Robotics Topic 2 – Things that think

Prepared for the module team by Jeffrey Johnson and Tony Hirst, updated by David McDade, Jon Rosewell and Daniel Gooch

1 Introduction

Welcome to Topic 2. Having very briefly looked at how machines move and sense in Topic 1, we will now look in more detail at the implications of what it means to add intelligence to machines.

The sense–act robot model provides us with a good starting point for thinking about robotic systems. The concept behind this model is that machines receive information about their environment and surroundings which then determines how they will act. We will then expand on this model with the sense–think–act model, which differs from the sense–act model in that robots act *deliberatively*; that is, they 'think' about what to do before performing an action. We will also look at other consequences of machines having to learn, other cognitive aspects such as reasoning and conflict resolution, and also the more physical attributes of robots.

This topic follows the same pattern of work you met in Topic 1.

1.1 Learning outcomes

By the end of Topic 2 you should be able to:

- explain how the sense–act model can be applied to robots
- explain the place of cognition in the sense–think–act model
- explain how new facts can be deduced from existing facts
- explain what it means for a robot to plan and learn its actions.

2 Sensing, acting and thinking

2.1 The sense–act model

The sense–act model provides a basic concept for the operation of robots. It can be described in terms of 1. *perception* (sensing) and 2. *actuation* (doing) subsystems

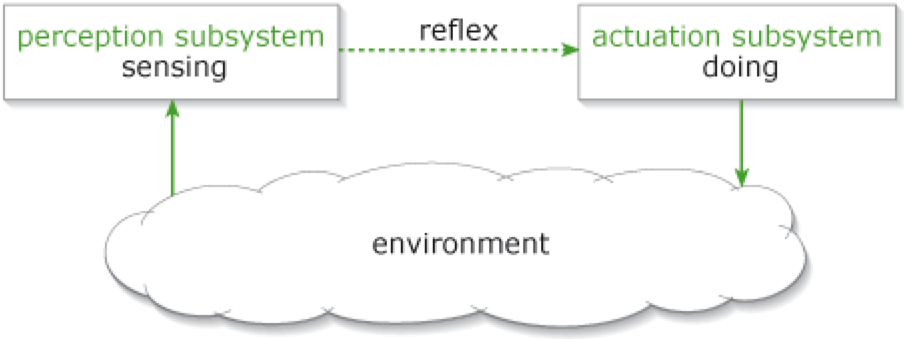


Figure 2.1 The sense–act model

Show description

1. **The perception subsystem**

   People commonly consider themselves as having five basic senses (sight, sound, touch, taste and smell) which form part of what is referred to as our perception subsystem. These senses receive information from a range of sensory organs in our bodies – namely the eyes, ears, touch receptors, taste buds and olfactory receptors. Correspondingly, a robot receives sensory information about its physical environment from *sensors*. Robot designers can choose from a wide range of sensors; you will come across several of them in this block.

2. **The actuation subsystem**

   The actuation subsystem is the means by which the machine interacts with its environment, generally changing its relationship with the environment (e.g. moving to a new location), and sometimes changing its environment (such as picking up an object). A special part of the actuation subsystem, known as the *mobility subsystem*, is used to move the robot.

Whenever a robot does something in, or to, its environment, we say that it has performed an *action*. An example of an action is when a robot moves itself from one location to another. Alternatively, another action may be moving an object with a manipulator arm. Using the sensors of its perception subsystem, the robot detects these changes, which starts the cycle all over again.

Reflexes and reaction

Sometimes we perform an action or series of actions without thinking, such as pulling one's hand away from very hot water or jumping out of the path of a speeding bicycle. These so-called *reflex actions* can be surprising since we do not need to consciously think about doing them. Reflex actions are 'hard wired' into parts of our bodies; a sensor input passes directly to the actuators without being processed in the brain. This saves time, and possibly our life, in dangerous situations. Reflex actions such as these fit the sense–act model described above. Robots are said to be built to the sense–act model if they react directly to sensor information.

2.1.1 Grey Walter's tortoises

The neuroscientist W. Grey Walter from the Burden Neurological Institute in Bristol was a pioneer in developing sense–act robots. Between 1948 and 1949, Grey Walter programmed the first robots capable of using perception and actuation; he called them Elmer and Elsie.

Grey Walter is increasingly recognised as a visionary in the development of robotics. Many of the approaches he advocated over 70 years ago remain close to the cutting edge of current research. Figure 2.2 shows a version of Grey Walter's tortoise robot, Elsie, with a modern transparent shell. The robot is driven and steered by the front wheel. The robot has a light sensor and the shell is connected to a touch sensor.

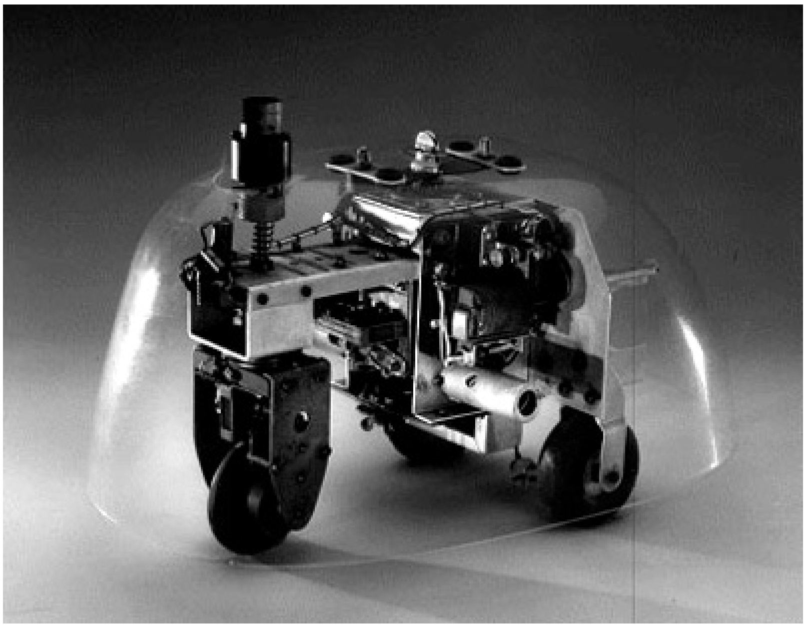


Figure 2.2 Grey Walter's tortoise robot, Elsie

Show description

In a 1948 radio interview, Grey Walter discussed future plans for his tortoises:

> I've been planning to make just such a plaything and when it is finished I challenge anyone to tell whether or not it is living, without prolonged observation. In shape it would be rather like a tortoise. [You] could tell that it disliked cold, damp weather, because it would moan pitifully at temperatures below 40°, and humidities above 80%. What is more, it would run towards any source of warmth, but would avoid great heat and bright lights. In the evening, it would come out from under the sofa and sit by the fire, or nestle against your leg, but it would scuttle back to its hiding place at a loud sound or sudden movement and could only be enticed out again by a low whistle. When the door bell rang, it would potter out into the hall to look at the visitor, but would have an aversion to men in blue and a liking for silk stockings. It would recognise its master's voice and learn a few simple tricks, such as coming to heel, and begging. At intervals it would become petulant and whine and would nose around for an electric power socket into which it would plug its antennae to recharge its batteries. While feeding in this way it would resent interference by yelping and giving the intruder a mild electric shock. It would need an occasional drink of distilled water, and might I'm afraid sometimes make little messes in corners. It would rest for part of the day, when alone, and would only be active in company when it would tend to emit a rhythmic purring sound; but if you teased it by too many contradictory stimuli it would become neurotic and sulk for days, or even break down altogether. The most difficult thing to incorporate in our mechanical pet [would be] the power to reproduce.
>
> There is no doubt that even relatively simple machines often display unpredicted properties, and do useful things that their inventors never expected them to do, as well as useless ones of course. This sort of accidental originality is sometimes quite eerie and awe-inspiring.
>
> (*Mind and Machines*, 1948)

Dr Owen Holland of Essex University, a leading expert on Grey Walter, writes:

> An idea that had interested Grey Walter for some time, and that he probably first heard from Craik, was that a control system operating in a complex world would have to be able to represent that complexity internally … It was clear that the brain itself, with its vast number of neurons, was enormously complex, and so it was easy to assume that it was capable of representing the external world adequately because of its sheer size. From the beginning, Grey Walter realised the impossibility of building a device with anything like the number of functional units that the brain appeared to possess, and so he began to wonder if there was any way in which internal complexity could be achieved using relatively small numbers of components. By 1948 he thought he had found a suitable method, and he had begun to wonder if the brain used this method, rather than relying on its immense number of components.
>
> (Holland, 2001)

Grey Walter's idea focused on connections in the brain. He observed that whilst the brain has a very large number of cells (known as *neurons*), it has a much greater number of connections between neurons: by some estimates, a human brain contains 100 billion (thousand million) neurons, but it contains 100 trillion (million million) connections. This led him to the idea of connecting light and touch sensor outputs via some intermediate circuitry to the motor actuator controlling the tortoise robots. In this way he demonstrated that the behaviour of the machine depends on the way its sensors and actuators are connected. His robots therefore perfectly fit the sense–act model of behaviour.

Activity 2.1 Watch

Watch this clip about Grey Walter's robots.

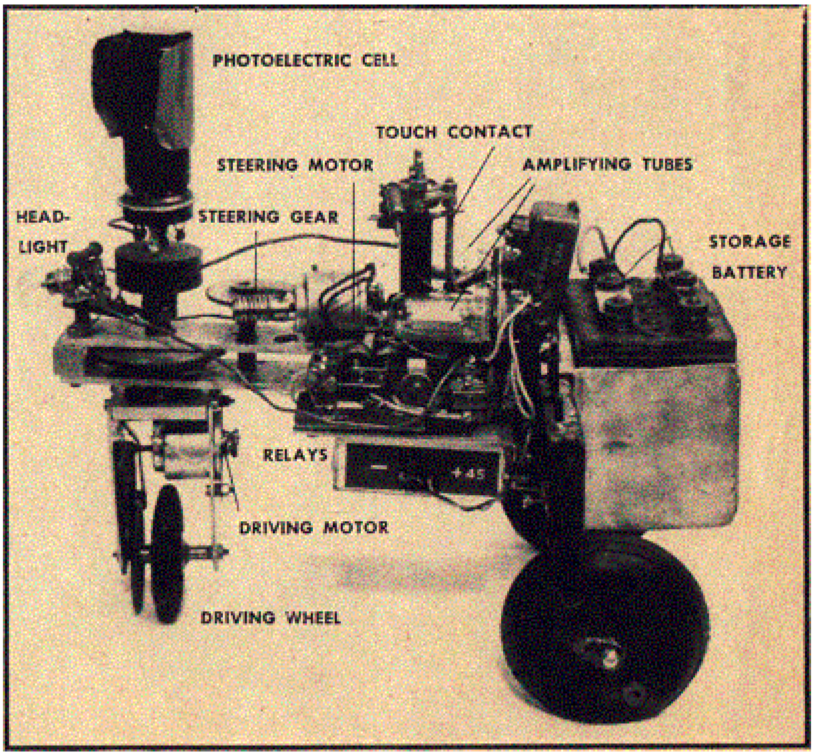


Figure 2.3 An annotated view of Grey Walter's tortoise

Show description

The construction of Grey Walter's tortoises can be seen in Figure 2.3. Figure 2.4 shows how the sensors were connected to the motors. In scan mode the light sensor and front wheel would rotate as the robot moved. On detecting a light source, the robot would move towards it. When the robot got too close to the light source it would be dazzled, and a reflex would make it move away again. The diagram shows how the sensors pass through 'neural' circuitry to control the front wheel speed and direction.

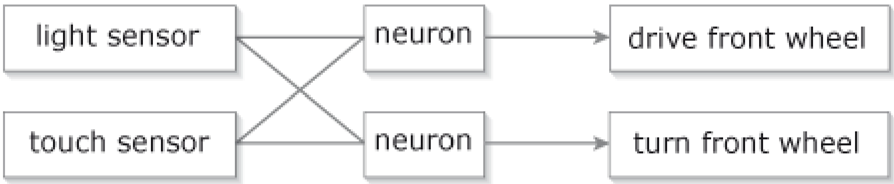


Figure 2.4 Connections between sensors and actuators in Grey Walter's tortoise

Show description

2.1.2 Braitenberg's vehicles

In the 1980s, the Italian researcher Valentino Braitenberg was exploring the sense–act model without being aware of Grey Walter's work, which was by then more than 30 years old. In his book, *Vehicles: Experiments in Synthetic Psychology* (Braitenberg, 1984), Braitenberg presented a series of so-called 'vehicles' which would use the sense–act model to interact with the world. Whilst Braitenberg described how these vehicles would behave, he did not actually build any robots; however, subsequent researchers have created simulated or real versions of his vehicles to demonstrate the rich behaviour that can be achieved from a very simple control system.

One of Braitenberg's basic ideas is illustrated in Figure 2.5. A light sensor is wired directly to each of two motorised wheels. The idea is that the more intense the sensor reading, the faster the wheel to which it is attached rotates. Thus, if the intensity is the same for both sensors, the vehicle moves forward.

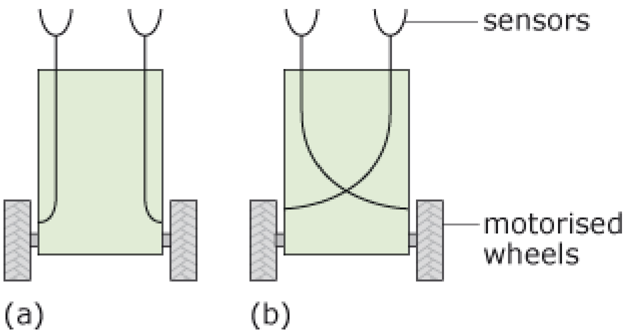


Figure 2.5 Braitenberg vehicles: (a) light avoider (b) light seeker

Show description

However, if there is a bright light to the left of robot (a) then it turns away from the light because the left wheel rotates faster than the right wheel. Similarly, if there is a bright light to the right of the robot then it turns away from the light because the right wheel rotates faster than the left. So, robot (a) has lightavoiding behaviour.

If there is a bright light to the left of robot (b) then it turns towards the light because the right wheel rotates faster than the left wheel. Similarly, if there is a bright light to the right of the robot then it turns towards the light because the left wheel rotates faster than the right wheel. So robot (b) has light-seeking behaviour.

The behaviour of Braitenberg's robots clearly emerges from the way the sensors and the motors are connected. There is no logical process or reasoning of the form '*if left sensor is illuminated more than right sensor then power the left wheel more than the right wheel*', even though this behaviour is 'embodied' within the robot. What is not explicitly represented anywhere are the goals '*turn towards light*' or '*avoid light*'. In some sense, the robot 'just does it', because it has been wired that way.

> **Activity 2.2 Read (optional)**
>
> If you would like to know more about Braitenberg's vehicles, and you have time, then take a look at this review of Braitenberg's vehicles by Dr Michael Dawson at the University of Alberta, Edmonton, Canada (Dawson, no date).

Sense–act in nature

These sorts of behaviours are found in animals: many moths are attracted to light whilst a woodlouse avoids bright light. This response by movement of an animal towards or away from a stimulus is called *taxis* (pronounced 'tacksiss'). More specifically, movement towards light is called *positive phototaxis*, and movement away from light is called *negative phototaxis*.

It does not take a great leap to realise that evolution could develop such mechanisms. If an animal happens to be 'connected up' in a way that better suits its environment, then it has a better chance of surviving long enough to reproduce than less-fortunate compatriots. The physical connections that gave rise to those reflexes pass to the next generation. In one of the most exciting areas of current robotics, researchers are looking at ways of 'co-evolving' sensors, controllers and actuators to exploit the rich, and often unpredictable, behaviours that can be obtained from certain relationships between these subsystems.

2.2 The cognition subsystem

In contrast to hard-wired reflexes, *reactive behaviour* requires a learning process. An example of a reactive behaviour in our everyday life is that an experienced car driver will automatically apply the brakes in a dangerous situation. Although the driver's reactions are now automatic, this behaviour initially had to be learned. Learning teaches the driver how to identify a dangerous situation, to know the location of the brake pedal and how hard to depress it in order to safely bring the car to a halt.

Even though there is no general definition of intelligence, human intelligence is characterised by what we call 'thinking'. A more scientific term for thinking is *cognition*: the process of acquiring knowledge and understanding. The process by which humans combine knowledge and facts to deduce new facts forms an important part of human cognition and is known as *reasoning*. Likewise, it is possible to develop cognition in machines (although so far in a much more limited manner than in humans), so it is possible to apply human cognition concepts – including reasoning – to machines.

The part of an intelligent machine that performs cognition is unsurprisingly called its *cognition subsystem*, popularly thought of as the robot's brain – although, as you will see later, this is often a wildly misleading term. The cognition subsystem has the important job of deciding what to do next and telling the various actuators in the actuation subsystem what they should be doing. The cognition subsystem allows more complex decisions to be made than are possible with reflex actions or reactions. This requires more information to be processed.

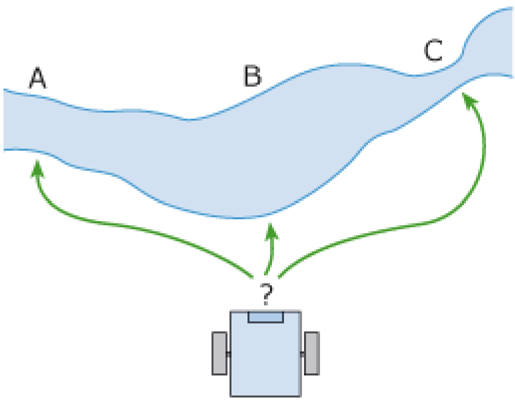
The cognition subsystem will be characterised by the way a robot uses *knowledge*, including factual information about the robot and its environment, both of which are stored in its memory. The cognition subsystem is often referred to as an *information processing system*. It takes information about the world and processes it to produce new information. This new information may then initiate some particular action or actions via the actuation subsystem.

A robot's cognition system is often engaged in solving problems. For example, a robot may have the constant problem of '*what shall I do next?*' in the context of '*how can I achieve my goal?*'

SAQ 2.1

Suppose the robot below has a goal of crossing a river. It must decide at which of three places it can cross safely. Try to identify suitable crossings for each of the four following scenarios:

1. if it uses the rule '*cross at the narrowest point*'?
2. if it knows that the crossing at A is shallow?
3. if it has seen similar crossings to C that are dangerous?
4. if it has always in the past crossed safely at B?



Show description

Reveal answer

Classic problems in artificial intelligence

Over the last sixty years, basic reasoning capabilities of artificial intelligence (AI) systems have been tested using a number of ‘toy problems’, some of which you may have heard of, including the ‘Towers of Hanoi’, the ‘missionaries and cannibals problem’ and the ‘Bridges of Königsberg’.

Study note – Toy problems

A toy problem is a simplified problem. In the field of artificial intelligence, they can either be used to explain an aspect of artificial intelligence, or to explore an aspect of AI in great detail. Toy problems usually simplify or ignore many aspects of the real world. For example, a toy problem used in robotics is stacking blocks on top of one another. In the toy version, the blocks are all the same shape, all made of hard materials and they sit in a nice neat world. Only when the robot programmer is happy that their program can solve the toy problem would they consider going on to try the robot stacking different-size objects made of different materials in a messy environment.

The reality of operating intelligently in the real world is, of course, another matter. In this topic you will begin to see the enormous gap between the theoretical possibilities of AI algorithms and robotics, and the realities of implementation.

Activity 2.3 Research (optional)

If you are interested, you can use a web search to research some of the toy problems mentioned above. Think about what sorts of rules might allow a computer program – or a robot – to solve these problems.

2.3 The sense–think–act model

In Section 2.1, the perception subsystem of a robot was defined to be the sensors (which collect sensory information) and the actuation subsystem (defined as the means by which the machine interacts with its environment). In this section we will add another possibility – one where the robot is capable of ‘thinking’ about its actions before performing them.

The *sense–think–act* model shown in Figure 2.6 extends the sense–act model we discussed earlier by placing the ‘cognition’ subsystem between the perception and actuation subsystems.

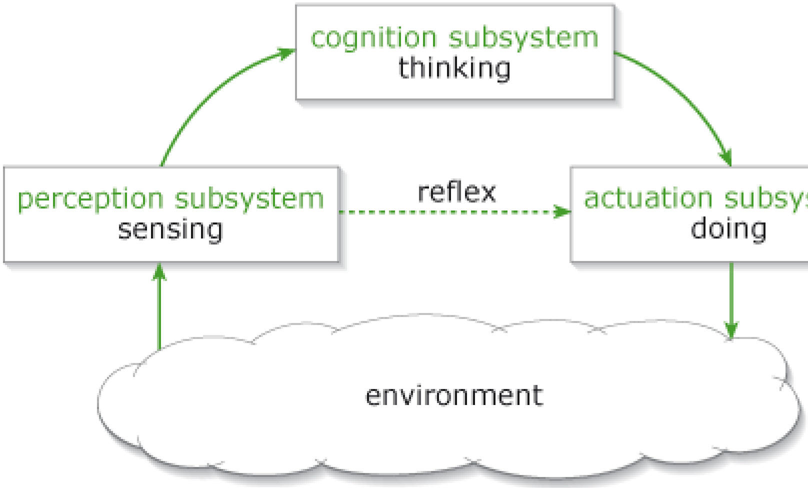


Figure 2.6 The sense–think–act model

Show description

Many of the pioneering experiments in AI algorithms during the 1950s placed a high premium on replicating the reasoning and planning actions of the cognition subsystem. This approach was driven by the new field of computer science, where computers were seen as logical machines able to perform complex sequences of calculations. It is now time for us to explore cognition in greater detail.

3 Thinking and learning

The ability to think about thinking may distinguish humans from all other animals. It has certainly engaged the attention of philosophers over many centuries. Rather than get bogged down in these philosophical details, we will take a practical attitude towards 'thinking' in machines. Our own robots, and other robots you are likely to come across, are very unlikely to have an identity crisis about their existence. Can they think at all? We'll review this question during this topic.

Figure 3.1 Who? What? Where? Why me?

Show description

As SAQ 2.1 (in Section 2.2) showed, the cognition system of a human or an artificially intelligent robot could use several different types of 'thinking'. We consider some of them in more detail in this section.

3.1 Approaches to learning

In a human-designed robot it is easy to see that someone has 'designed' the behaviour embodied into the robot. In the spirit of Braitenberg's vehicles, I could design and make a robot with a touch sensor that will follow the edge of a wall; such a robot can find its way out of a maze. A mouse shows similar behaviour: it will run with its whiskers in contact with a wall. Why does it do this? One reason might be that it keeps the mouse in cover and out of the reach of predators.

Over many generations, mice will vary in the tendency to either stay close to a wall or wander into the open; those that stick to the wall might be less vulnerable to predation and so survive and reproduce better than those who stray away from the wall. By this process of natural selection, the mouse species will evolve into one that shows this characteristic behaviour. Similar ideas have been used by AI researchers. Rather than designing behaviour, they simulate the process of natural selection among different behaviours, looking for the behaviour that works best. This technique, of evolving a solution using a 'genetic algorithm', is one we will come back to in a later topic.

This, of course, implies that we have designed our robot to embody rules to deal with all eventualities it will encounter. This is a difficult task; an alternative approach is to design robots that can learn. *Artificial neural networks* (commonly just called neural nets) offer one approach to training robots. Neural nets use software (and sometimes hardware) to mimic some aspects of the connectivity found in animal brains. Before it can be used, a neural network must be trained by a human. We will discuss neural nets in more detail in Section 4.1.

However, animals can learn without being taught by a human. For example, a young chick will first peck indiscriminately at anything, but learns to peck at seeds rather than stones. When the chick pecks at a seed, it is ‘rewarded’ by obtaining food and this *reinforces* the correct behaviour of pecking only at seeds. *Reinforcement learning* is a possible strategy for creating a robot that can learn.

It is clear that there are many different models for what the cognition system of a robot could be like. Reasoning and deduction are important parts of human thinking, and also of artificial intelligence. We will now look at how robots can represent facts and rules which they use to deduce appropriate behaviour.

3.2 Reasoning

Although it is difficult to define and measure intelligence, the ability to reason is certainly a major factor. Reasoning is the process of deducing new facts from those that are already known.

Suppose I have observed that my dog always wags her tail when she eats her dinner. I can use this observation to create a ‘hypothesis’ about my dog which is:

> *If my dog is eating dinner, then my dog is wagging her tail.*

This hypothesis is made of two parts:

Condition	Deduction
If my dog is eating dinner,	then my dog is wagging her tail.

Now suppose I’m working late and call my wife to apologise. My wife tells me that she has given my dinner to the dog.

I can write down what I know my dog is doing in the following way:

> *My dog is eating dinner.*

The fact that ‘*My dog is eating dinner*’ satisfies the conditional part of my hypothesis which allows me to conclude (or deduce) that my dog is wagging her tail.

SAQ 3.1

1. If the light sensor on my moving robot measures less than 30%, then it will stop. The light sensor on my robot measures 27%. Which of the following can you deduce?

   i. My robot will stop.
   ii. My robot will turn to the left.
   iii. My robot will continue moving.

2. If the left touch sensor on my robot is pressed *and* the right touch sensor on my robot is pressed, then my robot will go backwards. Both the touch sensors on my robot are pressed. Which of the following can you deduce?

   i. My robot will go forwards.
   ii. My robot will go anticlockwise.
   iii. My robot will stop.
   iv. My robot will go backwards.

3. If the temperature of my car engine is greater than 85 °C, then the engine is in danger. The temperature of the engine is 90 °C. What can you deduce about the engine?

Reveal answer

In SAQ 3.1, instead of saying '*the engine is in danger*' it would be more precise to say that '*the engine is in danger is true*', but most of us don't tag on the words '*is true*' to everything we say. This is just one illustration of the subtlety of everyday human language. However, it is necessary to be much more explicit where machines are involved. To that end, it is helpful to define what we mean by a 'fact'.

3.3 Facts

Let's start off by saying that a 'fact' has two parts. The first part conveys the meaning, and the second states whether the fact is true or not. For example, the following statements are all facts.

The Queen of England is a woman is true.

The Queen of England is a man is false.

The Queen of England is a robot is false.

We will call the first part of the fact a *phrase*. The second part of the fact, ‘*is true*’ or ‘*is false*’, is called its *truth value*.

SAQ 3.2

Split each of the following facts into a phrase and a truth value.

- The headquarters of The Open University is based in London is false.
- About 200,000 people study with the OU each year is true.
- The OU national office for Northern Ireland is in Netherwallop is false.
- Some Members of Parliament have OU degrees is true.

Reveal answer

We can define ‘fact’ more precisely:

i. a phrase is any collection of words
ii. a fact is a phrase that can be recognised and be given a truth value.

This allows us to avoid a long digression into mathematical logic and complex philosophical discussions about ‘well-formed’ phrases and ‘meaning’. So far as we are concerned, if a robot can recognise the phrase and give it a truth value, then that phrase has meaning to the robot.

SAQ 3.3

Deduce how this robot is moving from the following two facts.

1. The left robot wheel is going forward is true.
2. The right robot wheel is going forward is false.

Reveal answer

3.4 Rules

In addition to facts, logical deductions need rules of inference. Rules have a very particular structure or form; for example, returning to my hungry dog:

if my dog is eating her dinner is true

then my dog is wagging her tail is true

The first fact, *My dog is eating her dinner is true*, is often called the *conditional* or *antecedent* part of the rule.

The second fact, *My dog is wagging her tail is true*, is often called the *consequent* part of the rule.

SAQ 3.4

a. What are the antecedent and consequent facts in these rules?

   - **if** the engine is hot is true
   - **then** the engine may fail is true

b. What are the antecedent and consequent facts in these rules?

   - **if** the oil is low is true
   - **then** the engine is hot is true

c. Using these two rules, what two deductions can you make about the engine if the fact the oil is low is true?

Reveal answer

3.5 Conflict resolution

Human beings use hundreds of rules in their everyday lives, often without realising it. The sheer number of rules means that it is inevitable that there are occasions where rules clash or conflict with each other. For example, here are two rules:

if it_is_Sunday = TRUE **then** visit_Aunty = TRUE

if the_roof_is_leaking = TRUE **then** fix_it = TRUE

Suppose that in your collection of rules you have the facts:

it_is_Sunday = TRUE

and

the_roof_is_leaking = TRUE

It is not possible to visit Aunty and fix the roof at the same time, so what do you do? All but the most indecisive people have methods for resolving conflicts. For example, some rules may have priority over others. In this case fixing the roof might take priority. This would require the rule:

if the_roof_is_leaking = TRUE **then** fix_it_immediately = TRUE

which would mean fixing the roof takes priority and visiting Aunty would have to wait.

What happens when a robot has conflicting rules? For example, consider the rules:

if light_sensor_>_50 = TRUE **then** turn_left = TRUE

if touch_sensor_pressed = TRUE **then** turn_right = TRUE

when the list of facts contains:

light_sensor_>_50 = TRUE

and

touch_sensor_pressed = TRUE

In this case, the robot requires a *conflict-resolution* strategy in order to decide which rule has the highest priority. Typical conflict-resolution strategies include one or more of the following:

- the priority given to the rules
- the order in which the rules are written
- which rules have not recently ‘fired’ (been used).

A rule that *could* have fired, but did not, is said to have been *triggered*. That is, its conditional fact was known to be true, and in principle it *could* have fired, but it was prevented from doing so by the firing of a higher-priority rule.

SAQ 3.5

As you probably guessed, my dog has quite a personality, and at one year old she is rather a handful. Even so, there are rules that seem to govern her behaviour. One of these is:

> **Rule 1**. If my_dog_is_on_the_sofa = TRUE then my_dog_is_happy = TRUE

However, there’s another rule:

> **Rule 2**. If my_dog_is_on_the_sofa = TRUE then my_dog_is_told_off = TRUE

Observing her demeanour when this rule fires, I think there is yet another rule:

> **Rule 3**. If my_dog_is_told_off = TRUE then my dog_is_happy = FALSE

a. Is there any conflict in the rules that apply to my dog?
b. Do you think that animals can have conflict-resolution strategies?
c. Do you think that dogs are capable of reasoning, as suggested here?

Reveal answer

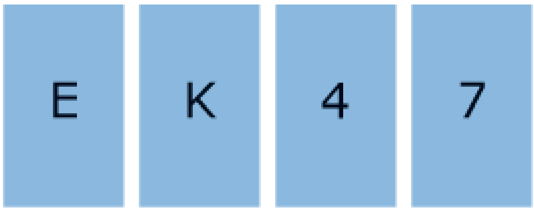
3.6 Is reasoning built into our brains?

Some biologists believe that the rules for reasoning are ‘hard-wired’ into our brains. The evidence for this includes the way that almost every human being seems to accept the legitimacy of this way of reasoning. Another remarkable feature of humans is that almost all human societies use language to communicate facts and ideas. The linguist Noam Chomsky put forward a widely accepted theory that the ability to learn and use language – any language – is built into our brains. Language enables us to represent and manipulate abstract concepts in our heads – it is what enables us to reason.

Do you think that the ability to reason logically is built into our brains? Try the experiments in SAQs 3.6 and 3.7 which test people’s reasoning powers.

SAQ 3.6

Each of these cards have a letter on one side and a number on the other side.



Show description

Which cards would you turn over to test the rule: ‘***If*** *a card has a vowel on one side,* ***then*** *it has an even number on the other*’?

Reveal answer

SAQ 3.7

Consider another set of cards. These have a city on one side and a mode of transport on the other. The mode of transport on each card is the only way that you travel to the city on the other side of the card. Which cards would you turn over to test the rule: ‘*Every time I go to Paris I travel by plane*’?

Show description

Reveal answer

Experiments have been conducted using these cards. Researchers found that only 12% of people answered the rather abstract rule in the first set of cards correctly, compared to 60% for the second set, which uses associations with the real world. These (and other findings) have been used to argue that logic is not hard-wired into our brains; if it were, then there should be no difference in the results of the two experiments. That there is a difference suggests that humans use knowledge and experience about the world to answer the question rather than relying on built-in logic.

One of the difficulties of all this for roboticists is that we don’t really understand how our brains work, and it is very difficult to make robots with brains and modes of reasoning like our own.

4 Machine intelligence

Machine learning is a field of computer science dating back to 1959. In that year, Arthur Samuel published a paper in the *IBM Journal of Research and Development* titled ‘Some Studies in Machine Learning Using the Game of Checkers’. Samuel had set out to prove that machine learning would allow a computer to become a better player of the board game checkers (known as draughts in the UK), to a point where it could be a superior player to its programmer. Samuel’s checker player had the ability to learn without being explicitly programmed. The key feature of machine learning is the concept of *self-learning*.

Although machine learning is initially dependent on computer programming, it also applies statistical modelling to detect patterns and improve performance based on data and information about its performance so far. Samuel demonstrated that machines don’t require direct *input commands* to perform particular tasks, but rather *input data*.

Put simply, data will be fed into the machine, an algorithm will be selected, settings will be configured (and adjusted if necessary) and the machine will then be instructed to conduct its analysis. Through a process of repeated trial and error, the machine will start to decipher patterns found in the data and create a data model which the computer will use to predict future values. The machine will then try to replicate the human decision-making process by formulating decisions based on experience.

Machine learning incorporates many different types of statistical algorithms and it is a constant challenge for those working in the field of machine learning to choose an appropriate algorithm, or combination of algorithms, to do the right job.

Away from the complexities of the inner workings of machine learning, we now find that machine learning is starting to have an effect on many aspects of our lives and is something that we now take for granted. In Topic 4 we will look at other automated technologies and how they have become ubiquitous in our day-to-day lives.

Activity 4.1 Watch

Machine learning is heavily based on data. Watch the following clip from the BBC *Click* programme (broadcast November 2018) about how some companies are generating huge amounts of annotated data using human labour.

A computer programmed with a sufficiently large number of rules *can* appear to be intelligent by combining the results of applying many rules one after the other. A large number of artificial intelligence systems were developed in the 1980s and 1990s which used rules to perform tasks such as playing chess, diagnosing diseases in humans and plants, identifying valuable geological resources and assisting financial companies in making lending decisions.

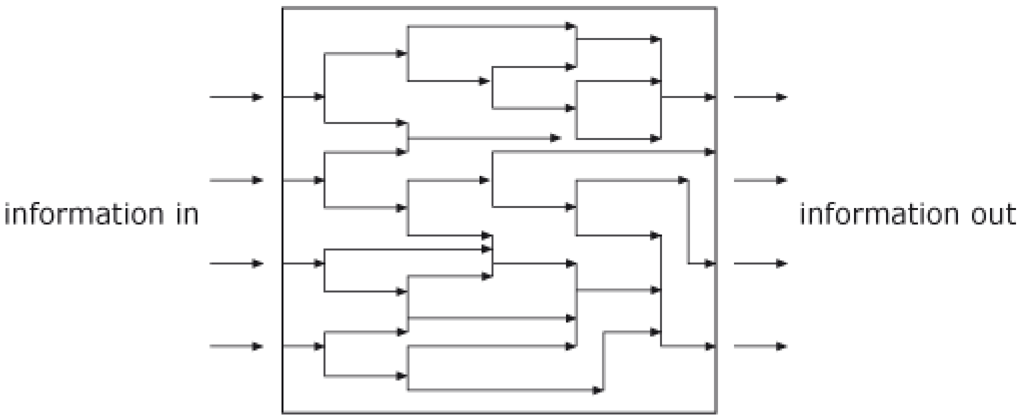


Figure 4.1 Processing information by if-then decisions

Show description

However, whilst these rules-based intelligences could sometimes outperform human experts in specific areas, they were much less successful confronting more generalised problems. Nor could they easily be moved between different areas of specialisation: for example, an intelligent system working in diagnosing pre-birth medical conditions could not be reused to diagnose diseases in soybeans without extensive modifications.

4.1 Neural networks

Neural nets are a form of machine learning and are the main alternative to the logic-based calculations that you met in Section 3.

Inspired by biology

Neural nets were first described in the 1940s and the first implementations appeared about a decade later. They are inspired by the structure of the human brain which, as you learned earlier, is made up from specialised cells called neurons. This isn't a biology module, so the only major aspect of biological neurons you need to know is that they communicate with one another through links known as *synapses*. A single neuron may be linked to many other neurons. Electrical signals pass along these links, causing some neurons to become more active (and creating yet more activity) and some less active. The exact method by which neurons store and process information in our brains is still a subject of active research, but we know enough to attempt to replicate this process inside computers.

The design of a neural net

An artificial neural net is a simplified model of a brain. Here, neurons can be represented either in software or in hardware. Artificial neural nets are arranged into layers. The input layer receives data from elsewhere – for instance, in a robot the inputs might come from sensors. There is also an output layer, which passes the decision of the neural net to another part of the system – so again, in our robot the output layer might talk to the robot's motors. Some neural nets have one or more 'hidden layers' between the input and output layers; these perform additional processing and greatly improve the performance of the net. The exact numbers of neurons, layers and the connections between them vary greatly between neural nets.

Figure 4.2 is a diagram of a typical neural network showing the input and output layers of neurons as well as a single hidden layer. The lines between neurons show the links from one layer to another.

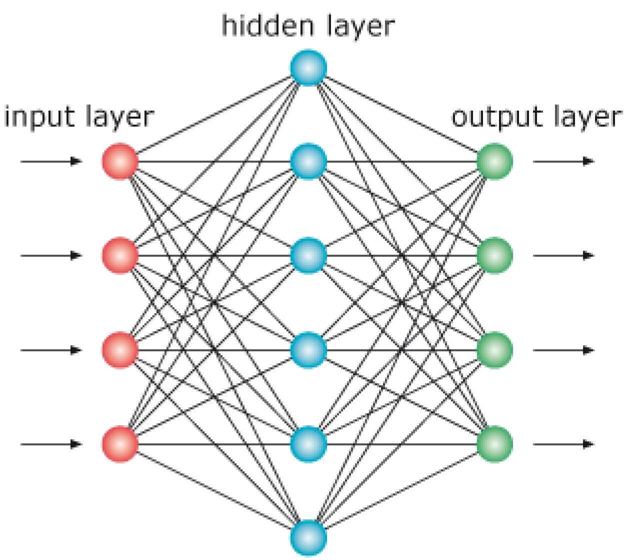


Figure 4.2 An artificial neural network

Show description

A single artificial neuron receives one or more inputs (usually numbers) which are processed to produce an output. A simple process might be for the neuron to add up all of its input values and compare it to a so-called 'threshold' value. If the total is greater than the threshold, then the neuron is activated and it produces an output.

SAQ 4.1

The artificial neuron described above has three inputs and a threshold value of 1.

a. The three inputs are 0.1, 0.6 and 0.7. Will the neuron be activated?
b. The inputs are 0.2, 0.2 and 0.2. Will the neuron be activated?

Reveal answer

Each of the output links from a neuron has a 'weight' which can be thought of as a measure of that link's importance. When a neuron's output is passed over a link, its value is multiplied by the weight of that link.

SAQ 4.2

The output value from a neuron is 1. The neuron has three outputs with weights of 0.7, 0.25 and 0.9. What are the weighted output values for each of these links?

Reveal answer

These weighted outputs form the inputs to the next layer of neurons in the net. This process is repeated until the output layer is reached.

Training a neural net

A neural net needs to be 'trained' before it can be put to use. This training uses large sets of known inputs and outputs called 'training data'. Initially, when the inputs are given to the neural net, the outputs are wrong. The errors – the differences between the desired outputs for a particular input pattern and the actual outputs that are generated from that input pattern – are used to change the weights in the network. This process may be repeated tens of thousands of times with the weights between neurons constantly being adjusted until the output of the network matches that of the training set.

During training, the networks do not always learn to give the correct outputs for a given set of inputs. For example, the data may contain inconsistencies, with two training items having the same inputs but different outputs. In this case, no network could ever be trained to give the 'right' answer, since there are two different answers.

Study note – Training and energy consumption

The process of training neural nets with very large data sets can take many hundreds of hours of processor time on high-performance computers and is enormously energy intensive. One estimate (Strubell *et al.*, 2019) is that training a neural net to process human language consumed five times as much energy as a typical family car would use in its lifetime!

In this way, the logical information processing box seen in Figure 4.2 can be replaced by a neural information processing box, illustrated in Figure 4.3.

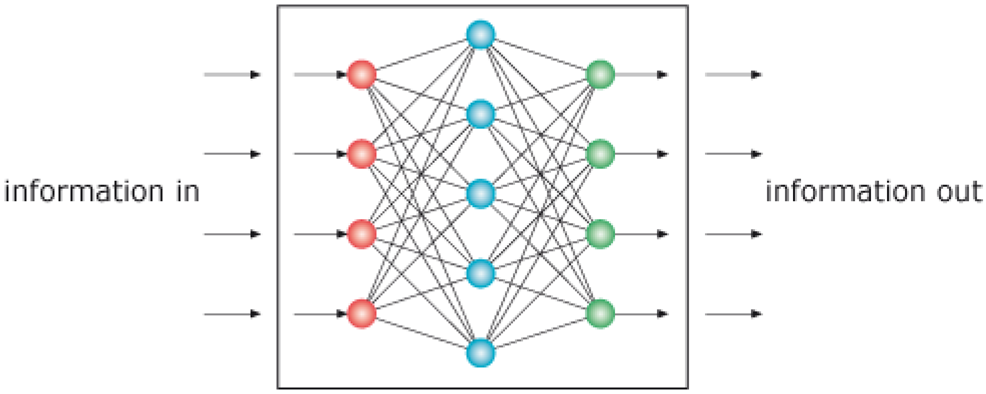


Figure 4.3 Processing information by an artificial neural network

Show description

4.1.1 Using neural nets

The power of a neural network lies in the fact that, once trained on a range of training data, the network can respond correctly to inputs it has never seen before. This is called *generalisation* and represents a major difference with rules-based intelligences which struggle to produce useful results when confronted with unfamiliar data.

There are many kinds of neural computing, some of which are very powerful. Neural nets are widely used in areas as diverse as recognising objects from images, translating documents, speech recognition and filtering spam emails. However, be warned that most artificial neural systems are as different from real brains as are the processors inside a computer. It is easy to hype up neural networks, but they remain far removed from the biological neurons that underpin human intelligence.

The following cartoon illustrates the difficulty we have in implementing the neural information processing box and getting ‘consciousness’ to appear. The expression ‘*cogito, ergo sum*’, means ‘*I think, therefore I am*’. It is attributed to the seventeenth-century philosopher, mathematician and soldier René Descartes. The joke here is that the only way the operators know what’s going on inside this huge machine is by the messages it prints out. The suggestion is that this machine is thinking all kinds of thoughts unknown to its creators and has its own autonomous intelligence to grapple with the conundrum of its own existence – in Latin!

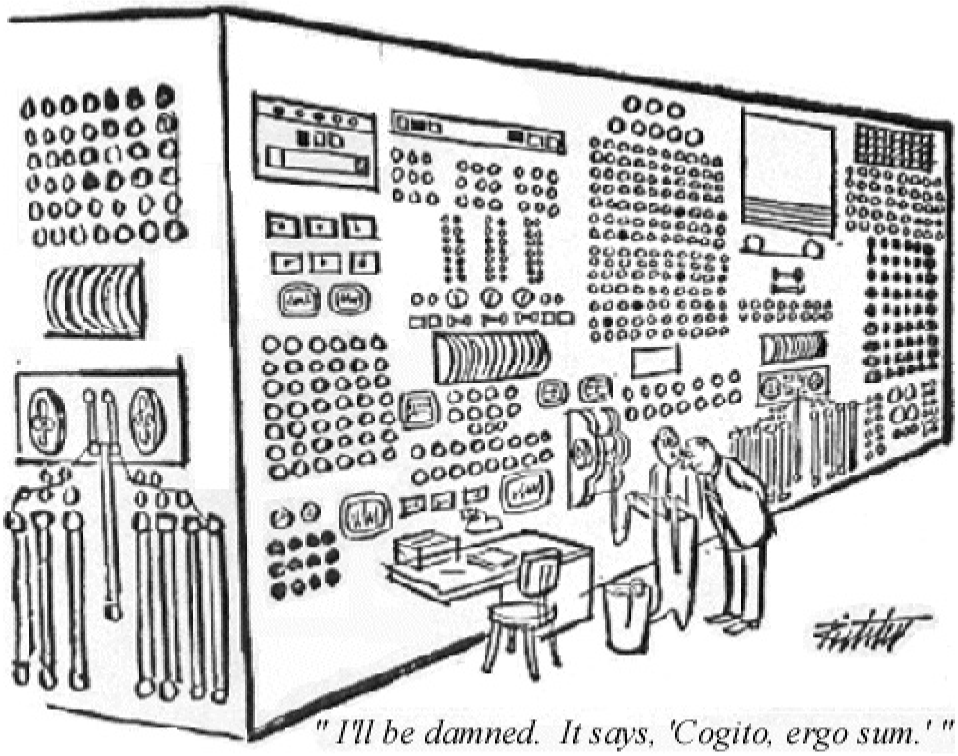


Figure 4.4 “I’ll be damned. It says, ‘Cogito, ergo sum.’”

Show description

4.1.2 Deep learning

Deep learning is a subset of machine learning that tries to emulate the workings of the brain using methods based on artificial neural networks. As humans, we learn from experience. In the same way, deep learning algorithms will perform tasks over and over, adjusting each time in order to improve performance. There are many ‘layers’ of information that can be processed by deep learning algorithms (hence ‘deep’). In the same way that humans use thought to figure out problems, deep learning can also learn to solve complex tasks.

Deep learning differs from machine learning in that machine learning relies mainly on structured data, whereas deep learning algorithms have the ability to process information through many layers. Nowadays, deep learning also tends to be more popular as it needs very little ‘manual’ adjustment and can naturally take advantage of increases in computation as well as data. Deep learning has applications in many areas and has recently been responsible for breakthroughs in areas including image processing, audio and speech processing and pharmaceutical research.

If you are interested in finding out more about deep learning – and have the time – we recommend this set of articles on the history of neural nets and deep learning (Kurenkov, 2015).

4.2 The Turing test

Machine intelligence is hard to define, but there have been many attempts to design tests that would reveal just how intelligent machines are (or indeed – aren’t).

Alan Turing was one of the greatest British mathematicians and scientists of the twentieth century. He made many profound contributions to the fledgling discipline of computer science from the 1930s to 1950s. Turing worked at Bletchley Park during the Second World War, where he helped crack the Enigma code used by Germany to send military signals. His pre-war work directly inspired the team that built Colossus – the world's first programmable, digital computer which was used to crack the most secure military communications used by Nazi Germany. Following the war, Turing went on to design computers at both the University of Cambridge and the University of Manchester. He worked in areas as diverse as computer chess (without actually having a computer to play it!), biology and artificial intelligence. If you have time, you can read more at a website all about Turing (Hodges, no date).

An imitation game

Even before anyone in the world had built a working computer, Turing was interested in the extent to which machines might be able to mimic intelligent human behaviour. He proposed a thought experiment that has since come to be known as the Turing test.

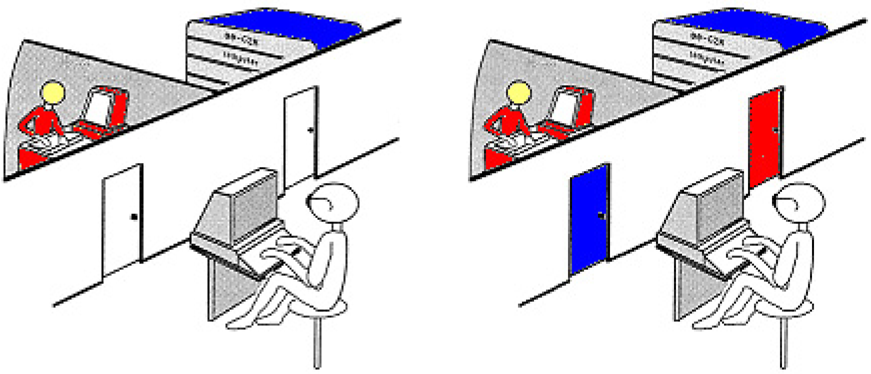


Figure 4.5 The Turing test: is there a person or a computer behind the door?

Show description

Turing first imagined an imitation game in which a man and a woman are sent to separate rooms. A questioner, who cannot see the participants, asks both of them questions to try to identify which is the man and which is the woman. The man and the woman type their replies to the questions and can answer the question any way they choose – including being free to lie.

After introducing the problem, Turing then supposed that either the man or the woman could be replaced by a computer:

> We now ask the question, 'What will happen when a machine takes the part of A in this game?' Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman?

With the same restrictions as before, the questioner now has to decide which of the two is the computer. If the questioner can't tell the difference, then the computer is deemed to have passed the test.

Study note – AI winters and summers

You may have noticed and been surprised by just how long ago some of the fundamental aspects of AI systems were first described. The 1950s and 1960s saw many AI breakthroughs (perhaps best characterised by the villainous HAL 9000 computer in the 1968 film *2001: A Space Odyssey*).

This early period of optimism is known as the ‘First AI Summer’ and it came to an abrupt end when many of these early successes proved unable to cope with more general problems. The summer was followed by an ‘AI Winter’ during which the number of researchers, and the amount of funding for AI, fell dramatically. This winter lasted until 1983 when another ‘AI Summer’ arrived, largely driven by well-funded Japanese researchers working on the hugely ambitious (and ultimately unsuccessful) Fifth Generation Computer Project. This Second AI Summer ended in the mid 1990s when it became clear that rule-based intelligent systems were inadequate for solving complex tasks.

The Second AI Winter lasted until the early twenty-first century when advances in hardware and software produced the current wave of machine learning applications. Perhaps wary of previously unfulfilled expectations, this rapid growth is sometimes called an ‘AI Spring’. How long it will last is anyone’s guess.

4.3 Intelligence and language

In 1980, the philosopher John Searle raised concerns that the Turing test does not actually test understanding, only the ability of a person or a computer to manipulate symbols. Searle created his own thought experiment known as ‘The Chinese Room’.

Here, we have to imagine that there is a closed room with a slot through which messages can be passed. A person who can read and write Chinese text is invited to ask a question. They would write it down using Chinese characters and push their query through the slot. A little while later, a perfectly understandable reply, also written in Chinese, would emerge from the slot. From the questioner’s point of view they appear to have had a conversation with a Chinese-speaking intelligence inside the room. According to the rules laid down by Turing, the person (or is it a computer?) has passed the Turing test.

Searle’s experiment continued. The door would be unlocked. Inside, the questioner would find a second person who they quickly realise cannot understand a single character of Chinese writing! The person in the room is surrounded by lots of books all written in Chinese and has been given a set of instructions which they can understand.

The instructions (a program) tell the person in the room to search the books for the characters in the question (an input) and then tell them how to find a matching response in the Chinese texts. The person would then copy down the second set of Chinese characters and push it through the slot (an output).

From the point of view of the questioner outside, the Chinese Room appears to be intelligent – it is behaving as if it understands questions written in Chinese. But when it is opened up, it is clear that the person in the room is simply manipulating apparently meaningless symbols according to a set of instructions. If a person, or indeed a computer, cannot understand the symbols it is processing, then can they be said to be intelligent?

What is understanding?

Searle's Chinese Room raises an important question: what do we mean when we say that a human 'understands language'? Even simple words such as 'moon' or 'ball' are just arbitrary symbols: the sound of the spoken word or the shape of the written one has no relation to the object itself. For two humans to agree what they are talking about, they must attach the word to sensory experiences. For example, the word 'moon' corresponds to seeing a big white thing in the night sky. This is called *symbol grounding* and is thought to be central to the development of language. The human brain cannot compute language on its own: it has to be connected to the world in a real body, with real senses. Thus the 'meaning' of the symbols that are manipulated inside our heads must somehow be 'grounded' in reality.

One of the big challenges in robotics is to get robots to understand symbols such as the words we use in speech. This is difficult enough for physical objects such as 'moon' or 'ball', but the challenge is even greater for the many words in everyday language that refer to complex and abstract things. For example, the words 'shopping', 'motoring' and 'gardening' may evoke complex thoughts. The problem of designing robots that can understand human language is therefore still a long way from being solved.

However, there are signs that AI-based language skills are improving in certain contrived situations. For example, in February 2011, IBM's Watson supercomputer competed against two human competitors in the American game show *Jeopardy*, and won, as reported by the BBC (2011).

4.4 Good old-fashioned AI

There is 'artificial intelligence' and there is 'Good Old-Fashioned Artificial Intelligence' (GOFAI).

Steve Grand was something of a maverick among robotics researchers. Working at home rather than in an academic or industrial research lab, he set out to create a robot of his own. He wrote about his experiences in his book *Growing up with Lucy: How to Build an Android in Twenty Easy Steps* (Grand, 2003). In the introduction he writes that GOFAI 'set out with the explicit aim of creating human-like intelligence but got it all horribly wrong and wasted the best part of 50 years'.

Grand was referring to early AI researchers who concentrated on 'symbolic' processes such as the logic and reasoning discussed in Section 3 – an approach he dismissed as GOFAI. Grand was one of the first of a growing number of researchers who are tackling AI using non-symbolic approaches based on biological analogies such as the neural nets described in Section 4.1, or statistical approaches backed by massive computational power and data resources.

Grand suggested that humans, as conscious beings, don't actually live in the real world, and that most of the time we live in a virtual world inside our heads. What do you think? What kind of virtual world could exist inside a robot's head? How could you know what is in the mind of a robot?

In robotics there is a tension between what we strive to achieve and what we manage to achieve. The popular view of roboticists is that they all want to build machines with human-like intelligence, whereas many robots are little more than computing boxes on wheels with primitive sensing mechanisms. This tension is also true of AI systems in general. A long-running joke among some AI researchers is that AI stands for 'Almost Invented'.

GOFAI is a symptom of interpreting 'intelligence' as the ability to do numerical and logical calculations. In this, symbols play a key role. Explicit manipulation of symbols and reasoning are important for some things, such as a legal argument, a newspaper article or a mathematical proof. But mapping sensor input onto symbols can be very difficult in practice.

The problem of symbol grounding is a hard one to solve. GOFAI can be criticised for paying too much attention to symbolic manipulation, and not enough to sensors and abstracting useful information from them. For example, one of the problems that GOFAI still has to solve is how to extract information from visual scenes in order to generate useful symbolic descriptions of them.

4.5 Strong AI and weak AI

In the artificial intelligence world, there is a considerable range of opinion about the extent to which machines will be able to replicate human intelligence. For the sake of convenience, AI researchers often classify themselves as belonging to one of two camps. These distinctions were first made by John Searle, whose Chinese Room was introduced in Section 4.3.

1. **Strong AI** advocates believe that it will eventually be possible to build machines that are capable of demonstrating human levels of intelligence and, ultimately, a full range of emotions, and even consciousness. Their arguments are based on the claim that there can be nothing mystical going on in our brains, and that fundamentally our bodies and brains are just machines, albeit made out of biological 'wetware'.
2. **Weak AI** proponents take issue with these rather grandiose claims and suggest that machines will only ever be able to *act* as if they had natural intelligence. Although a robot might appear to be able to understand what is going on in a conversation, for example, this would just be an illusion. Again, there is nothing mystical going on – just a computer processing information. There is no emergence of feeling or consciousness – it's just machinery.

Of course, researchers are a long way away from creating conscious machines. As we have mentioned, there is still the challenge (particularly in the GOFAI tradition) of just how to implement an intelligent reasoning process between sensory inputs and behavioural outputs.

Human reasoning can involve long and complicated sequences of deductive steps. Each of these has to be legitimate. The joke here is that all the complex reasoning and symbolic manipulation fails because an essential step in the chain is missing. No matter how good the sensors and actuators are in a robot, if the cognition subsystem between them cannot cope with long and complicated deductive chains of reasoning, then the robot will not be able to perform intelligent tasks like humans can.

5 A layered model of behaviour

In this topic we have grappled with what it means for humans and machines to be considered ‘intelligent’. To be pragmatic, we can sidestep the issue and simply demand that a robot ‘does the right thing’ in its environment, whether due to ‘intelligence’ or not.

5.1 Differing approaches to the same task

Earlier, we considered the sense–act model and saw that simple animals and robots such as Grey Walter’s Elsie can ‘do the right thing’ even without a cognition subsystem. They do this by relying on hard-wired behaviours; in the case of animals produced by millions of years of evolution and in the case of robots, created by its designer. For example, a simple vacuum-cleaning robot can work perfectly using the sense–act model: it simply moves randomly about the room, backing away when it bumps into an obstacle. Left long enough, the robot will clean all of the floor without needing a complex plan of operation or even a map of the room.

The sense–think–act model with its cognition subsystem has potential for more flexible and complex behaviour in the face of an unpredictable world. But complex decision making requires both good sensory information and a ‘model’ of the world, and this inevitably increases the complexity of a robot. Consider a more complex vacuum-cleaning robot that covers the floor systematically; such a robot will require a model of its world. This model might include a map of the room for route planning, which implies that the robot knows both its own location and that of obstacles. The robot will also need to ‘know’ that some types of object such as walls are fixed and that others such as chairs and cats may not always be found in the same place.

The second robot would be much more complex than the first, in both its design and programming. It would also require considerably more processing power than the sense–act robot. Complexity increases further as we try to anticipate all of the potential situations in which the robot must act intelligently.

An alternative approach might be to design a robot with the potential to learn. After all, most human behaviour is learned, and many animals searching for food learn by trial and error or by the example of others. So instead of designing and programming a robot to deal with many eventualities, perhaps it would be better to design a robot that was able to learn from its experience.

One aspect of learned behaviours in humans is that they can become *reactive*. Reactions are low-level, sensor–motor actions that can be learned or improved over time. Once learned, they become automatic and no longer require conscious thought: examples include walking, riding a bike or driving a car, which we can ‘do without thinking’. Reactive behaviours are similar to reflexes – they are driven by sensory input, quick-reacting and efficient – and so could be described by the sense–act model, but they are more adaptable since they can be relearned or unlearned.

5.2 Introducing layers

A fruitful approach for robot behaviour is to combine these models in different layers. The lowest layer would consist of reflexes; these are fast, initiated directly by sensor input, and carried out with no cognitive processing. Like human reflexes, these would be important for the safety of the robot or safeguarding people around the robot, for example in collision detection and obstacle avoidance.

Above the reflex layer would be a further layer of reactive behaviours. These would be a set of pre-programmed or learned packages of actions that, once started, require no further cognition. Reactive behaviours respond to sensor input and require only 'local' processing, so they are quick and efficient. An example of a simple reactive behaviour is a robot that follows a line marked on the floor which can be described as '*if over the line then keep going straight ahead; otherwise turn towards the line*'.

The topmost layer of the hierarchy is concerned with planning and strategy. This is where cognition occurs, dealing with the high-level goals and using appropriate reactive behaviours to carry them out. A parts-delivery robot in a factory might use this layer to plan a route to various parts of an assembly line, then delegate the task of reaching each destination to the underlying line-following reactive behaviour.

6 Putting it all together

We started this topic by introducing two abstract models of robot behaviour: the sense–act model followed by the sense–think–act model. These models suggest that robots need perception, cognition and actuation subsystems; indeed, it is possible to identify such subsystems in most robots.

6.1 Parts of a robot

Real physical robots also require some other important subsystems, such as a source of power and a 'body' to hold things together. We can produce a high-level description of the organisation or architecture of a robot as shown in Figure 6.1.

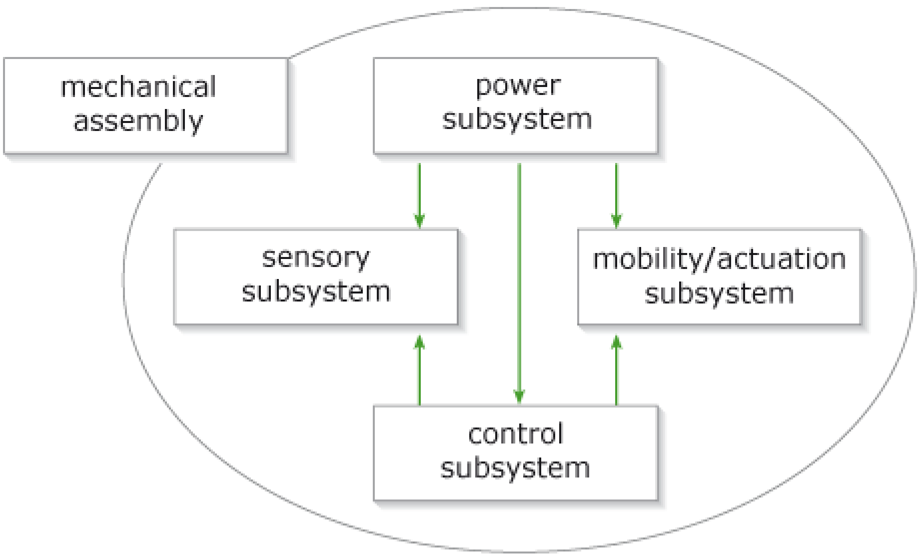


Figure 6.1 The architecture of a robot

Show description

1. The *sensory subsystem* provides the robot with information about the state of the world, for example through light sensors, bumper-activated touch switches, or range-finding ultrasound sensors. It is the physical counterpart to the perception subsystem of either the sense–act or sense–think–act behavioural models.

2. The *mobility/actuation subsystem* is what moves the robot around in the world – wheels or legs, for example, as well as the motors that actuate them.
3. The *control subsystem* is the physical counterpart of the cognition subsystem of the behavioural model. It is typically made up of one or more computer processors and will be responsible for using its own control plans along with sensory input information to decide what the robot should do. It will also need to tell the mobility/actuation subsystem to move the robot around in the world or otherwise perform useful actions.
4. The *power subsystem* provides power to the robot, not only to the motors or other actuators to make the robot move, but also to power the control and sensory subsystems. The power subsystem includes the power source (e.g. batteries) as well as power management and monitoring.
5. Finally, the *mechanical assembly* is the robot's body and is what holds everything together.

Particular robot designs implement these subsystems in different ways; however, this architectural description allows us to think about robot design in general terms as well as understand the specific requirements that lead to particular design decisions.

7 Practical activities

Activity 7.1 Robot practical activities

Follow the instructions on this week of the study planner and carry out the activities.

8 Summary

In this topic we have focused on how robots might 'think'. This is in the context of gathering information from sensors and controlling actuators in order to interact with the environment. To some extent we have gone beyond the sense–think–act model introduced at the beginning of the topic by considering how robots might engage in more abstract forms of thought.

We have also briefly considered the neural network approach to processing information as well as the concepts of machine learning and deep learning. These are very different from the rules-based approaches used by many earlier AI systems, but are still artificial and a long way from biological neural systems. Arguably, Grey Walter's tortoises come closest to the biological model.

In this topic we have looked at:

- sensing surroundings using the sense–act model
- how new facts can be deduced from existing facts using reasoning
- what it means for a robot to learn and plan its actions

The main conclusion to be drawn from this topic is how differently machines and humans think.

Where next

This is the end of Topic 2.

Topic 3 looks at human–robot interaction, and the relationships between people and robots.

References

BBC (2011) ‘IBM’s Watson supercomputer crowned Jeopardy king’, *BBC News: Technology*, 17 February. Available at: www.bbc.co.uk/news/technology-12491688 (Accessed: 4 May 2020).

Braitenberg, V. (1984) *Vehicles: Experiments in Synthetic Psychology*. Cambridge, MA: MIT Press.

Dawson, M.R.W. (no date) ‘Dawson’s Margin Notes on Braitenberg’s Vehicles’, *University of Alberta*. Available at: www.bcp.psych.ualberta.ca/~mike/Pearl_Street/Margin/Vehicles/index.html (Accessed: 3 February 2013).

Grand, S. (2003) *Growing up with Lucy: How to Build an Android in Twenty Easy Steps*. London: Weidenfeld & Nicholson.

Hodges, A. (no date) *Alan Turing*. Available at: https://www.turing.org.uk/ (Accessed: 4 May 2020).

Holland, O. (2001) ‘From the imitation of life to machine consciousness’, *Evolutionary Robotics. From Intelligent Robotics to Artificial Life – International Symposium*. Tokyo, Japan, 18–19 October 2001. London: Springer-Verlag, pp. 1–37. doi:10.1007/3-540-45502-7.

Kurenkov, A. (2015) *A ‘Brief’ History of Neural Nets and Deep Learning*. Available at: http://www.andreykurenkov.com/writing/ai/a-brief-history-of-neural-nets-and-deep-learning/ (Accessed: 4 May 2020).

Mind and Machines (1948) BBC West of England Home Service, 5 February.

Strubell, E., Ganesh, A. and McCallum, A. (2019) ‘Energy and Policy Considerations for Deep Learning in NLP’, *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence, Italy. doi:10.18653/v1/P19-1355 (Accessed: 4 May 2020).

Further Reading

Kelly, J. and Whitfield, P. (1994) *The Robot Zoo: a Mechanical Guide to the Way Animals Work*. Nashville, TN: Turner Publishing.

Searle, J. (1980) ‘Minds, brains and programs’, *Behavioral and Brain Sciences*, 3(3), pp. 417–457.

Acknowledgements

Grateful acknowledgement is made to the following sources:

Figure 2.2: Taken from: http://www.theoldrobots.com/ElmerElsie.html

Figure 2.3: Taken from: https://old.computerra.ru/vision/667023/

Figure 4.4: Taken from: https://opiniaocentral.wordpress.com/2017/02/27/computatio-ergo-sum/