



CHAPTER OUTLINE

LEARNING OBJECTIVES

INTRODUCTION

PERCEPTION AND ILLUSION

- When seeing goes wrong
- Theories of perception
- Spotting the cat in the grass
- Explaining after-effects

MAKING SENSE OF THE WORLD

- Grouping and segmentation
- Visual search – or finding the car
- How do we know what we see?
- Seeing without knowing

SEEING WHAT WE KNOW

- Perception or hallucination?
- Resolving visual ambiguity
- Tricks of the light
- Non-visual knowledge and perceptual set

PERCEPTUAL LEARNING

- How training influences performance
- Top-down mechanisms

FINAL THOUGHTS

SUMMARY

REVISION QUESTIONS

FURTHER READING

Learning Objectives

By the end of this chapter you should appreciate that:

- the efficiency of our perceptual mechanisms is acquired over many years of individual learning experience;
- as perceptual knowledge grows and accumulates, it enables ever more efficient interpretation of the stimuli that impinge upon our sensory receptors;
- our almost effortless ability to perceive the world around us is achieved by a number of inter-connected regions in the brain;
- corresponding to these anatomical connections, there are continual interactions between conceptually driven processes and the current sensory inflow from the environment;
- usually this two-way interaction improves the operation of the system, but sometimes this is not the case, and illusions occur;
- the study of illusions and the effects of brain injury provide valuable information about the mechanisms of perception;
- there is an important interaction between psychological and biological studies of perception.

INTRODUCTION

Which is the harder of these two problems?

1. Work out the square root of 2018 in your head.
2. As you walk past your neighbour's garden, decide whether the cat in the long grass is moving or not.

For most people, the answer is the square root problem. Indeed, the second problem does not seem like a problem at all. We usually make such judgements effortlessly every day without thinking about them. But when it comes to saying *how* we might solve the problems, the order of difficulty probably reverses.

Think of a number smaller than 2018 and multiply it by itself. If the answer is greater than

2018, choose a smaller number and try again. If this answer is smaller than 2018, choose a larger number (but smaller than the initial number you thought of), and so on. It is possible to program a computer to find the square root of any number by following rules like this.

The goal of people who study perception is to discover the rules that the brain uses to solve problems like that demonstrated by the 'cat in the garden' example. Although we have made some progress towards this goal, it is much easier to program a machine to find a square root than to program it to see. Perhaps one third of the human brain is devoted to seeing, which not only demonstrates that it must be difficult, but also perhaps explains why it seems to be so easy.

PERCEPTION AND ILLUSION

WHEN SEEING GOES WRONG

One way of uncovering the processes of seeing is to look at the circumstances in which they go wrong. For example, returning to the 'cat in the garden' problem, suppose that, when you were walking past your neighbour's garden, you were on your way to the station. When you arrived, you boarded your train, then you waited, and waited . . . At last you sighed with relief as the train started to move . . . but then the back of the train on the adjacent track went past your carriage window and you saw the motionless platform opposite. Your train had not moved at all, but your brain had interpreted the movement of the other train – incorrectly – as caused by your own movement, not that of the object in the world.

Why are we fooled?

How does your brain decide what is moving in the world and what is not? What can we discover from the train experience about how seeing works?

As we look at a scene full of stationary objects, an image is formed on the retina at the back of each eye (see chapter 7). If we move our eyes, the image shifts across each retina. Note that all parts of the image move at the same velocity in the same direction. Similarly, as we look through the window of a moving train, but keep our eyes still, the same thing happens: our entire field of view through the window is filled with objects moving at a similar direction and velocity (though the latter varies with their relative distance from the train).

In the first case, the brain subtracts the movements of the eyes (which it knows about, because it caused them) from the motion in the retinal image to give the perception of its owner being stationary in a stationary world. In the second scenario, the eyes have not moved, but there is motion in the retinal image. Because of the coherence of the scene (i.e. images of objects at the same distance moving at the same velocity), the brain (correctly) attributes this to movement of itself, not to that of the rest of the world.

To return to the situation in which we may be fooled by the movement of the other train into thinking that our train is moving – notice that, although the visual information produced by the two situations (your train stationary, other train moving, or vice versa) is identical, other sensory information is not. In principle, the vestibular system can signal self-motion as your train moves. However, slow acceleration produces only a weak vestibular signal, and this (or its absence, as in the present case, if we are in fact stationary) can often be dominated by strong visual signals.

Of course, objects in the world are not always stationary. But objects that do not fill the entire visual field cause patterns of movement which are piecemeal, fractured and unpredictable. One object may move to the right, another to the left, and so on, or one object may move to partially obscure another. So lack

of coherence in the pattern of motion on the retina suggests the motion of objects, instead of (or as well as) motion of the observer.

Think back to what happened as you were walking past your neighbour's garden. The patterns of movement in the retinal images caused by the movements of your body and your eyes were mostly coherent. The exceptions were caused by the movements of the long grasses in the breeze and the tiny movements of the cat as it stalked a bird, which were superimposed on the coherent movements caused by your own motion.

The visual system needs to detect discrepancies in the pattern of retinal motion and alert its owner to them, because these discrepancies may signal vital information such as the presence of potential mates, prey or predators (as in the case of the cat and the bird). Indeed, when the discrepancies are small, the visual system exaggerates them to reflect their relative importance.

Contrast illusions and after-effects

Some further examples of perceptual phenomena that result from this process of exaggeration are shown in the *Everyday Psychology* box. These are known collectively as simultaneous contrast illusions. In each case the central regions of the stimuli are identical, but their surrounds differ. Panel A (figure 8.1) lets you experience the simultaneous tilt illusion, in which vertical stripes appear tilted away from the tilt of their surrounding stripes. Panel B shows the luminance illusion: a grey patch appears lighter when surrounded by a dark area than when surrounded by a light area. Panel C shows the same effect for colour: a purple patch appears slightly closer to blue when surrounded by red, and closer to red when seen against a blue background. There is also an exactly analogous effect for motion, as well as for other visual dimensions such as size and depth.

Suppose your train finally started and travelled for some time at high speed while you gazed fixedly out of the window. You may have noticed another movement-related effect when your train stopped again at the next station. Although the train, you, and the station platform were not physically moving with respect to each other, the platform may have appeared to drift slowly in the direction in which you had been travelling.

This is another case of being deceived by the mechanisms in our nervous systems. This time what is being exaggerated is the difference between the previously continuous motion of the retinal image (produced by the train's motion) and the present lack of motion (produced by the current scene of a stationary platform), to make it appear that the latter is moving. Such effects are known as successive contrast illusions, because visual mechanisms are exaggerating the difference between stimuli presented at different times in succession (compared with simultaneous contrast illusions, in which the stimulus features are present at the same time).

A famous example of this effect is the 'waterfall illusion', which has been known since antiquity, although the first reliable description was not given until 1834 (by Robert Addams: see Mather et al., 1998). If you gaze at a rock near a waterfall for 30–60 seconds and then transfer your gaze to a point on the banks of the waterfall, you will notice a dramatic upward movement of the

after-effect change in the perception of a sensory quality (e.g. colour, loudness, warmth) following a period of stimulation, indicating that selective adaptation has occurred

banks, which lasts for several seconds before they return to their normal stationary appearance. Because the first stimulus induces an alteration in the subsequently viewed stimulus, this and

other similar illusions are often known as *after-effects*.

Several further examples of successive contrast are given in the *Everyday Psychology* section of this chapter. In each case the adapting field is shown in the left-hand column and the test field is shown on the right. Now look at figure 8.2. Panel A lets you experience the tilt after-effect, in which vertical stripes appear tilted clockwise after staring at anti-clockwise tilted stripes, and vice versa.

Panel B offers the luminance after-effect: after staring at a dark patch, a grey patch appears lighter, and after staring at a white patch the grey patch appears darker. Panel C shows the colour after-effect: after staring at a red patch a yellow patch appears yellow-green, and after staring at a green patch a yellow patch appears orange.

Like the simultaneous contrast illusions, these after-effects demonstrate that the visual system makes a comparison between stimuli when calculating the characteristics of any stimulus feature.

These illusions are not just for fun, though. They also give us vital clues as to how we see, hear, touch, smell and taste under normal circumstances. Indeed, there are three general theories about how we perceive, and these illusions help us to decide between them.

Everyday Psychology

Illusions and after-effects

In everyday life you may encounter visual illusions and after-effects that you are unaware of. Although we rarely encounter these illusions and after-effects in their pure form, they are important components of our world and our visual experience of it. They also help us to understand how we see, hear, touch, smell and taste under normal circumstances.

The central parts of figure 8.1 are identical in each panel.

- A The tilt illusion: both central circles are filled with vertical stripes, but they appear tilted in the opposite direction to the stripes in the surrounding regions.
- B The luminance contrast effect: both central squares are of identical physical luminance, but the one on the lighter background appears darker than the other.
- C A colour illusion: both central panels are the same shade of purple, but the one on the red background appears bluer than the other.

You have probably experienced these contrast phenomena unconsciously in your everyday life, where the perception of an object is influenced by its surroundings. In these examples you become conscious of these contrast effects because they are presented to you simultaneously – you cannot ‘ignore’ them!

In each panel in figure 8.2, the right-hand pair of stimuli is the test field and the left-hand pair is the adapting field. Before adapting, gaze at the central black dot between the upper and lower stimuli in each test pair and note that the latter appear identical. Then adapt by gazing at the spot (or in A by running your eyes slowly back and forward over the horizontal bar) between the upper and lower adapting stimuli. After about 30 seconds, switch your gaze back to the spot between the two test stimuli, and note any difference in their appearance.

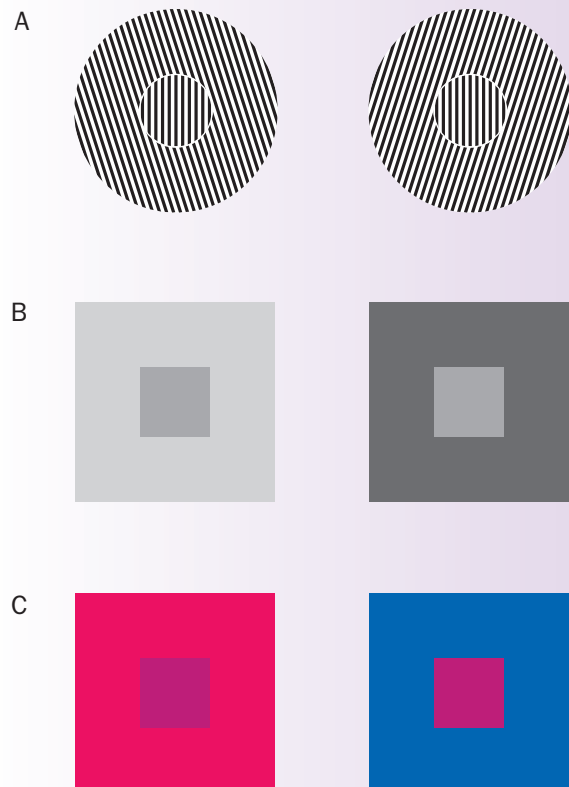
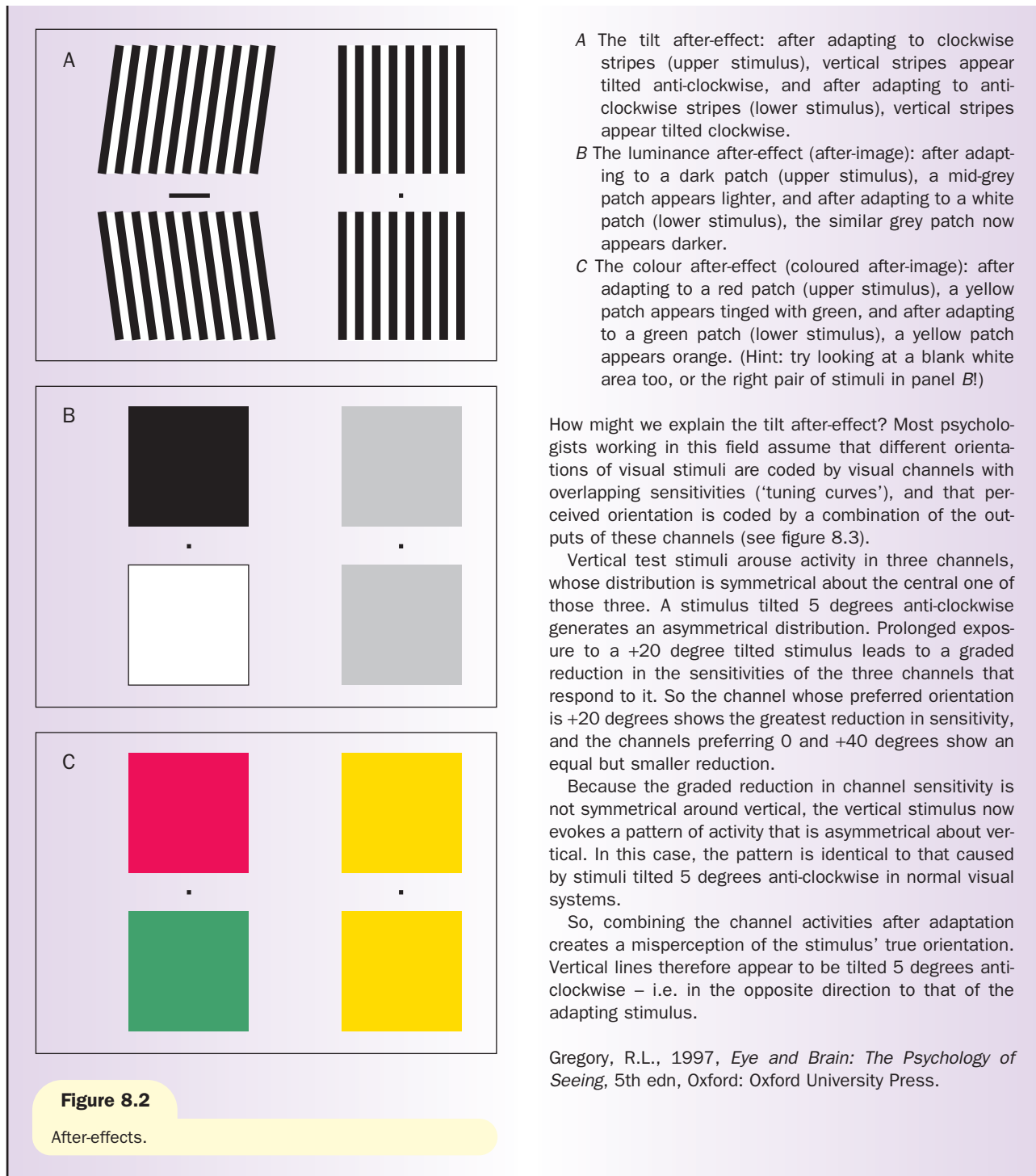


Figure 8.1

Simultaneous contrast illusions.



A The tilt after-effect: after adapting to clockwise stripes (upper stimulus), vertical stripes appear tilted anti-clockwise, and after adapting to anti-clockwise stripes (lower stimulus), vertical stripes appear tilted clockwise.

B The luminance after-effect (after-image): after adapting to a dark patch (upper stimulus), a mid-grey patch appears lighter, and after adapting to a white patch (lower stimulus), the similar grey patch now appears darker.

C The colour after-effect (coloured after-image): after adapting to a red patch (upper stimulus), a yellow patch appears tinged with green, and after adapting to a green patch (lower stimulus), a yellow patch appears orange. (Hint: try looking at a blank white area too, or the right pair of stimuli in panel B!)

How might we explain the tilt after-effect? Most psychologists working in this field assume that different orientations of visual stimuli are coded by visual channels with overlapping sensitivities ('tuning curves'), and that perceived orientation is coded by a combination of the outputs of these channels (see figure 8.3).

Vertical test stimuli arouse activity in three channels, whose distribution is symmetrical about the central one of those three. A stimulus tilted 5 degrees anti-clockwise generates an asymmetrical distribution. Prolonged exposure to a +20 degree tilted stimulus leads to a graded reduction in the sensitivities of the three channels that respond to it. So the channel whose preferred orientation is +20 degrees shows the greatest reduction in sensitivity, and the channels preferring 0 and +40 degrees show an equal but smaller reduction.

Because the graded reduction in channel sensitivity is not symmetrical around vertical, the vertical stimulus now evokes a pattern of activity that is asymmetrical about vertical. In this case, the pattern is identical to that caused by stimuli tilted 5 degrees anti-clockwise in normal visual systems.

So, combining the channel activities after adaptation creates a misperception of the stimulus' true orientation. Vertical lines therefore appear to be tilted 5 degrees anti-clockwise – i.e. in the opposite direction to that of the adapting stimulus.

Gregory, R.L., 1997, *Eye and Brain: The Psychology of Seeing*, 5th edn, Oxford: Oxford University Press.

THEORIES OF PERCEPTION

The serial model

It is natural to assume that sensory processing proceeds through a series of stages. Obviously, the sense organs first transduce the

stimulus (convert it from one form of energy to another – see chapter 7). In the case of vision, further processing then occurs in the retina before the results of the analysis are sent up the optic nerve, to the thalamus, and then to the primary visual cortex. In other sensory modalities, the signals pass to their own 'primary' sensory areas of cerebral cortex for interpretation (see chapters 3 and 7).

serial model the assumption that perception takes place in a series of discrete stages, and that information passes from one stage to the next in one direction only

parallel processing perceptual processing in which it is assumed that different aspects of perception occur simultaneously and independently (e.g. the processing of colour by one set of neural mechanisms at the same time as luminance is being processed by another set)

recurrent processing occurs when the later stages of sensory processing influence the earlier stages (top-down), as the output of a processing operation is fed back into the processing mechanism itself to alter how that mechanism subsequently processes its next input

For all sensory modalities, there are then several further stages of processing which occur within the cortex itself. Indeed, as much as one half of the cortex is involved purely in perceptual analysis (mostly in vision). At each stage, further work takes place to analyse what is happening in the environment. Because several such steps are involved, this way of understanding perception as a sequence of processes is known as the *serial model*.

But the serial model is now known to be inadequate, or at least incomplete. So it has been replaced, or at least modified, firstly by the *parallel processing* model and then, most recently, by the *recurrent processing* model.

The parallel processing model

According to the parallel processing model, analysis of different stimulus attributes, such as identity and location, proceeds simultaneously along different pathways, even from the earliest stages. For example, the fact that there are cones (of three types, maximally sensitive to different wavelengths of light) and rods in the retina (see chapter 7) is evidence for multiple mechanisms that extract information in parallel from the retinal image.

The recurrent processing model

The recurrent model emphasizes that the effects of a stimulus on the higher centres of the brain not only influence our subjective perception but also feed back down to modulate the 'early' stages of processing. 'Higher' stages of processing are taken to be those that exist anatomically further away from the sensory receptors, and are also those with more 'cognitive' as opposed to primarily 'sensory' functions, i.e. where learning, memory and thinking enter into the processing. As we shall see, a substantial amount of evidence has now accumulated indicating that the influence of these higher functions can be seen at almost all stages of sensory analysis, thereby casting serious doubt on the existence of sharp divisions between serial stages of sensation, perception and cognition. First, however, let us look at evidence for the parallel processing model.

SPOTTING THE CAT IN THE GRASS

Selective adaptation

An important early stage of vision is finding out which bits of the retinal image correspond to what kinds of physical thing 'out there' in the world. Our visual system first needs to discover the locations of objects, their colours, movements, shapes, and so on. This process can be demonstrated by the technique of selective *adaptation*.

adaptation decline in the response of a sensory or perceptual system that occurs if the stimulus remains constant

Whenever we enter a new environment, our sensory systems adjust their properties quite rapidly (over the course of a few seconds), optimizing their ability to detect any small change away from the steady background conditions. This is because interesting and important stimuli are usually ones that deviate suddenly in some way from the background (such as a tiger jumping out from behind a tree). Remember the cat in the grass: its tiny movements had to be extracted from the pattern of coherent movement on the retina produced by your movements as you walked past.

By staring at something for a time (selective adaptation), we produce an unchanging pattern of stimulation on one region of the retina, and the visual system starts to treat this as the steady background, and lowers its sensitivity to it. When we stop staring at this same location, it takes a while for our vision to return to normal, and we can notice during this period of compensation that the world looks different. These differences represent the after-effects of adaptation.

This whole process of adaptation is described as selective because only some perceptual properties are affected. The adaptations are restricted to stimuli similar to the one that has been stared at.

Many kinds of visual after-effect have been discovered (as we can see in Everyday Psychology). These clear and robust phenomena are not confined to vision, but are found in touch, taste, smell and hearing also. For example:

1. After running your fingers over fine sandpaper, medium sandpaper feels coarser (and vice versa).
2. After listening to a high tone for a while, a medium tone appears lower.
3. Musicians often build their music to a loud and cacophonous crescendo just before a sudden transition to a slow, quiet passage, which then seems even more mellow and tranquil than it otherwise would.
4. Holding your hand under running cold (or hot) water before testing the temperature of baby's bath water will lead you to misperceive how comfortable the water will be for the baby. This is why you are always advised to test the temperature with your elbow.
5. After eating chocolate, orange juice tastes more tart.
6. When we enter a dark room, it takes a few minutes for our receptors to adapt, and we begin to notice things that had been simply too faint to activate those receptors at first.

EXPLAINING AFTER-EFFECTS

It can be helpful to think of an object (or visual stimulus) as having a single value along each of several property dimensions. For example, a line's orientation could be anywhere between -90 and $+90$ degrees with respect to vertical. And an object's colour could be anywhere between violet (shortest visible wavelength) and red (longest visible wavelength).

The general rule that describes perceptual after-effects is that adapting to some value along a particular dimension (say $+20$ degrees from vertical) makes a different value (say 0 degrees) appear even more different (say -5 degrees). For this reason, these phenomena are sometimes called negative after-effects. The after-effect is in the opposite direction (along the stimulus dimension) away from the adapting stimulus, rather than moving the perceived value towards that of the adapting stimulus.

What do these effects tell us about how perceptual systems encode information about the environment?

The existence and properties of channels

One implication of after-effects is that different features, or dimensions, of a stimulus are dealt with separately. Each dimension is, in turn, coded by a number of separate mechanisms, often called

channel transmits a restricted range of sensory information (e.g. in the case of colour, information about a restricted range of wavelengths, but no information about the movement or orientation of the stimulus)

channels, which respond selectively to stimuli of different values along that particular dimension. Each channel responds in a graded fashion to a small range of neighbouring values of the stimulus dimension. So several channels

respond to any given stimulus, but to differing extents. The channel that most closely processes (i.e. is most selective for) the stimulus will give the greatest output, channels selective for nearby stimuli will give a lesser output, and so on. For example, different channels may selectively code for different angles of orientation of visual stimuli, from horizontal round to vertical.

This enables us to give a simple explanation of after-effects, illustrated in this chapter using the tilt after-effect (Blakemore, 1973). Perception depends not on the output of any single channel, but on a combination of the outputs of all the active channels (see chapter 7 for a related discussion). This is because a given level of activity in any single channel might be caused by a weak (say, low contrast) stimulus of its optimal type (such as a vertical line for a channel that responds best to vertical lines) or an intense (high contrast) stimulus away from the optimal (such as a line tilted 20 degrees). So the output of a single channel on its own is ambiguous.

For the sake of simplicity, we will look at the relationship between just five channels (see figure 8.3), although in practice there are many more. In panel A, each bell shaped curve ('tuning curve') represents the activity in one channel produced by lines of different orientations. One channel responds most strongly to vertical lines (the channel whose tuning curve is centred on

0 degrees), and progressively less strongly to stimuli further and further from that optimal orientation of line (either clockwise or anti-clockwise). Another channel has the same degree of selectivity but responds best to lines tilted to the right by 20 degrees. A third channel is similar but 'prefers' (or is 'tuned' to) tilt in the opposite direction from vertical (-20 degrees). The orientations over which these latter two channels respond overlap, so they respond weakly but equally to zero tilt (vertical stimuli), as shown in panel B. Finally, we include two outermost channels, which respond best to 40 degrees ($+40$ deg) clockwise and 40 degrees anti-clockwise (-40 deg.). These two channels do not respond at all to vertical lines. This system of channels can signal orientations which do not correspond to the preferred orientation of any single channel. Panel C shows the pattern of activation produced by a line tilted 5 degrees anticlockwise. Compared with activity produced by a vertical line, activity in the -20 degree channel has increased and that in the other two channels has decreased.

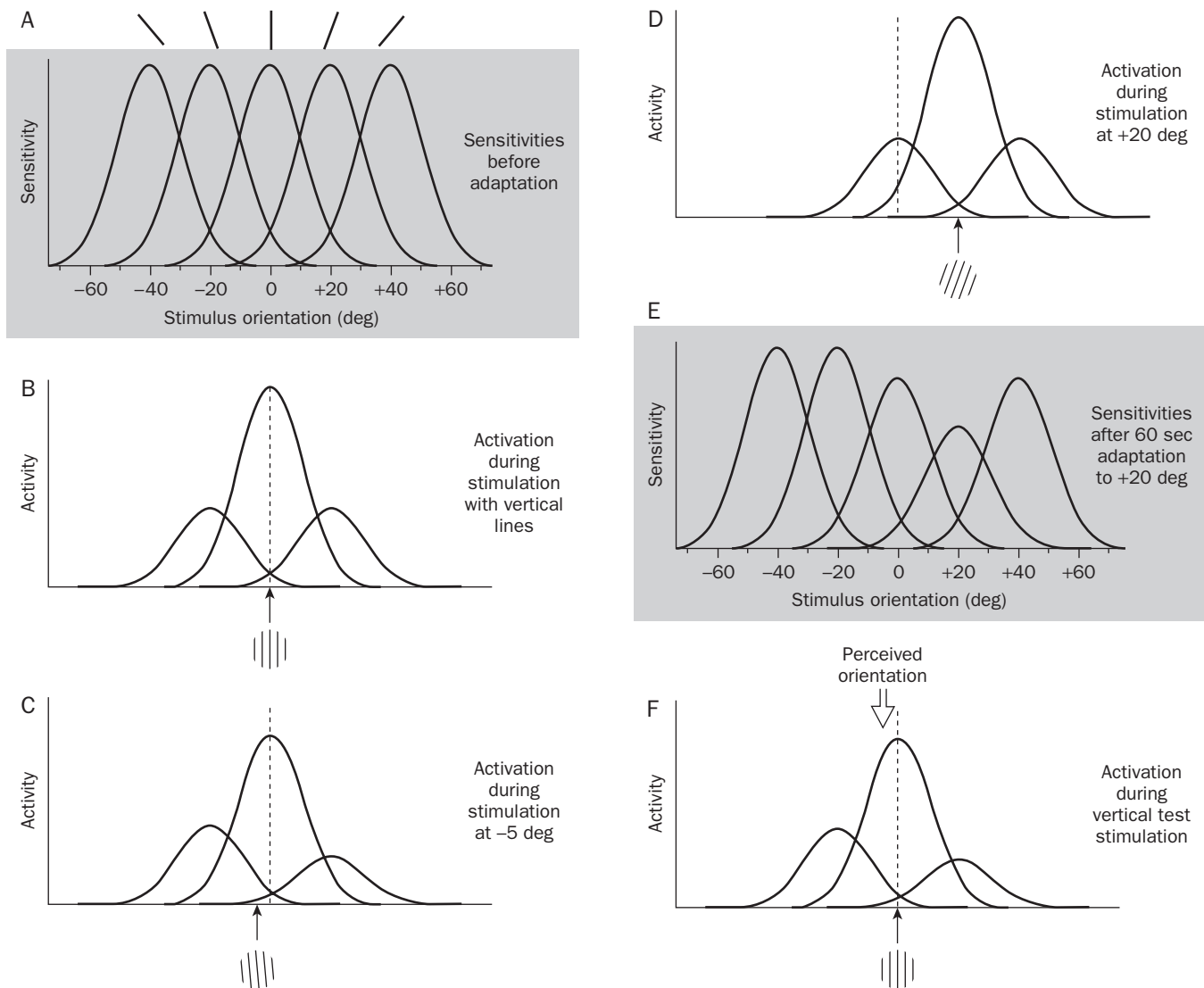
How is the information from all these channels combined when a visual stimulus is presented? There is likely to be a process that combines the activities across all channels, weighted according to the level of activity in each channel. Such a process finds the 'centre of gravity' of the distribution of activity. The centre of gravity (in statistical terms, the weighted mean) corresponds to the perceived orientation of the stimulus.

The tilt after-effect

During prolonged stimulation, the activity in the stimulated channels falls – in other words, channels 'adapt'. This fall is proportional to the amount of activity, so adaptation is greatest in the most active channels. After the stimulus is removed, recovery occurs slowly. We can see the effects of adaptation by presenting test, or 'probe', stimuli in the period shortly after the adapting stimulus has been removed. For example, think back to the waterfall illusion: when you gaze at a waterfall and then transfer your gaze to a point on the banks of the waterfall, you notice an apparent dramatic upward movement of the banks.

So we can explain the tilt after-effect as follows. Initially, all channels have equal sensitivity (as in panel A, figure 8.3). During presentation of a vertical stimulus, the distribution of active channels is symmetrical about zero, so the perceived orientation corresponds to the actual stimulus orientation – i.e. vertical (panel B, figure 8.3). A stimulus that falls between the optimal values of two channels is also seen veridically (that is, true to its actual orientation) by taking the centre of gravity of the activity pattern; this is how we see, for example, a small degree of tilt away from vertical (panel C, figure 8.3). With stimuli tilted 20 degrees clockwise, the active channels are also symmetrically distributed and have a centre of gravity at 20 degrees, so perception is again veridical (panel D, figure 8.3).

But during a prolonged presentation of such a stimulus (for, say, 60 seconds), the 20 degree channel adapts and its sensitivity declines. The reduction in each channel's sensitivity is proportional to the amount that it is excited by the stimulus, so the 0 degree and 40 degree channels are also adapted and have become

**Figure 8.3**

An explanation for after-effects.

less sensitive due to the presentation of this stimulus tilted 20 degrees clockwise, although to a smaller extent than the 20 degree channel. (The two channels that respond best to anti-clockwise tilts are not adapted at all.) The effects on sensitivity in the channel system of adapting to +20 deg stimulus are shown in panel E, figure 8.3. Sensitivity is reduced most in the +20 deg channel, and to a lesser but equal extent in the 0 and +40 deg channels.

What happens when we present a test stimulus whose tilt is zero (panel F, figure 8.3)? The -20 degree channel will give a small output, as normal, because the stimulus is away from the channel's optimal orientation, although within the range of tilts to which it is sensitive. But the output of the +20 degree channel will be even smaller, not only because the stimulus is not optimal for the channel, but also because the channel's sensitivity has

been reduced by the prior adaptation to a 20 degree stimulus. So the -20 degree channel will clearly be more active than the +20 degree channel, although its normal optimal is equally far from the vertical orientation of the stimulus. The distribution of activity across channels will therefore be asymmetrical, with its mean shifted towards negative tilts. So, after adaptation to a +20 deg stimulus, the pattern of activity in the channel system produced by a vertical test stimulus will be identical to that produced before adaptation by a -5 deg stimulus (compare panels C and F, figure 8.3). So the observer's percept is of a tilt at 5 degrees to the left. Finally, as the channel's sensitivities return to normal after adaptation, so the apparent orientation of the test bar changes back to vertical.

This general idea can explain other after-effects too, such as those for luminance and colour, for texture, pitch, and so on.

Localization and inter-ocular transfer

Other characteristics of the underlying mechanisms can also be inferred from the properties of after-effects. For example, visual after-effects are usually confined to the adapted region of the visual field. So staring at a small red patch does not change the perceived colour of the whole visual field but only of a local region. This point is emphasized in figures 8.1 and 8.2, which demonstrate that opposite after-effects can be induced simultaneously above and below the fixation point in each panel.

inter-ocular transfer the adaptation or learning that occurs when a training stimulus is inspected with one eye and a test stimulus is subsequently inspected with the other eye

In addition, most visual after-effects show *inter-ocular transfer*. This means that if the observer stares at a stimulus with only one eye, the tilt and other after-effects can be experienced not only with the adapted eye but also with the corresponding retinal region in the other eye, which is not adapted. These two properties suggest that such after-effects are mediated by mechanisms that are linked to a particular region of the visual field and can be accessed by both eyes. In other words, they suggest that the mechanisms underlying these after-effects are located centrally (i.e. within the brain) after information conveyed from the two eyes has converged, rather than peripherally (i.e. within each eye or monocular pathway).

Neurophysiologists recording the electrical activity in single nerve cells in the visual systems of cats and monkeys have discovered that in area V1 (the cortical area where information from the eyes first arrives – see figure 8.10, below), many neurons have properties that would enable them to mediate visual after-effects. Different neurons in V1 respond to the orientation, size, direction of motion, colour and distance from the animal of simple stimuli such as bars or gratings.

Many of the neurons in V1 are binocular, meaning their activity can be changed by stimuli presented to either eye. And they are linked to particular and corresponding places on each retina, which means that a stimulus has to fall within a particular region (receptive field) on one or both retinas to affect them. Neurons in V1 also, as you would expect in a mechanism which mediates the tilt after-effect, adapt to visual stimulation, so their response to a stimulus declines over time with repeated presentation (Maffei et al., 1973).

The localized receptive fields and binocular characteristics of these neurons correlate very well with the perceptual characteristics of after-effects described above. Although adaptation occurs in other visual cortical areas, the neurons in area V1 are prime candidates for the mechanisms that underlie visual after-effects in people.

One implication of this account of early visual processing is that the images of complex objects (trees, houses, people) are initially analysed by mechanisms that respond to their local physical characteristics and have no connection with the identity of the objects themselves. From the point of view of a neuron in V1, the vertical blue edge moving to the left might as well belong to a train as to the shirt of the frustrated passenger who has just missed it and is running along the platform in desperation after it.

In other words, the visual system appears initially to decompose the scene into its constituent parts and to analyse these separately (i.e. in parallel).

MAKING SENSE OF THE WORLD

GROUPING AND SEGMENTATION

The patches of light and shade that form a retinal image are produced by a world of objects. The task of the visual system is to represent these objects and their spatial relationships. An important step towards this goal is to work out which local regions of the retinal image share common physical characteristics, and which do not. These processes are known as grouping and segmentation, respectively.

Many of the stimulus attributes that give visual after-effects, and are probably encoded at an early stage of cortical processing, are also important in segmentation and grouping. Figure 8.4 shows a display used in a classic study by Beck (1966), who presented his observers with three adjacent patches of texture. Their task was to decide which of the two boundaries between the three regions was most salient, or prominent. They chose the boundary between the upright and tilted Ts, even though, when presented with examples of single texture elements, they said that the reversed L was more different from the upright T than was the tilted T.

This suggests that similarities and differences in orientation between elements of different textures, rather than their perceived similarity when presented in isolation, govern whether elements of different types are grouped or segregated. Segmentation and grouping can also be done on the basis of motion (Braddick, 1974), depth (Julesz, 1964, 1971) and size (Mayhew & Frisby, 1978) as well as colour and luminance. Figure 8.5 shows (A) an array of randomly positioned identical vertical lines, in which no sub-regions or boundaries appear, and (B) the same array but with a sub-set of the lines coloured red. The region of red texture now appears as a figure against a background of black lines. In (C) the figure is defined by making the same sub-set of lines

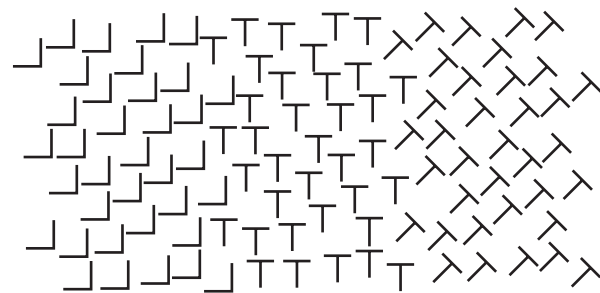


Figure 8.4

Grouping and segmentation by orientation.

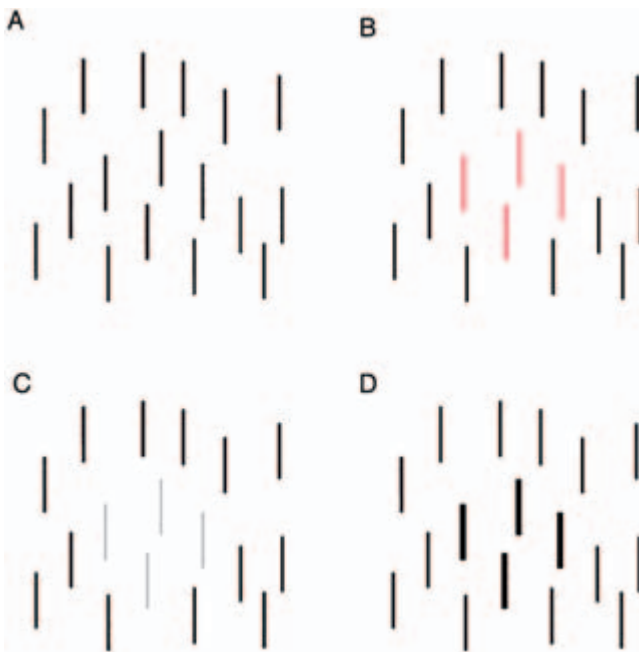


Figure 8.5

Grouping and segmentation by colour, luminance and size, showing how similarities and differences in physical characteristics are powerful cues to seeing isolated elements as belonging together or as separate.

as in (B) of a different luminance rather than a different colour from those forming the background. And in (D) the figure is defined by making the local elements of a different width to those in the background.

In addition to the nature of the elements within a display, their spatial arrangement can also contribute to grouping. Figure 8.6 shows how proximity can interact with shape. In panel A, the equi-spaced circular dots can be grouped perceptually either in rows or in columns. The dots are all physically identical and their vertical and horizontal separations are the same, so there is no reason for one or the other possible grouping to be preferred. This ambiguity may be resolved so that the elements are grouped as columns either by reducing the vertical separation of the dots (panel B), or by changing the shapes of alternate columns of dots (panel C).

Gestalt psychologists a group of German psychologists (and their followers) whose support for a constructionist view of perception has been enshrined in several important principles, such as ‘the whole (in German, *Gestalt*) is more than the sum of the parts’

The *Gestalt psychologists* first drew attention to effects of this type and attributed them to the operation of various perceptual laws (though they were really re-describing the effects rather than explaining them). Figure 8.6B illustrates the Law of Proximity and figure 8.6C the Law of Similarity.

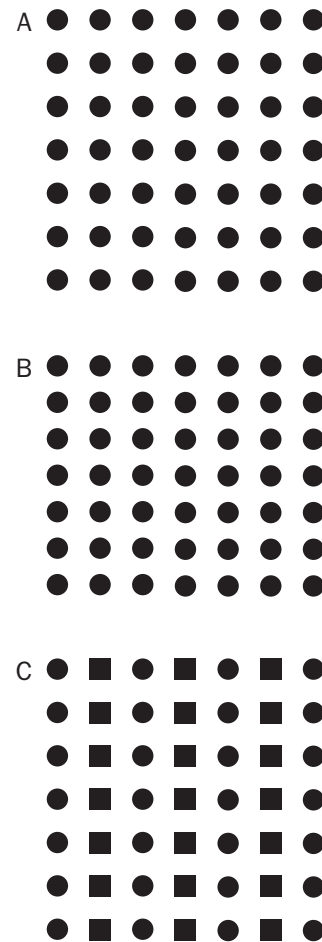


Figure 8.6

Grouping by proximity and similarity.

VISUAL SEARCH – OR FINDING THE CAR

Look for a tilted line in figure 8.7A. Carrying out a *visual search* for a target in an array of distractors (in the present case, vertical lines) is effortless and automatic: the target practically pops out from the array. In the same way, the region of texture formed from tilted Ts in figure 8.4 seems to stand out at first glance from the other two textures.

visual search a type of experiment in which the observer typically has to report whether or not a target is present among a large array of other items (distractors)

Parallel search

Performance on visual search tasks is often measured by the time it takes to complete the search. Psychologists then examine the effects of varying the nature of the difference between target and distractors, and the number of distractors. When the target differs

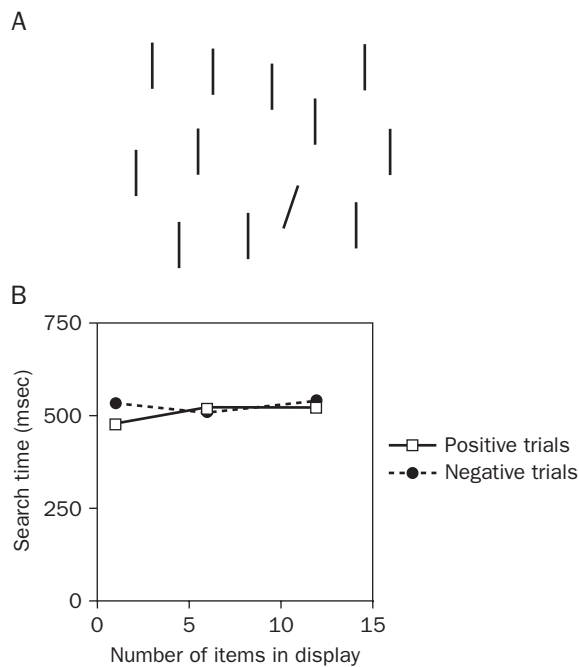


Figure 8.7

Pop-out and parallel visual search. Source: Based on Treisman and Gormican (1988).

from the distractors on only a single feature (such as tilt), the search time involved in making a decision whether or not a target is present is about the same whatever the number of distractors, and whether or not there is a target in the array (see figure 8.7B; Treisman & Gormican, 1988). In positive trials the

target is present in the display, whereas in negative trials the target is absent in the display. This pattern of performance is described as *parallel search*, as items from all over the display are analysed separately and simultaneously.

parallel search a visual search task in which the time to find the target is independent of the number of items in the stimulus array because the items are all processed at the same time (in parallel)

As well as tilt, stimulus dimensions on which target/distractor differences allow parallel search include luminance (Gilchrist et al., 1997), colour (Treisman & Gelade, 1980), size (Humphreys et al., 1994), curvature (Wolfe et al., 1992) and motion (McLeod et al., 1988). This list of features is very similar to those that give after-effects and govern grouping and segmentation.

conjunction search visual search for a unique conjunction of two (or more) visual features such as colour and orientation (e.g. a red tilted line) from within an array of distractors, each of which manifests one of these features alone (e.g. red vertical lines and green tilted lines)

Conjunction and serial search

Search tasks of this type can be contrasted with a second type, *conjunction search*, in which the target/distractor difference is not based on a

single feature, but on conjunctions of features. For example, the target might be a vertical red line in an array of vertical blue lines and tilted red lines (see figure 8.8A). In this scenario, search time for the target is not constant, but instead rises with the number of distractors. The observer apparently searches through the display serially, scanning each item (or small group of items) successively (*serial search*).

serial search a visual search task in which time to find the target increases with the number of items in the stimulus display, suggesting that the observer must be processing items serially, or sequentially

This kind of task might arise in real life when you have forgotten the location of your car in a large car park. You have to find a blue Ford amongst an array of cars of many makes and colours, where, for example, red Fords and blue Volkswagens are the distractors. The target does not pop out, but finding it requires effortful attentive scrutiny (Treisman & Gormican, 1988). When search times are compared for scenes in which a target is or is not present, the times rise with the number of visible items, but they rise twice as steeply when there is no target (see figure 8.8B). This is probably because, when there is a target present (which can occur anywhere in the visual display), on average, the observer has to scan half the items in the display to find it. When there is no target, on the other hand, the observer has to scan all the items in the display in order to be sure that no target is present.

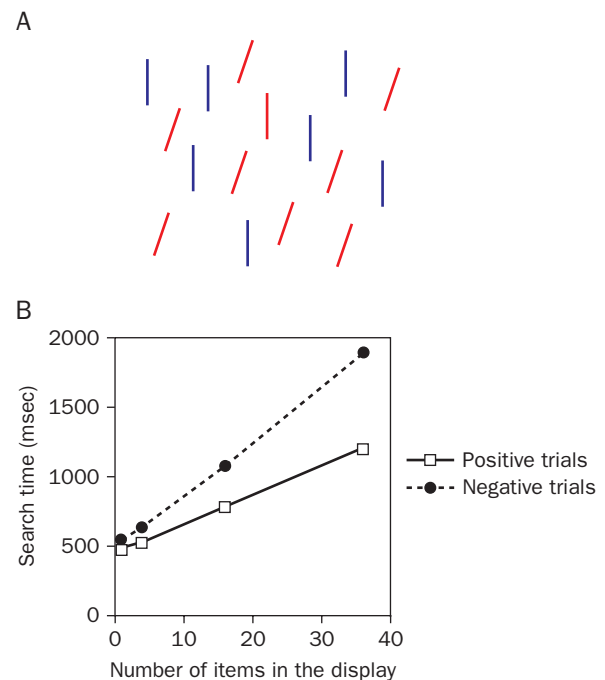


Figure 8.8

Conjunction search. Source: based on Treisman & Gormican (1988).

Feature integration theory

feature integration theory different features of an object (e.g. colour, orientation, direction of motion) are thought to be analysed separately (and in parallel) by several distinct mechanisms, and the role of attention is to 'glue together' these separate features to form a coherent representation

Based on findings from parallel search and conjunction search tasks, Treisman and colleagues put forward a theory – the *feature integration theory* – which sought to explain the early stages of object perception. This theory has become very influential (Treisman & Gelade, 1980; Treisman & Schmidt, 1982). These authors suggested that the individual features that make up an object (its colour, motion, orientation, and so on) are encoded separately and in parallel by pre-attentive cognitive mechanisms. However, in order to perceive a whole object, the observer needs to 'glue together' (or integrate) these separate features, using visual attention.

One interesting prediction from the theory (which has been borne out by experiments) is that, if attention is diverted during a conjunction search task, there would be nothing to hold the features of an object together, and they could then change location to join inappropriately with features of other objects. For example, if observers are distracted by requiring them to identify two digits during the presentation of a display like that in figure 8.9, they often report seeing dollar signs, even though the S and the straight line which make up the sign are never in the same

illusory conjunctions perceptual phenomena which may occur when several different stimuli are presented simultaneously to an observer whose attention has been diverted (e.g. the perception of a red cross and a green circle when a red circle and a green cross are presented)

location. It is as though, pre-attentively, the S and the parallel lines are 'free-floating' and are able to combine to present objects that are not physically in the display. These so-called *illusory conjunctions* provide support for feature integration theory (Treisman & Schmidt, 1982; Treisman, 1986).

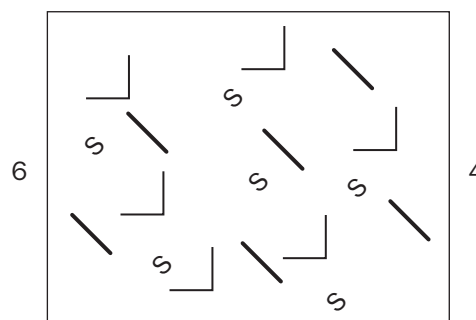


Figure 8.9

'Illusory conjunctions' occur when features from one object are erroneously combined with features from another object. Source: Treisman (1986).

How do we know what we see?

Treisman's ideas suggest that image features like colour and motion are analysed separately at an early stage of visual processing. As we shall see, this is consistent with evidence from anatomical and physiological studies of the visual system, and studies of humans with certain kinds of brain damage.

Up to now we have discussed visual neurons as *feature detectors*, responding best to certain aspects of the retinal image, such as the orientation or direction of movement of an edge. But recent studies suggest that, rather than forming part of a single homogeneous visual system, the feature detectors are embedded in several different sub-systems, in which information is processed separately, at least to some extent.

feature detector a mechanism sensitive to only one aspect of a stimulus, such as red (for the colour dimension) or leftwards (for direction of motion) and unaffected by the presence or value of any other dimension of the stimulus

Magno and parvo cells

The rods and cones in the retina function in dim and bright light, respectively. The cones are of three types, which are selective to different, if overlapping, ranges of light wavelength. The information from the cones is reorganized in the retina to give green–red and blue–yellow opponent channels (see chapter 7).

There is, in addition, a group of large retinal cells alongside the smaller colour-opponent cells. These large cells respond to the difference between the luminances (of any wavelength) in their centre and surrounding regions. They could be described as black–white opponent channels. The large cells are known as the *magno* or *M cells*, contrasting with the colour-sensitive

magno (M) cell a large cell in the visual system (particularly, the retina and lateral geniculate nucleus) that responds particularly well to rapid and transient visual stimulation

Pioneer

Anne Treisman (1935–), one of the pioneers of the empirical study of selective attention, went on to develop the influential feature integration theory. This theory suggests that attention involves the binding of feature information about an object across a network of parallel processing mechanisms, each of which handles separate and distinct features of the object (such as its colour, motion and orientation).

parvo (P) cell a small cell in the visual system (particularly, the retina and lateral geniculate nucleus) that responds particularly well to slow, sustained and coloured stimuli

field sizes, but in being more sensitive to movement and to black–white contrast. M and P cells both receive inputs from both cones and rods, but M cells do not distinguish between the three cone types and so respond positively to light of any wavelength, whether dim or bright.

The motion properties of M cells are exceptionally important. They respond to higher frequencies of temporal flicker and higher velocities of motion in the image than P cells do. Indeed M cells signal transients generally, while the P channels deal with sustained and slowly changing stimulus conditions. For example, a dim spot of white light switched on or off seems to appear or disappear suddenly, whereas a dim spot of coloured light seems to fade in or out gradually (Schwartz & Loop, 1983). This supports the hypothesis that different flicker/motion sensations accompany activation of M and P channels.

Most famously, Livingstone and Hubel (1987) ascribed colour sensations to P cell activity, motion and distance (depth) to M cell activity, and spatial pattern analysis to a combination of both. This tripartite scheme was based on a reorganization of the retinal information that subsequently occurs in the cerebral cortex. The optic nerve carries signals to a pair of nuclei near the centre of the brain called the LGN (lateral geniculate nuclei), and from there the signals are sent on to the primary visual cortices (area V1) at the back of the brain (see figure 8.10). There are perhaps 100 million cells in each of the left and right areas V1, so there is plenty of machinery available to elaborate on the coded messages received from the retina.

Cortical pathways

In the cortex, the general flow of information runs vertically – that is, to cells in other layers above and below the activated cells.

Pioneer

David Hubel's (1926–) discovery, with Torsten Wiesel, of the orientation tuning of cells in the primary visual cortex initiated an entire industry investigating how the visual scene can be encoded as a set of straight-line segments. Their theory also became a cornerstone for serial processing models of visual perception. Later, though, with Margaret Livingstone, he supported the theory that visual features are processed in parallel streams stemming from magno and parvo cells in the retina.

parvo or *P cells* (the names are taken from the Latin words for 'large' and 'small' respectively). The M cells differ from the P cells not only in their lack of colour selectivity and their larger receptive

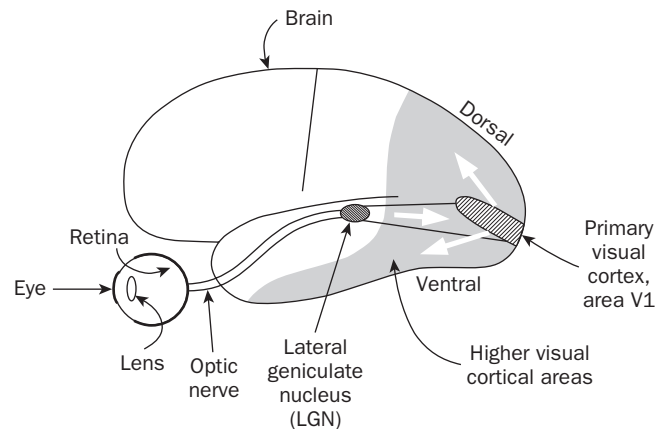


Figure 8.10

The early stages of the neural pathways for analysing the visual stimulus. The lateral geniculate nucleus (LGN) and the primary visual cortex (area V1) pass information on to the 'higher' areas of cortex (shown shaded). The latter can be divided into the dorsal and the ventral streams leading to the parietal and the inferior temporal lobes respectively.

The cortex contains *columns* of cells, which respond to similar properties of the stimulus and lie alongside other columns that respond to different aspects or features of the world.

The earlier work of Hubel and Wiesel (1968) emphasized this vertical organization. They discovered that, unlike the retina and LGN, where neurons respond best to spots of light, many cortical neurons respond best to straight lines or edges. Some cells respond best to vertical lines (figure 8.11), others to diagonals, others to horizontals, and so on for all orientations around the clock.

There is a very fine-grained, high-resolution representation of image-edge orientation at this stage of sensory processing. Moreover, each cell is sensitive only to lines in a relatively small area of the retinal image – the cell's receptive field. The cells are also selective for the spacing between parallel lines (the spatial frequency), and in many cases also for the direction of stimulus movement, the colour of the stimulus and its distance.

These cortical cells form the basis for the tilt after-effect and the other after-effects described above. The activity in these cells probably also underlies our perception of orientation, motion, etc. Even if we knew nothing about the neural organization of the visual system, we could suggest the existence of mechanisms with some of the properties of these cortical neurons, which we would infer from the properties of visual after-effects. However, the further evidence that has been obtained by researchers regarding these brain mechanisms gives us greater confidence in their actual reality, and shows how psychology and neurophysiology can interact to form a satisfyingly interlocking pattern of evidence.

column a volume of cells stretching the entire depth of the cerebral cortex, which all have some physiological property in common (e.g. the preferred orientation of the bar or edge stimulus to which they respond, in the case of a column in the primary visual cortex)

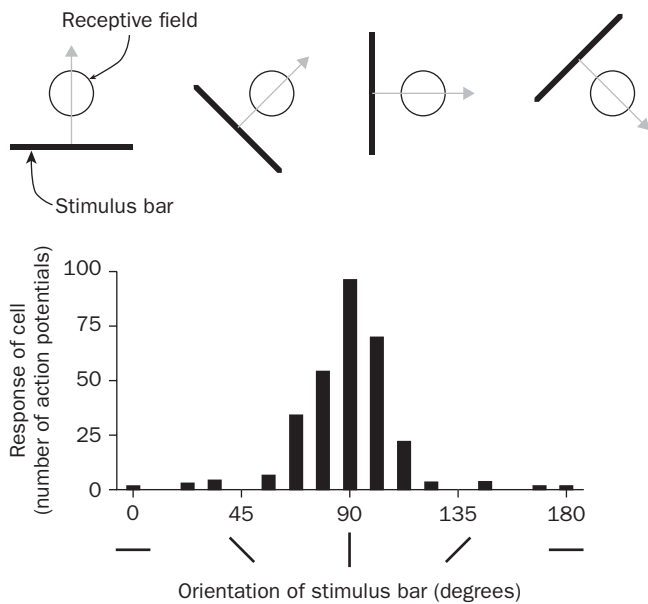


Figure 8.11

Orientation tuning of a typical nerve cell in area V1. In each of the four diagrams at the top, a circle indicates the receptive field. A bar stimulus is moved across the receptive field at different angles in each diagram. At the bottom is plotted the magnitude of the response of the cell to the bar at each of these angles and to other angles around the clock. The cell is highly selective for the orientation of the bar, responding only to near-vertical bars.

Pioneer

Horace Barlow (1921–) is a physiologist whose insights into the possible relationships between perception and neural activity have guided much thinking in the field. Barlow is especially well known for his discussion of whether we possess ‘grandmother cells’. These are single neurons whose activity would reflect the presence of an elderly female relative. More generally, do we possess cells within our brain that respond selectively to very specific familiar visual experiences in our environment, such as the sight of our car, our house or our grandmother?

Serial versus parallel theories of perception

The research on visual cortical neurons was at first thought to support serial hierarchical theories of perception (Selfridge, 1959), in which perception is thought to proceed in a sequence of stages, starting at the retina and ending (presumably) somewhere in the cortex, with information flowing in just one direction.

Such frameworks can be called ‘hierarchical’ because a unit in each successive stage takes input from several units in the preceding stage. This kind of organization could be likened to the

Catholic church, in which several parish priests report to a bishop, several bishops to a cardinal, and several cardinals to the pope. In the same way, general features of the retinal image, such as lines, were thought to be extracted by early visual processing, while whole complex objects were recognized later in the sequence by the analysis of combinations of these features. For example, the capital letter ‘A’ contains a horizontal line and two opposite diagonals, the letter ‘E’ contains three horizontals and a vertical, and so on. These letters can therefore be defined with respect to a combination of their elementary perceptual features. Representations of corners, squares, and then three-dimensional cubes, were thought to be built up by combining the outputs of these early feature detectors to form more complex object detectors in ‘higher’ regions, such as the cortex of the inferior temporal lobe.

However, more recently there has been an increasing emphasis on the parallel organization of the cortex (Livingstone & Hubel, 1987). So in V1, M and P cell signals (projected from the magno and parvo components of the retina, respectively) arrive in different layers of the cortex. These messages are processed in V1 and are then carried by axons out of V1 and into several adjacent regions of the cortex, called V2, V3 and V5 (see figure 8.12). In V2, Livingstone and Hubel argued that the M and P signals are kept separate in different columns of cells. Consistent with our previous discussion these columns represent information about motion and distance (magno system) and colour (parvo system), respectively.

This theory became complicated by Livingstone and Hubel’s description of activity in a third type of column in V2, where the cells receive converging input from the magno and parvo systems. They suggested that these columns are used for spatial pattern analysis. However there are problems with this scheme. For example, Livingstone and Hubel claimed that images in which the different regions are red and green, but all of the same brightness appear flat. They attributed this to the insensitivity of cells in the magno/depth system to differences purely in hue, which are detected primarily by the parvo system. Quantitative studies,

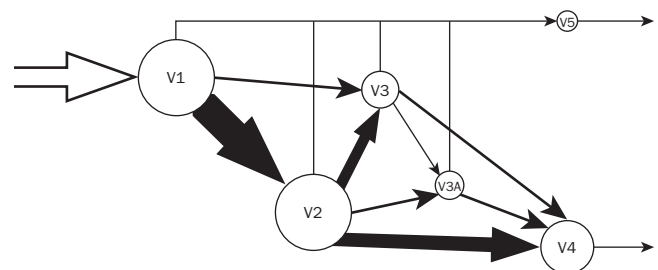


Figure 8.12

Ascending connections between some of the major areas of visual cortex. The size of each circle is proportional to the size of the cortical area it represents, and the width of each arrow indicates the number of nerve axons in the connection. Although the arrows show information flowing principally from left to right in the diagram, hence from ‘lower’ to ‘higher’ areas, in fact most of the connections are reciprocal, which is to say information passes in both directions along them. Source: Adapted from Lennie (1998); Felleman and van Essen (1991).

however, found that perceived depth is not reduced at all in such images (Troscianko et al., 1991). It appears, then, that depth percepts can be derived from both magno and parvo information, though not necessarily equally well at all distances (Tyler, 1990).

In fact, there are many more visual areas in the cerebral cortex than are shown in figure 8.12. Some two dozen or so have now been discovered by neuroanatomists and by brain imaging studies (see chapter 3). The functions of these areas are still being studied intensively by physiologists and psychologists, and we do not yet have the complete picture.

Zeki (1993) has put forward the most influential theory of cortical visual functioning. According to this scheme, area V3 is important for analysing stimulus shape from luminance or motion cues, V4 is important for the perception of colour and for recognising shape from colour information, and V5 is critical for the perception of coherent motion. But this theory is still controversial. Recent physiological studies have found fewer differences between the properties of the various cortical areas, emphasizing that many areas co-operate in the performance of any given task. For example, Lennie (1998) points out that most information flow in the brain is from V1 to V2 to V4, and that area V4 is not specialized for colour in particular, but for finding edges and shapes from any cue or feature. Lennie argues that only the small stream through V5 is specialized, to monitor image motion generated by self-movement of the body and eyes (optic flow). This would therefore be the area activated in the illusion of self-motion we experience when the other train moves, as described at the very beginning of this chapter.

SEEING WITHOUT KNOWING

Destruction of small parts of the cortex, for example after stroke, tumour, surgery or gunshot wounds, can result in bizarre and unexpected symptoms.

Colour and motion awareness and the strange phenomenon of blindsight

In the syndrome known as cerebral achromatopsia, for example, patients lose all colour sensations, so the world appears to be in shades of grey (see Sacks, 1995, for a good example, and Zeki, 1993, for historical details). If the damage is restricted to a small portion of the lower surface of the occipital lobes, the loss of colour vision can occur without any other detectable anomaly: visual acuity is normal, as are depth perception, shape understanding, and so on.

Recently, another syndrome has been associated with damage to a lateral part of the occipital lobe: akinetopsia. Someone with akinetopsia loses motion awareness, so that visual stimuli all look stationary even when they are moving. These patients notice if there is a change of stimulus location, but there is no sense of pure motion occurring between the two successive locations (Zihl et al., 1983).

Syndromes like this support the theory that humans possess many specialized processing areas, as do monkeys and other

primates. These specialisms contrast with the general loss of subjective vision that follows lesions of the primary visual cortex, area V1. This has been strikingly demonstrated by rare cases of damage to V1 in one hemisphere of the brain. Vision is then affected in one half of visual space, so if your right visual cortex is damaged and you look straight ahead, everything to the left of you is in some way visually absent or missing.

Interestingly, though, there are some visual stimuli that can still evoke behavioural responses in the 'blind' half of the visual field. For example, if you hold a stick in the blind field and ask the person, 'Am I holding this stick vertically or horizontally?' they will say, 'What stick? I can't see anything over there at all.' So you say, 'Well, I am holding a stick, so please guess what the answer is.' Amazingly, these patients will answer correctly most of the time, and much more often than they would by chance guessing. Their behavioural responses to large visual stimuli, including the location, motion and orientation, presented in the blind half of the visual field will be correct more than nine times out of ten. They cannot respond to the fine details of the scene, and they cannot initiate movements towards stimuli they have not been told are there, but something remains of their previous visual capacities within the blind half of the field. This phenomenon has been termed 'blindsight' (Weiskrantz et al., 1974). It has been of great interest in recent studies on how subjective awareness of the visual world arises (e.g. Zeki, 1993; Weiskrantz, 1997).

The ventral and dorsal streams

Leading away from area V1, a distinction is generally made between two broad streams of parallel visual processing (see figures 8.10 and 8.12 above). These were initially known as the 'what' and the 'where' stream, but there has been some dispute over the exact role of the latter, since some researchers believe it is also involved in the visual control of movements (the 'how' stream), not simply in locating objects. Partly for this reason, the streams have since become known as the 'ventral' and 'dorsal' streams, emphasizing their (uncontroversial) anatomical locations, not their more controversial functional roles.

The ventral stream takes mainly parvo retinal input from V1 and flows towards the inferotemporal cortex, where cells respond to the sight of whole, complex three-dimensional objects (or at least to the constellations of features that characterize these objects). Damage to this stream impairs object recognition and knowing what objects are for (Milner & Goodale, 1995; Newcombe et al., 1987). This stream includes a specialized area that deals selectively with face recognition (see chapter 7), and which is damaged in the syndrome called prosopagnosia (as in the example of the man who mistook his wife for a hat: Sacks, 1985).

The dorsal stream, in contrast, takes magno input and runs into the parietal lobe. It deals with locating objects and with sensorimotor coordination, mostly occurring subconsciously. Damage to the parietal lobe can hamper the ability to grasp something with the hand or post a letter through the slot in a mailbox (Milner & Goodale, 1995). With right parietal lesions particularly, it becomes difficult to recognize objects from unusual points of view (such as a bucket from above), rotate an object mentally,

read a map, draw, use building blocks, and pay attention to spatial locations especially on the left side of space (a phenomenon known as spatial 'neglect'; Robertson & Halligan, 1999).

In summary, these different lines of evidence strongly support the idea of parallel processing. However, they do not explain why our behaviour is not a bundle of reflex reactions to sensory stimuli. In the next sections, we consider the role of different types of cognitive knowledge in perception.

SEEING WHAT WE KNOW

Various kinds of knowledge about the world can be shown to influence perception. One class of perceptual processes seems to reflect an assessment of what it is that particular features of the stimulus are most likely to represent.

PERCEPTION OR HALLUCINATION?

What do you see in Figure 8.13? Most observers perceive an inverted 'whiter-than-white' triangle with clearly defined edges filling the space between the black discs, each with a sector removed. This inverted triangle is illusory, since the white paper on which it is perceived is of the same luminance as that outside the triangle. It is as though, faced with the incomplete black discs and line corners, the visual system makes the best bet – that this particular configuration is likely to have arisen through an overlying object occluding complete black discs, and a complete outline triangle. In other words, since the evidence for an occluding object is so strong, the visual system creates it.

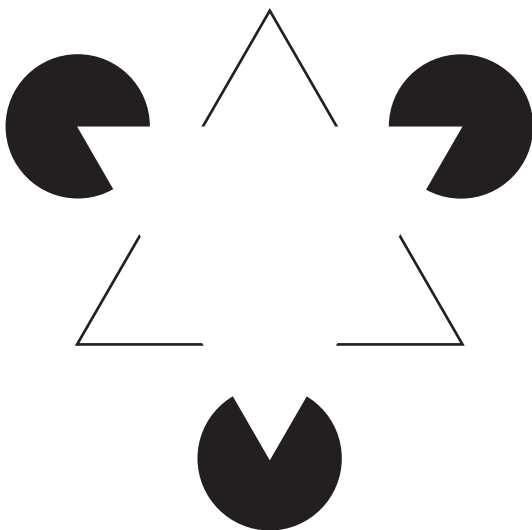


Figure 8.13

A Kanizsa figure. Most observers report perceiving a white triangle whose corners are defined by the 'cut-outs' on the black discs and whose edges touch the free ends of the Vs.

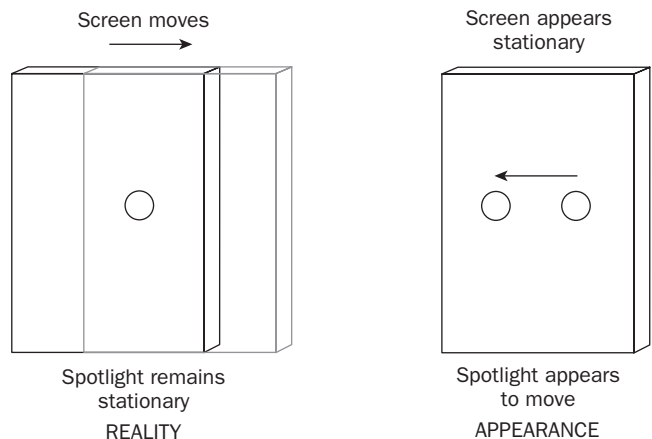


Figure 8.14

Schematic representation of induced movement. A stationary spot is projected on a screen which moves from side to side. Source: Adapted from Gregory (1997).

A rather different example is depicted in figure 8.14. A small spot is projected onto a large frame or screen, which is then moved. What the observer sees is the spot moving on a stationary screen. Again, this appears to reflect an assessment of relative probabilities. Small foreground objects are more likely to move than large background objects, and so this is what the observer sees.

RESOLVING VISUAL AMBIGUITY

What happens when alternative probabilities are about equal? The outline (Necker) cube in figure 8.15 appears to change its orientation spontaneously. Sometimes the lower square face of the cube appears nearer, and sometimes the upper square face. This reflects the absence of depth information from shading, perspective or stereopsis (3-D vision based on differences in the visual information received by each eye) that would normally reveal the orientation of the cube. Faced with two equally good interpretations, the visual system oscillates between them.

But why does our visual system fail to generate a single stable percept, which is veridical (i.e. matches the characteristics of the scene exactly), namely a flat drawing on a flat sheet of paper? The reason that the brain chooses to interpret the scene as 'not flat' seems to reflect the power or salience of the depth cues provided by the vertices within the figure.

Further evidence for this comes from the fact that we can bias the appearance of the cube by changing our point of view. So if you fixate the vertex marked 1 in figure 8.15, the lower face will seem nearer. Fixate the vertex marked 2 and the upper face tends to appear nearer. But why does this happen? Again, the answer brings us back to probabilities. When we fixate a particular vertex, it is seen as protruding (i.e. convex) rather than receding (concave). This is probably because convex junctions are more

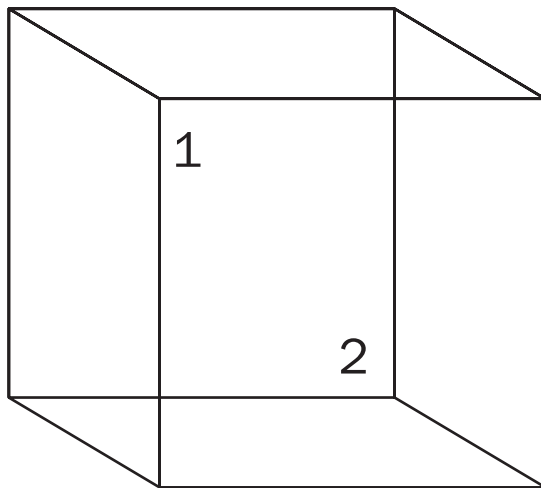


Figure 8.15

The Necker cube. This wire-frame cube is perceptually ambiguous or multi-stable, so that sometimes the upper square face appears nearer and sometimes the lower square face appears nearer.

likely in the real world. To be sure, you will see concave corners (for example, the inside corners of a room), but most concave corners are hidden at the back of an object and therefore outnumbered by convex corners at the front. You can easily test this out by simply counting how many of each type you can see from where you are sitting now.

TRICKS OF THE LIGHT

Another powerful example of the effects of knowledge in perception is illustrated in figure 8.16. In the upper half of the figure, the circular blobs, defined by shading, appear convex, whereas in the lower half they appear concave. Rotating the page through 180 degrees reverses the effect. The blobs that appeared concave now appear convex, and vice versa (Ramachandran, 1995). Why?

Notice that the pattern of shading of the blobs is ambiguous. In the upper part of the figure, it could be produced if protruding blobs were illuminated from above, or if receding blobs were illuminated from below (and vice versa, for the lower half of the figure). Yet we tend to perceive them as protruding blobs illuminated from above. This is because our visual system tends to 'assume' (on the basis of previous probabilities) that objects in our world are lit from above (as they are in natural surroundings by our single sun), and this assumption governs the perception of ambiguous shading. Presumably, someone who lived on a planet where the only illumination came from luminous sand on the planet's surface would see the blobs on the upper part of figure 8.16 as receding and the blobs on the lower part as protruding.

If the gradient of shading is switched from vertical to horizontal (figure 8.17), then all the blobs, whether on the top or

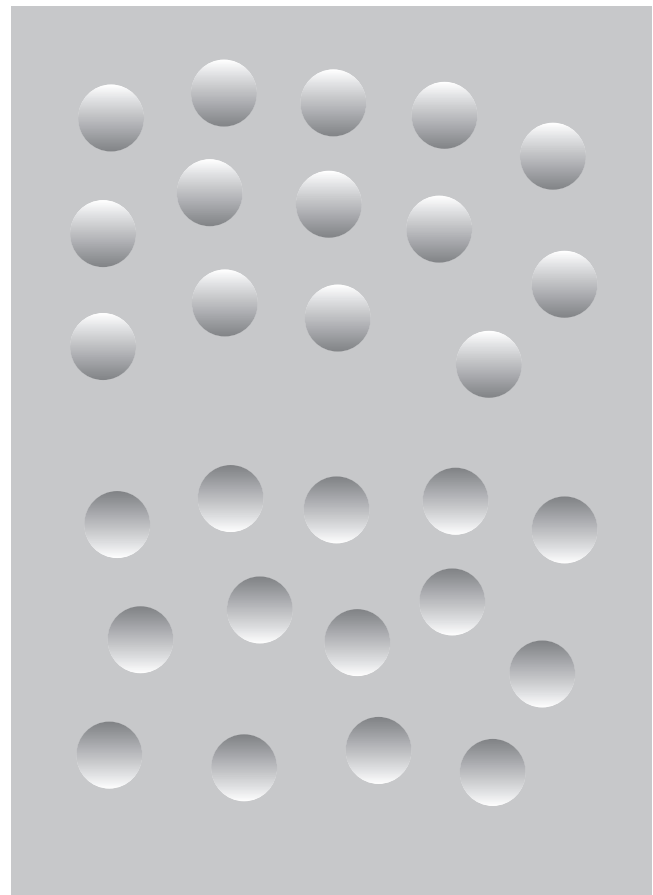


Figure 8.16

Shape from shading. The blobs in the upper half of the figure appear to protrude, and those in the lower half appear to recede.

bottom, tend to be seen as protruding. This suggests that, once the direction of illumination is clearly not vertical, it tends to be ignored. Instead, another assumption dominates perception, namely that ambiguous blobs protrude (the same assumption about vertices that governs perception of the Necker cube).

Although the assumption that objects are lit from above by a single light source is important, it does not always govern our perceptions, even when it is clearly applicable. Gregory (1997) has pointed out that it may be defeated by other knowledge about very familiar objects – in particular, human faces. Gregory drew attention to the fact that the hollow mask of a face does not usually appear hollow (figure 8.18A). Instead, the receding nose appears to protrude. It is only when the mask is viewed from a short distance that stereoscopic depth information (i.e. information from both eyes) is able to overcome the 'assumption' that noses always protrude.

What would happen if this assumption about noses were to conflict with the assumption that objects are lit from above? When the rear of a hollow mask is lit from below, the nose appears to protrude and to look as though it is lit from above, in line with both assumptions (figure 8.18B). But when the lighting

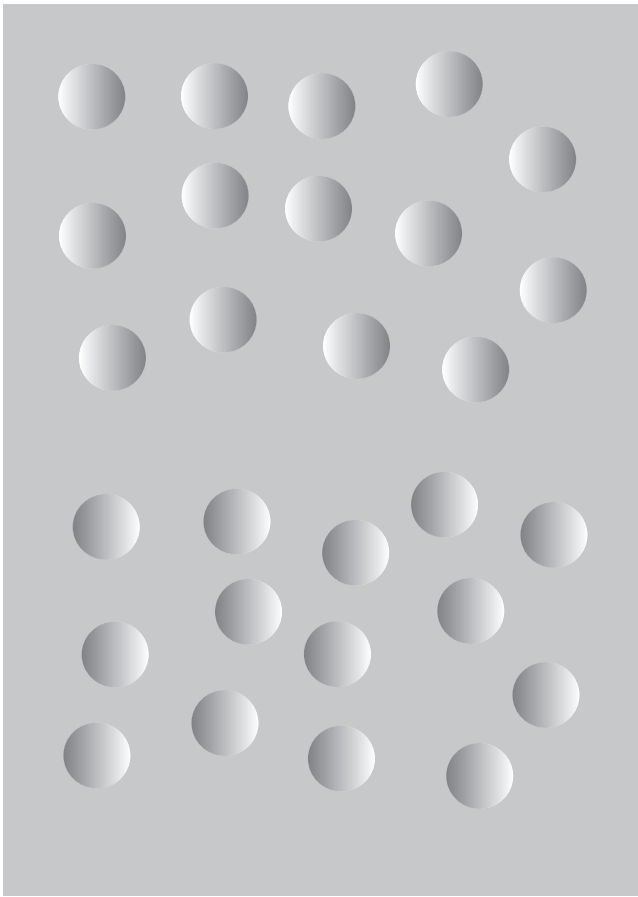


Figure 8.17

Shape from shading with lighting from the side.

Pioneer

Richard Gregory (1923–) is a well-known supporter of cognitive constructionist approaches to understanding perception. Originally trained in philosophy as well as psychology, he has summarized and reviewed much experimental evidence (some of which he has provided himself) for the ‘intelligence’ of the visual system in interpreting its input, and related this ‘top-down’ view of perception to its philosophical context. His books, especially *Eye and Brain*, have fired generations of students with an enthusiasm for the study of perception

is from above, the nose still appears to protrude, even though it also appears to be lit from below (figure 8.18C). Clearly the assumption that noses protrude is stronger than the assumption that objects are usually lit from above. This is probably because we have no day-to-day experience of non-protruding noses, but

we occasionally experience objects lit from below by reflected or artificial light.

NON-VISUAL KNOWLEDGE AND PERCEPTUAL SET

Perceptual assumptions about lighting and noses are probably common to all humans. But there are other kinds of knowledge affecting perception which depend on linguistic, graphic and other cultural conventions.

The central symbol in figure 8.19A is perceived as ‘B’ if the vertical set of symbols is scanned, and as ‘13’ if the horizontal set of symbols is scanned. Similarly, the central letter in the two words shown in figure 8.19B is perceived as an ‘H’ when reading the first word, and as an ‘A’ when reading the second. Such effects depend on knowledge of a particular set of alpha-numeric conventions and of the graphology of the English language (and so would presumably not be experienced by someone who spoke and wrote only Arabic). They illustrate that non-visual knowledge can be important in visual perception.

There are other situations in which the role of past experience and verbal clues become apparent. Figure 8.20A shows a dappled pattern of light and shade, which at first glance may appear meaningless. But consider the clues ‘leaves and a Dalmation dog’, and you will probably see the dog nosing among the leaves almost instantly. Similarly, the pictures in figure 8.20B have been transformed into black blocks and black lines, so that the identity of the objects they represent may not be obvious. But again, verbal clues such as ‘elephant’, ‘aeroplane’ or ‘typewriter’ are often sufficient for the observer to identify the objects.

Interestingly, once you perceive the Dalmation and the elephant, it is impossible to look at the pictures again without seeing them. These effects are sometimes described as examples of *perceptual set*: the verbal clues have somehow ‘set’, or programmed, the individual to interpret or perceptually organize ambiguous or impoverished stimuli in a certain way.

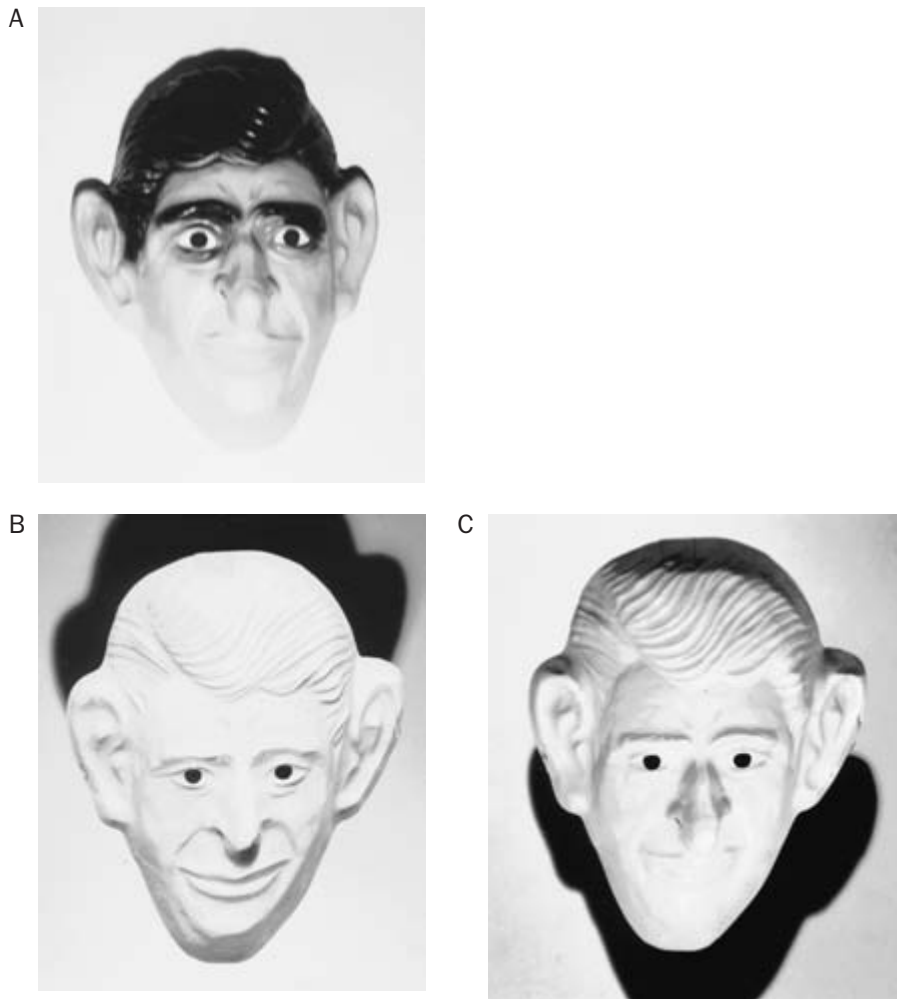
PERCEPTUAL LEARNING

HOW TRAINING INFLUENCES PERFORMANCE

Although the role of knowledge and assumptions in perception is now quite clear, the detailed ways in which past experience influences perception are less clear.

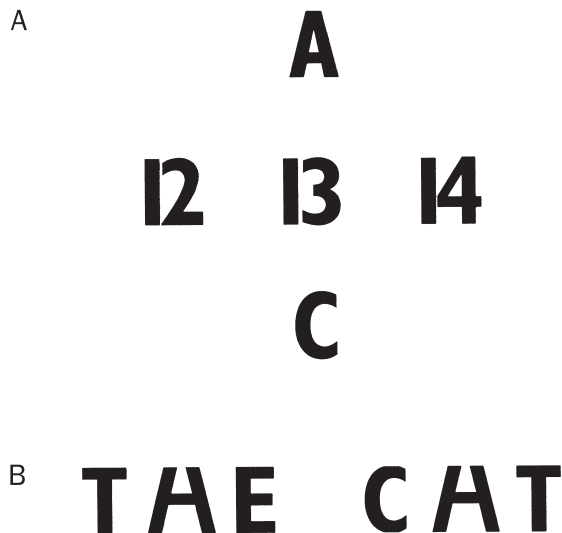
Recently, experimenters have begun to examine these questions by studying how training can influence performance on apparently simple visual tasks, such as judging whether the lower line in figure 8.21 is offset to the left or right of the upper line (a *vernier acuity* task). Humans can discern the direction of very tiny offsets,

vernier acuity the ability to see very small differences in the alignment of two objects, which becomes particularly obvious when the objects are close to one another

**Figure 8.18**

Opposing assumptions: 'light comes from above' vs 'noses protrude'.

(A) A hollow mask seen from the front and lit from below. (B) The same hollow mask seen from behind, with the lighting coming from below. (C) The same hollow mask seen from behind, but with the lighting coming from above.



but can improve even more with practice, though this may require thousands of presentations (Fahle & Edelman, 1993). The nature of the learning can be studied by measuring the extent to which it transfers from the training stimulus to other stimuli and conditions. Thus if, after training, the vernier stimulus is rotated through 90 degrees, performance on the new task is no better than it was at the start of the experiment. Similarly, performance falls if observers are trained on one retinal location and tested on others, or trained using one eye and tested on the other.

Figure 8.19

Effects of linguistic knowledge on the perception of objects.

(A) How the central symbol is read depends on whether one is scanning from left to right or from top to bottom. (B) How the central symbol in each cluster is read depends on the surrounding symbols.

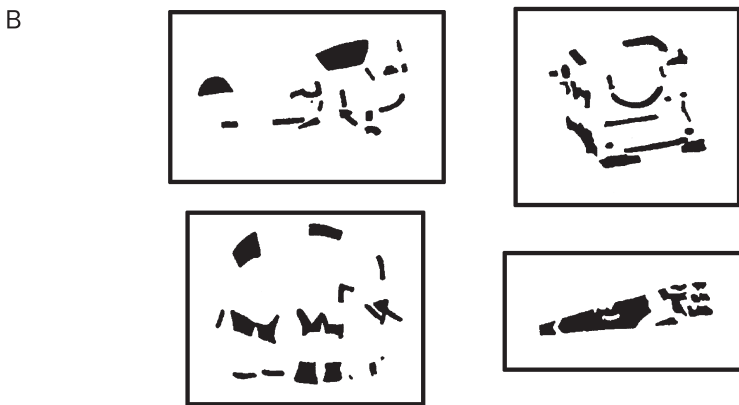


Figure 8.20

What are the hidden objects?



Findings like these suggest that some of the training occurs at a site where the neurons are driven by one eye, receive input from restricted regions of the retina, and are orientation-specific. Fahle (1994) speculated that the learning might reflect changes occurring in orientation-specific neurons in V1, some of which are monocular (driven by only one eye). Others have questioned the extent and nature of the specificity of learning, and suggested that there might be a general as well as a stimulus-specific component to the observed learning effects (e.g. Beard et al., 1995). This general component might reflect, for example, a change in the ability to direct attention to particular regions of the visual

Figure 8.21

In a vernier acuity task, the observer has to decide whether the lower line is to the left or to the right of the upper line. Performance on this task can improve dramatically with practice.

Research close-up 1

Can people learn to modify their visual search strategies?

The research issue

In a typical laboratory visual search task, the subject searches an array of items for a pre-defined target, pressing one key if the target is present, and a second key if the target is absent. Some visual search tasks seem easy and effortless, regardless of the number of items in the display. So the tilted line in figure 8.7A 'pops out' immediately, and search time to find the tilted target stays the same, regardless of the number of vertical distractors in the visual array, as though all items are being processed simultaneously (parallel search).

Other tasks are harder: for example it requires scrutiny of each item successively to find the vertical red line in figure 8.8A (serial search). Ellison and Walsh (1998) asked whether the mechanisms underlying these two types of search were fixed and immutable, or whether participants could learn to search in parallel for targets that initially required serial search.

Design and procedure

Observers sat in front of a computer screen. On each presentation, a number of items (which on half the trials included a target, and on the other half did not) appeared on the screen. The number of items that appeared in the search array was systematically varied. Observers carried out search tasks repeatedly in several training sessions spread out over several days (so that each subject experienced more than 2000 presentations of the search array).

One group of participants was trained on target search tasks that were initially serial (i.e. in which target search time increased with the number of distractors), and the other group on tasks that were initially parallel (i.e. in which target search times did not change with the number of distractors).

Results and implications

The experimenters found that, as training proceeded, performance on the initially serial task became parallel, so that search times no longer increased with the number of distractors. But this only occurred for those presentations on which a target was present. Although search time was reduced by training when there was no target present, it still rose with the number of distractors.

The performance of the group who were trained on parallel search tasks also improved with training. When they were tested on serial search tasks after training on parallel tasks, their performance was better when there was a target present than if they had previously had no training. But for presentations on which there was no target present, they were worse than if they had not been trained.

So, training can lead to a change in search strategies, but this does not seem to be a generalized improvement in performance. In parallel search tasks, the observers may learn to distribute their attention more evenly over all regions of the display, so that they can respond quickly to the signal from pre-attentive mechanisms wherever the target occurs, and to the absence of a signal when there is no target. When transferred after training to a serial task, this strategy will allow faster detection of items likely to be a target, which then receive attentional scrutiny. But it may delay the serial scrutiny of items needed to eliminate non-targets.

Conversely, during training on serial tasks, observers may develop across the course of testing new templates (or feature detectors) for the very *conjunctions* for which they are searching. In other words, they may acquire new feature detectors as a consequence of their experience. Activity in these feature detectors then seems to allow parallel search when a target is present (i.e. after participants have received a sufficient amount of training with this target). But these new detectors must be different in some way from those that underlie the usual parallel search, found without training, since they do not confer an advantage when a target is absent.

Ellison, A., & Walsh, V., 1998, 'Perceptual learning in visual search: Some evidence of specificities', *Vision Research*, 38, 333–45.

field. This idea receives further support from studies into visual search conducted by Ellison and Walsh (1998).

Different types of visual search not only have different behavioural characteristics, but also depend on different brain regions. So some patients with attention deficits (due to damage to the part of the brain where the temporal, parietal and occipital lobes of the brain join) may be able to perform normally on

feature search tasks but are markedly impaired in conjunction search tasks (Arguin et al., 1993). Also, Ashbridge et al. (1997) used transcranial magnetic stimulation (TMS) to study the role of

feature search visual search for a unique feature such as a particular colour or orientation (e.g. a red spot) in an array of distractors defined by different features along the same visual dimension (e.g. green spots)

different brain regions in visual search. In this technique, a strong magnetic field is applied briefly to the surface of a localized region of the skull, temporarily disrupting neural activity in the underlying brain region. These researchers found that stimulation of the right parietal lobe did not affect initially parallel searches, but did affect initially serial searches. Moreover, a related study found that right parietal stimulation did not affect initially serial searches once they had become parallel through training. But when the observers were switched to another task, which they initially had to perform serially, right parietal stimulation could disrupt search again (Walsh et al., 1998).

Walsh et al. (1998) suggest that the right parietal lobe may

template an internally stored representation of an object or event in the outside world, which must be matched with the pattern of stimulation of the sensory systems before identification, recognition or naming of that object or event can occur

be involved in setting up new *templates* in the temporal lobe for processing conjunctions of, say, colour and form. Once the learning is complete, the right parietal lobe no longer plays a role in the task and so stimulating this region no longer impairs performance.

TOP-DOWN MECHANISMS

Perceptual processes that are concerned solely with sensory input are often called 'bottom-up'. But perception also depends on 'top-down' processes, which reflect our personal goals and past experience.

'Bottom-up' processes are governed only by information from the retinal image. 'Top-down' is a vaguer concept, since it is not clear where the 'top' of the visual pathway is or what it does. But 'top-down' certainly involves the voluntary components of perception, such as moving the eyes. For example, when we discussed conjunction search, earlier in the chapter, remember the observer moving their attention around a display to search for a target (such as a tilted red line or a blue Ford car). This kind of deployment of attention to locate a target is generated internally rather than externally, and is therefore considered to be 'top-down' (in contrast to, say, the sudden appearance of an object in peripheral vision, which will capture the observer's attention and gaze automatically).

Additional support for this idea comes from studies showing that selective adaptation phenomena can be affected by changes in attention. As we saw earlier in this chapter, adaptation can occur at relatively early stages of visual processing, perhaps including V1.

A major finding of the anatomical studies we discussed earlier, in the section *Serial versus parallel theories of perception*, is that almost all the connections between the visual areas of cortex (e.g. figure 8.12) are reciprocal. In other words, information passes not only serially up the system but also backwards, from 'higher' regions, down towards (but not reaching) the sense organs. For

example, just as area V1 projects to V2, so area V2 also sends messages to V1.

How might these reverse connections mediate the perceptual functions that involve top-down influences? The idea that attention to different aspects of the world is mediated by top-down connections is supported by several recent brain scanning studies indicating that relevant regions of the visual cortices alter their activity levels when the person is attending (Kastner & Ungerleider, 2000; Martínez et al., 2001). The idea is that 'higher' parts of the brain decide what to concentrate on, causing messages to be sent back down to prime the relevant parts of the visual cortex. This facilitates cell responses to expected stimuli and improves cell selectivity (tuning), so there are now increased differences in the output of a cell when it is tested with its preferred and some non-preferred stimuli (e.g. Doshier & Lu, 2000; Lee et al., 1999; Olson et al., 2001). It has been further noted that even the LGN can be affected when attention changes (O'Connor et al., 2002).

Another idea is that perceptual learning, recognition and recall depend upon these top-down connections. The hippocampus is important in laying down new long-term memories (see chapter 3). Feedback connections from the hippocampus to the cortex, and within the cortex, appear to be responsible for building these new memories into the fabric of the cortex (Rolls, 1990; Mishkin, 1993; Squire & Zola, 1996). Physiological studies of cells in area V1 of the monkey support theories (Gregory, 1970; Rock, 1983) that memory for objects interacts with the early, bottom-up stages of sensory processing. So the selectivities of the cells in V1 change in the first few hundred milliseconds after a stimulus is presented (Lamme & Roelfsema, 2000; Lee et al., 1998). As activity reaches the 'higher' visual centres, it activates neural feedback, which reaches V1 after a delay. The latency of this feedback is caused by the limited conduction velocity of the messages along the nerve axons (see chapter 3) and by the time taken to process the information in the 'higher' cortical areas.

Recent studies of practice on perceptual tasks indicate that the learning triggered by these feedback projections is so specific for the relevant stimuli that it can only be taking place in the 'early' processing areas of the visual cortex (Ahissar & Hochstein, 2000; Fahle, 1994; Lee et al., 2002; Sowden et al., 2002). Moreover, scans taken of observers' brains when they are recalling or imagining a visual scene show activation of the same early areas of visual cortex that are activated during stimulus presentation itself (Kosslyn et al., 1993; Le Bihan et al., 1993).

As indicated by this discussion, the old division between sensory and cognitive processing by early and higher neural centres has recently been replaced by a new dynamic model. Incoming sensory information interacts with task-relevant knowledge, acquired during the development of the individual concerned, and has been built into the neural network structures in several different cortical areas. Acting together, these influences create an integrated and dynamic representation of the relevant aspects of the environment (e.g. Friston & Price, 2001; Hochstein & Ahissar, 2002; Lamme & Roelfsema, 2000; Schroeder et al., 2001).

Research close-up 2

How early in visual processing does attention operate?

The research issue

In our interactions with the environment, we attend to some things and ignore others. An important question has been at what level of processing this selection operates. For example, it might be that, at one extreme, only a small part of the visual input receives full perceptual processing ('early selection'), or, at the other, that all visual stimuli receive full processing, and selection occurs when a response has to be chosen ('late selection').

This research addressed the question in a novel way, using a visual *after-effect* – the motion after-effect (MAE). To obtain such an after-effect, the observer fixates, for at least 30 seconds, a small spot in the centre of an adapting display of randomly scattered dots drifting, say, to the left. When the physical motion is stopped, the now stationary dots appear to drift slowly to the right (the MAE). The strength of the MAE is often measured by timing how long it lasts – i.e. how long before the stationary dots appear to stop moving.

Physiological studies have shown that motion-sensitive neurons in the first stages of cortical processing (area V1) reduce their activity when repeatedly stimulated. It seems likely, then, that the selective adaptation thought to underlie the MAE occurs early in visual processing. Can it be affected by variations in attention?

Design and procedure

In a groundbreaking study, Chaudhuri (1990) showed that diverting attention from the moving pattern during adaptation could reduce the duration of the MAE. To do this, he modified the usual adapting display by using as the fixation point a letter or digit whose identity changed about once a second. During adaptation, the participants' attention could be diverted from the moving dots by getting them to strike a computer key when a numeral rather than a letter appeared.

Chaudhuri measured MAE duration after diverting attention in this way, and compared it with MAE duration after participants' adaptation to the same display but without the requirement for them to report the letter/digit characters. In another condition, participants were required to strike a key when the colour of the moving dots changed (in this condition participants fixated a stationary unchanging point).

Results and implications

In the conditions in which attention had been diverted from the drifting dots during adaptation by the letter/digit discrimination task, the durations of the MAEs on the stationary stimulus fields that followed were significantly shorter. But MAE durations were not affected by the requirement to attend to the colour of the moving dots while fixating a stationary unchanging point.

So, a secondary task during adaptation does not affect the MAE, if this involves attention to the adapting stimulus (as when participants are required to discriminate the colours of the moving dots). But when the secondary task diverts attention away from the moving dots (for example, by requiring participants to discriminate changing letter/digit characters at the fixation point), subsequent MAEs are weaker. So attention can affect a process, namely selective adaptation, that is thought to occur at an early stage of vision. This finding supports the early selection account of attentional processing.

Chaudhuri, A., 1990, 'Modulation of the motion after-effect by selective attention', *Nature*, 344, 60–2.

FINAL THOUGHTS

The original idea about perception was that perceptual systems are organized serially, with perception proceeding in a series of ordered stages. As we have seen, various experimental and observational techniques have been devised to study different stages of perception, including selective adaptation, perceptual grouping and segmentation, and target search. However, there are at least two reasons why the serial model of perception is too limited:

1. Physiological studies of single neurons in animals and clinical studies of humans with brain damage suggest that perception is performed by sub-systems, which, for example, analyse colour and motion separately and in parallel.
2. In addition to the 'bottom-up', stimulus driven, processing suggested by the serial model, we have seen that 'top-down', conceptually driven processes are important in perception. These top-down processes apply knowledge to information conveyed by the senses, based on assumptions about the nature of the world and objects in it. They can select various aspects of the perceptual input for processing, depending on current attentional demand, and can influence the development of early perceptual mechanisms. Our perceptual experience depends on complex interactions between the sensory input and various types of stored knowledge of the world.

Summary

- We have seen in this chapter that our almost effortless ability to perceive the world around us is in fact achieved by a mass of complicated machinery in the brain.
- The efficiency of our perceptual mechanisms is acquired over many years of individual learning experience, which continues through adulthood.
- As perceptual knowledge grows and accumulates, it enables ever more efficient interpretation of the stimuli that impinge upon our sensory receptors.
- There is continual, recurrent interaction between our knowledge base ('top-down', conceptually driven) and the current sensory inflow from the environment ('bottom-up', stimulus driven).
- Repetitive occurrences of the same environmental phenomena can change the very fabric of the perceptual-sensory systems.
- Bottom-up processing occurs both serially and in parallel. Top-down influences are demonstrable both psychophysically and physiologically.
- Psychological investigations of perceptual phenomena and biological studies of the neural hardware interact with each other to provide a deep understanding of perception.

REVISION QUESTIONS

1. What causes illusions, and why? What does this tell us about the mechanisms of 'normal' perception?
2. How is information coded in the visual system?
3. If the visual system breaks the visual scene up into its various features, how does it keep track of actual objects in the real world?
4. If I am looking for something in the world, how does my visual system find it?
5. How are the features belonging to the same object linked together?
6. How does knowing what you expect (or want) to see influence visual processing?
7. How does knowledge of what the world is like aid (or harm) our interpretation of what we are seeing now?
8. What brain mechanisms could underlie top-down influences on perceptual functions?
9. What happens when the brain machinery underpinning different elements of visual perception is damaged?

FURTHER READING

Bruce, V., Green, P.R., & Georgeson, M.A. (2003). *Visual Perception: Physiology, Psychology and Ecology*. 4th edn. Hove: Psychology Press. An up-to-date and research-orientated textbook on vision, linking ecology (i.e. how the environment and lifestyle of a species influence its perceptual processes) with physiology and psychophysics.

Goldstein, E.B. (2001). *Sensation and Perception*. 6th edn. Pacific Grove, CA: Wadsworth. A very clear introductory text, covering all the senses.

Gregory, R., Harris, J., Heard, P., & Rose, D. (eds) (1995). *The Artful Eye*. Oxford: Oxford University Press. Chapters on many of the topics introduced here, with links to the fascinating interface between art and perception.

Sekuler, R., & Blake, R. (2001). *Perception*. 4th edn. New York: McGraw-Hill. Clearly written textbook covering hearing and other senses. The sections on motion and depth perception are especially good.

Wade, N.J., & Swanston, M.T. (2001). *Visual Perception. An Introduction*. 2nd edn. Hove: Psychology Press. A very readable introduction to the phenomena and mechanisms of vision, particularly the historical background to our modern understanding of what vision is and what it does.

Zeki, S. (1993). *A Vision of the Brain*. Oxford: Blackwell.

The development of the parallel processing model in vision, from the point of view of its central proponent, integrating anatomical, philosophical, perceptual, physiological and clinical studies.