Chapter 4

Perception

LEARNING OBJECTIVES

After studying this chapter you should be able to do the following:

- ☑ Perception is the process by which the cognitive system constructs an internal representation of the outside world. To do so, it uses simple input data from the environment (sensations) and processes these data with rules.
- ☑ Perception does not produce a copy of reality. Instead, it is an active process that allows us to make sense of the world around us.
- ☑ Perception does not always provide a veridical rendition of reality: we can see things that are present, we can see objects that are physically impossible, and we can see two different animals in one single drawing. Studying visual illusions is an important way to understand human perception.
- ☑ Although the image on the retina is two-dimensional, the brain is able to construct three-dimensional representations of scenes, by using monocular and binocular cues.
- ☑ According to Marr's and Biederman's theories, human perception uses basic three-dimensional shapes to recognise objects.
- ☑ The brain uses different areas for recognising faces than for recognising other objects.

Introduction

Released in March 1999, the movie *The Matrix* directed by Andy and Larry Wachowski became an instant hit. The plot was simple yet amazing. By deceiving our perceptual system, computers take the control of our mind and rule mankind. The matrix is the program and interface imposing a non-existent reality to our senses. It makes us believe that we are living in a free world; in fact, the virtual reality is just a mere simulation that keeps us in psychological prisons. Humans are used as batteries. The question posed by the movie is whether we can escape our own perception.

Perception is the only way to know the world around us. If perception fails and does not account for what is going on outside, are we going to notice it? The answer provided by the Wachowski brothers is a 'no'. In the matrix, the hero *Neo* (Keanu Reeves) is shown the way out of the matrix by people who are not under the control of virtual reality. The movie can be understood metaphorically. Do we perceive what is in there? To what extent are we deceived by our senses? Is there anything to do to correct the mistakes that perception sometimes commits in capturing the world? All these questions, implicit in the movie, are addressed explicitly in this chapter.

We start by showing that the aim of the perceptual system is to inform the mind about reality. Then, we describe the senses that serve this purpose. By analysing the processes that create percepts, we show why perception is not a mirror of reality but is a reconstructed representation, much like a painting, with its beauties and flaws. The relationship between the objective intensity of physical stimuli and their perceived intensity constitutes the field of psychophysics, which is presented in the second part of this chapter. From this section onwards, we focus on visual perception. Visual perception is the kind of perception we rely on most often. Evolution has shaped what was at the beginning a mere light detector to a remarkably complex perceptual system. Focusing on vision will allow us to demonstrate the key characteristics of perception in a pictorial and intuitive way.

The third part details the basic mechanisms that our perceptual system uses to group together meaningful pieces of information so as to form basic chunks of information. We will see that these laws, known as the Gestalt laws of perception, are applied automatically to the bottom-up processing of perceived objects.

The fourth part shows how our cognitive apparatus builds a three-dimensional view of the world by analysing the two-dimensional images provided by the retina. In the remainder of the chapter, we show how visual illusions highlight some specific aspects of the perceptual system by pointing to its weaknesses. We also present two central theories aiming to explain visual recognition. Marr (1982) proposed a stage view of object recognition, and Biederman (1987) proposed a theory emphasising that our perceptions are the result of binding together basic visual shapes. Finally, we consider face perception. Faces are the most common visual objects that we meet in our everyday life. The study of face perception provides an insight about how we perceive objects with complex, changing visual features.

What is perception?

Sensing the world

Our environment changes continuously and rapidly. Although most of the events are of no relevance, a few are potentially either harmful or beneficial. Our species has evolved perceptual mechanisms to select such relevant information, mechanisms that are shared to a considerable extent with those used by non-human primates. We do not have a direct access to what is happening in the environment – rather, we collect cues that are used to build an internal representation of the external world. This chapter is concerned with how we figure out what is going on in the outside world.

Consider Doolittle's painting 'Pintos' (1979), shown in Figure 4.1. How many horses are hiding in this Figure? The painting is interesting in that both the horses and the background are

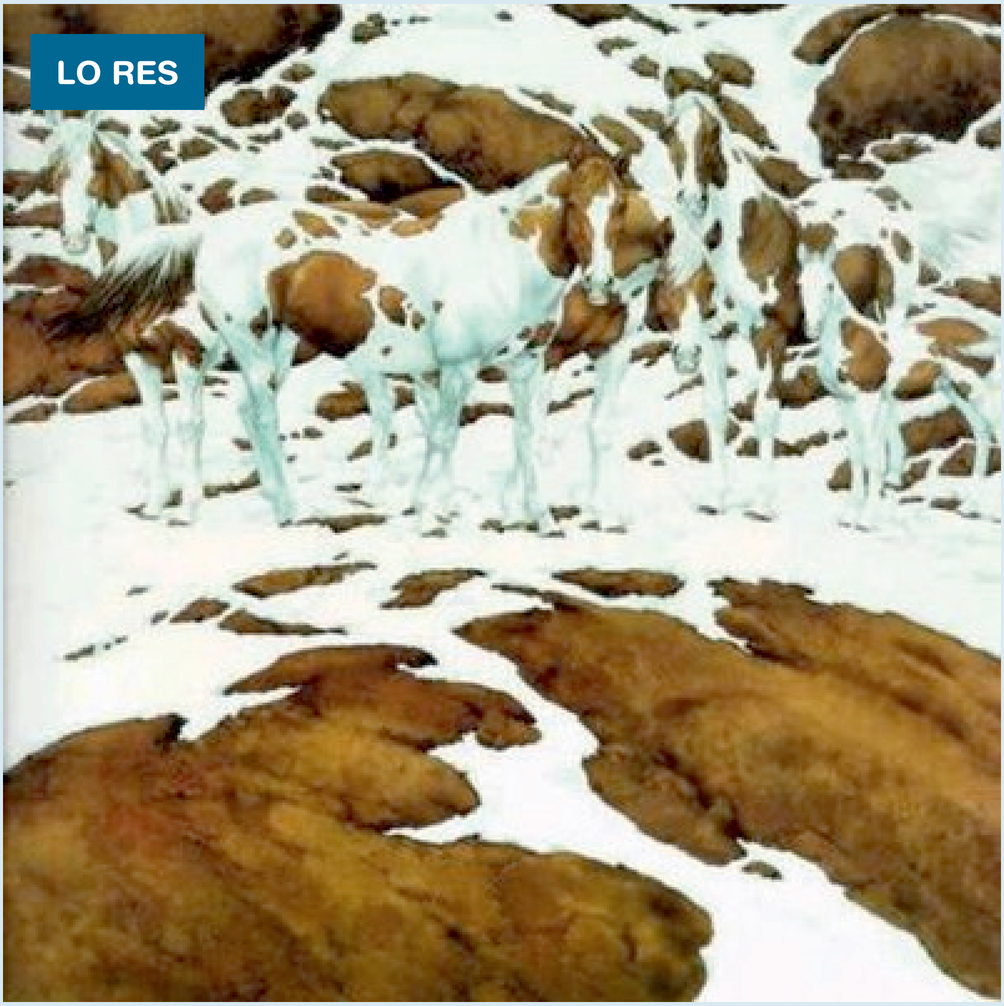

Figure 4.1 Doolittle's painting 'Pintos'

white and brown. It is likely that you needed time to find out that five horses are embedded in the image. The fact that you needed some time illustrates that perception is an active, complex process, and not just a passive system registering the external world. It might fail to organise percepts correctly, so that you do not perceive a visual scene as it is in reality. For instance, in Doolittle's painting, one might see only four horses instead of five. One crucial point, to which we shall return in this chapter, is that **representations** of reality are not the same are reality: a page of this book is obviously too small to host five horses! We can see something that does not exist (horses) and we can fail to see something that in fact exists (a drawing on a page of paper).

Doolittle's painting illustrates what visual perception is all about. Obviously, our perception of the world is not limited to visual input. Many other types of inputs such as sounds and odours provide useful information. This relates to what is known as 'sensory modalities'.

Box 4.1 RESEARCH CLOSE-UP: **Oscillatory processes**

An oscillatory process is determined by the following equation $y = A \sin (k2\pi)$. A stands for amplitude (how far from zero the oscillations go). For example, sea waves may range from a few millimetres to dozens of metres, and this would be indicated by different values of A. The coefficient k stands for the number of oscillations per second; it is called frequency and is noted in hertz (Hz). The inverse of k is $p = 1/k$, and is the wavelength. Another parameter which does not appear in the above equation is wave speed. On the sea, the waves may travel more or less fast. As for electromagnetic radiations (including visible light), the speed is around 300 000 km/s!

Electromagnetic radiations are classified according to their wavelength. Visible lights are radiations of wavelength in the range between 470 and 700 nanometres (nm; a nanometre is one millionth of a millimetre). Thus, the eye perceives only a fraction of the radiation range. It is worth having a brief look at the other wavelengths, and Table 4.1 shows the relationship between wavelength and type of electromagnetic radiation.

What does this table mean? That most of the electromagnetic reality is out of the reach of our perceptual system. For example, we need complex devices to detect X-rays, a more intense form of 'light'. We also need a device to capture infrared rays. Our perceptual system captures a narrow window within the range of electromagnetic radiations. As this is the fraction of light we are sensitive to, our senses make us believe that this is reality. For example, you perceive your pen as red because it reflects the light mostly in the frequency corresponding to red. But you are unable to see whether your pen is hot (since you touched it an instant ago) or whether it is cold. If you were sensitive to frequencies indicating heating radiations, you could know whether your pen is hotter that the objects around it just by looking at it. If your eyes were sensitive to much shorter wavelengths than it is in reality, you could see X-rays and thus scan inside everyone's body!

Table 4.1 Wavelength. The spectrum of electromagnetic radiations

Wavelength (in metres)	Class	Visible
10^{-12} m	Gamma rays	No
10^{-10} m	X-rays	No
400 – 700 10^{-9} m	Light	Yes
10^{-6} m	Infrared	No
10^{-3} m	Radar	No
10^{-0} m	FM Radio	No

A **sensory modality** is a dedicated subset of the nervous system that responds to specific physical inputs. For instance, audition is the detection and interpretation of air motions or vibrations, and balance detects the direction of gravity and corrects the body posture accordingly so that we can keep upright. Sensory systems inform us not only about many aspects of the world around us (**exteroception**), but also about what happens inside our body (**proprioception**).

There are important limits in what the senses can detect. Consider for instance the fact that lights that are too intense force us to close our eyes or that motions that are too slow cannot be perceived. These limits imposed upon our ability to capture the physical aspects of the outside world narrow the window of the portion of reality we are sensitive to (see Box 4.1). For this reason, sensory systems are said to be 'bounded'. These boundaries, in turn, force the perceptual system to make some assumptions about how events are connected in order to build a representation of the outside world. Unfortunately, the assumptions can be proven wrong in some circumstances, leading to perception being deceived by the senses (see the section on illusions below). Table 4.2 lists the most common human senses and the type of physical entities they detect. (The exact number of sensory modalities is debated.)

Each physical property is captured by specific sensory neurons, which are called **sensors** or **receptors**. For instance, photoreceptors are neurons sensitive to light. The role of the sensors is to translate physical stimulation into signals that are interpretable by the brain, a process that is called **transduction**. After transduction has taken place, the brain interprets the neural signals to build a representation. The percepts built up by neural analysis are not necessarily related to the outside world. While most of the senses presented in Table 4.2 are concerned with events occurring outside the body, many senses such as those related to arterial pressure and heart rhythm are concerned with events occurring inside the body. Such senses, which are grouped in the category of proprioception, are less accessible to consciousness. Still, through automatic controls, they play a key role in the regulation of the body. The senses related to the outside world are the focus of this chapter, since they concern how we understand and interact with our environment, including other people through social interactions.

Sensory systems receive information from their dedicated sets of receptors. The role of the receptors is to bring into the system the basic elements of perception that combine to form a **percept** (i.e. the mental representation of what is perceived). These basic elements are referred to as **sensations**, while more elaborated material is referred to as **perceptions**. Neither sensations nor perceptions can exist on their own. Both processes are entangled, and it is sometimes

Table 4.2 Senses. The most common human senses

Vision	The ability to detect electromagnetic radiations within a narrow band of frequencies (visible light)
Audition	The perception of air vibrations
Taste	The detection of various chemical compounds in food and liquids, by a variety of receptors. Each receptor is in charge of detecting the presence of specific molecules
Smell	Similar to taste, but detects chemical compounds in the air
Touch	The perception of pressure on skin
Equilibrioception	The perception of balance; it is based on the detection of the direction of gravity
Thermoception	The perception of temperature
Nociception	The perception of pain. Yes! This is an independent sense served by its own specific set of receptors

difficult to determine where sensations end and perceptions begin. It is important to realise that multiple specialised sensory systems participate in constructing an internal representation of reality.

Bottom-up and top-down processes in perception

Our brain processes the information collected by our senses in order to construct percepts. If this processing were perfect, we would just capture an exact copy of the external world. However, this is not the case. Information can be ambiguous, and often the brain uses previous knowledge to remove ambiguity. For example, try to read the words presented in Figure 4.2 as fast as possible.

When reading this, did you notice that the H and A had the same symbol?

It is likely that you read the words 'the cat'. Have you noticed that the same symbol was used to represent an A and an H? Somehow, your brain managed to decide that it was an A in one case and an H in the other, and not vice-versa. The visual features constituting the letters were captured by the visual receptors. To interpret the signals, they were then forwarded to increasingly complex levels of processing (see Chapter 16 for details). This kind of information processing, starting from basic sensory information up to more conceptual information, is called **bottom-up processing** or **stimulus-driven processing**.

Another mode of processing – called **top-down processing** or **concept-driven processing** – is also at work: high levels of cognition control and regulate events affecting lower levels of cognition. In the example of Figure 4.2, high-level cognition and the knowledge that 'the' and 'cat' are words helped perception to decide which word was read. Another interesting example of top-down processing is offered by the 'word-superiority effect', which we will describe in Chapter 9. In a nutshell, the effect shows that it is easier to recognise a letter when it is part of a word than when it is presented individually. As a final example of top-down processing, consider search, where higher levels of cognition direct perception through attention (see Chapter 5). When you are looking for a pen, you process objects one after another until an object matches your target.

Figure 4.2 Word recognition

Top-down processes may use context, expectations and knowledge to structure the information sent by bottom-up processes. You will see in the next chapter how knowledge influences attentional processes. Here, we focus on the bottom-up processing aspect of perception; that is, how the basic sensory building blocks are put together to make a percept. You should be aware that the influence of top-down processes is limited in that they cannot correct some assumptions made by bottom-up processes. For example, as we live in a three-dimensional world, our visual system has been designed by evolution to process a three-dimensional space; as a consequence, it automatically interprets ambiguous data in a way that satisfies the constraints of a three-dimensional world, and there is not much that top-down processes can do to change this interpretation.

Consider the Ponzo illusion in Figure 4.3. The top horizontal segment seems longer than the bottom one, but in fact the two segments are of equal length. One standard way to explain this illusion is to consider it as a normal interpretation in a three-dimensional world. Imagine that you are standing on the tracks of a railway, looking at objects placed at different distances. If an object closer to you has the same apparent size than another object located farther, this implies that the object located farther is bigger. This interpretation is automatically generated by the perceptual system without any possibility to influence it. Even though you *know* that the top segment has the same length as the bottom segment, you cannot correct your perception that it is bigger. This is clear evidence that the process of reconstructing reality is under the influence of knowledge. The perceptual system uses a set of assumptions to interpret data; when one or several of these assumptions are incorrect, an illusion occurs. The fact that the perceptual system makes assumptions or hypotheses about what to expect in the environment is a crucial aspect of the cognitive system (Gregory, 1980). Perception is not an isolated process but is under the influence of knowledge.

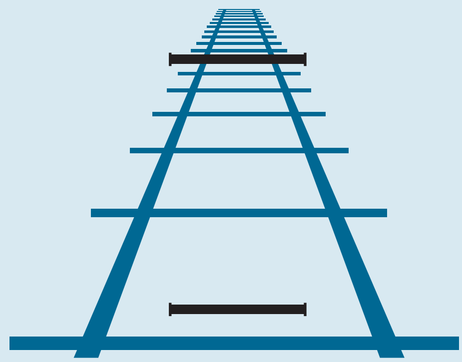

Figure 4.3 The Ponzo illusion

The psychophysics approach

As we have seen, perception is the product of intertwined processes. Figure 4.4 illustrates how these processes influence the final percept. Which one of the two squares in the middle of the two large squares is brighter? On a scale from 1 (white) to 10 (black), try to estimate the level of grey for these two squares.

In fact, the two squares have the same intensity of grey. (If you are not convinced, cut a piece of paper so that you can see only the two central squares). This illusion is a startling example that perception does not mirror reality. It actually distorts it. Psychophysicists –psychologists interested in the relationship between physical stimuli and their perception – have addressed three central questions about how perception captures reality.

The first question has been to determine the **detection threshold** of the receptors. In other words, what is the minimal strength of a stimulation so that we can notice its presence?

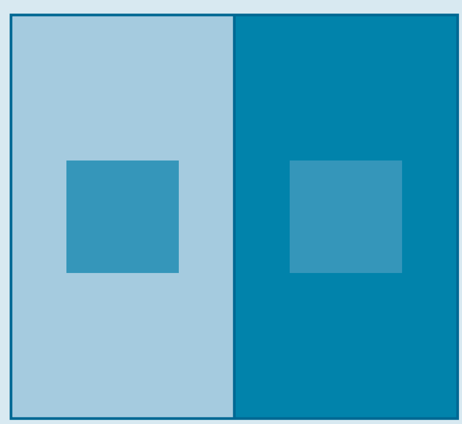

Figure 4.4 Two dark squares
Pay attention to the squares in the middle. Which one is brighter than the other?

This minimum value is called the *absolute threshold* of the receptor. Each class of receptor is characterised by a different absolute threshold. For example, consider the sensation of pressure on the skin. What is the lowest-intensity stimulus that can be detected at a given location? Below the absolute threshold, our sensors are unable to detect the presence of a physical stimulus. Sensors also have a maximum level of detection. For example, we cannot hear sounds of very high frequency (ultrasounds). Thus, sensors have windows of responsiveness (i.e. windows between minimum and maximum levels of detection).

The second question concerns the minimal difference in amplitude that can be detected between two stimuli. Of particular interest is the issue of **just noticeable difference**. How sensitive are the receptors to variations in magnitude? For example, you apply pressure on your skin. How much additional pressure should you apply to notice a difference? Such a difference relates to what is known as the *relative threshold*. In general, receptors are not sensitive to any variation but only to variations within a certain range.

The last question – the so-called question of **scaling** – is about quantifying the relationship between stimulus intensity and subjective perception. Here, the questions are of the following kind: how much must the magnitude of a physical stimulus increase so that you have the subjective impression that the stimulus is twice as strong in intensity? The relationship between physical intensity and subjective intensity is not linear. Perception does not merely reproduce the external world: a light twice as intense (i.e. the physical amplitude is doubled) does not lead to the subjective perception of a light twice as bright (i.e. subjective perception). (See Box 4.2 for more detail about scaling.)

Reality, therefore, can be deformed when captured by our sensory modalities. How did humans manage to survive and even dominate earth using such (apparently) limited information? To address this issue, we need to have a closer look at the visual system. This will illustrate how a sensory modality, by picking up only a few samples of the external world, is nevertheless able to inform us about key events. In the next section we briefly survey the rules used by the cognitive system to organise two-dimensional information.

Two-dimensional visual information (Gestalt theory)

As we have seen in Chapter 2, Gestalt psychology emerged at the beginning of the twentieth century. Its leaders, such as Wertheimer, Köhler, and Koffka, viewed perception as the process by which object form is reached after self-organisation of basic elements – 'the whole is more than the sum of its parts'. By emphasising self-organisation, Gestalt psychology was in opposition to the traditional **reductionist** approach that science normally uses to understand phenomena. It is mainly known for developing the so-called 'Gestalt laws of perception' (see Figure 4.6), which describe how the mind coordinates the perception of several basic parts to generate a whole percept.

For Gestalt psychologists, perceptual organisation may be reduced to the *principle of Prägnanz* (*Prägnanz* means conciseness or simplicity in German). This principle states that we organise the perceptual input so that we perceive the simplest and most stable forms. However vague this definition, this principle is the central assumption of the Gestalt theorists and the starting point for several laws. The *law of proximity* states that perception clusters objects according to their proximity (Figure 4.6, Panel A). Thus, if a set of objects are close to each other and separated

Box 4.2 RESEARCH CLOSE-UP: **Stevens's law**

Fechner (1860/1966) was among the first to address the question of scaling, which really boils down to studying how much distortion our senses impose on physical stimulations. He showed that, when the magnitude of the stimulus increases multiplicatively, psychological impression increases only linearly. Fechner's law is not precise and is also not valid for some sensory modalities. However, the fact that perceived intensity increases less rapidly than physical intensity is an important result. Stevens (1957) estimated that the relationship linking the brightness of light (stimulus intensity) to perceived brightness (perceived intensity) is characterised by a proportional increase with a ratio between 0.3 and 0.5 (see Figure 4.5).

Steven's law links stimulus intensity to perceived intensity. The law links stimulus intensity (here in arbitrary units) to perceived intensity (also in arbitrary units). The bold line shows that perceived intensity increases only as a power function of stimulus intensity, and not as a linear function (showed by the dashed line). Thus, perception departs from a mirror image of reality

The dashed line shows what a linear relation between intensity and perception would look like. The bold line plots the physical – psychological relationship as determined by Stevens's equation; note that the difference between two perceived intensities is not the same as the difference between two physical magnitudes, as it is with the linear function. Let us consider the four points A, B, C and D in the graph. The points are located at equal distance on the scale of intensity (the physical stimulus). However, the progression of the perceived brightness is different: at each step, the difference in perceived brightness diminishes. Stevens's law is a powerful demonstration that perception applies transformations to the stimuli from their very detection.

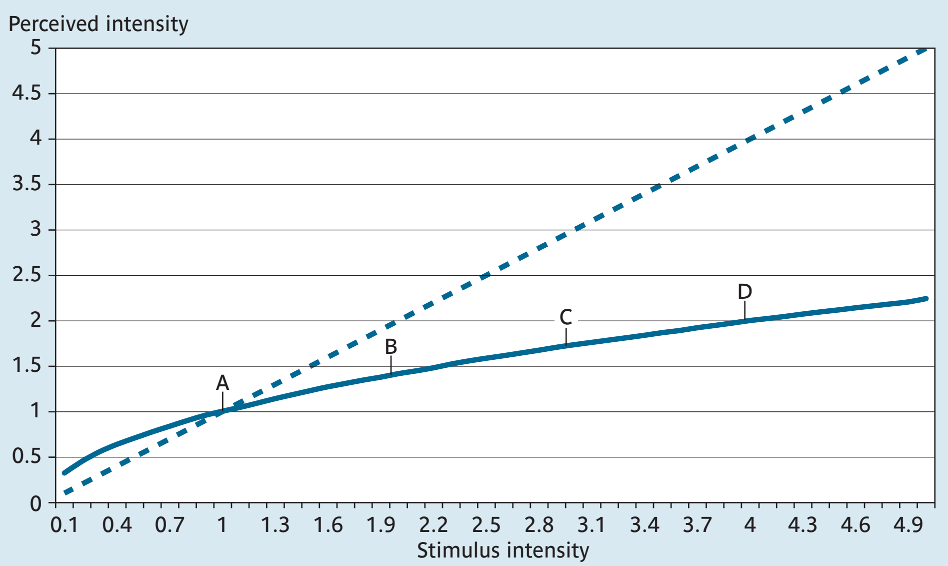

Figure 4.5 Psychophysics of brightness

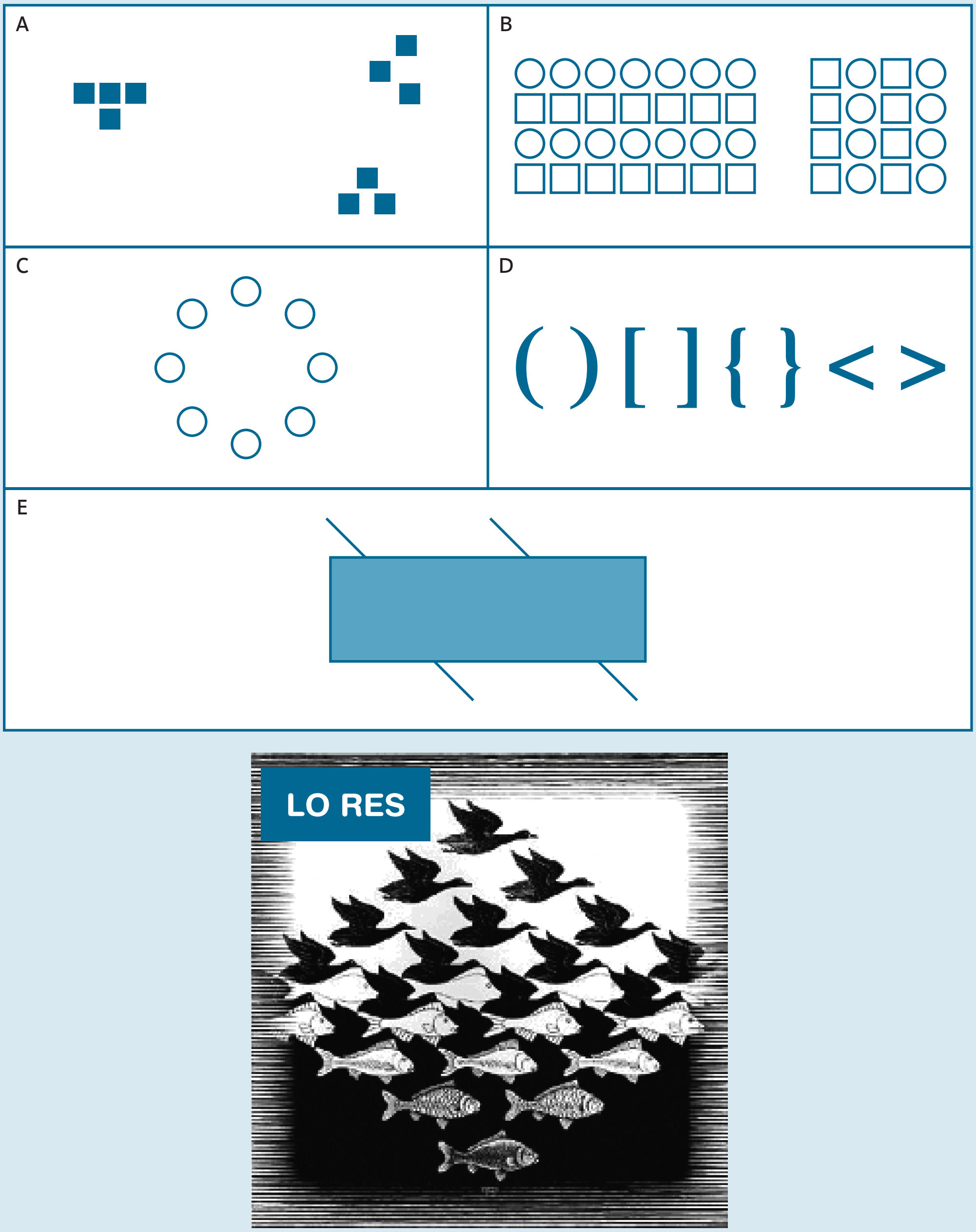

Figure 4.6 Gestalt laws

Drawings showing how the Gestalt laws organise perception. Panel A shows the law of proximity. Panel B shows how similarity affects perception. Panel C offers an illustration of the law of closure. Panel D shows how symmetry influences grouping of perceptual elements. Panel E illustrates the law of good continuity. Finally, the drawing at the bottom shows an artistic use of law of figure-ground segregation.

Source: Escher, *Sky and Water*, 1938.

tinuity tricked your perception! Objects presented in spatial sequences are clustered together to form lines. The law of continuity highlights the fact that we use contours of objects to form perceptual units.

Finally, the *law of figure–ground segregation* states that perception tends to structure the visual field into two parts: a figure and a ground. The bottom panel in Figure 4.6 illustrates figure–ground segregation. Whether you are looking at the birds or the fish, the figure is the focus of attention and the ground recedes to the back of perception and attention. What is striking in Escher's drawing is that you can focus on either the birds or the fish in the air–water frontier. However, if you follow the figure from top to bottom, the perception is forced to see birds at the top, then either birds or fish in the middle, and finally fish as you reach the bottom. It is very difficult to force perception to see the birds between the last rows of fish. Another standard example of figure–ground segregation is offered by **bistable percepts** – stimuli that can be perceived as two different objects. In the Rubin illusion (see Figure 4.10 below), you can switch at will between the perception of a vase or of two faces.

Gestalt theory has been very good in describing the organising principles that guide perceptual processes. However, little or no explanation has been put forward to explain the existence of these laws (see Box 4.3).

The third dimension: depth

The Gestalt laws illustrate how we organise simple visual patterns into larger perceptual structures. However, they do not address the question as to how we localise these percepts in our three-dimensional world. To realise the importance of three-dimensional representations, imagine trying to assess the distance of the cars coming into a junction, or finding your way in a new building. The perception of the third dimension is crucial, for the obvious reason that we

Box 4.3 RESEARCH CLOSE-UP: **The advance of science**

Science is a difficult subject matter. Before they come to the correct conclusion, scientists are often mistaken. For example, it took centuries to have a basic understanding of the laws of gravity. It had also been a long journey before scientists came to understand the basic brain processes that underlie perception. The French philosopher Descartes was one of the first to give it a try. Unfortunately, the science and technology of the seventeenth century did not allow him to untie this knot. Three centuries later, the Gestalt School, so successful at the time in explaining two-dimensional perception, also attempted to explain perception with neural mechanisms. Wolfgang Köhler postulated that percepts emerge from the electrical field generated by our neurons. The pattern of electrical activity was supposed to mirror the topological features of the percept. If this explanation was correct, then disturbing the electrical field – for example, by placing metallic implants in the brain – should perturb the electrical activity and thus impair perception. This hypothesis was put to the test by Karl Lashley and colleagues (Lashley *et al.*, 1951) in experiment using monkeys. The results showed that the metallic implants did not alter performance. The conclusion is that percepts are not the result of a patterned field of electric activity.

live in a three-dimensional world. Because the eyes occupy different locations in space, they have two different images of a scene, and these images are two-dimensional. The question is then to understand how we perceive the world in three dimensions using two two-dimensional images.

To perceive depth, our visual system uses both monocular cues (i.e. cues from one eye) and binocular cues (i.e. cues from both eyes). Three monocular cues are of particular importance with respect to the relative distance between objects. **Occlusion** refers to the fact that an object partly hidden by another object must be behind it. **Texture gradient** refers to the fact that objects, such as walls, have a specific texture that changes with distance. **Motion parallax** refers to the fact that, when you are moving, objects that are closer seem to move faster than objects that are far away. Next time you travel by car or train, notice how the objects close to the road or the railway seem to pass by very quickly, while the buildings at the horizon seem almost motionless.

Two binocular cues are particularly important: **binocular disparity** and **convergence**. The eyes have a slightly different view of the world. The difference between the images is called disparity (see Figure 4.7). To have an idea of how much retinal images are affected by disparity, try the following. Position a finger at 30 cm from the nose. Look at the finger with the left eye only and then with the right eye only. You will notice that, depending on which eye is open, the finger is in front of a different part of the visual scene – the so-called *shift in the image*. The difference between the two retinal images provides information about distance and thus about depth: the farther the object, the smaller the shift. The other cue that is used to estimate depth is convergence, which the coordinated and inwards movement is made by the two eyes to focus on a near object. The information provided by convergence is independent of the image on the retina: the brain is informed of the gaze direction of each eye by the muscles controlling the position of the eyes.

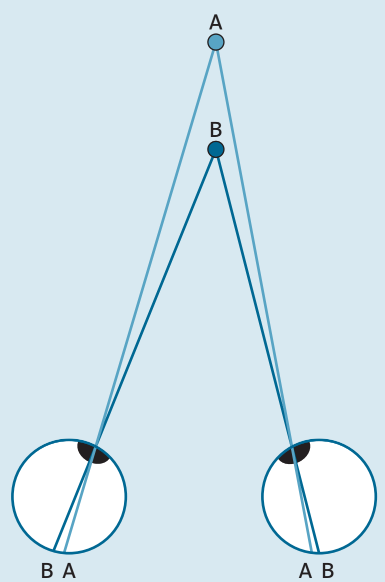

Figure 4.7 Binocular disparity
Here the eyes are focusing on two different objects, A and B. A is located farther than B. You can see that the image of both objects in one retina will mirror the image of the other retina.

Thought question 4.1 **Is one eye enough?**
Considering the fact that binocular cues are employed to reconstruct the third dimension used by two-eyed people to navigate in their environment, how would you explain the fact that people with vision in only one eye can navigate safely in a three-dimensional world?

When the visual system is deceived: visual illusions

Our visual system has been shaped by evolution to rapidly detect potential threats or opportunities. Although it is well adapted to the kinds of natural environments in which it has evolved, it can be confused surprisingly

Box 4.4 IN FOCUS: **Watching the ghost**

If people report seeing a ghost, you are likely to think they are drunk or under the influence of drugs. Yet it is possible to see things that do not exist. The mind sometimes plays little tricks on us and makes us believe that we see something while there is actually nothing there in reality. From today on, you can tell your friends that you can see things that do not exist, and that you can even show them these things if they do not believe you! Consider Figure 4.8. You can clearly see a white equilateral triangle, which is brighter than the background. Please, feel free to follow the contours of the triangle. By doing so, you will notice that the triangle has actually no contours and, in fact, does not exist. The illusion occurs because of the way information is processed bottom-up and cannot be corrected by top-down processes; in this case, the brain tends to complete disconnected visual elements so that they form a whole (Gestalt law of closure). Although you now know that this is an illusion, you still perceive the bright white triangle!

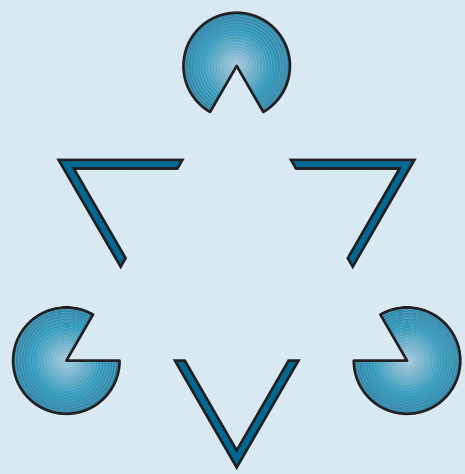

Figure 4.8 Kanizsa triangle

easily. **Visual illusions** are physical stimuli that deceive our perception. It is useful to study visual illusions because they provide clues about how our cognitive system processes information. By understanding how stimuli cause illusions, we can identify which processes have failed, and why they have failed. Let us have a look at several famous visual illusions that illustrate some aspects of our cognitive system particularly well.

The conclusion from the Kanizsa illusion is that those who report having seen things that are impossible should not be automatically considered as drunk or on speed; rather, they may just have been tricked by their perceptual system. They have really experienced something that is not there, just as you can see a triangle that is not drawn!

A striking example of the influence of knowledge on perception is provided by Figure 4.9. Something is hiding in this figure. What could it be? Take a few seconds and try to find out.

What is hiding in the picture is a Dalmatian dog. If you have not found it, have a second look. It is likely that you will find it now. Knowing that there is a Dalmatian in there helps you to find it. As another example of the influence of knowledge on our understanding of the environment, consider watching the North Pole star. Knowing that it marks the direction of the North Pole (as it is approximately aligned with the Earth's axis of rotation), you can now 'see' where north is and thus by inference where east, west, and south are. If you see the same star without knowing that it is the North Pole star, you just see a pale dot in the sky. What a change in the perception of the whole environment is made by just adding a piece of knowledge to a star!

Figure 4.9 Dalmatian dog
Looking for something?

We have just considered the influence of knowledge on the perception of a known object. Another interesting case is where one single visual stimulus can elicit two different percepts. This raises an intriguing question, not remote from science fiction: can we control whether we can perceive the same stimulus as one or another object? A positive answer would be evidence for a clear influence of our knowledge and control processes on how the perceptual system organises information. Look at the left panel of Figure 4.10 for a few seconds.

Do you see a rabbit or a duck? If you see the rabbit, try to see the duck! Once you can see both, you can train yourself to switch your perception from the duck to the rabbit and vice versa. You can control what you see! As already mentioned, that kind of stimulus is called a **bistable percept**. It is interesting to note that, in the Rubin illusion (right panel of Figure 4.10), the switch in perception between the vase and the face can be identified by neural recordings (Hasson *et al.*, 2001). Two more important points about the pictures in Figure 4.10. First, you initially saw either the rabbit or the duck. The reason why you see one rather than the other is unknown. Second, you cannot see the two animals at the same time, even though the visual features necessary for imagining them together are present. Once a visual feature is considered to be part of one animal, it cannot be part of the other, and this precludes any simultaneous perception. Perception has to make a choice!

Figure 4.10 Rabbit – duck – Rubin
Two examples of perception switch: the rabbit-duck figure and the Rubin illusion.

Visual illusions shed light on some interesting features of perception. So far, we have considered illusions related to objects that actually exist in our environment (for example, Rubin's vase illusion). But we can also see objects that are physically impossible, as shown by Figure 4.11. These examples illustrate the extent to which our perception is automatic and how conditioned we are by our senses. Just like in the movie *The Matrix*, we are limited by our ability to capture meaningful events in the external world. We are bounded by our perceptual system to such an extent that it is actually surprising that we can see beyond our nose's.

Thought question 4.2 Four-dimensional objects

We have seen that it was possible to trick the mind so that it perceives impossible three-dimensional objects. Is it possible to perceive four-dimensional objects?

Perceptual illusions have sometimes led to curious debates in popular science. A famous example, which illustrates how defiant we should be with respect to our own perception, is the case of the *Face on Mars*, discussed by the astronomer Carl Sagan (1996). When the Viking orbiter was approaching the planet Mars in 1976, it took many photographs among which one was quite disconcerting (see Figure 4.12(a)). As you can see, one of the rocks really looks like a face. Unsurprisingly

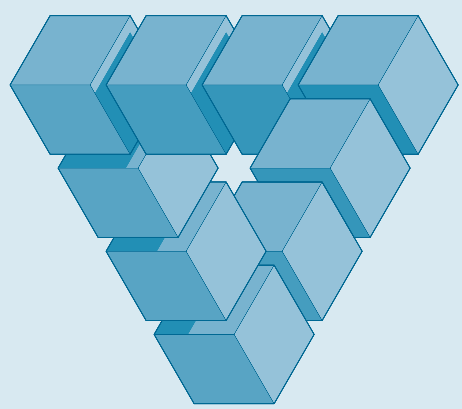

Figure 4.11 Tricking the mind: an impossible object (by Escher)

Figure 4.12(a) Mars rock from afar
Surface of Mars from afar. The black rings indicate the rock that looks like a face.

Figure 4.12(b) Close-up view of Mars rock
Region of interest on the surface of Mars. Where has the face gone?

but also unfortunately, this rock has been interpreted by some as a sign of extraterrestrial intelligence. Since then, astronomers had a hard time making people understand that this was not proof of intelligence on Mars. In his book, Sagan hoped that a new and closer photograph would disprove this speculation. It actually took more than 20 years before a better photograph, much closer to the surface, could be taken. It clearly showed that the face was just an effect of shadows (see Figure 4.12(b)).

Thought question 4.3 UFOs
Many people have reported seeing UFOs (Unidentified Flying Object). Assuming that that these people did not hallucinate, use what you have learnt in this chapter to account for these reports.

Perception is cognition

The purpose of perception is to recognise objects so that actions can be selected and carried out. In most cases, to recognise an object is to check whether or not we know the object under consideration. For instance, we have to know Paul to recognise him as Paul. Another example will illustrate that recognition is not a mere bottom-up perceptual process. Let us travel back in time and ask a person of the thirteenth century to identify a computer and to tell us what it is used for. It is likely that we will not get any answer. The person does see the computer (showing that perception works) but does not know what it is. The process of object recognition is at work all the time: recognising the computer with which you work, the chair on which you sit, the remote control to change channels, the letters in the book you are reading, and so on. To identify an object is to put into consciousness not only the recognised object but also what we know about it. The process is central in that it retrieves information about the environment and allows us to structure the world as sets of potentially useful or harmful objects. The information recognised is then evaluated with respect to the task at hand in order to assess its utility (see Chapter 10).

The question that psychologists have addressed is how the recognition of complex objects can be made so quickly. We now describe two theories that were developed to account for object recognition in general. Later, when we talk about face recognition, we consider more specific theories that are limited to only one class of objects – in that case, faces.

Marr's theory

Marr's (1982) objective was to develop a computer model of object recognition. As we have seen in Chapter 1 and will discuss further in Chapter 15, developing computer models helps us understand the consequences of theories. Marr's model describes the process of object recognition in three bottom-up stages. In the first stage, the visual information is used to generate an image of the scene in the outside world. The image represents the contours of objects. The contours are derived from averaging the light intensity of small regions. Such a representation is called a *grey-scale representation*. The contours are used to identify the edges and surfaces belonging to objects. The process determines which edges go together. Following the rules put forward by the Gestalt school, edges are grouped together to form *tokens*. As a result of this process, the representation consists of sets of tokens. This representation is known as the *full primal sketch*. Two more steps have to be carried out before recognition is completed. The first step is to incorporate the third dimension. This is done by adding a range map to the full primal sketch. The *range map* adds information about distance to every point of the two-dimensional (2D) representation. It is assumed that the perceiver uses the depth cues we described earlier to allocate a given distance and angle to each object. Distance and angle are measured from the point of view of the perceiver. This representation of the world (called **egocentric representation**) is dependent on the perceiver. If the perceiver moves, the visual angle changes and so does the representation. At this stage of processing, the representation is called the *2½ dimensional sketch*. The last step is to match the target object with an internal representation. To do so, the perceptual system has to move from view-dependent to view-independent perception (also called **allocentric perception**). Figure 4.13 illustrates these steps.

The four panels of the figure are a rough approximation of the result of each processing stage in Marr's theory. The top-left panel shows the input on the retina. The image has been reformatted using larger pixels to simulate the fact that the retina has a very high, but limited, number of cells for encoding a visual scene. This first image is the raw material from which the perceptual system will extract the meaningful information and eventually recognise a plush rabbit. The top-right image shows the primal sketch. This grouping of visual elements is done in a 2D representation – that is, it is impossible to distinguish in this image what is figure (the rabbit) from what is ground (the rose on the wall). The third image, bottom-left panel, corresponds to the state of processing when the image is set in 2½ dimensions. In this image we can see which tokens belong to which objects and also how the surfaces relate to each other. Now we can see that the drawing of the rose is behind the rabbit. The last image, bottom right, shows the result of the processing. A full three-dimensional image (even though on a 2D paper!) which clearly

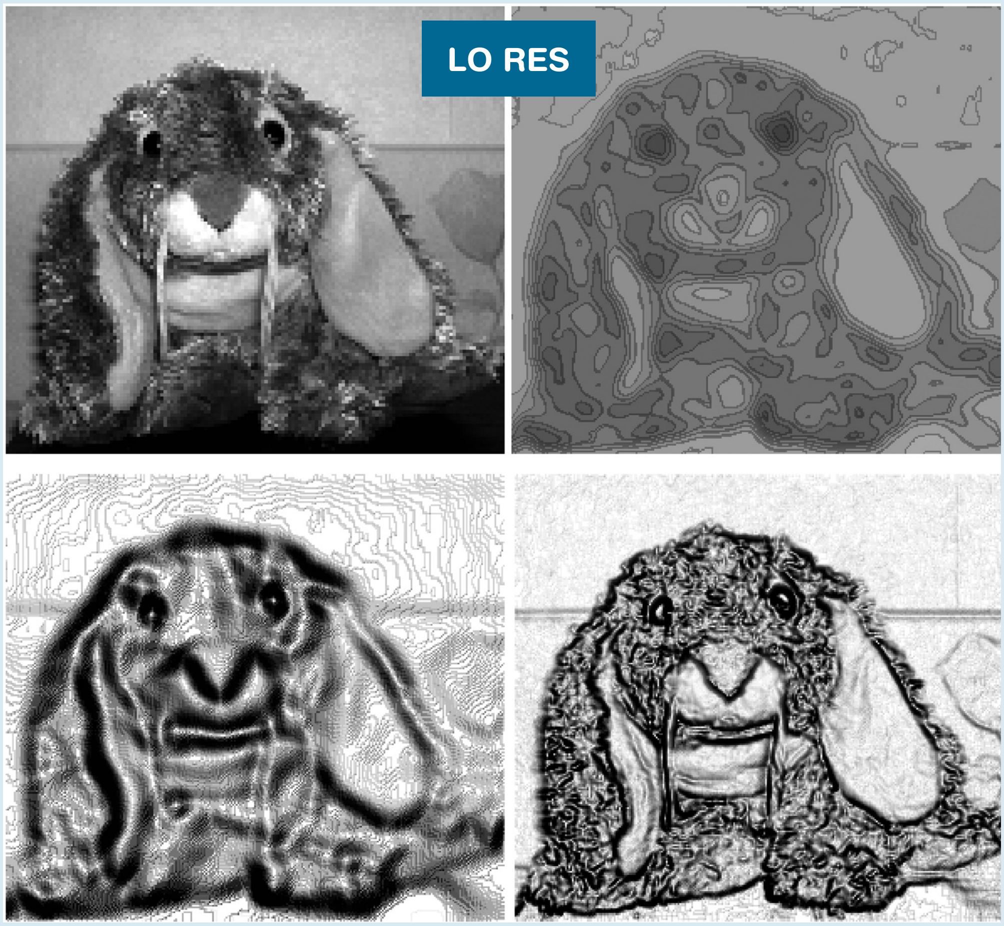

Figure 4.13 Plush rabbit
A plush rabbit illustration of Marr's theory.

shows the relationships between the various elements of the image, providing instantaneous information about the size and distance of each and every part of the visual scene.

Among the key ideas put forward by Marr, the idea of using primitives to represent objects has proved very valuable. Marr proposed that these primitives are *cylinders*. Accordingly, objects are perceived as aggregates of cylinders (see Figure 4.14). This idea of coding objects as a set of primitives has been used to develop a second important model of object recognition, which we will consider next.

Recognition by component theory

The recognition-by-component theory, developed by Irving Biederman (1987), posits the existence of 36 basic three-dimensional geometric shapes that constitute a visual alphabet used to represent objects. Biederman introduced the term **geon** (standing for geometric ions) to refer to

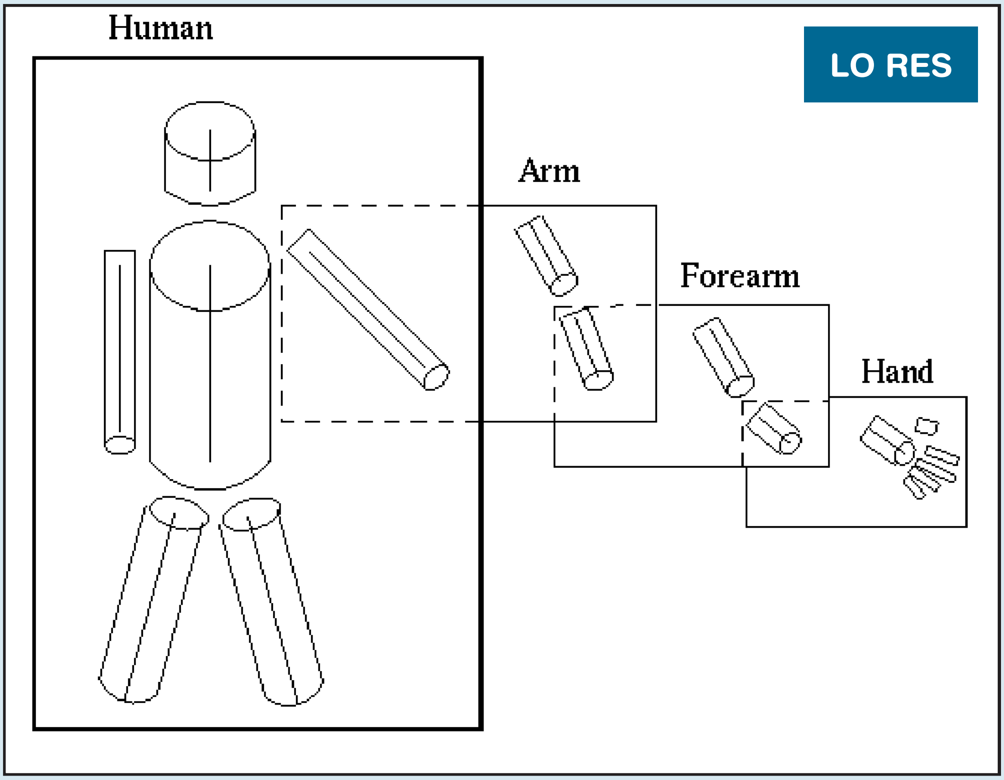

Figure 4.14 Marr's theory
A computational view of Marr's recognition theory.
Source: Adapted from Marr, 1982.

such basic shapes. Geons are a simplified representation of perceived objects as they code only for their gross features.

Figure 4.15 illustrates how geons (left panel) combine to code objects (right panel). According to Biederman, objects can be coded with a set of well-chosen geons, regardless of their complexity. The fact that we can recognise the objects presented in the right panel supports Biederman's view, to some extent. Indeed, if we could not recognise geons, we would not recognise the objects represented in these drawings!

Most of Biederman's model is concerned with the identification of geons. As with Marr's model, the first stage in processing the visual image consists of analysing luminance, texture and colour in order to yield a basic representation of objects. The basic representation is made of contours as in drawings. In the second stage, the properties of the regions (as delineated by contours) forming the image serve to identify which surface is part of which object. The outcome of this analysis determines the geons. In a third phase, the arrangements of geons are matched against a representation in memory to identify the object. Partial matches might occur but they do not lead to recognition. It is clear from this cascade of events and analyses that Biederman views the process of recognition as mostly bottom up. (See Box 4.5 for empirical support for Biederman's theory.)

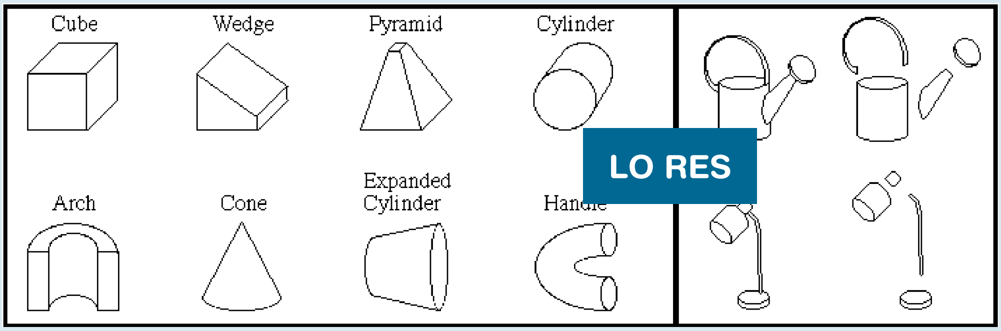

Figure 4.15 Geons

Geons and objects made of geons. Adapted from a cyberbook 'avian visual cognition' edited by Robert J. Cook which is available at http://www.pigeon.psy.tufts.edu/avc/toc.htm. The images are an adaptation of the pages of Kimberly Kirkpatrick section object recognition at http://www.pigeon.psy.tufts.edu/avc/kirkpatrick/default.htm

The reason why Biederman's model has been so important is that it provides theoretical arguments, backed up with empirical evidence, to solve a classic problem in psychology. Decades ago, a debate was raging about how objects are coded in the brain. Two opposite views of recognition were put forward. One part of the scientific community was arguing that objects were coded as single units in the brain. According to this theoretical standpoint, your grandmother would be recognised by one single neuron. The theory positing a one-to-one correspondence between neurons and objects is often referred to as the 'grandmother cell' theory (Gross, 2002). This hypothesis leads to a questionable conclusion: it implies that the number of objects we can encode is limited by the number of neurons at our disposal. But what do the free neurons do until they are allocated to encode an object? They cannot just wait in a queue, since biological systems have to minimise energy expenditure: evolution does not tolerate unused resources. On the other hand, if we believe that these neurons are allocated to carry out some other task, then this implies that any visual learning implies a loss of performance in another task, which is rather improbable.

The second view, which is also the current one, is that objects are coded as sets of visual properties and those neurons encode these properties. The same neuron coding for a particular visual shape may be involved in the encoding of several objects insofar as the objects share this shape. Hebb (1949) introduced the concept of a **cell assembly** where neurons code for objects organised as sets of properties. Later, this view was adapted to the visual cortex with theories of recognition such as Marr's and Biederman's, which state that objects are coded as groups of visual, easy-to-recognise visual features. The importance of Biederman's geon theory is then to have promoted this view of perception as the processing of groups of visual features.

Thought question 4.4 **Building the perception of new objects**

Biederman put forward the theory stating that we recognise known objects by assembling basic geometric items (geons). How can we perceive new objects according to this theory?

Box 4.5 CLASSICAL EXPERIMENT: **Hayworth and Biederman's experiment**

If objects are coded as sets of basic geons, then there should be a stage where geons are identified before being combined to form a recognisable object. Hence, eliminating some of the geons from a 2D drawing should impair recognition. This hypothesis was tested by Hayworth and Biederman (2006).

Hayworth and Biederman designed two classes of probe stimuli and three classes of target stimuli (see Figure 4.16). In the first class of probe stimuli, part of the line drawings of two-dimensional images were deleted (*Local Feature Deleted condition*). The entire image of the object was thus altered but all the geons were still present in the image. For the other probe stimuli, they just deleted one entire fraction of the object (presumably coded by one geon) and thus obtained images where entire parts of the objects were missing (*Parts Deleted condition*). Three conditions were used for the target images. In the identical condition (I condition), the images were a copy of the probe images. In the complement condition (C condition), the images consisted of the missing parts of the probe images. In the different exemplar condition (DE condition), the images consisted of other images that were altered with the same process. Hayworth and Biederman presented participants with target-probe pairs and asked them to tell whether the reference and the target were identical, complementary or different images. It turned out that deleting an entire geon made the task much more difficult, as indicated by accuracy (percentage correct).

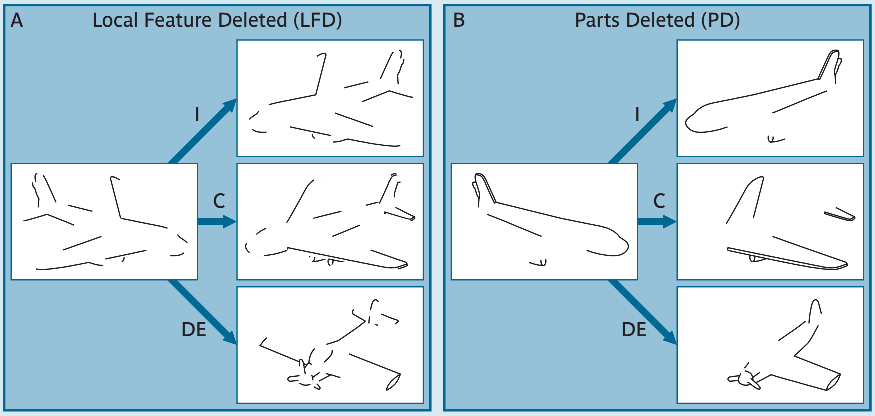

Figure 4.16 Hayworth and Biederman

Images from Hayworth and Biederman's experiment (Fig. 1, p. 4025). A minor alteration of all the geons yields images of the Local Feature Deleted condition. Getting rid of a part of an object (supposedly an entire geon) yields images for the Parts Deleted condition.

Complex objects: faces

Why are faces that attractive for researchers? Well, they are some of the objects with which we interact most often. Moreover, they convey emotions and as such are important for social communication. What is striking is that we can recognise the face of someone we know whatever the angle, lighting and the expression displayed. The variability of traits has little influence on our ability to process a face.

Face perception involves several stages (Liu *et al.*, 2002). Visual features are assembled according to the processes described in the previous sections. Then, the brain computes features such as gender and provides an estimate of age. Three lines of evidence support the hypothesis that specific processes deal with faces. First, unlike many familiar objects (such as a pair of scissors), faces are more difficult to recognise when they are upside down. Second, photographs of faces are more difficult to recognise than familiar objects when they are presented as negatives. The third line of evidence is offered by data from neuroimaging techniques. Using fMRI, Kanwisher and co-workers have found that a specific region of the brain called the 'fusiform face area' responds preferentially to faces (Kanwisher *et al.*, 1997). Similarly, single-neuron recordings in monkeys suggest that specific neurons respond to faces (Desimone, 1991). It is not to say that one given neuron is associated with one face, but rather that a given neuron preferentially responds to some facial features. Thus, a face is coded by a pattern of activation in the neurons concerned with face processing.

Not everybody agrees with the role of the fusiform face area. According to an alternative view, the fusiform face area is a visual area specialising in the visual processing of objects we are familiar with (Gauthier *et al.*, 1999). Faces are the objects we interact most with, so it makes sense that this part of the fusiform gyrus is activated when recognising faces. Thus, rather than being the product of a specialised part of the cortex, our ability to discriminate faces would be the result of becoming expert in face perception, much like expert chess players discriminate chess positions at a glance (see Chapter 12).

Figure 4.17 Attractive girl
A girl's face.

Facial expressions are a very efficient way to diffuse one's feelings socially. A face is also an important factor in a person's attractiveness. A research field in cognitive evolutionary psychology is interested in identifying the features of a face that make it attractive. How pretty is the girl in Figure 4.17?

The odds are that you found the girl in the photograph good-looking. But actually, you will never meet this girl, because she does not exist. This face is the average face of 14 young women. Scientists have noticed that average faces are usually more attractive than individual faces (Rhodes, 2006). Thus, it seems that beauty is inversely proportional to cognitive demand. As you have seen numerous female faces since you were born, you have a template for the typical female face and what it should look like. The closer the percept to a prototype, the simpler the perceptual analysis. Thus, a typical female face imposes a lower cognitive demand and looks pretty.

You can realise the complexity involved in face processing by considering the model developed by Schweinberger and Burton (2003). In this model, no less than a dozen modules are cooperating to complete face recognition. Many of the modules are mapped into anatomical structures. For example, the *Person Identity* module is linked to the anterior temporal cortex, and face recognition units are in the fusiform gyrus. This modularity accounts for the fact that, if a brain region is damaged, the model can predict the type of syndrome that will appear. The model also has a component dealing with changes in skin conductance when recognising a face: that is, the model accounts for the fact that our entire body responds to the presence of a familiar face. This component is often associated with emotions, and its presence makes the point, once more, that the components of cognition are highly connected.

Like other models, Schweinberger and Burton's assumes that the visual features of a face are stored in a different brain region from the semantic features (e.g. who is the person?). Thus, this model makes two important points. First, faces are visually complex objects, as the perceiver should be able to recognise the face despite changes in the exact position of the visual features (e.g. eyebrows position vary with mood). Second, faces mobilise neural networks in different brain centres.

Chapter summary

The main lesson of this chapter is that perception is far from being a passive process. To take the visual system as an example, visual representations prove to be different from photographs of reality. The perceptual system captures aspects of the physical reality and *actively* reconstructs a representation of the world. The representation arises as a result of two streams of information. The first stream, called 'bottom-up', is concerned with reorganising the information brought about by the receptors. Step by step, and following the laws of Gestalt, the visual system links together basic perceptual elements to form basic objects. The second stream, called 'top-down', selects relevant information based on past experiences. As counter-intuitive as it may seem, our perception is under the influence of what we know. Both streams constitute the essence of the dynamics that help us make sense of the world.

Visual illusions demonstrate that the mechanisms used to perceive are not infallible: we do not perceive reality as it is. The most striking case is when perception misinterprets the input data and provides a representation of an otherwise physically impossible object. The end product of perception is the percept: a chunk of information used to make sense of the world. Perceiving a chair or a fork leads to a percept, but so does perceiving a dog or a face. Percepts are representations of reality on which the perceptual system focuses, regardless of their size or features. They are the chunks with which we think. The remainder of this book will explain how we learn chunks and use them when memorising objects or when thinking.

Thought question 4.5 **Exobiology, exoperception?**

In astronomy, exobiology refers to the possible biological features of animals from other planets. In the movie *Predator*, an extraterrestrial creature has an electronic device that enables it to capture a part of the spectrum of the electromagnetic radiations which is different to what humans use: it can see heat. What are the things that we can perceive but that it cannot perceive? Conversely, what are the things that it can detect but that would escape our perceptual world?

Further reading

Blake and Sekuler (2005) and Shiffman (2002) provide an in-depth discussion of all the aspects of perception we have introduced in this chapter. They add many interesting, but also more complex topics, such as colour perception. These advanced textbooks will be of interest to those who wish to carry out experiments in the field of perception. Tovée (2008) is a good introduction to the neuroscience of vision. It describes the main anatomical features relevant to the psychology of vision and explains how the neural signal is forwarded from the retina to higher regions of the brain, with an explanation of the steps involved. The book might seem technical at first sight but, once you overcome the biological technicalities, it really provides an insight about how the brain reconstructs an internal image of the external world. Arnheim (2004) is particularly interesting because it links basic research with real-world data. It explains the perception of visual art and uncovers part of its mysteries. A very good book, for learning and for fun!