could. The word *robot* is derived from the Czech word *robota*, meaning *labor*. The theme of *R.U.R.* is now a familiar one in science fiction, a revolt by the robots. Another science fiction writer, Isaac Asimov, introduced the word *robotics* to mean the science and technology of robots. Robotics is now an established field of research and requires a combination of many different disciplines, ranging from mechanical engineering and power systems to computer vision and machine learning.

The first industrial robot was a far cry from the humanoid robots imagined by Čapek and Asimov. In 1954, George Devol (B.16.1) invented a static, immobile machine with a programmable arm. His patent on a device for "programmed article transfer" issued in 1961 laid the foundation for the modern robotics industry (Fig. 16.1). In his patent application he wrote, "The present invention makes available for the first time a more or less general purpose machine that has universal application to a vast diversity of applications where cyclic digital control is desired."[4]

Devol coined the phrase *universal automation* to describe a robot that could be programmed to perform a variety of tasks, and he called his first product the Unimate. The first Unimate machine was sold to General Motors in 1960. General Motors installed it at the auto-body plant in Ewing Township, New Jersey, to lift and stack hot pieces of metal from a die-casting machine. Devol's next product was a robotic arm for spot welding. The early industrial robots did not look anything like the humanoid robots of science fiction; many consisted of little more than a mechanical arm. Other automobile companies soon followed General Motors's lead, using robots to do jobs that were tedious, hard, or dangerous for people.


B.16.1. George Devol (1912–2011) was the inventor of robotic arm. This programmable device became very successful and revolutionized the manufacturing industry.

Around twenty thousand robots are now sold each year in North America. Although the U.S. automotive industry is still the dominant sector, sales also grew in the life sciences and pharmaceutical industry. Japan leads the world with an installed base of several hundred thousand industrial robots. These modern industrial robots are far more sophisticated than their early ancestors. On the new Tesla electric car assembly line in California, for example, at each station there are up to eight robots, some more than eight feet tall, each with a single arm with multiple joints, and each capable of multiple functions, such as welding, riveting, and bonding different components (see Fig. 16.2). Since

Fig. 16.2. Robots handling the delicate operation of glass panel unloading.

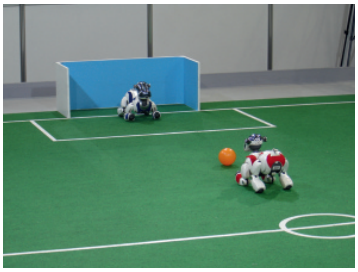


Fig. 16.3. Sony's doglike AIBO robots playing soccer at the 2005 RoboCup competition.

the introduction of the early industrial robots, the number and type of robots have increased dramatically. We will look at four examples: humanoid robots, robotic laboratories, driverless vehicles, and drones.

Japanese companies have pioneered the development of animal-like and humanoid robots. In the 1990s, after a suggestion by a Canadian researcher, Alan Mackworth, Japanese researchers in artificial intelligence (AI) started an annual soccer competition for robots called the Robot World Cup, or RoboCup (Fig. 16.3). The aim of the RoboCup was to promote robotics and AI research. Playing soccer requires the robots not only to move and act independently but also to collaborate and follow a team strategy to beat the opposing team. As the robots play, they have to process many different types of sensor input and make real-time decisions based on this input. In 1999, Sony Corporation produced the AIBO – Artificial Intelligence roBOt – a four-legged, doglike robot designed to serve as a household pet. Teams of AIBOs have regularly competed in the RoboCup.

In 2000, the Honda Motor Company produced a humanoid robot called ASIMO (Advanced Step in Innovative Mobility), an acronym chosen to give homage to Isaac Asimov (Fig. 16.4). The robot is about four feet high and can detect movements of objects and recognize distance and direction using two camera "eyes." ASIMO can also understand some voice commands and gestures, such as when a person offers to shake hands. In 2006, at the International Consumer Electronics Show in Las Vegas, Nevada, ASIMO demonstrated its ability to walk, run, and kick a football. Such experiments are not just research stunts. Because Japan has an aging population, robots of all sorts may serve as one possible way of assisting the elderly.

Fig. 16.4. Honda's ASIMO robot has appeared at conferences and toured the world since 2000.

The National Aeronautics and Space Association (NASA) uses robotic geologists for its exploration of the surface of Mars. The Mars Exploration Rovers – Spirit and Opportunity – landed on Mars in 2004 and examined rocks and soils to find out the role that water has played in the history of Mars. These robots could drive up to forty meters a day and carried a range of scientific instruments, including a panoramic camera, various types of spectrometers, magnets, a microscope, and an abrasion tool for scratching rock surfaces. Curiosity, a much more ambitious mobile robotic laboratory, successfully landed on Mars in August 2012 (see Fig. 16.5). The Curiosity rover is about three meters long and five times as heavy as the previous rovers. Unlike the earlier vehicles, Curiosity can gather samples of rocks and soil and distribute them to onboard analytical instruments. Its mission is to investigate whether conditions on Mars have ever supported microbial life.

Fig. 16.5. NASA rovers began the robotic exploration of Mars in 2004.

Fig. 16.6. Carnegie Mellon University's driverless vehicle Sandstorm competed in the 2004 and 2005 DARPA Grand Challenges.

The first Defense Advanced Research Projects Agency (DARPA) Grand Challenge, held in the United States in 2004, was a competition for driverless vehicles funded by DARPA. The goal was to successfully navigate a 150-mile course in the Mojave Desert in California. No team successfully completed the course in the first Grand Challenge. The vehicle that traveled farthest was Sandstorm, a converted Humvee built by a research team from Carnegie Mellon University. Sandstorm covered more than seven miles before it caught fire and ended up stranded on a rock. No prize was given that year, and the organizers scheduled a second event a year later. That time, five vehicles finished the course. The winner was Stanley, built by a team from Stanford University led by Sebastian Thrun. Stanley completed the course in about seven hours, closely followed by two entries from Carnegie Mellon, Sandstorm and Highlander, led by the roboticist Red Whittaker (Fig. 16.6). In 2007, DARPA organized a third driverless car competition, this time on a sixty-mile course called the Urban Challenge that required driving through inhabited areas. The robotic vehicles had to avoid other vehicles and obstacles in a crowded urban environment, obeying all traffic regulations. The challenge was won by Tartan Racing, a team from Carnegie Mellon University, driving a modified Chevy Tahoe SUV named Boss.

Fig. 16.7. The Northrop Grumman Global Hawk drone can fly at sixty thousand feet for flights as long as thirty hours.

The robotic vehicles most in the news are undoubtedly *unmanned aerial vehicles*, also called *drones* (Fig. 16.7). The military increasingly uses drones for surveillance and battlefield exploration. These military drones are typically large and expensive. Just as we saw the PC movement emerge from the hobbyist community, today we are seeing explosive growth of a low-cost "hobbyist" drone movement. A key ingredient for a drone is an autopilot. When autopilots were originally introduced in the 1930s, the control systems merely kept the aircraft level and flying on a preset course. Nowadays, autopilots can be used to automate the whole flight plan, as well as the takeoff and landing. What has changed in the last ten years is that all the components needed to construct an autopilot have become much smaller and cheaper. These devices include *gyroscopes* to measure rates of rotation; *magnetometers* to act as a digital compass; *barometric pressure sensors* to determine altitude; and

Fig. 16.8. The MeCam quadcoptor is a miniature drone that can follow and photograph you wherever you go and then upload the video images to your smart phone.

accelerometers to measure changes in motion. Chips with all these functionalities now cost less than $20. Similarly, the demand for smaller global positioning system chips to provide navigation systems in phones has brought the price down from thousands of dollars to as little as $10. Finally, the demand for better mobile phone cameras has led to the availability of cheap, powerful imaging chips. As a result, there is now a thriving do-it-yourself drone community. Hobbyist drone builders employ smart phone technology, including low-cost sensors, cameras, low-power processors, and batteries. Drone enthusiasts exchange information on DIYdrones.com, a website set up by Chris Anderson of *Wired* magazine (Fig. 16.8). The site lists a large number of nonmilitary, nonpolice uses of drones – including agriculture, search and rescue, home movies, coverage of sports events, environmental monitoring, and delivering medicines.

The Internet of Things

The growth of the Internet over the last thirty years has been dramatic. From connecting a few thousand computers at research centers, the Internet now connects billions of people through their personal computers, smart phones, and tablets. Yet this is only the first step in connectivity. We can now attach cheap electronic tags and sensors to objects and connect them to create an even larger global network called the "Internet of Things." The MIT engineer Kevin Ashton first used the term in 1999. In his original definition, he said:

> Today computers – and, therefore, the Internet – are almost wholly dependent on human beings for information. Nearly all of the roughly 50 petabytes (a petabyte is 1,024 terabytes [trillion bytes]) of data available on the Internet were first captured and created by human beings – by typing, pressing a record button, taking a digital picture or scanning a bar code.… The problem is, people have limited time, attention and accuracy – all of which means they are not very good at capturing data about things in the real world.… If we had computers that knew everything there was to know about things – using data they gathered without any help from us – we would be able to track and count everything, and greatly reduce waste, loss and cost. We would know when things needed replacing, repairing or recalling, and whether they were fresh or past their best. The Internet of Things has the potential to change the world, just as the Internet did. Maybe even more so.[5]

After the revolutions caused by the World Wide Web and mobile, networked, wireless devices, the Internet of Things represents the next disruptive technology on the horizon. With more than fifty billion objects predicted to be connected to the Internet by 2020, we will see a world in which everyday objects, such as books, cars, and refrigerators, can be interrogated for information. Some of the smart devices will not only use sensors to get information but also use *actuators*, devices that move or control things, to modify the environment. "Intelligent houses" will be able to check on their inhabitants; water drainage systems will know about storm threats and adjust accordingly; and businesses will actively monitor supply chains so that they no longer run out of stock or generate wasteful surpluses.

How will we be able to cope with the immense complexity of this new world? The computer entrepreneur Ray Ozzie has painted a vision of what he calls "a world of continuous services and connected devices":[6]

> To cope with the inherent complexity of a world of devices, a world of websites, and a world of apps and personal data that is spread across myriad devices and websites, a simple conceptual model is taking shape that brings it all together. We're moving toward a world of 1) cloud-based *continuous services* that connect us and do our bidding, and 2) appliance-like *connected devices* enabling us to interact with those cloud-based services.[7]

Cloud computing is a way of sharing computing resources by linking large numbers of computers and other devices over the Internet with massive data centers where huge amounts of information can be stored together with massive, on-demand, computational capacity. Websites will use these cloud resources to provide continuous services that are always available and can be scaled to meet any fluctuation in demand. These services will constantly gather and analyze data from both the real and online worlds. Users will interact with these services using "apps" – software applications – on a range of connected devices. Increasingly, as the Internet of Things grows, these devices will include many types of embedded systems, from webcams in our homes to sensors on our highways.

Strong AI and the mind-body problem

WARNING: In the remaining sections of this chapter, we enter into areas in which there is no clear consensus among researchers, computer scientists, neuroscientists, and philosophers. There are many different opinions and often little agreement even about definitions.

In their book *Artificial Intelligence*, Stuart Russell and Peter Norvig define the terms *weak AI* and *strong AI* as follows:

> The assertion that machines could act *as if* they were intelligent is called the *weak AI* hypothesis by philosophers, and the assertion that machines that do so are actually thinking (not just simulating thinking) is called the *strong AI* hypothesis.[8]

The proposal for John McCarthy's 1956 workshop that introduced the term *AI* confidently asserted that weak AI was possible, saying, "Every aspect of learning or any other feature of intelligence can be so precisely described that a machine can be made to simulate it."[9]

In their book, Russell and Norvig take the view that "intelligence is concerned mainly with rational action."[10] They introduce the idea of building intelligent systems in terms of *agents*, subsystems that can perceive their environment through sensors and can act on the environment through actuators. A rational agent is one that selects an action that maximizes its performance for every possible sequence of inputs. The agent can also learn from experience to improve its performance. Russell and Norvig identify different types of agents, including reflex agents, goal-based agents, and utility-based agents. *Reflex agents* respond only to their last input. *Goal-based agents* act to achieve a well-defined goal, and *utility-based agents* try to maximize some specific measure of performance. Such rational agent-based systems have had considerable success during the last twenty years in robotics, speech recognition, planning and scheduling, game playing, spam fighting, and machine translation, to list only a few examples. Because of such progress, Russell and Norvig declare:

> Most AI researchers take the weak AI hypothesis for granted, and don't care about the strong AI hypothesis – as long as their program works, they don't care whether you call it a simulation of intelligence or real intelligence.[11]

In spite of this very pragmatic approach from the majority of AI practitioners, intelligent machines have continued to be an active topic of discussion by philosophers since Alan Turing devised his Universal Turing Machine in 1950. Francis Crick, one of the discoverers of DNA, has said that the scientific study of the brain during the twentieth century has led to the acceptance of consciousness as a valid subject for scientific investigation. In his 1994 book *The Astonishing Hypothesis*, Crick suggests that "a person's mental activities are entirely due to the behavior of nerve cells, glial cells, and the atoms, ions and molecules that make up and influence them."[12] In other words, the human mind arises entirely from the actions of billions of neurons in the brain.

Ever since the days of Plato and Aristotle, philosophers have been concerned with the *mind-body* problem, which examines the relationship between mind and matter. René Descartes, in the seventeenth century, viewed the activity of thinking and the physical processes of the body as distinct – a philosophy known as *dualism*. By contrast, *monism* maintains that the mind and brain are not separate and that mental states are just physical states – a viewpoint sometimes described as *physicalism*.

Many philosophers and computer scientists are attracted to the idea of *functionalism*, in which a mental state is defined solely by its function – that is, its relation to sensory inputs, other mental states, and behavior. There are many varieties of functionalism, but we shall focus on Hilary Putnam's idea of *machine functionalism*, which makes an analogy between the states of a Turing machine and the mental states of the brain. As we have seen, the output of a Turing machine is determined by the initial state of the machine and the tape input. This is the basis of *computationalism*, the theory that mental states are just computational states and the transition from one mental state to another depends only on its inputs and is independent of the particular physical implementation. This viewpoint leads naturally to the question "Can machines think?" and to questions about strong AI.

The dominant trend in psychology in the first half of the twentieth century was an approach called *behaviorism*, championed by John Watson and B. F. Skinner. This movement maintained that psychology should be concerned only with observable behavior of people and animals and not with untestable, unobservable events that may or may not be taking place in their minds. Alan Turing's famous Turing Test, which we discussed in Chapter 13, is a behavioral test for intelligence. In a response to this type of intelligence test, in 1980 philosopher John Searle introduced his famous "Chinese room" experiment to show that behavior is not enough for understanding and strong AI. We introduced Searle's Chinese room experiment in Chapter 14 in the context of IBM's

Watson's victory on *Jeopardy!* In more detail, his thought experiment tests the thesis that human *cognition* – thinking, understanding, and feeling – is nothing more than computation. Searle's argument is very simple:

> Because it is impossible to know whether anyone else but myself cognizes (thinks, understands, feels) I cannot say whether that computer over there, successfully passing the Turing Test [TT] in Chinese, is really cognizing (i.e. understanding Chinese). However, because computation is implementation-independent – which means that every implementation of the same computer program should have the same properties, if the properties are truly just computational ones – then I, Searle, can become an implementation of the Chinese-TT-passing program too. Yet it is evident, even without doing the experiment, that if I did so, I would not be understanding Chinese: I would just be manipulating meaningless symbols, according to the formal rules I had memorized – squiggles and squoggles.… So whereas I still cannot say whether the TT-passing computer understands Chinese, I can say for sure that if it is understanding Chinese, it is not because it is implementing the right computer program. This is because I am implementing the same computer program, and I am definitely not understanding Chinese. So computationalism (strong AI) is false (or incomplete).[13]

The cognitive scientist Stevan Harnad (B.16.2) argues further that there is something else missing in the Chinese room, besides the question of Searle's understanding of Chinese:

> Not only does he not understand the meaning of the symbols he is manipulating, but he also cannot pick out their referents. If you ask him what "BanMa" ("zebra" in Chinese) means, he will not only say, correctly, that he has no idea (even though he has just got done stating, *in Chinese,* that "A BanMa looks like a striped horse"). But apart from the missing feeling of understanding, Searle *also cannot pick out the thing that BanMa refers to in the world*: the symbols are *ungrounded.* Grounding (for which you need more than computation – you need robotic sensorimotor interactions with the world, to learn what symbol refers to what object) is necessary, though not sufficient, for meaning. In addition, it also *feels like something* to mean (or understand) BanMa.[14]

B.16.2. Stevan Harnad was born in Budapest, Hungary. He currently holds a Canadian Research Chair in cognitive science at the Universite du Quebec in Montreal and is also professor of cognitive science at the University of Southampton, England. He has championed the need for a *Total Turing Test* for which the standard Turing Test is extended to include the computer's perceptual and manipulative abilities.

Harnad's introduction of the *symbol grounding problem* is based on the following argument. Computation is the manipulation of symbols based on the symbols' shapes, not their meanings. Computation alone does not and cannot connect symbols to their meanings, as it would have to do for computation to be cognition. Harnad believes that some of the symbols have to be grounded in the sensory and motor capacity to pick out their corresponding objects in the world. He has therefore proposed extending the traditional Turing Test to the *Total Turing Test*, which includes a test of the computer's perceptual and manipulative abilities that are not purely computational. From this point of view, Searle's Chinese room merely shows that computation alone is insufficient for cognition.

The British mathematician Roger Penrose has put forward another objection to strong AI and computationalism. Penrose's argument is based on the logician Kurt Gödel's finding that there are "nonalgorithmic truths," statements

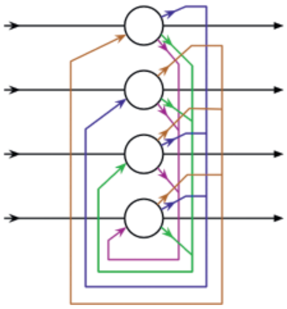


Fig. 16.9. A four-node Hopfield network with feedback loops.

that humans know to be true but that cannot be proved within a formal system based on a set of axioms. Penrose claims that this finding shows that computers, which can only operate by following algorithms, are therefore necessarily more limited than humans. This argument has been the subject of much debate by many people, including Turing. He observed that such results from mathematical logic could have implications for the Turing Test:

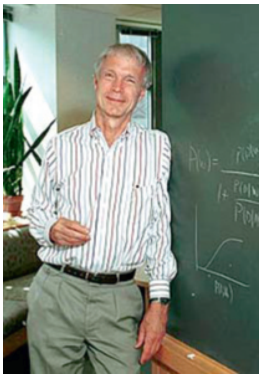
> There are certain things that [any digital computer] cannot do. If it is rigged up to give answers to questions as in the imitation game, there will be some questions to which it will either give a wrong answer, or fail to give an answer at all however much time is allowed for a reply.[15]

In the context of the Turing Test, the existence of such nonalgorithmic truths implies the existence of a class of "unanswerable" questions. However, Turing asserted that these questions are only a concern for the Turing Test if humans are able to answer the same questions.

Neural networks revisited

Rather than delve deeper into these hotly contested, largely philosophical issues, we shall look again at what the brain might tell us about intelligence and consciousness. We start with another look at neural networks. In the body, a neural network consists of interconnected nerve cells that work together, such as in the brain. In computer science, a neural network is a network of electronic components loosely modeled on the operation of the brain. As we have seen, the artificial neural networks (ANNs) described in Chapter 13 have successfully performed many pattern recognition tasks.

These ANNs, however, are very far from functioning like a realistic neural network in a living organism. Besides the huge difference in the numbers of neurons and connections, the primary element lacking is that of *feedback*, in which information is sent back into the system to adjust behavior. The favored method of training the ANN is *back propagation*, in which the initial output is compared to the desired output, and the system is adjusted until the difference between the two is minimized. However, the ANN was purely a *feed-forward* network that produced a specific output for each given set of inputs. In real brains, nerve cells not only feed forward but also send information back to other neurons.



B.16.3. John Hopfield was originally trained as a physicist but is most widely known for his research on ANNs. He was responsible for setting up the Computation and Neural Systems PhD program at Caltech and is now the Howard A. Prior Professor of Molecular Biology at Princeton.

An example of an ANN that allows feedback is the *Hopfield network* (Fig. 16.9), named after the multidisciplinary scientist John Hopfield (B.16.3). This network introduces bidirectional connections between the artificial neurons and assumes that the weights for each connection are the same in each direction. Such neural networks are able to function as *auto-associative memories* – that is, when a pattern of activity is presented to the network, the neurons and connections form a memory of this pattern. Even if you only input a part of the original pattern, the auto-associative memory can retrieve the entire original pattern. It is also possible to design these networks to store temporal sequences of patterns, capturing the order in time in which they occur. Feeding in only a part of this sequence generates the whole sequence, just as hearing the first few notes of a song brings back the whole song. The computer architect Jeff

Hawkins (B.16.4) has built on these ideas of neurons and memory and proposed an alternative model of the brain to the computationalist view discussed in the preceding text. We briefly outline some of the key ideas of his *memory-prediction theory* of intelligence.

B.16.4. Jeff Hawkins is a computer entrepreneur most known for his work on handheld computing devices such as the Palm Pilot and the Treo. He invented the handwriting character recognition system known as *Graffiti* for use with such devices. In addition to his successful career in the computer industry Hawkins has a deep interest in the function of the brain and wrote the book *On Intelligence* describing his *memory-prediction framework* of how the brain works.

Hawkins believes that any model of the brain and intelligence needs to incorporate neurons with feedback and be able to respond to rapidly changing streams of information. He focuses his attention on the architecture of the human cortex, the part of the brain responsible for higher functions, such as voluntary movement, learning, and memory. As we described in Chapter 15, the cortex is about 2.5 millimeters deep and is made up of six layers, each about as thick as a playing card. The cortex is estimated to contain around thirty billion neurons, each with thousands of connections making a total of more than thirty trillion synapses, the junctions at which nerve impulses pass from one neuron to another. Neurologists have found that the cortex consists of many different functional regions, each semiindependent and specialized for certain aspects of thought and perception. Each region is arranged in a hierarchy, with "lower" areas feeding information up the hierarchy and "higher" areas sending feedback back down toward the lower layers, although the terms *higher* and *lower* are not necessarily related to their physical arrangement in the brain. The lowest areas are the primary sensory areas, where sensory information arrives. The cortex has regions to process sensations from the eyes, the ears, and the skin and internal organs, and each region has its own hierarchies of regions. The cortex also has "association" areas where inputs from more than one sense can be combined. There is also a motor system in the frontal lobes of the brain that sends signals to the spinal cord and thus moves muscles. The hierarchies of all these sensory areas look very similar. This similarity led Vernon Mountcastle (B.16.5), a neuroscientist from Johns Hopkins University in Baltimore, to propose a model for the basic structure of the cortex in a paper titled "An Organizing Principle for Cerebral Function."

B.16.5. Vernon Mountcastle is Professor Emeritus at Johns Hopkins University. He is best known for his discovery of the columnar organization of the cerebral cortex in the 1950s. In 1978 he proposed that all parts of the cortex operate on a common principle based on these cortical columns.

In 1950, Mountcastle had discovered that the cortex was organized into vertical columns of neurons, with each column having a particular function. In 1978, he proposed that all parts of the cortex operate on a common principle, with the cortical column being the fundamental computational unit (Fig. 16.10). All the inputs from our primary sensory areas arrive at the cortex as patterns of partly chemical and partly electrical signals. We rely on our brains to make sense of this stream of data and to produce a consistent and stable view of the world. For example, several times a second, our eyes make sudden movements called *saccades*. With these saccades, the focus of our eyes moves around, locating interesting parts of the scene so that our brain can build up a three-dimensional model of what we are seeing (Fig. 16.11). Our impression of a stable world with objects and people moving in a continuous way is only possible because our brain has the processing capability to make sense of this continuous stream of changing retinal patterns (Fig. 16.12). Mountcastle speculated that all neurons in the cortex use the same basic algorithm to process the different input patterns arriving at the different sensory input regions – those for vision, hearing, language, motor control, touch, and so on. In other words, the brain processes patterns and constructs a model of the world that it then holds in memory made up of neurons and their synapses.

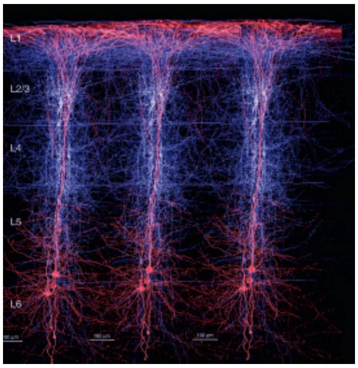


Fig. 16.10. Visualization of cortical columnar structure discovered by Vernon Mountcastle. Jeff Hawkins describes Mountcastle's 1978 paper proposing that all cortical columns operate on a common principle as the *rosetta stone* of neuroscience.

Up to now, we have considered Turing's behavioral test for intelligence as the basis for the computationalist model of the brain, which views the brain as a computer running programs. Hawkins makes two points in criticism of this viewpoint. First, there is what he calls the *input-output fallacy*: the behaviorist view that you present the brain with a given input and observe what output you get. In fact, when our brains receive input, we may indeed process that data and perform a visible action, but we do not necessarily respond that way. The input can just lead to thoughts that are not expressed in actions. Actions are optional, and this aspect of intelligence is not captured by a purely behavioral test. The second criticism concerns what Hawkins calls the *one hundred–step rule*. In the computer analogy of the brain, it is customary to contrast the one hundred billion neurons in a human brain with the few billion transistors on a chip. By contrast, a typical neuron takes about 5×10^{-3} seconds (5 milliseconds) to fire and reset compared to the cycle time for a modern chip which can be as short as 5×10^{-9} seconds, about a million times faster than a neuron. To account for the amazing power of our brain, given the relative slowness of the individual neurons, computationalists point to the fact that billions of neurons can be computing at the same time, as in a parallel computer that uses more than one CPU to execute a program, making it run faster. But consider the problem of looking at a photograph to determine whether there is a cat in the image. A human can pick out any cat in the photograph in less than a second. However, in that second, because neurons operate so slowly, the visual information entering the brain can only cross a chain of about a hundred neurons or so from the visual sensory input region. Thus the brain must "compute" its answer using only a tiny fraction of its billions of neurons. By contrast, to solve such a cat recognition problem on a digital computer would take many billions of steps.

Fig. 16.11. The human brain uses the tracks of saccades, sudden movements made by the eyes when observing a face, to build up a three-dimensional model of what the eyes see.

How can a brain perform a difficult recognition task in only a hundred steps that would take a supercomputer many billions of steps? Hawkins suggests:

> The answer is the brain doesn't "compute" the answers to problems; it retrieves the answers from memory. In essence, the answers were stored in memory a long time ago. It takes only a few steps to retrieve something from memory. Slow neurons are not only fast enough to do this, but they constitute the memory themselves. The entire cortex is a memory system. It isn't a computer at all.[16]

Fig. 16.12. A one-day-old baby, appropriately called Ada, is building up her picture of the world, three images per second, every second of her waking life.

Let us give one last example to show how the brain handles the task of catching a ball. Someone throws a ball toward you and less than a second later you catch it. If we want to program a robot arm to catch the same ball, the program requires an enormous amount of computation. First, you have to estimate the trajectory of the ball and calculate it numerically by solving Newton's laws of motion. This calculation tells you roughly where to position the robot arm to catch the ball. Because the first calculation of the ball's trajectory was only an estimate, the whole calculation needs to be repeated several times as the ball gets nearer. Finally, the fingers of the robotic arm need to be programmed to actually close around the ball when it arrives. To accomplish all this, a computer requires many millions of steps to solve the numerous mathematical equations involved. Yet our brain uses its neurons to catch the ball in only about a hundred steps. The brain clearly solves the problem a different way than relying on conventional computation. According to Hawkins, it uses memory:

> How do you catch the ball using memory? Your brain has a stored memory of the muscle commands required to catch a ball (along with many other learned behaviors). When a ball is thrown, three things happen. First, the appropriate memory is automatically recalled by the sight of the ball. Second, the memory actually recalls a temporal sequence of muscle commands. And third, the retrieved memory is adjusted as it is recalled to accommodate the particulars of the moment, such as the ball's actual path and the position of your body. The memory of how to catch a ball was not programmed into your brain; it was learned over years of repetitive practice, and it is stored, not calculated in your neurons.[17]

To account for the fact that the position of the ball needs to be constantly adjusted as the ball comes toward us, Hawkins uses the idea that the memories stored in the cortex are actually *invariant representations*. Artificial autoassociative memories can recall complete patterns when given only a partial image as input. But ANNs have a hard time recognizing a pattern if the pattern has been rescaled, rotated, or viewed from a different angle – a task our brains can handle with ease. If you are reading a book, you can change your position, rotate the book, or adjust the lighting, so that the visual input of the book to your brain can be constantly changing. Yet your brain knows that the book is the same, and its internal representation of "this book" does not change. The brain's internal representation is therefore called an *invariant representation*. The brain combines such invariant representations with changing data to make predictions of how to perform tasks, such as catching a ball.

Our understanding of the world is tied to our ability to make such predictions. Our brain receives a constant stream of patterns from the outside world, stores them as memories, and makes predictions by combining what it has seen before with the incoming stream of information. Hawkins says:

B.16.6. Daniel Dennett is a philosopher and cognitive scientist who has written popular books on evolution and consciousness – *Darwin's Dangerous Idea* and *Consciousness Explained*. He is codirector of the Center for Cognitive Studies at Tufts University in Massachusetts.

> Thus intelligence and understanding started as a memory system that fed predictions into the sensory stream. These predictions are the essence of understanding. To know something means that you can make predictions about it.[18]

This idea is the basis of Hawkins's *memory-prediction framework* of intelligence: "Prediction not behavior is proof of intelligence."[19] According to this view of intelligence, intelligent machines could be built that have just the equivalent of a cortex and a set of input sensors. There is no need to connect to the emotional systems of the other, older regions of the brain. Such intelligent systems will not resemble the humanoid robots of science fiction but would be able to develop an understanding of their world and make intelligent predictions. However, the technical challenges of building such systems in silicon still remain formidable, both in terms of the number of neurons required and their vast connectivity requirements.

B.16.7. Christof Koch was a professor of neuroscience at Caltech and, since 2011, the Chief Scientific Officer at the Allen Institute for Brain Science. During the 1990s, Koch collaborated with the late Nobel Prize recipient Francis Crick on the problem of consciousness as a scientifically addressable problem. He and Crick co-authored the 2004 book *The Quest for Consciousness: A Neurobiological Approach.*

Consciousness?

In their discussions of consciousness, philosophers often introduce the idea of a "zombie" as a category of imaginary human being. The philosopher Daniel Dennett (B.16.6) says:

> According to common agreement among philosophers, a zombie would be a human being who exhibits perfectly natural, alert, loquacious, vivacious behavior but is in fact not conscious at all, but rather some sort of automaton. The whole point of the philosopher's notion of a zombie is that you can't tell a zombie from a normal person by examining external behaviors.[20]

Philosophers also frequently introduce the idea of *qualia*, the plural of *quale*, into their discussions of consciousness. Neuroscientist Christof Koch (B.16.7) explains the concept of qualia as follows:

> What it feels like to have a particular experience is the quale of that experience: The quale of the color red is what is common to such disparate percepts as seeing a red sunset, the red flag of China, arterial blood, a ruby gemstone, and Homer's wine-dark sea. The common denominator of all these subjects is "redness." Qualia are the raw feelings, the elements that make up any one conscious experience.[21]

In his attempt to move the debate about consciousness from a philosophical level to a legitimate topic for scientific investigation, Koch introduces four different definitions of consciousness:

> A *commonsense definition* equates consciousness with our inner, mental life …
>
> A *behavioral definition* of consciousness constitutes a checklist of actions or behaviors that would certify as conscious any organism that could do one or more of them …
>
> A *neuronal definition* of consciousness specifies the minimal physiologic mechanisms required for any one conscious sensation …
>
> A *philosopher* [will] give you a fourth definition, "consciousness is what it is like to feel something."[22]

However, Dennett, in his book *Consciousness Explained*, takes a different approach and explicitly abandons the arguments and debates about qualia and avoids this concept in his own discussion of consciousness.

It is the ability to be self-aware that probably embodies what most people think is the essence of consciousness. Nevertheless, as we have seen, there is still a long way to go before computer scientists, cognitive scientists, and neuroscientists are ready to reach a consensus about strong AI, the mind-body problem, and consciousness. As philosopher Dennett has said, "Human consciousness is just about the last surviving mystery."[23]

Key concepts

- Humanoid robots
- Unmanned aerial vehicles

- Cloud computing
- The mind-body problem
- Strong AI
- Agents
- Functionalism
- Computationalism
- Behaviorism
- Symbol grounding problem
- Back propagation
- Auto-associative memories
- One hundred-step rule
- Consciousness