Chapter 15
Artificial Intelligence

Key Topics

Turing Test
Searle's Chinese Room
Philosophical Problems in AI
Cognitive Psychology
Linguistics
Logic and AI
Robots
Cybernetics
Neural Networks
Expert Systems

15.1 Introduction

The long-term[1] goal of artificial intelligence is to create a thinking machine that is intelligent, has consciousness, has the ability to learn, has free will and is ethical. The field involves several disciplines such as philosophy, psychology, linguistics, machine vision, cognitive science, mathematics, logic and ethics. Artificial intelligence is a young field, and the term was coined by John McCarthy and others in 1956. Alan Turing had earlier devised the Turing test as a way to test the intelligent behaviour of a machine. There are deep philosophical problems in artificial intelligence, and some researchers believe that its goals are impossible or incoherent. These views are shared by Hubert Dreyfus and John Searle. Even if

[1] This long-term goal may be hundreds or even thousands of years.

G. O'Regan, *A Brief History of Computing*, DOI 10.1007/978-1-4471-2359-0_15,

artificial intelligence is possible, there are moral issues to consider such as the exploitation of artificial machines by humans and whether it is ethical to do this. Weizenbaum[2] has argued that artificial intelligence is unethical.

One of the earliest references to creating life by artificial means is that of the classical myth of Pygmalion. Pygmalion was a sculptor who carved a woman out of ivory. The sculpture was so realistic that he fell in love with it and offered the statue presents and prayed to Aphrodite, the goddess of love. Aphrodite took pity on him and brought the statue (Galathea) to life.

There are several stories of attempts by man to create life from inanimate objects, for example, the creation of the monster in Mary Shelly's Frankenstein. The monster is created by an overambitious scientist who is punished for his blasphemy of creation (in that creation is for God alone). The monster feels rejected following creation and inflicts a horrible revenge on its creator.

There are stories of the creation of the golem in Prague dating back to sixteenth century. A golem was created from mud by a holy person who was able to create life. However, the life that the holy person could create would always be a shadow of that created by God. Golems became servants to the holy men but were unable to speak.

The most famous golem story involved Rabbi Judah Loew of Prague. He is said to have created a golem from clay on the banks of the Vltava River to defend the Jews of Prague from attack. The golem was brought to life by the Rabbi by incantations in Hebrew. The golem became violent over time and started killing people in Prague. The Rabbi destroyed the golem by erasing the first letter of the word '*emet*' from the golem's forehead to make the Hebrew word '*met*', meaning death.

The story of the golem was given a more modern version in the Czech play 'Rossum's Universal Robots'. This science fiction play by Capek appeared in Prague in 1921. It was translated into English and appeared in London in 1923. It contains the first reference to the term 'robot', and the play considers the exploitation of artificial workers in a factory. The robots (or androids) are initially happy to serve humans but become unhappy with their existence over a period of time. The fundamental questions that the play is considering are whether the robots are being exploited, and, if so, is this ethical, and what should be the response of the robots to their exploitation. It eventually leads to a revolt by the robots and the extermination of the human race.

The story of Pinocchio as a fictional character first appeared in 1883 in the 'Adventures of Pinocchio' by Carlo Collodi. He was carved as a wooden puppet by a woodcarver named Geppetto, but he dreamt of becoming a real boy. Pinocchio's nose became larger whenever he told a lie.

[2] Weizenbaum was a psychologist who invented the ELIZA program. This program simulated a psychologist in dialogue with a patient. He was initially an advocate of artificial intelligence but later became a critic.

15.2 René Descartes

René Descartes was an influential French mathematician, scientist and philosopher. He was born in a village in the Loire valley in France in 1596 and studied law at the University of Poitiers. He never practised as a lawyer and instead served Prince Maurice of Nassau in the Netherlands. He invented the Cartesian coordinate system that is used in plane geometry and algebra. In this system, each point on the plane is identified through two numbers: the x-coordinate and the y-coordinate (Fig. 15.1).

He made important contributions to philosophy and attempted to derive a fundamental set of principles which can be known to be true. His approach was to renounce any idea that could be doubted. He rejected the senses since they can deceive and are not a sound source of knowledge. For example, during a dream, the subject perceives stimuli that appear to be real, but these have no existence outside the subject's mind. Therefore, it is inappropriate to rely on one's senses as the foundation of knowledge.

He argued that a powerful 'evil demon or mad scientist' could exist who sets out to manipulate and deceive subjects, thereby preventing them from knowing the true nature of reality. The evil demon could bring the subject into existence including an implanted memory. The question is how one can know for certain what is true given the limitations of the senses. The '*brain in the vat thought experiment*' is a more modern formulation of the idea of an evil spirit or mad scientist. A mad scientist could remove a person's brain from their body and place it in a vat and connects its neurons by wires to a supercomputer. The computer provides the disembodied brain with the electrical impulses that the brain would normally receive. The computer could then simulate reality, and the disembodied brain would have conscious experiences and would receive the same impulses as if it were inside a person's skull. There is no way to tell whether the brain is inside the vat or inside a person.

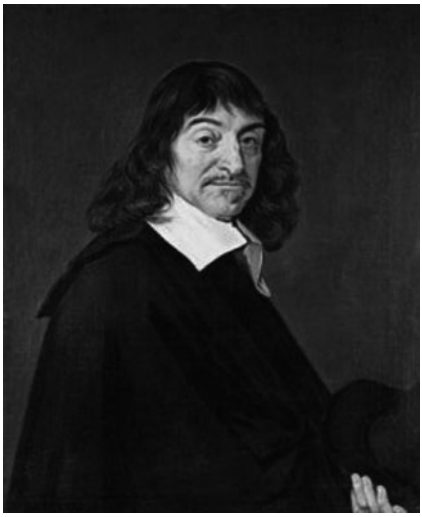

Fig. 15.1 René Descartes

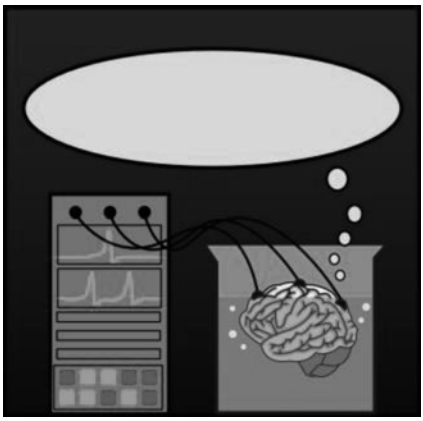

Fig. 15.2 Brain in a VAT thought experiment

That is, at any moment, an individual could potentially be a brain connected to a sophisticated computer program or inside a person's skull. Therefore, if you cannot be sure that you are not a brain in a vat, then you cannot rule out the possibility that all of your beliefs about the external world are false. This sceptical argument is difficult to refute.

The perception of a 'cat' in the case where the brain is in the vat is false and does not correspond to reality. It is impossible to know whether your brain is in a vat or inside your skull; it is therefore impossible to know whether your belief is valid or not (Fig. 15.2).

From this, Descartes deduced that there was one single principle that must be true. He argued that even if he is being deceived, then clearly he is thinking and must exist. This principle of existence or being is more famously known as 'cogito, ergo sum' (I think, therefore I am). Descartes argued that this existence can be applied to the present only, as memory may be manipulated and therefore doubted. Further, the only existence that he sure of is that he is a '*thinking thing*'. He cannot be sure of the existence of his body as his body is perceived by his senses which he has proven to be unreliable. Therefore, his mind or thinking thing is the only thing about him that cannot be doubted. His mind is used to make judgements and to deal with unreliable perceptions received via the senses.

Descartes constructed a system of knowledge from this one principle using the deductive method. He deduced the existence of a benevolent God using the ontological argument. He argues [Des:99] that we have an innate idea of a supremely perfect being (God), and that God's existence may be inferred immediately from the innate idea of a supremely perfect being:

1. I have an innate idea of a supremely perfect being (i.e. God).
2. Necessarily, existence is a perfection.
3. Therefore, God exists.

He then argued that since God is benevolent that he can have some trust in the reality that his senses provide. God has provided him with a thinking mind and does

not wish to deceive him. He argued that knowledge of the external world can be obtained by both perception and deduction, and that reason or rationalism is the only reliable method of obtaining knowledge. His proof of the existence of God and the external world are controversial.

Descartes was a *dualist,* and he makes a clear *mind-body* distinction. He states that there are two substances in the universe: mental substances and bodily substances. The mind-body distinction is very relevant in AI, and the analogy of the human mind and brain is software running on a computer.

This thinking thing (*res cogitans* or mind/soul) was distinct from the rest of nature (*res extensa*) and interacted with the world through the senses to gain knowledge. Knowledge was gained by mental operations using the deductive method, where starting from the premises that are known to be true, further truths could be logically deduced. Descartes founded what would become known as the Rationalist school of philosophy where knowledge was derived solely by human reasoning. The *analogy of the mind in AI would be an AI program running on a computer* with knowledge gained by sense perception with sensors and logical deduction.

Descartes believed that the bodies of animals are complex living machines without feelings. He dissected (including vivisection) many animals for experiments. Vivisection has become controversial in recent years. His experiments led him to believe that the actions and behaviour of non-human animals can be fully accounted for by mechanistic means, and without reference to the operations of the mind. He realised from his experiments that a lot of human behaviour (e.g. physiological functions and blinking) is like that of animals in that it has a mechanistic explanation.

Descartes was of the view that well-designed automata[3] could mimic many parts of human behaviour. He argued that the key differentiators between human and animal behaviour were that humans could adapt to widely varying situations and also had the ability to use language. The use of language illustrates the power of the use of thought, and it clearly differentiates humans from animals. Animals do not possess the ability to use language for communication or reason. This, he argues, provides evidence for the presence of a soul associated with the human body. In essence, animals are pure machines, whereas humans are machines with minds (or souls).

The significance of Descartes in the field of artificial intelligence is that the Cartesian dualism that humans seem to possess would need to be reflected amongst artificial machines. Humans seem to have a distinct sense of 'I' as distinct from the body, and the 'I' seems to represent some core sense or essence of being that is unchanged throughout the person's life. It somehow represents personhood, as distinct from the physical characteristics of a person that are inherited genetically.

[3] An automaton is a self-operating machine or mechanism that behaves and responds in a mechanical way.

The challenge for the AI community in the longer term is to construct a machine that (in a sense) possesses Cartesian dualism. That is, the long-term[4] goal of AI is to produce a machine that has awareness of itself as well as its environment.

15.3 The Field of Artificial Intelligence

The origin of the term 'artificial intelligence' is in work done on the proposal for Dartmouth Summer Research Project on Artificial Intelligence. This proposal was written by John McCarthy and others in 1955, and the research project took place in the summer of 1956.

The success of early AI went to its practitioners' heads, and they believed that they would soon develop machines that would emulate human intelligence. They convinced many of the funding agencies and the military to provide research grants as they believed that real artificial intelligence would soon be achieved. They had some initial (limited) success with machine translation, pattern recognition and automated reasoning. However, it is now clear that AI is a long-term project. Artificial intelligence is a multidisciplinary field and includes disciplines such as:

- Computing
- Logic and philosophy
- Psychology
- Linguistics
- Neuroscience and neural networks
- Machine vision
- Robotics
- Expert systems
- Machine translation
- Epistemology and knowledge representation

The British mathematician, Alan Turing, contributed to the debate concerning thinking machines, consciousness and intelligence in the early 1950s [Tur:50]. He devised the famous 'Turing test' to judge whether a machine was conscious and intelligent. Turing's paper was very influential as it raised the idea of the possibility of programming a computer to behave intelligently.

Shannon considered the problem of writing a chess program in the late 1940s and distinguished between a brute force strategy where the program could look at every combination of moves or a strategy where knowledge of chess could be used to select and examine a subset of available moves. The ability of a program to play

[4] This long-term goal may be hundreds of years as there is unlikely to be an early breakthrough in machine intelligence as there are deep philosophical problems to be solved. It took the human species hundreds of thousands of years to evolve to its current levels of intelligence.

Fig. 15.3 John McCarthy
(Courtesy of John McCarthy)

chess is a skill that is considered intelligent, even though the machine itself is not conscious that it is playing chess.

Modern chess programs have been quite successful and have advantages over humans in terms of computational speed in considering combinations of moves. Kasparov was defeated by the IBM chess program 'Deep Blue' in a six-game match in 1997. The result of the match between Kasparov and X3D Fritz in 2003 was a draw.

Herbert Simon and Alan Newell developed the first theorem prover with their work on a program called 'Logic Theorist' or 'LT' [NeS:56]. This program could independently provide proofs of various theorems[5] in Russell's and Whitehead's Principia Mathematica [RuW:10]. It was demonstrated at the Dartmouth conference and showed that computers had the ability to encode knowledge and information and to perform intelligent operations such as solving theorems in mathematics.

John McCarthy proposed a program called the Advice Taker in his influential paper 'Programs with Common Sense' [Mc:59]. The idea was that this program would be able to draw conclusions from a set of premises, and McCarthy states that a program has common sense if it is capable of automatically deducing for itself 'a sufficiently wide class of immediate consequences of anything it is told and what it already knows' (Fig. 15.3).

The Advice Taker uses logic to represent knowledge (i.e. premises that are taken to be true), and it then applies the deductive method to deduce further truths from the relevant premises.[6] That is, the program manipulates the formal language

[5] Russell is said to have remarked that he was delighted to see that the Principia Mathematica could be done by machine, and that if he and Whitehead had known this in advance, they would not have wasted 10 years doing this work by hand in the early twentieth century. The LT program succeeded in proving 38 of the 52 theorems in Chap. 2 of Principia Mathematica. Its approach was to start with the theorem to be proved and to then search for relevant axioms and operators to prove the theorem.

[6] Of course, the machine would somehow need to know what premises are relevant and should be selected for to apply the deductive method from the many premises that are already encoded.

(e.g. predicate logic) and provides a conclusion that may be a statement or an imperative. McCarthy envisaged that the Advice Taker would be a program that would be able to learn and improve. This would be done by making statements to the program and telling it about its symbolic environment. The program will have available to it all the logical consequences of what it has already been told and previous knowledge. McCarthy's desire was to create programs to learn from their experience as effectively as humans do.

The McCarthy philosophy is that commonsense knowledge and reasoning can be formalised with logic. A particular system is described by a set of sentences in logic. These logic sentences represent all that is known about the world in general and what is known about the particular situation and the goals of the systems. The program then performs actions that it infers are appropriate for achieving its goals. That is, commonsense[7] knowledge is formalised by logic and commonsense problems are solved by logical reasoning.

15.3.1 Turing Test and Strong AI

Turing contributed to the debate concerning artificial intelligence in his 1950 paper on computing, machinery and intelligence [Tur:50]. Turing's paper considered whether it could be possible for a machine to be conscious and to think. He predicted that it would be possible to speak of machines thinking, and he devised a famous experiment that would allow a computer to be judged as a conscious and thinking machine. This is known as the 'Turing test'. The test itself is an adaptation of a well-known party game which involves three participants. One of them, the judge, is placed in a separate room from the other two: one is a male and the other is a female. Questions and responses are typed and passed under the door. The objective of the game is for the judge to determine which participant is male and which is female. The male is allowed to deceive the judge whereas the female is supposed to assist.

Turing adapted this game by allowing the role of the male to be played by a computer. The test involves a judge who is engaged in a natural language conversation with two other parties, one party is a human and the other is a machine. If the judge cannot determine which is machine and which is human, then the machine is said to have passed the 'Turing test'. That is, a machine that passes the Turing test must be considered intelligent, as it is indistinguishable from a human. The test is applied to test the linguistic capability of the machine rather than the audio capability, and the conversation is limited to a text-only channel.

[7] Common sense includes basic facts about events, beliefs, actions, knowledge and desires. It also includes basic facts about objects and their properties.

Turing's work on 'thinking machines' caused a great deal of public controversy as defenders of traditional values attacked the idea of machine intelligence. It led to a debate concerning the nature of intelligence. There has been no machine developed, to date, that has passed the Turing test.

Turing strongly believed that machines would eventually be developed that would stand a good chance of passing the 'Turing test'. He considered the operation of 'thought' to be equivalent to the operation of a discrete state machine. Such a machine may be simulated by a program that runs on a single, universal machine, that is, a computer.

Turing's viewpoint that a machine will one day pass the Turing test and be considered intelligent is known as '*strong artificial intelligence*'. It states that a computer with the right program would have the mental properties of humans. There are a number of objections to strong AI, and one well-known rebuttal is that of Searle's Chinese room argument [Sea:80].

15.3.1.1 Strong AI

The computer is not merely a tool in the study of the mind, rather the appropriately programmed computer really *is* a mind in the sense that computers given the right programs can be literally said to *understand* and have other cognitive states [Searle's 1980 Definition].

15.3.1.2 Weak AI

Computers just *simulate* thought, their seeming understanding is not real understanding (just as-if), their seeming calculation is only as-if calculation, etc. Nevertheless, computer simulation is useful for *studying* the mind (as for studying the weather and other things).

The Chinese room thought experiment was developed by John Searle to refute the feasibility of the strong AI project. It rejects the claim that a machine has or will someday have the same cognitive qualities as humans. Searle argues that brains cause minds and that syntax does not suffice for semantics.

A man is placed into a closed room into which Chinese writing symbols are input to him. He is given a rulebook that shows him how to manipulate the symbols to produce Chinese output. He has no idea as to what each symbol means, but with the rulebook, he is able to produce the Chinese output. This allows him to communicate with the other person and appear to understand Chinese. The rulebook allows him to answer any questions posed, without the slightest understanding of what he is doing or what the symbols mean.

1. Chinese characters are entered through slot 1.
2. The rulebook is employed to construct new Chinese characters.
3. Chinese characters are outputted to slot 2.

The question 'Do you understand Chinese?' could potentially be asked, and the rulebook would be consulted to produce the answer 'Yes, of course' despite of the fact that the person inside the room has not the faintest idea of what is going on. It will appear to the person outside the room that the person inside is knowledgeable on Chinese. The person inside is just following rules without understanding.

The process is essentially that of a computer program which has an input, performs a computation based on the input and then finally produces an output. Searle has essentially constructed a machine which can never be mental. Changing the program essentially means changing the rulebook, and this does not increase understanding. The strong artificial intelligence thesis states that given the right program, *any* machine running it would be mental. However, Searle argues that the program for this Chinese room would not understand anything and that therefore the strong AI thesis must be false. In other words, Searle's Chinese room argument is a rebuttal of strong AI by showing that a program running on a machine that appears to be intelligent has no understanding whatsoever of the symbols that it is manipulating. That is, given any rulebook (i.e. program), the person would never understand the meanings of those characters that are manipulated.

That is, just because the machine acts like it knows what is going on, it actually only knows what it is programmed to know. It differs from humans in that it is not aware of the situation like humans are. It suggests that machines may not have intelligence or consciousness, and the Chinese room argument applies to any Turing equivalent computer simulation.

There are several rebuttals of Searle's position,[8] and one well-known rebuttal attempt is the 'System Reply' argument. This reply argues that if a result associated with intelligence is produced, then intelligence must be found somewhere in the system. The proponents of this argument draw an analogy between the human brain and its constituents. None of its constituents have intelligence, but the system as a whole (i.e. the brain) exhibits intelligence. Similarly, the parts of the Chinese room may lack intelligence, but the system as a whole is intelligence.

15.4 Philosophy and AI

Artificial intelligence includes the study of knowledge and the mind, and there are deep philosophical problems (e.g. the nature of mind, consciousness and knowledge) to be solved.

[8] I don't believe that Searle's argument proves that strong AI is impossible. However, I am not expecting to see intelligent machines anytime soon.

Early work on philosophy was done by the Greeks as they attempted to understand the world and the nature of being and reality. Thales and the Miletians[9] attempted to find an underlying principle that would explain the nature of the world. Pythagoras believed that mathematics was the basic principle, and that everything (e.g. music) could be explained in terms of number. Plato distinguished between the world of appearances and the world of reality. He argued that the world of appearances resembles the flickering shadows on a cave wall, whereas reality is in the world of ideas[10] or forms, in which objects of this world somehow participate. Aristotle proposed the framework of a substance which includes form plus matter. For example, the matter of a wooden chair is the wood that it is composed of, and its form is the general form of a chair.

Descartes had a significant influence on the philosophy of mind and AI. Knowledge is gained by mental operations using the deductive method. This involves starting from premises that are known to be true and deriving further truths. He distinguished between the mind and the body (Cartesian dualism), and the analogy of the mind is an AI program running on a computer with sensors and logical deduction used to gain knowledge.

British empiricism rejected the Rationalist position and stressed the importance of empirical data in gaining knowledge about the world. It argued that all knowledge is derived from sense experience. It included philosophers such as Locke, Berkeley[11] and Hume. Locke argued that a child's mind is a blank slate (tabula rasa) at birth and that all knowledge is gained by sense experience. Berkeley argued that the ideas in a man's mind have no existence outside his mind [Ber:99], and this philosophical position is known as idealism.[12] David Hume formulated the standard empiricist philosophical position in 'An Enquiry concerning Human Understanding' [Hum:06] (Fig. 15.4).

[9] The term 'Miletians' refers to inhabitants of the Greek city state Miletus which is located in modern Turkey. Anaximander and Anaximenes were two other Miletians who made contributions to early Greek philosophy approx 600 B.C.

[10] Plato was an idealist, that is, that reality is in the world of ideas rather than the external world. Realists (in contrast) believe that the external world corresponds to our mental ideas.

[11] Berkeley was an Irish philosopher (not British). He was born in Dysart castle in Kilkenny, Ireland; educated at Trinity College, Dublin; and served as bishop of Cloyne in Co. Cork. He planned to establish a seminary in Bermuda for the sons of colonists in America, but the project failed due to lack of funding from the British government. Berkeley University in San Francisco is named after him.

[12] Berkeley's theory of ontology is that for an entity to exist, it must be perceived, that is, '*Esse est percipi*'. He argues that 'It is an opinion strangely prevailing amongst men, that houses, mountains, rivers, and in a world all sensible objects have an existence natural or real, distinct from being perceived'.

This led to a famous Limerick that poked fun at Berkeley's theory. 'There once was a man who said God; Must think it exceedingly odd; To find that this tree, continues to be; When there is no one around in the Quad'.

The reply to this Limerick was appropriately: 'Dear sir, your astonishments odd; I am always around in the Quad; And that's why this tree will continue to be; Since observed by, yours faithfully, God'.

Fig. 15.4 George Berkely. Bishop of Cloyne

Fig. 15.5 David Hume

Hume argued that all objects of human knowledge may be divided into two kinds: *matters of fact* propositions that are based entirely on experience or *relation of ideas* propositions that may be demonstrated (such as geometry) via deduction reasoning in the operations of the mind. He argued that any subject[13] proclaiming knowledge that does not adhere to these empiricist principles should be committed to the flames[14] as such knowledge contains nothing but sophistry and illusion (Fig. 15.5).

[13] Hume argues that these principles apply to subjects such as Theology as its foundations are in faith and divine revelation which are neither matters of fact nor relations of ideas.

[14] 'When we run over libraries, persuaded of these principles, what havoc must we make? If we take in our hand any volume; of divinity or school metaphysics, for instance; let us ask, *Does it contain any abstract reasoning concerning quantity or number?* No. *Does it contain any experimental reasoning concerning matter of fact and existence?* No. Commit it then to the flames: for it can contain nothing but sophistry and illusion'.

Kant's Critique of Pure Reason [Kan:03] was published in 1781 and is a response to Hume's theory of empiricism. Kant argued that there is a third force in human knowledge that provides concepts that cannot be inferred from experience. Such concepts include the laws of logic (e.g. modus ponens), causality and so on, and Kant argued that the third force was the manner in which the human mind structures its experiences. These structures are called categories.

The continental school of philosophy included thinkers such as Heidegger and Merleau-Ponty who argued that the world and the human body are mutually intertwined. Heidegger emphasised that existence can only be considered with respect to a changing world. Merleau-Ponty emphasised the concept of a body-subject that actively participates both as the perceiver of knowledge and as an object of perception.

Philosophy has been studied for over two millennia, but to date, very little progress has been made in solving its fundamental questions. However, it is important that it be considered as any implementation of AI will make philosophical assumptions, and it is important that these be understood.

15.5 Cognitive Psychology

Psychology arose out of the field of psychophysics in the late nineteenth century with work by German pioneers in attempting to quantify perception and sensation. Fechner's mathematical formulation of the relationship between stimulus and sensation is given by

$$S = k \log I + c$$

The symbol S refers to the intensity of the sensation, the symbols k and c are constants and the symbol I refers to the physical intensity of the stimulus. Psychology was defined by William James as the science of mental life.

One of the early behavioural psychologist's was Pavlov who showed that it was possible to develop a conditional reflex in a dog. He showed that it is possible to make a dog salivate in response to the ringing of a bell. This is done by ringing a bell each time before meat is provided to the dog, and the dog therefore associates the presentation of meat with the ringing of the bell after a training period.

Skinner developed the concept of conditioning further using rewards to reinforce desired behaviour and punishment to discourage undesired behaviour. Positive reinforcement helps to motivate the individual to behave in the desired way, with punishment used to deter the individual from performing undesired behaviour. The behavioural theory of psychology explains many behavioural aspects of the world. However, it does not really explain complex learning tasks such as language development.

Merleau-Ponty[15] considered the problem of what the structure of the human mind must be in order to allow the objects of the external world to exist in our minds in the form that they do. He built upon the theory of phenomenology as developed by Hegel and Husserl. Phenomenology involves a focus and exploration of phenomena with the goal of establishing the essential features of experience. Merleau-Ponty introduced the concept of the body-subject which is distinct from the Cartesian view that the world is just an extension of our own mind. He argued that the world and the human body are mutually intertwined. The Cartesian view is that the self must first be aware of and know its own existence prior to being aware of and recognising the existence of anything else.

The body has the ability to perceive the world, and it plays a double role in that it is both the subject (i.e. the perceiver) and the object (i.e. the entity being perceived) of experience. Human understanding and perception is dependent on the body's capacity to perceive via the senses and its ability to interrogate its environment. Merleau-Ponty argued that there is a symbiotic relationship between the perceiver and what is being perceived, and he argues that as our consciousness develops, the self imposes richer and richer meanings on objects. He provides a detailed analysis of the flow of information between the body-subject and the world.

Cognitive psychology is a branch of psychology that is concerned with learning, language, memory and internal mental processes. Its roots lie in Piaget's child development psychology and in Wertheimer's Gestalt psychology. The latter argues that the operations of the mind are holistic and that the mind contains a self-organising mechanism. Holism argues that the sum of the parts is less than the whole and is the opposite of logical atomism[16] developed by Bertrand Russell. Russell (and also Wittgenstein) attempted to identify the atoms of thought, that is, the elements of thought that cannot be divided into smaller pieces. Logical atomism argues was that all truths are ultimately dependent on a layer of atomic facts. It had an associated methodology whereby by a process of analysis, it attempted to construct more complex notions in terms of simpler ones.

Cognitive psychology was developed in the late 1950s and is concerned with how people understand, diagnose and solve problems, as well as the mental processes that take place during a stimulus and its corresponding response. It argues that solutions to problems take the form of rules, heuristics and sudden insight, and it considers the mind as having a certain conceptual structure. The dominant paradigm in the field has been the *information processing model*, which considers the mental processes of thinking and reasoning as being equivalent to software

[15] Merleau-Ponty was a French philosopher who was strongly influenced by the phenomenology of Husserl. He was also closely associated with the French existentialist philosophers such as Jean-Paul Sartre and Simone De Beauvoir.

[16] Atomism actually goes back to the work of the ancient Greeks and was originally developed by Democritus and his teacher Leucippus in the fifth century B.C. Atomism was rejected by Plato in the dialogue the Timaeus.

running on the computer, that is, the brain. It has associated theories of input, representation of knowledge, processing of knowledge and output.

Cognitive psychology has been applied to artificial intelligence from the 1960s, and some of the research areas include:

- Perception
- Concept formation
- Memory
- Knowledge representation
- Learning
- Language
- Grammar and linguistics
- Thinking
- Logic and problem solving

It is clear that for a machine to behave with intelligence, it will need to be able to perceive objects in the physical world. It must be able to form concepts and to remember knowledge that it has already been provided with. It will need an understanding of temporal events. Knowledge must be efficiently represented to allow easy retrieval for analysis and decision making. An intelligent machine will need the ability to produce and understand written or spoken language. A thinking machine must be capable of thought, analysis and problem solving.

15.6 Linguistics

Linguistics is the theoretical and applied study of language, and human language is highly complex. It includes the study of phonology, morphology, syntax, semantics and pragmatics. Syntax is concerned with the study of the rules of grammar, and the syntactically valid sentences and phrases are formed by applying the rules of the grammar. Morphology is concerned with the formation and alteration of words, and phonetics is concerned with the study of sounds and how sounds are produced and perceived as speech (or non-speech).

Computational linguistics is an interdisciplinary study of the design and analysis of natural language processing systems. It includes linguists, computer scientists, experts in artificial intelligence, cognitive psychologists and mathematicians.

Early work on computational linguistics commenced with machine translation work in the United States in the 1950s. The objective was to develop an automated mechanism by which Russian language texts could be translated directly into English without human intervention. It was naively believed that it was only a matter of time before automated machine translation would be done.

However, the initial results were not very successful, and it was realised that the automated processing of human languages was considerably more complex. This led to the birth of a new field called computational linguistics, and the objective of this field is to investigate and develop algorithms and software for

processing natural languages. It is a subfield of artificial intelligence and deals with the comprehension and production of natural languages.

The task of translating one language into another requires an understanding of the grammar of both languages. This includes an understanding of the syntax, the morphology, semantics and pragmatics of the language. For artificial intelligence to become a reality, it will need to make major breakthroughs in computational linguistics.

15.7 Cybernetics

The interdisciplinary field of cybernetics[17] began in the late 1940s when concepts such as information, feedback and regulation were generalised from engineering to other systems. These include systems such as living organisms, machines, robots and language. The term '*cybernetics*' was coined by Norbert Wiener, and it was taken from the Greek word 'κυβερνητη' (meaning steersman or governor). It is the study of communications and control and feedback in living organisms and machines to ensure efficient action.

The name is well chosen as a steersman needs to respond to different conditions and feedback while steering the boat to achieve the goal of travel to a particular destination. Similarly, the field of cybernetics is concerned with the interaction of goals, predictions, actions, feedback and responses in all kinds of systems. It uses models of organisations, feedback and goals to understand the capacity and limits of any system.

It is concerned with knowledge acquisition through control and feedback. Its principles are similar to human knowledge acquisition where learning is achieved through a continuous process of feedback from parents and peers which leads to adaptation and transformation of knowledge rather than its explicit encoding.

The conventional belief in AI is that knowledge may be stored inside a machine, and that the application of stored knowledge to the real world in this way constitutes intelligence. External objects are mapped to internal states on the machine, and machine intelligence is manifested by the manipulation of the internal states. This approach has been reasonably successful with rule-based expert systems, but it has made limited progress in creating intelligent machines. Therefore, alternative approaches such as cybernetics warrant further research. Cybernetics views information (or intelligence) as an attribute of an interaction, rather than something that is stored in a computer.

[17] Cybernetics was defined by Couffignal (one of its pioneers) as the art of ensuring the efficacy of action.

15.8 Logic and AI

Mathematical logic is used in the AI field to formalise knowledge and reasoning. Commonsense reasoning is required for solving problems in the real world, and therefore McCarthy [Mc:59] argues that it is reasonable for logic to play a key role in the formalisation of commonsense knowledge. This includes the formalisation of basic facts about actions and their effects, facts about beliefs and desires and facts about knowledge and how it is obtained. His approach allows commonsense problems to be solved by logical reasoning.

Its formalisation requires sufficient understanding of the commonsense world, and often the relevant facts to solve a particular problem are unknown. It may be that knowledge thought relevant may be irrelevant and vice versa. A computer may have millions of facts stored in its memory, and the problem is how to determine the relevant facts from its memory to serve as premises in logical deduction.

McCarthy influential 1959 paper discusses various commonsense problems such as getting home from the airport. Other examples are diagnosis, spatial reasoning and understanding narratives that include temporal events. Mathematical logic is the standard approach to express premises, and it includes rules of inferences that are used to deduce valid conclusions from a set of premises. It provides a rigorous definition of deductive reasoning showing how new formulae may be logically deduced from a set or premises.

Propositional calculus associates a truth value with each proposition and includes logical connectives to produce formulae such as $A \Rightarrow B$, $A \wedge B$ and $A \vee B$. The truth values of the propositions are normally the binary values of *true* and *false*. There are other logics, such as three-valued logic or fuzzy logics that allow more than two truth values for a proposition. Predicate logic is more expressive than propositional logic and includes quantifiers and variables. It can formalise the syllogism 'All Greeks are mortal; Socrates is a Greek; Therefore, Socrates is mortal'. The predicate calculus consists of:

- Axioms
- Rules for defining well-formed formulae
- Inference rules for deriving theorems from premises

A formula in predicate calculus is built up from the basic symbols of the language. These include variables, predicate symbols such as equality; function symbols, constants logical symbols such as $\exists$, $\wedge$, $\vee$ and $\neg$ and punctuation symbols such as brackets and commas. The formulae of predicate calculus are built from terms, where a *term* is defined recursively as a variable or individual constant or as some function containing terms as arguments. A formula may be an atomic formula or built from other formulae via the logical symbols.

There are several rules of inference associated with predicate calculus, and the most important of these are modus ponens and generalisation. The rule of modus ponens states that given two formulae p, and $p \Rightarrow q$, we may deduce q. The rule of generalisation states that given a formula p that we may deduce $\forall(x)p$.

McCarthy's approach to programs with common sense has been criticised by Bar-Hillel and others on the grounds that common sense is fairly elusive and the difficulty that a machine would have in determining which facts are relevant to a particular deduction from its known set of facts. However, logic remains an important area in AI.

15.9 Computability, Incompleteness and Decidability

An algorithm (or procedure) is a finite set of unambiguous instructions to perform a specific task. The term 'algorithm' is named after the Persian mathematician Al-Khwarizmi. The concept of an algorithm was defined formally by Church in 1936 and independently by Turing. Church defined computability in terms of the lambda calculus, and Turing defined computability in terms of the theoretical Turing machine. These formulations are equivalent.

Formalism was proposed by Hilbert as a foundation for mathematics in the early twentieth century. A formal system consists of a formal language, a set of axioms and rules of inference. Hilbert's program was concerned with the formalisation of mathematics (i.e. the axiomatisation of mathematics) together with a proof that the axiomatisation was consistent. The specific objectives of Hilbert's program were to:

- Develop a formal system where the truth or falsity of any mathematical statement may be determined
- A proof that the system is consistent (i.e. that no contradictions may be derived)

A proof in a formal system consists of a sequence of formulae, where each formula is either an axiom or derived from one or more preceding formulae in the sequence by one of the rules of inference. Hilbert believed that every mathematical problem could be solved and therefore expected that the formal system of mathematics would be complete (i.e. all truths could be proved within the system) and decidable, that is, that the truth or falsity of any mathematical proposition could be determined by an algorithm. However, Church and Turing independently showed this to be impossible in 1936, and the only way to determine whether a statement is true or false is to try to solve it.

Russell and Whitehead published *Principia Mathematica* in 1910, and this three-volume work on the foundations of mathematics attempted to derive all mathematical truths in arithmetic from a well-defined set of axioms and rules of inference. The questions remained whether the Principia was *complete* and *consistent*. That is, is it possible to derive all the truths of arithmetic in the system, and is it possible to derive a contradiction from the Principia's axioms?

Gödel's second incompleteness theorem [Goe:31] showed that first-order arithmetic is incomplete, and that the consistency of first-order arithmetic cannot be proved within the system. Therefore, if first-order arithmetic cannot prove its own consistency, then it cannot prove the consistency of any system that contains first-order arithmetic. These results dealt a fatal blow to Hilbert's program.

15.10 Robots

The first use of the term 'robot' was by the Czech playwright Karel Capek in his play '*Rossum's Universal Robots*' performed in Prague in 1921. The word 'robot' is from the Czech word for forced labour. The theme explored is whether it is ethical to exploit artificial workers in a factory and what response the robots should make to their exploitation. Capek's robots were not mechanical or metal in nature and were instead created through chemical means. Capek rejected the idea that machines created from metal could think or feel.

Asimov wrote several stories about robots in the 1940s including the story of a robotherapist.[18] He predicted the rise of a major robot industry, and he also introduced a set of rules (or laws) for good robot behaviour. These are known as the three Laws of Robotics, and a fourth law was later added by Asimov (Table 15.1).

The term 'robot' is defined by the Robot Institute of America as:

Definition 15.1 (Robots). *A re-programmable, multifunctional manipulator designed to move material, parts, tools, or specialized devices through various programmed motions for the performance of a variety of tasks.*

Joseph Engelberger and George Devol are considered the fathers of robotics. Engelberger set up the first manufacturing company 'Unimation' to make robots, and Devol wrote the necessary patents. Their first robot was called the 'Unimate'. These robots were very successful and reliable and saved their customer's (General Motors) money by replacing staff with machines. The robot industry continues to play a major role in the automobile sector.

Robots have been very effective at doing clearly defined repetitive tasks, and there are many sophisticated robots in the workplace today. These are industrial manipulators that are essentially computer-controlled 'arms and hands'. However, fully functioning androids are many years away.

Robots can also improve the quality of life for workers as they can free human workers from performing dangerous or repetitive jobs. Further, it leads to work for robot technicians and engineers. Robots provide consistently high-quality products

Table 15.1 Laws of robotics

Law	Description
Law 0	A robot may not injure humanity or, through inaction, allow humanity to come to harm
Law 1	A robot may not injure a human being or, through inaction, allow a human being to come to harm, unless this would violate a higher order law
Law 2	A robot must obey orders given by human beings, except where such orders would conflict with a higher order law
Law 3	A robot must protect its own existence as long as such protection does not conflict with a higher order law

[18] The first AI therapist was the ELIZA program produced by Weizenbaum in the mid-1960s.

and can work tirelessly 24 h a day. This helps to reduce the costs of manufactured goods thereby benefiting consumers. They do not require annual leave but will, of course, from time to time require servicing by engineers or technicians. However, there are impacts on workers whose jobs are displaced by robots.

15.11 Neural Networks

The term 'neural network' refers to an interconnected group of processing elements called nodes or neurons. These neurons cooperate and work together to produce an output function. Neural networks may be artificial or biological. A biological network is part of the human brain, whereas an artificial neural network is designed to mimic some properties of a biological neural network. The processing of information by a neural network is done in parallel rather than in series.

A unique property of a neural network is fault tolerance; that is, it can still perform (within certain tolerance levels) its overall function even if some of its neurons are not functioning. Neural network may be trained to learn to solve complex problems from a set of examples. These systems may also use the acquired knowledge to generalise and solve unforeseen problems.

A biological neural network is composed of billions of neurons (or nerve cells). A single neuron may be physically connected to thousands of other neurons, and the total number of neurons and connections in a network may be enormous. The human brain consists of many billions of neurons, and these are organised into a complex intercommunicating network. The connections are formed through axons[19] to dendrites,[20] and the neurons can pass electrical signals to each other. These connections are not just the binary digital signals of *on* or *off*, and, instead, the connections have varying strength which allows the influence of a given neuron on one of its neighbours to vary from very weak to very strong.

That is, each connection has an individual weight (or number) associated with it that indicates its strength. Each neuron sends its output value to all other neurons to which it has an outgoing connection. The output of one neuron can influence the activations of other neurons causing them to fire. The neuron receiving the connections calculates its activation by taking a weighted sum of the input signals. Networks learn by changing the weights of the connections. Many aspects of brain function, especially the learning process, are closely associated with the adjustment of these connection strengths. Brain activity is represented by particular patterns of firing activity amongst the network of neurons. This simultaneous

[19] These are essentially the transmission lines of the nervous system. They are microscopic in diameter and conduct electrical impulses. The axon is the output from the neuron, and the dendrites are input.

[20] Dendrites extend like the branches of a tree. The origin of the word dendrite is from the Greek word (δενδρον) for tree.

cooperative behaviour of a huge number of simple processing units is at the heart of the computational power of the human brain.[21]

Artificial neural networks aim to simulate various properties of biological neural networks. They consist of many hundreds of simple processing units which are wired together in a complex communication network. Each unit or node is a simplified model of a real neuron which fires[22] if it receives a sufficiently strong input signal from the other nodes to which it is connected. The strength of these connections may be varied in order for the network to perform different tasks corresponding to different patterns of node firing activity. The objective is to solve a particular problem, and artificial neural networks have been applied to speech recognition problems, image analysis and so on.

The human brain employs massive parallel processing whereas artificial neural networks have provided simplified models of the neural processing that takes place in the brain. The largest artificial neural networks are tiny compared to biological neural networks. The challenge for the field is to determine what properties individual neural elements should have to produce something useful representing intelligence.

Neural networks are quite distinct from the traditional von Neumann architecture which is based on the sequential execution of machine instructions. The origins of neural networks lie in the attempts to model information processing in biological systems. This relies more on parallel processing as well as on implicit instructions based on pattern recognition from sense perceptions of the external world.

The nodes in an artificial neural network are composed of many simple processing units which are connected into a network. Their computational power depends on working together (parallel processing) on any task, and computation is related to the dynamic process of node firings rather than sequential execution of instructions. This structure is much closer to the operation of the human brain and leads to a computer that may be applied to a number of complex tasks.

15.12 Expert Systems

An expert system is a computer system that contains domain knowledge of one or more human experts in a narrow specialised domain. It consists of a set of rules (or knowledge) supplied by the domain experts about a specific class of problems and allows knowledge to be stored and intelligently retrieved. The effectiveness of the expert system is largely dependent on the accuracy of the rules provided, as incorrect inferences will be drawn with incorrect rules. Several commercial expert systems have been developed since the 1960s.

Expert systems have been a success story in the AI field. They have been applied to the medical field, equipment repair and investment analysis. They employ

[21] The brain works through massive parallel processing.

[22] The firing of a neuron means that it sends off a new signal with a particular strength (or weight).

Table 15.2 Expert systems

Component	Description
Knowledge base	The knowledge base is represented as a set of rules of form (IF condition THEN action)
Inference engine	Carries out reasoning by which expert system reaches conclusion
Explanatory facility	Explains how a particular conclusion was reached
User interface	Interface between user and expert system
Database/memory	Set of facts used to match against IF conditions in knowledge base

a logical reasoning capability to draw conclusions from known facts as well as recommending an appropriate course of action to the user.

An expert system consists of the following components: a knowledge base, an inference engine, an explanatory facility, a user interface and a database (Table 15.2).

Human knowledge of a specialty is of two types: namely, public knowledge and private knowledge. The former includes the facts and theories documented in text books and publications, whereas the latter refers to knowledge that the expert possesses that has not found its way into the public domain. The latter often consists of rules of thumb or heuristics that allow the expert to make an educated guess where required, as well as allowing the expert to deal effectively with incomplete or erroneous data. It is essential that the expert system encodes public and private knowledge to enable it to draw valid inferences.

The inference engine is made up of many inference rules that are used by the engine to draw conclusions. Rules may be added or deleted without affecting other rules, and this reflects the normal updating of human knowledge. Out of date facts may be deleted and are no longer used in reasoning, while new knowledge may be added and applied in reasoning. The inference rules use reasoning which is closer to human reasoning. There are two main types of reasoning with inference rules, and these are backward chaining and forward chaining. Forward chaining starts with the data available and uses the inference rules to draw intermediate conclusions until a desired goal is reached. Backward chaining starts with a set of goals and works backwards to determine if one of the goals can be met with the data that is available.

The expert system makes its expertise available to decision makers who need answers quickly. This is extremely useful as often there is a shortage of experts and the availability of an expert computer with in-depth knowledge of specific subjects is therefore very attractive. Expert systems may also assist managers with long-term planning. There are many small expert systems available that are quite effective in a narrow domain. The long-term goal is to develop expert systems with a broader range of knowledge. Expert systems have enhanced productivity in business and engineering, and there are several commercial software packages available to assist.

Several expert systems (e.g. Mycin, Colossus and Dendral) have been developed. Mycin was developed at Stanford University in the 1970s. It was written in LISP and was designed to diagnose infectious blood diseases and to recommend appropriate antibiotics and dosage amounts corresponding to the patient's body weight. It had a knowledge base of approximately 500 rules and a fairly simple

inference engine. Its approach was to query the physician running the program with a long list of yes/no questions. Its output consisted of various possible bacteria that could correspond to the blood disease, along with an associated probability that indicated the confidence in the diagnosis. It also included the rationale for the diagnosis and a course of treatment appropriate to the diagnosis.

Mycin had a correct diagnosis rate of 65%. This was better than the diagnosis of most physicians who did not specialise in bacterial infections. However, its diagnosis rate was less than that of experts in bacterial infections who had a success rate of 80%. Mycin was never actually used in practice due to legal and ethical reasons on the use of expert systems in medicine. For example, if the machine makes the wrong diagnosis, who is to be held responsible?

Colossus was an expert system used by several Australian insurance companies. It was used to help insurance adjusters assess personal injury claims, and helped to improve consistency, objectivity and fairness in the claims process. It guides the adjuster through an evaluation of medical treatment options, the degree of pain and suffering of the claimant and the extent that there is permanent impairment to the claimant, as well as the impact of the injury on the claimant's lifestyle. It was developed by Computer Sciences Corporation (CSC).

Dendral (Dendritic Algorithm) was developed at Stanford University in the mid-1960s, and it was the first use of artificial intelligence in medical research. Its objectives were to assist the organic chemist with the problem of identifying unknown organic compounds and molecules by computerised spectrometry. This involved the analysis of information from mass spectrometry graphs and knowledge of chemistry. Dendral automated the decision-making and problem-solving process used by organic chemists to identify complex unknown organic molecules. It was written in LISP, and it showed how an expert system could employ rules, heuristics and judgement to guide scientists in their work.

15.13 Review Questions

1. Discuss Descartes and his Rationalist philosophy and his relevance to artificial intelligence.
2. Discuss the Turing test and its relevance on strong AI. Discuss Searle's Chinese room rebuttal arguments. What are your own views on strong AI?
3. Discuss the philosophical problems underlying artificial intelligence.
4. Discuss the applicability of logic to artificial intelligence.
5. Discuss neural networks and their applicability to artificial intelligence.
6. Discuss expert systems.

15.14 Summary

Artificial intelligence is a multidisciplinary field, and its branches include logic and philosophy, psychology, linguistics, machine vision, neural networks and expert systems.

Turing believed that machine intelligence was achievable, and he devised the 'Turing test' to judge whether a machine was intelligent. Searle's Chinese room argument is a rebuttal of strong AI and aims to demonstrate that a machine will never have the same cognitive qualities as a human even if it passes the Turing test.

McCarthy proposed programs with commonsense knowledge and reasoning formalised with logic. He argued that human-level intelligence may be achieved with a logic-based system.

Cognitive psychology is concerned with cognition and some of its research areas include perception, memory, learning, thinking and logic and problem solving. Linguistics is the scientific study of language and includes the study of syntax and semantics.

Artificial neural networks aim to simulate various properties of biological neural networks. They consist of many hundreds of simple processing units which are wired together in a complex communication network. Each unit or node is a simplified model of a real neuron which fires if it receives a sufficiently strong input signal from the other nodes to which it is connected. The strength of these connections may be varied in order for the network to perform different tasks corresponding to different patterns of node firing activity. Artificial neural networks have been successfully applied to speech recognition problems and to image analysis.

An expert system is a computer system that allows knowledge to be stored and intelligently retrieved. It is a program that is made up of a set of rules (or knowledge). These rules are generally supplied by the domain experts about a specific class of problems. They include a problem-solving component that allows analysis of the problem to take place, as well as recommending an appropriate course of action to solve the problem.