2

Theories and explanations of perception

Introduction

In this chapter we will examine key theories of perception. Two important approaches to the problem of perception come from almost opposite directions. For example, some psychologists feel that perception is direct (e.g. Gibson) and all the information needed is contained in the visual display. Others believe that the brain uses past experience and other influences to construct a version of reality (e.g. Gregory).

Other theorists (e.g. Neisser) have attempted to reconcile these opposing views, and yet others have taken an **artificial intelligence** approach using knowledge about computer programs to help explain perceptual processes (e.g. Marr's computational theory).

The central question which needs to be addressed is: How do we perceive the world around us so quickly and generally so accurately? Researchers have taken two main approaches to this question and they can be broadly divided into two categories:

- **Bottom-up theories**
- **Top-down theories**

Top-down and bottom-up processing

Top-down and bottom-up approaches have been applied to virtually every aspect of cognition including perception. The terms are used to refer to the different methods of interpreting sensory data and they come from the **information processing approach** to the study of areas like memory, attention, perception, etc. This approach constructs models of the mind that are similar to the flow charts used by computer programs, and it sees the human brain as a machine which manipulates information through a series of processing stages.

Bottom-up processing

This is processing which begins with an analysis of sensory inputs. It is based on properties of the stimulus such as the distribution of light and dark areas or the arrangement of lines and edges in the visual scene.

The information which is acquired from these sensory inputs is then transformed and combined until we have formed a perception. The information is transmitted upwards from the bottom level (the sensory input) to the higher, more cognitive levels. This kind of processing is also called 'data-driven processing' because the information (i.e. the data) received by the sensory receptors determines (drives) perception.

So, according to this idea, we observe an object (e.g. a chair) and the visual system extracts simple, low-level features like vertical and horizontal lines. These simple features are then combined into more complex, complete shapes like legs, seat and back, and we finally perceive a set of integrated shapes which we recognise as a chair.

Top-down processing

Top-down processing is the reverse of bottom-up and is used to describe the higher, more cognitive influences on perception. It is based on the idea that sensory information from the retina is insufficient to explain how we interpret visual information. We also

need to use our stored knowledge about the world in order to make sense of the visual input. This higher-level information works downwards from the top in order to influence the way in which we interpret sensory inputs. This kind of processing is also called 'concept-driven processing' because prior knowledge (stored mental concepts) comes from the top to determine (drive) interpretation of sensory input at the bottom.

Consider the following example:

Gregory claims that perception is a dynamic, constructive process

You should have no difficulty in reading this sentence as 'Gregory claims that perception is a dynamic, constructive process'. However, look carefully at the writing again. You will see that the 'cl' at the beginning of 'claims' and the 'd' at the beginning of 'dynamic' are identical, i.e. the image that falls on your retina will be exactly the same in both cases. Similarly, the 'n' in 'constructive' is the same as the 'u' in 'constructive'. The reason that you perceived them differently is that you read them within a context and that context influenced your interpretation of the written script. This is an example of top-down processing.

Now that we have explained the differences between bottom-up and top-down processing, you may be asking how it is that we can recognise a chair without reference to stored knowledge. Surely we only understand the concept of a chair because of our experience of seeing chairs in the past and our experience of sitting on them. All theorists, including bottom-up, acknowledge that there has to be some matching process between sensory information and stored mental representation in order for final identification (naming) to take place. We can only know the word 'chair' because it is stored in memory. The difference is that data-driven theorists assume that the matching process itself operates in a bottom-up direction until a match is found. The concept-driven theorists, on the other hand, assume that stored knowledge is required throughout the matching process. In other words, the question is whether our visual system can recognise a chair solely from a bottom-up analysis of individual features like four legs, differentiating it from other objects with four legs (e.g. a table, a dog, etc.), or whether our knowledge and experience with chairs in terms of factors, such as where they are likely to

be found and the different shapes they can be, helps us in a top-down direction to recognise the object.

In practice, there is often an overlap between bottom-up and top-down processing. No single theory that takes an extreme view on the use of the two processing approaches can explain *all* the evidence we have from perceptual studies. It seems likely that we use both top-down and bottom-up processing in our everyday life. Whether we use one approach more than the other would seem to depend on the viewing conditions. We will discuss this in more detail when we look at individual theories.

Matlin and Foley (1992) have suggested that there are three reasons why our perceptions are a reasonably accurate mirror of the real world.

- Stimuli are rich in information.
- Human sensory systems are effective in gathering information.
- Concepts help shape our perceptions.

Palmer (1975) carried out a study to investigate the interaction of top-down and bottom-up strategies. He showed participants various drawings which they were asked to identify (see Figure 2.1). The sketchy line drawings in Figure 2.1(a) are very difficult to identify out of context. However, when they are put into the context of a face (Figure 2.1(b)), even though the drawing is quite unlifelike, the squiggles are easily identifiable as facial features. In recognising Figure 2.1(b) we are using both bottom-up and top-down processing. We recognise the face as a whole because we recognise the parts, but we would not be able to recognise the parts without the context of the whole. In Figure 2.1(c) we are able to recognise the features out of context (i.e using bottom-up processing) because they are rich in detail.

Gibson's theory of direct perception

Description

James Gibson maintained that perception is a direct process. He firmly believed that there is enough rich sensory information in the patterns of light reaching the eyes – he called this the **optic array**
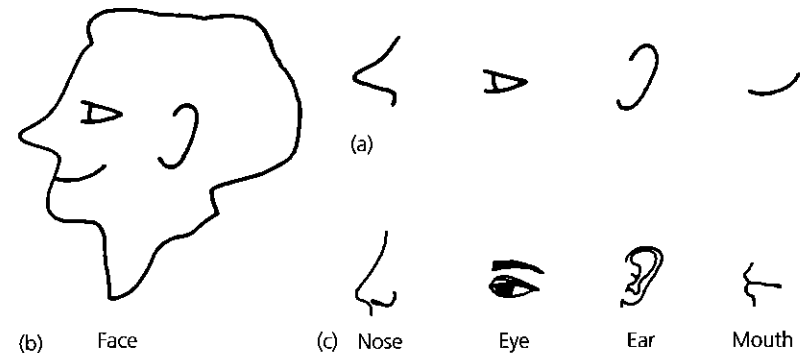
Figure 2.1 Some of the drawings used by Palmer (taken from Palmer, 1975)

– for recognition to take place without recourse to higher cognitive processes. This places his theory very firmly in the bottom-up camp. He was critical of the methods used by top-down theorists such as Gregory because he felt that they were artificial and ambiguous. Gibson was much more interested in perception as it occurs in the natural environment and, for this reason, his theory is sometimes known as an **ecological theory**.

Gibson's theory is quite complex and was formulated over a period of more than 30 years. We will focus here on the key aspects of the theory. Goldstein (1999) has suggested that there are four main principles:

- The proper way to describe a stimulus is not in terms of the retinal image but in terms of the optic array.
- The important information for perception is created by the movement of the observer.
- The key element of the optic array is invariant information (i.e. information that remains constant as the observer moves).
- It is the invariant information which leads directly to perception.

The optic array

Gibson felt that the starting point for perception should be the optic array (the structure or pattern of the light in the environment). An observer perceives objects, surfaces and textures in the visual

environment because of the way the light rays reaching him or her are structured by the objects. This light structure is extremely complex because of the myriad rays that are converging on the observer from all parts of the scene.

The importance of movement

The real importance of the optic array for Gibson was not so much the structure at any one time but in how the structure changes as the observer moves. He called this the *ambient* optic array, which can be described as follows. Imagine you are sitting down at one side of a room facing a window. There is a low table between you and the window. Outside the window is a tree. When you stand up the optic array changes. The standing observer now has some new information about the environment; for example, he or she can now see behind the table and can see the tree from a new angle.

'Ambient' means 'surrounding', so Gibson was describing how most of our perception occurs as we move relative to our environment. Even if we are sitting down, we still move our heads to look around us. Gibson was interested in the information contained in the ambient optic array. He believed that a basic property of this information is that it is invariant, i.e. it remains constant even when the observer changes position or moves through the environment.

Invariant information from the environment

There are several sources of invariant information identified by Gibson. We will mention three of them:

- **texture gradient**
- **flow pattern**
- **horizon ratio**

Texture gradient occurs when a textured piece of ground like a pebbled beach or a grassy field is viewed from an angle. The individual elements (e.g. the pebbles or the blades of grass) are seen as being packed closer and closer together as the distance increases. Gibson called this invariant information because, as you walk across the beach or field, texture gradient continues to provide information

about depth and distance – the further elements continue to look more densely packed. Consider an everyday example. Have you ever planned a picnic and looked out for a smooth area of grass? You start at the edge of the field and walk towards an appealing spot, only to find as you approach it that the grass is just as rough as it was at the edge of the field. It looks better further on however, so you trudge on, only to find that your new target is just as bad. This is an example of texture gradient making you see distant patches of grass as smoother and more densely packed.

A flow pattern

This is created as elements in the environment flow past a moving observer. If you look out of a train window as it travels through the countryside, you experience the rapid passing of objects like houses close to the railway line but the much slower movement of trees further away in the distance. This is an example of **motion parallax** which is a **depth cue** (discussed in Chapter 3), but Gibson emphasised the flow of the whole visual field rather than the relative movement of isolated objects.

Gibson investigated **optic flow patterns** (OFPs) particularly in the context of pilots' experiences in taking off and landing. When a pilot is approaching a landing strip, the point towards which he is aiming appears to remain motionless while the rest of the visual environment appears to move away from that point. These OFPs serve to provide pilots with clear, unambiguous information about their direction, speed and altitude.

Horizon ratio

This is the proportion of an object that is above the horizon divided by the proportion below. The horizon ratio principle states that when two objects of the same size are standing on a flat surface, their horizon ratio will be the same.

In Figure 2.2 all the telegraph poles have the same horizon ratio so we know that they are the same size. The tree, on the other hand, has a horizon ratio that is larger than the telegraph pole, so we know that the tree is taller than the telegraph pole. The horizon ratio is invariant so, even though the image of the telegraph pole itself may become
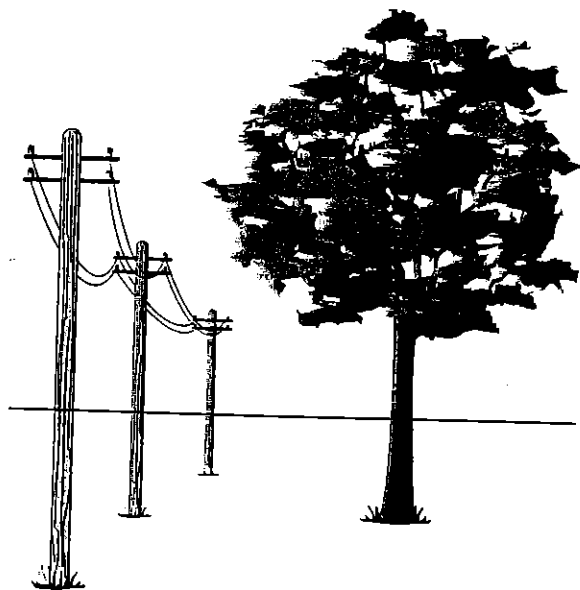
Figure 2.2 **An example of the horizon ratio**

larger on the retina as the observer moves towards it, the proportion of the pole that is above and below the horizon remains constant.

Direct perception

According to Gibson, this invariant information in the environment leads directly to perception.

Gibson seems to be able to account for our ability to locate objects spatially within the visual context but you may be wondering how he accounts for our ability to attach meaning to what we see. How do we realise, for example, that the object we are looking at is a chair and that we use it for sitting on? Gibson's answer is that visual perception does not occur in a vacuum and that we always find ourselves in a rich context which includes our:

- physical situation (e.g. in a classroom, on a train, etc.)
- psychological state (e.g. pleased, sad, angry, etc.)
- physiological state (e.g. highly aroused, thirsty, tired, etc.)

When we combine our physical and psychological states with our constantly changing optic arrays, we are enabled to recognise not only what the object is but what it does. Gibson called this the **affordance** of the object; for example, a cup affords drinking and a chair affords sitting down. The affordance chosen by the observer will depend on the factors mentioned above. Someone who is thirsty will perceive the affordance of a glass as for drinking. Someone who has just been given a bunch of flowers might see the affordance of the glass as a container/vase.

Evaluation

Visual perception is a very fast and accurate process. As soon as you open your eyes, the environment is perceived instantly. Studies where information is presented for brief periods of time indicate that some time is needed for processing but the processing (although variable) is measured in milliseconds. Direct perceptual processes would, by their very nature, be fast and accurate and this is the case. There would also be evolutionary pressure for perceptual systems to develop fast response times, as slow responses would make individuals more prone to predation.

Gibson himself paid little attention to physiological mechanisms but recent studies lend some support to Gibson's ideas. For example, it has been shown, at least in primates, that there are neurons in the **extrastriate cortex** which respond only to complex stimuli such as faces (Bruce *et al.*, 1981). There are also neurons which learn from visual experience to perceive specific forms (Logothetis and Pauls, 1995) and others which allow us to perceive properties of the environment as remaining constant even when we move about or view stimuli under different lighting conditions (Tovee *et al.*, 1994). It seems, then, that the physiologists may be discovering neurons which account for some of the direct perceptions which Gibson described.

Lee and Lishman (1975) conducted a study in which they used a specially built swaying room. They found that adults who were placed in the room usually made slight unconscious adjustments in order to avoid falling over. This kind of study tends to support Gibson's belief in the importance of movement in perception.

However, there are problems with his theory. The idea that meaning can be perceived directly (affordance) is one of the weaker

aspects of the theory. Human beings function in a cultural environment where knowledge of the use of objects is learned. Indeed, knowledge of the use of many objects is directly taught and would not necessarily appear simply by affordance. Bruce and Green (1990) have suggested that Gibson's concept of affordance may explain the visually guided behaviour of insects which have no need for a conceptual representation of their environment, but that is inadequate for explaining human perception.

Another problem for direct perception theories is the existence of visual illusions. Illusions demonstrate that the visual system can be inaccurate. Inaccuracies should not arise if perception is direct and relies on invariant and unambiguous properties of the optical array. Gibson believed that experiments using illusions were carried out in very artificial situations which had no relationship to the real world.

A clear demonstration that people tend to interpret situations from previous knowledge is provided by the Ames room (Figure 2.3). People view the person as being smaller and taller rather than being closer or further away. This is because rooms are usually box-shaped with right-angles. If the situation is arranged as in the Ames room the illusion works because of our assumption about room shapes and the fact that the room looks normal. This is significant because it shows that past experience and assumptions influence perception.

Gibson answered this criticism by asserting that our perception of the size of the two people in the room is based on how they fill the distance between the top and bottom edges of the room. Since one figure fills the entire space and the other takes up only a small part of it, we perceive the first figure as taller.

However, his explanations of how we experience perceptual illusions is not altogether convincing and does not account for all the experimental evidence in this area. His failure to account for the fact that we do not always perceive the world accurately remains one of the major weaknesses of his theory.

Constructivist theories

Background

The **constructivist** approach began over a hundred years ago with Helmholtz (1821–94), who believed that perception was based on a
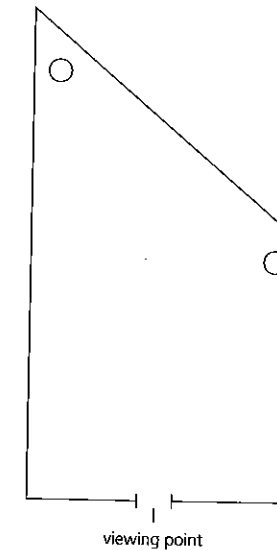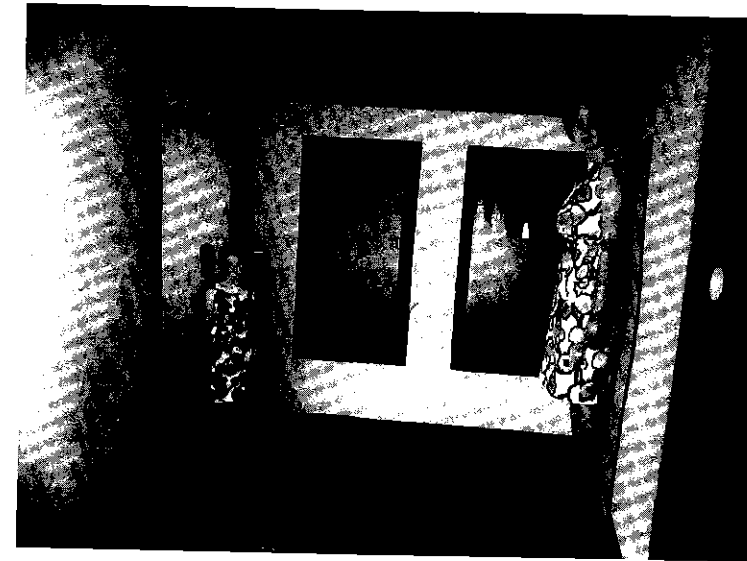

Figure 2.3 The Ames room. Photograph reproduced with permission from Eastern Counties Newspapers Limited

process of inference. He argued that, on the basis of the sensations we receive, we draw conclusions about the nature of the object or event that the sensations are most likely to represent. Because we make these inferences so quickly and without apparent awareness, he described the process as unconscious.

Modern constructivists suggest that the stimuli we receive from the environment are frequently ambiguous and have no clear-cut interpretation. This means that the observer has to solve the problem (or construct the best guess) as to the identity of the stimulus. In other words, the observer has to use indirect, top-down processes to make sense of the sensory input. Think back again to the example given on p.15, of

Gregory claims that perception is a dynamic, constructive process

You read that sentence easily and without any conscious awareness of solving a problem. You read the 'd' in dynamic as a 'd', even though exactly the same sensory input had just been interpreted a fraction of a second before as a 'cl' when it occurred in the word 'claims'. The only explanation for this would seem to be that you were processing an ambiguous stimulus input (handwriting) and having to find the most likely interpretation. 'Clynamic' is not a word whereas 'dynamic' is not only a proper English word but one that makes sense in the context of this particular sentence.

Some critics of the constructivist approach have suggested that such problem solving cannot take place without conscious aware-ness. However, this is completely untrue. Computers can carry out extremely complex mathematical problem solving that requires logical processing and yet machines have no consciousness.

Although the top-down processing involved in perception is believed to be largely unconscious and instantaneous, perception is seen to be *indirect* because information has to be processed at a level beyond the sensory level in order to be recognised accurately.

There are certain assumptions that all modern constructivist theorists share. Eysenck and Keane (1995) have suggested three shared assumptions:

1. Perception is an active and constructive process involving more than the direct registration of sensations.

2. Perception occurs indirectly as the end-product of the interaction between the stimulus input and the internal hypotheses, expecta-tions and knowledge of the observer. Motivational and emotional factors can also play a part in this perceptual processing.
3. Perception is influenced by individual factors and this means that errors will sometimes be made, leading to inaccurate perceptions.

Gregory's theory

Gregory acknowledged the importance of Gibson's work in the area of perception, particularly with regard to his ideas about texture gradient and motion parallax. However, he could not accept Gibson's overall conclusion that perception occurs directly with no intervening higher cognitive processing.

Gregory wrote in his book *Eye and Brain* (1990, p. 219):

> The sense organs receive patterns of energy, but we seldom see merely patterns: we see objects. A pattern is a relatively meaningless arrangement of marks, but objects have a host of characteristics beyond their sensory features. They have pasts and futures; they change and influence each other, and have hidden aspects which emerge under different conditions.

Gregory believed that the information supplied to the sensory organs is frequently impoverished and lacks sufficiently rich detail for perception to take place. Instead, it is used as the basis for making best guesses about the nature of the external stimuli. For Gregory, perception involves a dynamic search for the best interpretation of the available data – a process he called hypothesis testing. A study conducted by Pomerantz and Lockhead (1991) supports the idea that top-down processes can influence perception. They briefly showed participants visual stimuli which they then asked them to identify.

Participants who were shown a single stimulus like the one in Figure 2.4(a) usually reported having seen a circle. However, when shown the two circles in Figure 2.4(b), they were much more likely to see the breaks as significant and to report that they looked more like the letters 'c' and 'u' respectively. This is an interesting finding and is probably best explained in terms of hypothesis testing or best guessing. When just one circle is presented, observers 'see' it as
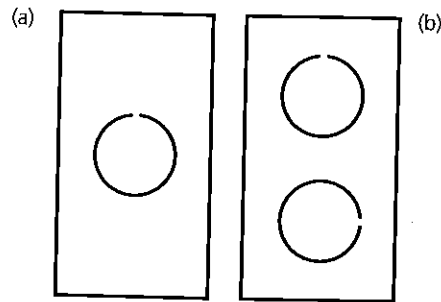
Figure 2.4 **Examples of stimuli similar to those used by Pomerantz and Lockhead (1991)**

a circle because they are used to seeing circles drawn by hand that are unintentionally imperfectly executed. However, when two circles are presented, both with an identical break but one which occurs in different positions, observers are more likely to conclude that the breaks, far from being unintentional, are highly significant and that they serve to distinguish between the two circles. When asked to describe the stimuli after presentation, participants seem to exaggerate the breaks (i.e. the circles become a 'c' and a 'u') because their hypothesis has led them to see the two circles as quite different from one another.

Gregory believed that individuals do not need much sensory data in order to formulate hypotheses. In a fairly recent article (1996), he cited a study by Johannson (1975) that seemed to support this idea. Johannson placed between ten and twelve small lights at points on a model's body (shoulders, elbows, knees, hips, wrists and ankles) and filmed the person moving around a darkened room. Participants viewing the film could make nothing of the apparently meaningless pattern of the lights if the model stayed still, but were instantly able to identify a 'walking person' once the model started to move.

Gregory himself was particularly interested in perceptual errors and made extensive use of visual illusions in his research. We will look at visual illusions in more detail in Chapter 3, but we need to consider them briefly here in order to understand Gregory's views.

Look carefully at the drawing called the Necker Cube in Figure 2.5. Concentrate hard on the drawing and try not to let your eyes stray. You will probably find that the cube suddenly seems to jump

and presents itself in a new orientation. It might take you a while to experience this but, once it has happened, you will find that the cube continues to jump backwards and forwards between the two orientations. Gregory explained this by saying that the drawing is ambiguous. At first sight, most people test the hypothesis that the drawing represents a cube resting on a flat surface (e.g. a table). However, there is no surrounding context in the picture (i.e. no table is drawn), and so it is suddenly possible to see an alternative interpretation of the line drawing, namely that it is a cube mounted on a wall that is coming out towards the viewer. In the absence of either a wall or a table, however, the picture offers no clue as to which of the interpretations is most plausible so the viewer switches between the two. Gregory believed that, in the kind of viewing conditions that exist in everyday life, there will normally be enough contextual information to remove any ambiguity and to lead to the confirmation of a single hypothesis.
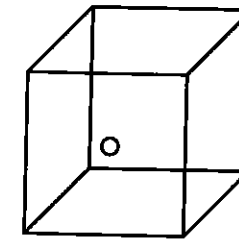

Figure 2.5 **The Necker Cube**

Gregory used other visual illusions to illustrate how we go beyond the information given in order to form perceptions; these will be discussed in Chapter 3. You will need to read the relevant section in that chapter in order to have a full understanding of Gregory's contribution to constructivist theory.

Perceptual set

Allport (1955), another constructivist theorist, introduced the concept of **perceptual set**. His thesis was that perceptual bias affects attention. Predispositions in the perceptual system make some stimuli stand out more from the background information arriving

from the senses. So, for example, if you are particularly interested in cars, you are more likely to notice makes and models than someone who sees cars merely as a means of travelling from A to B. This theory is directly relevant to Gregory's constructivist ideas because it views perception as an active process involving information processing and interpretation.

Sets have a wide range of functions. They are affected by motivation, emotion, past experience and expectations and serve to make perception more efficient. This is achieved because sets reduce the choice between alternatives. This means that predisposition towards a stimulus will make any choice quicker than considering all the alternatives. Coren *et al.* (1987) demonstrated this by presenting an image like the one in Figure 2.6.
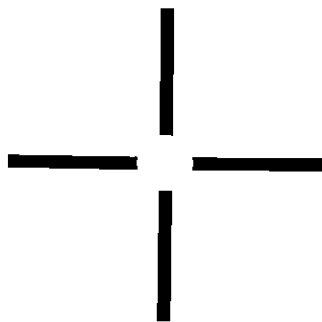

Figure 2.6 A stimulus figure similar to the one used by Coren *et al.* (1987)

Take a look at this yourself and see if you can detect a shape in the middle of the configuration of lines. Coren *et al.* found that, if participants were told in advance that they might see a circle, they tended to report back that a circle had been seen. If, on the other hand, they were told that they might see a square, they tended to report either that they had seen a square or no figure at all. Participants who were given no indication of what they were likely to see tended to report having seen either a circle or no figure at all. This seems to show that the circle is the dominant percept and that a square is only likely to be seen if a mental **set** can be induced first.

Expectancy can serve as a short cut to the interpretation of stimuli and aids planning and effective functioning in the environment.

Set has been found to be involved with many psychological variables such as motivation, emotion, context and beliefs. The effect of factors such as these on perception will be discussed in more detail in Chapter 5.

Evaluation

There are some problems with a very strong constructivist theory of perception. One clear hurdle is why people tend to see the world in a similar way if every person constructs their own perceptual model.

A further problem is that most people see the world correctly most of the time. Gibson believed that laboratory studies of perception were highly artificial and that illusions did not occur in the real world. This is not entirely so since illusions do occur in the real world, but not as frequently as Gregory's theory would predict, i.e. we are not easily misled. Many studies in this area involve presenting information that is fragmented and ambiguous, which means that people will use prior knowledge to try and understand what they are seeing because they have little else to go on. Presentations are typically also brief which reduces the scope for bottom-up processing. These factors suggest that constructivist explanations are magnified by the experimental situation and may be much reduced with normal rich environmental stimulation.

Gregory, unlike Gibson, has been quite successful in explaining why people experience perceptual illusions. However, his explanations cannot account for the fact that we continue to perceive an illusion such as the Müller-Lyer, even when we know it is an illusion. The Müller-Lyer (discussed in Chapter 3) is a drawing of two parallel straight lines, one with fins at each end pointing outwards and one with fins at each end pointing inwards (see Figure 3.7(a)). The straight lines are equal in length, but viewers experience the powerful illusion of one line being longer than the other. Even when viewers are given a ruler to check that the lengths are identical, they persist in the impression that the lines look different. This is difficult to explain and implies that the hypothesis is incapable of being modified in the light of experience. There is a related problem in that the hypothesis is supposed to be the best guess. Consider the Ames room (Figure 2.3). This is a specially constructed room which gives the illusion of having a normal, square construction but which has, in fact, a sloping

rear wall. To an observer looking into the room, two people placed at either corner of the far wall will look very different in size. This poses a puzzle for the observer, who believes (wrongly) that the two people are the same distance away. Ittelson (1952) arranged for the two people in the room to walk along the back wall and pass each other. Because one of the two people is actually moving away from the observer and the other one is moving towards her, the observer will experience the retinal image of one person getting smaller and the other getting larger. This retinal size information usually gives information about distance but, in this case, the observer believes that the people are at the same distance from her. A bizarre and unlikely hypothesis for explaining this problem is to say that the two individuals are changing in size. A much more sensible hypothesis is to guess that there is something strange about the construction of the room. Surprisingly, very few observers draw the more appropriate conclusion.

Synthesis theory

Background

There are some similarities between the direct and constructivist positions. They both acknowledge, for example, that:

- Visual perception depends on light reflected from stimuli in the environment.
- Perception cannot occur in the absence of a physiological system to support it.
- Perception is an active process even though the two theoretical positions see the activity involved rather differently. For constructivists like Gregory, this is embodied in the notion of the perceiver as a hypothesis tester. For Gibson, the perceiver acts as a map-reader rather than a passive camera.
- Perception can be influenced by learning.

However, there are also differences and a central disagreement, as we have seen, is about the relative contributions of bottom-up and top-down processes. This may, however, be largely a reflection of the different experimental methods used by the two types of theorist.

Gibson tended to work in natural situations where viewing conditions were optimal. In these conditions, bottom-up processing probably has more impact. Gregory, on the other hand, used mainly impoverished or ambiguous visual stimuli where there is little scope for pure bottom-up processing. It seems likely, therefore, that in most circumstances a combination of the two is probably needed.

Neisser's analysis-by-synthesis model

Neisser (1976) tried to reconcile the direct and constructivist positions by proposing a cyclic model of perception. He acknowledged that we are more likely to recognise objects quickly if they appear in a situational context. Human perceivers, according to Neisser, start out with certain expectations about the kinds of things they are likely to encounter in a given context. Perception, according to this view, is not a linear, one-way process with an input that leads progressively to a single interpretation. Neisser sees it instead as an active, cyclic process in which the viewer has to check and re-check input against expectations. Figure 2.7 shows a schematic representation of Neisser's model.
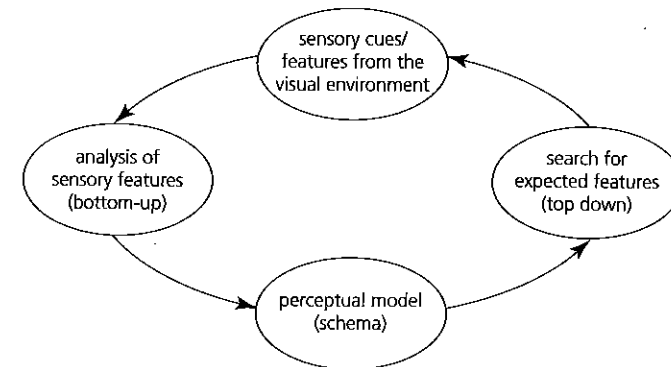

Figure 2.7 **Neisser's model**

Neisser believed that perception involves a series of processes.

Preliminary sampling There are pre-attentive processes (i.e. they occur automatically and unconsciously)

which produce a preliminary represen-tation of sensory data. This is bottom-up processing.

Direction

If the preliminary stage indicates some important stimulus, then attention is directed at it. The observer now uses schemata (packages of stored information about previous experiences) to help build a perceptual model (a mental representation of likely objects or events). This is top-down processing. The observer then compares this with the preliminary representation created at the first stage. This is called the inter-mediate representation and is the product of the interaction of bottom-up and top-down processing.

Modification

If the comparison with sensory data produces a match with the perceptual model, then this model can be accepted as the final perception. However, if the correspondence between the perceptual model and the sensory data is not perfect, the perceptual model will have to be revised until a perfect match is found.

Neisser called his model an analysis-by-synthesis theory of percep-tion. The synthesis involves generating a perceptual model based on past knowledge and experience which helps form perceptions in a top-down direction. The analysis involves analysing sensory data in order to extract relevant information about elements in the environment and this is passed up the system in a bottom-up direction.

Evaluation

There is no doubt that Neisser's theory is intuitively appealing. It combines perceptual hypotheses made on the basis of prior know-ledge or schemas with the extraction of sensory cues from the environment. The perceptual process is seen as a continuous active interaction between top-down and bottom-up processing and this seems highly likely.

However, there are some problems with the theory. A broad criticism is that the theory is too vague and does not specify exactly how schemas interact with the sensory data. It describes what we do but not how or why we do it. It is not clear exactly where and at what point the perception actually occurs in the cyclic process. Imagine that you go into your garden and see a black shape under a bush. Past experience gives rise to the perceptual model that this is your dog and you search for further dog-like features such as big paws and tail (bottom-up processing). However, you find no such dog-like features so have to abandon your first perceptual model and generate a new one. Perhaps it could be a black rubbish sack that has blown under the bush so you look for features such as the string tied round the top and the shiny surface of the plastic. You find these features and so your second perceptual model is confirmed. The question arises, though: At what point does a 'perception' emerge? Do you actually experience the shape in the first instance as your dog and then change your mind to see it as a sack, or do you only see the sack when the perceptual model is finally confirmed? And what if the black shape is neither of these two things but turns out to be a turkey that has escaped from a farm further down the road? You have never experienced a turkey in your garden before and so you are unlikely to have generated this perceptual model. Neisser does not explain how we experience totally unexpected perceptions.

Another criticism is that there is insufficient reference to the biological processes that might underlie the cyclic process of inter-actions between perceptual models and sensory data.

Computational theory

Background

The computational approach is a branch of artificial intelligence (AI) which involves designing computer systems to carry out cog-nitive tasks. Some researchers in this field design computers which can carry out perceptual tasks in a practical situation (e.g. computers that can detect faulty cells in a blood sample). Others develop computer programs which simulate mental processing in human beings. We will look at a computational theory that has attempted to estab-lish a set of rules and procedures which govern vision. Theorists

working in this area see visual perception as a problem and their aim is to provide a solution to the problem. They do this by providing a theoretical analysis of the solution and by describing the algorithms (problem-solving procedures) which work out the solution. AI researchers usually then run a computer program that mimics these algorithms. If the program actually works, it demonstrates empirically that this is a feasible explanation of human visual processing. Note that it does not necessarily *prove* that it is the correct way but only that it is a possibility. Marr, whose work made an important contribution to the computational approach, did not test out all of his proposals in this way and so some of them remain theoretical proposals.

Marr's model of visual perception

If you think back to Neisser's theory, you will recall that it can be criticised for failing to specify the processes which underlie the interactions between sensory input and stored knowledge. Marr (1982) argued that a comprehensive theory of visual perception should include at least three levels of explanation. These are:

- *The computational level.* This specifies the job that the visual system must do, i.e. its function.
- *The algorithmic level.* This is concerned with the detailed processes involved in perception.
- *The hardware level.* This is concerned with the neuronal mechanisms underlying perceptual processing operations.

Marr's approach is largely bottom-up, although there is some room for top-down processing in the later stages of perceptual processing. However, this aspect of his theory is less well defined than his description of the earlier stages of processing. The computational approach acknowledges the role of knowledge in perception but it is of a more general nature than the specific knowledge that the constructivist approach sees as necessary. General and basic knowledge about the laws of physics and geometry is used to analyse a complex scene into separate objects and shapes. Marr believed that the visual system carries out various mathematical computations about intensity changes in the image but, at the same time, takes into account what Marr calls the natural constraints in the world (basic properties of the environment). The computational approach is

complex and highly technical and it is beyond the scope of this book to describe it in detail. We will present a summary of the main elements of the theory.

Marr believed that object recognition was a central feature of vision and he concentrated on this aspect of perception. According to his theory, perception begins with the retinal image and then proceeds via a series of stages. At each stage, the image is transformed into a more complex representation of the input. Marr described each stage in terms of the basic elements (or primitives) characteristic of that stage.

There are four main stages:

Grey level description	The intensity of light is measured at each point in the image.
The primal sketech	This is the result of the initial stage of computations but we do not see it at this stage. Before conscious perception can occur, we have to process the information contained in the primal sketch. We group together primitives of similar size and shape at this stage to form structures and outline shapes.
2.5-D sketch	At this stage, a 'picture' of the world begins to emerge. It is no longer a straight image because it now contains additional information. It provides depth cues such as shading, texture gradient and motion. It describes only the current visible surfaces of the scene (excluding unseen surfaces such as those hidden behind other objects) and will change if the scene is viewed from a different angle. For this reason, it is called viewpoint-dependent.
3-D model representation	At this stage, the viewpoint-dependent descriptions are converted into object-

centred description. The three-dimensional shapes of objects and their spatial interrelationships are perceived. Objects that are obscured from view will be represented in the three-dimensional model (e.g. the hidden leg of a chair). At this stage, prior knowledge may influence perception (e.g. you have to know that a chair is likely to have a fourth leg even if it is not currently visible).

Evaluation

Marr's theory has been very important in stimulating theoretical and empirical research. A fairly fundamental question is whether Marr's model works. Marr and Hildreth (1980) developed a program which analysed images successfully to the raw primal sketch stage. However, the fact that the program works does not necessarily mean that human perceptual systems operate in the same way.

There is some support for the model from neurophysiological studies, but further algorithms have been generated by other computational theorists which appear to correspond more closely to results from human perceptual experiments (see e.g. Watt and Morgan, 1984).

Marr stressed the importance of providing a complete explanation at three levels. However, if you think about the function of visual perception (computational level) it is not clear-cut. There are several functions (e.g. balance, navigation, object recognition, etc.) and they may all require a separate computational model. Marr concentrated principally on object recognition.

Much of the detailed work that Marr carried out was focused on the bottom-up strategies associated with the early stages of processing. In other words, he was most concerned with the processing steps before object recognition actually occurs. When he described the final stage where top-down processes are thought to occur, his ideas become much more sketchy and unconvincing.

Summary

In this chapter, we have looked at various theories of perception and distinguished between bottom-up and top-down processing. Gibson's theory rests on the assumption that perception is a direct process. He believed that we can extract sufficient information from the sensory stimulus to experience accurate perception. His theory accounts for some of the available data and has some support from neurological studies, but it cannot explain our experience of visual illusions.

Constructivist theories place more emphasis on the role of learning and experience, and advocates of this approach, such as Gregory, suggest that perception is akin to hypothesis testing. Such theories can account for our interpretation of ambiguous stimuli but have been criticised for artificial experimental techniques. Neisser attempted to provide a reconciliation between the direct and constructivist positions with his synthesis theory. This theory is quite attractive but is descriptive rather than explanatory. A more detailed model is the computational theory put forward by Marr. This theory has been very important in advancing our understanding of object recognition, but it is limited. All the theories discussed have strengths and weaknesses but there is, as yet, no single theory to account for all that is known about human perception.

Draw up a table listing the main strengths and weaknesses of the theories covered in this chapter.

Review exercise

Further Reading

Eysenck, M.W. (1993) *Principles of Cognitive Psychology*, Hove: Lawrence Erlbaum Associates. A straightforward account of perception which gives a solid base for further reading.

Eysenck, M.W. and Keane, M.T. (1995) *Cognitive Psychology: A Student's Handbook (3rd edn)*, Hove: Lawrence Erlbaum

1 The Enactive Approach to Perception: An Introduction

The theory of the body is already a theory of perception.
—M. Merleau-Ponty

1.1 The Basic Idea

The main idea of this book is that perceiving is a way of acting. Perception is not something that happens to us, or in us. It is something we do. Think of a blind person tap-tapping his or her way around a cluttered space, perceiving that space by touch, not all at once, but through time, by skillful probing and movement. This is, or at least ought to be, our paradigm of what perceiving is. The world makes itself available to the perceiver through physical movement and interaction. In this book I argue that all perception is touch-like in this way: Perceptual experience acquires content thanks to our possession of bodily skills. *What we perceive* is determined by *what we do* (or what we know how to do); it is determined by what we are *ready* to do. In ways I try to make precise, we *enact* our perceptual experience; we act it out.

To be a perceiver is to understand, implicitly, the effects of movement on sensory stimulation. Examples are ready to hand. An object looms larger in the visual field as we approach it, and its profile deforms as we move about it. A sound grows louder as we move nearer to its source. Movements of the hand over the surface of an object give rise to shifting sensations. As perceivers we are masters of this sort of pattern of sensorimotor dependence. This mastery shows itself in the thoughtless automaticity with which we move our eyes, head and body in taking in what is around us. We spontaneously crane our necks, peer, squint, reach for our glasses, or draw near to

get a better look (or better to handle, sniff, lick or listen to what interests us). The central claim of what I call *the enactive approach* is that our ability to perceive not only depends on, but is constituted by, our possession of this sort of sensorimotor knowledge.[1]

One implication of the enactive approach is that only a creature with certain kinds of bodily skills—for example, a basic familiarity with the sensory effects of eye or hand movements, and so forth—could be a perceiver.[2] This is because, in effect, perceiving is a kind of skillful bodily activity. It may also be that only a creature capable of at least some primitive forms of perception could be capable of self-movement. Specifically, self-movement depends on perceptual modes of self-awareness, for example, proprioception and also 'perspectival self-consciousness' (i.e., the ability to keep track of one's relation to the world around one).[3]

A second implication of the enactive approach is that we ought to reject the idea—widespread in both philosophy and science—that perception is a process *in the brain* whereby the perceptual system constructs an *internal representation* of the world. No doubt perception depends on what takes place in the brain, and very likely there are internal representations in the brain (e.g., content-bearing internal states). What perception is, however, is not a process in the brain, but a kind of skillful activity on the part of the animal as a whole. The enactive view challenges neuroscience to devise new ways of understanding the neural basis of perception and consciousness.[4] I return to this controversial topic in chapter 7.

This idea of perception as a species of skillful bodily activity is deeply counterintuitive. It goes against many of our preconceptions about the nature of perception. We tend, when thinking about perception, to make vision, not touch, our paradigm, and we tend to think of vision on a photographic model. You open your eyes and you are given, at once, a sharply focused impression of the present world in all its detail. On this view, the relation between moving and perceiving is only instrumental. It is like the relation between the lugging around of a camera and the resulting picture. The lugging is preliminary to and disconnected from the photograph itself. And so with perceiving. By moving yourself, you can come to occupy a vantage point from which, say, better to see your goal. And then, having seen your goal, you can better decide what to do. But the seeing, and the moving, have no more to do with each other than the photograph and the schlepping of the camera, or the boxer's left hook, and the

training that preceeded it. Which is to say, they have a lot to do with each other, but the relation is nonconstitutive: The effectiveness of the punch is strictly independent of how the boxer learned to do it, and the qualities of the picture are independent of how the camera ended up where it was.

Susan Hurley (1998) has aptly called this simple view of the relation between perception and action the input-output picture: Perception is input from world to mind, action is output from mind to world, thought is the mediating process. If the input-output picture is right, then it must be possible, at least in principle, to disassociate capacities for perception, action, and thought. The main claim of this book is that such a divorce is not possible. I doubt that it is even truly conceivable. All perception, I argue, is intrinsically active. Perceptual experience acquires content thanks to the perceiver's skillful activity. I also argue—but I don't turn to this until late in the book (chapter 6)—that all perception is intrinsically thoughtful. Blind creatures may be capable of thought, but thoughtless creatures could never be capable of sight, or of any genuine content-bearing perceptual experience.[5] Perception and perceptual consciousness are types of thoughtful, knowledgeable activity.

My aim in this initial chapter is to set out the book's central themes.

1.2   A Puzzle about Perception: Experiential Blindness

For those who see, it is difficult to resist the idea that being blind is like being in the dark. When we think of blindness this way, we imagine it as a state of blackness, absence and deprivation. We suppose that there is a gigantic hole in the consciousness of a blind person, a permanent feeling of incompleteness. Where there could be light, there is no light.

This is a false picture of the nature of blindness. The longterm blind do not experience blindness as a disruption or an absence. This is not because, as legend has it, smell, touch and hearing get stronger to compensate for the failure to see (although this may be true to some degree; see Kaufman, Théoret, and Pascual-Leone 2002). It's because there is a way in which the blind do not experience their blindness at all. Consider, you are unable visually to discern what takes place in the room next door, but you do not experience this inability as a gaping hole in your visual awareness. Likewise, you don't encounter the absence of the sort of olfactory information that would be present to a bloodhound as something missing in