This page intentionally left blank
An Introduction to the Visual System
An Introduction to the Visual System

Second edition

Martin J. Tovée
Newcastle University
This book is dedicated to my wife Esther, and to our children Charlotte and James.
1 Introduction

A user's guide? 1
Brain organisation 2
Why is the cerebral cortex a sheet? 4
Cortical origami 6
Does connectivity predict intelligence? 7
Analysis techniques: mapping the brain 8
Structural imaging 8
Functional imaging techniques: PET and fMRI 10
What is the relationship between blood flow and neural activity? 12
The resolution problem 13
Measuring brain activity in real time: MEG and EEG 14
Transcranial magnetic stimulation (TMS) 15
Summary of key points 16

2 The eye and forming the image

What is the eye for? 18
Light 18
The structure of the eye 19
Focusing the image 25
The development of myopia 26
Clouding of the lens (cataracts) 28
Photoreceptors 28
Transduction 30
The calcium feedback mechanism 31
Signal efficiency 32
The centre-surround organisation of the retina 33
Light adaptation 36
Duplicity theory of vision 37
Sensitivity, acuity and neural wiring 40
Summary of key points 41

3 Retinal colour vision

Why do we need more than one cone pigment? 44
Trichromacy 44
The genetics of visual pigments 47
The blue cone pigment 53
Rhodopsin and retinitis pigmentosa 54
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>11</strong></td>
<td>Brain and space</td>
<td>164</td>
</tr>
<tr>
<td>The final frontier</td>
<td>164</td>
<td></td>
</tr>
<tr>
<td>Oculomotor cues</td>
<td>164</td>
<td></td>
</tr>
<tr>
<td>Interposition</td>
<td>165</td>
<td></td>
</tr>
<tr>
<td>Relative size</td>
<td>166</td>
<td></td>
</tr>
<tr>
<td>Perspective</td>
<td>166</td>
<td></td>
</tr>
<tr>
<td>Motion parallax</td>
<td>168</td>
<td></td>
</tr>
<tr>
<td>Stereopsis</td>
<td>168</td>
<td></td>
</tr>
<tr>
<td>The neural basis of three-dimensional space representation</td>
<td>169</td>
<td></td>
</tr>
<tr>
<td>The problem of visual neglect</td>
<td>170</td>
<td></td>
</tr>
<tr>
<td>The neural basis of neglect</td>
<td>172</td>
<td></td>
</tr>
<tr>
<td>Summary of key points</td>
<td>174</td>
<td></td>
</tr>
<tr>
<td><strong>12</strong></td>
<td>What is perception?</td>
<td>175</td>
</tr>
<tr>
<td>Putting it all together</td>
<td>175</td>
<td></td>
</tr>
<tr>
<td>Neuronal oscillations</td>
<td>175</td>
<td></td>
</tr>
<tr>
<td>How else to solve the problem</td>
<td>178</td>
<td></td>
</tr>
<tr>
<td>What is perception?</td>
<td>180</td>
<td></td>
</tr>
<tr>
<td>Change blindness</td>
<td>180</td>
<td></td>
</tr>
<tr>
<td>Perceptual rivalry</td>
<td>182</td>
<td></td>
</tr>
<tr>
<td>The illusion of perception</td>
<td>185</td>
<td></td>
</tr>
<tr>
<td>Summary of key points</td>
<td>185</td>
<td></td>
</tr>
</tbody>
</table>

**References**: 187

**Index**: 210

The colour plates are to be found between p. 88 and p. 89.
Introduction

A user’s guide?

The aim of this book is to provide a concise, but detailed account of how your visual system is organised and functions to produce visual perception. There have been a host of new advances in our understanding of how our visual system is organised. These new discoveries stretch from the structural basis of the visual pigments that capture light to the neural basis of higher visual function.

In the past few years, the application of the techniques of molecular genetics have allowed us to determine the genetic and structural basis of the molecules that make up the photopigments, and the faults that can arise and produce visual deficits such as colour blindness, night blindness and retinitis pigmentosa. Careful analysis has also allowed the changes in cell chemistry that convert the absorption of light by the photopigment into a neural signal to be understood. The use of functional imaging techniques, in concert with more traditional techniques such as micro-electrode recording, have made it possible to understand how visual information is processed in the brain. This processing seems to be both parallel and hierarchical. Visual information is split into its different component parts such as colour, motion, orientation, texture, shape and depth, and these are analysed in parallel in separate areas, each specialised for this particular visual feature. The processed information is then reassembled into a single coherent perception of our visual world in subsequent, higher visual areas. Recent advances have allowed us to identify which areas are performing these functions and how they interact with one another.

Many of the new advances have come from new experimental techniques such as magnetoencephalography (MEG) and functional magnetic resonance imaging (fMRI), which allow direct, non-invasive measurement of how the human visual system functions. In this introductory chapter, I will firstly discuss the gross structure of the brain and then some of the new methods used to determine the function of different brain areas. To understand vision, we must
understand its neural basis, and how this shapes and limits our perception.

Brain organisation

The mammalian cortex is a strip of neurons, usually divided into six layers. It varies in thickness from about 1.5 to 4.5 mm in humans, and this is not very different even for the very small cerebral hemispheres of the rat, where the thickness is about 1–2 mm. The most conspicuous difference is that the surface area increases enormously in higher animals. For example, the surface area ranges from 3–5 cm² per hemisphere in small-brained rodents to 1100 cm² in humans. To accommodate this increase in surface area within the confines of the skull, the cortex is thrown into a series of ridges (gyri), and furrows (sulci) (see Figure 1.1). In humans, about two-thirds of the cortex is buried in the sulci. The cortex is divided into four main lobes: the occipital lobe, the temporal lobe, the parietal lobe and the frontal lobe. These lobes are then subdivided into different functional areas.

Looking at the brain in detail, we find that it has an incredibly complex structure. It contains around $10^{11}$ neurons, which have more than $10^{15}$ synapses and at least 2000 miles of axonal connections (Young & Scannell, 1993). Fortunately, for those of us who wish to make sense of how the brain works, there are several rules of organisation that simplify our task. Firstly, neurons with similar patterns of connections and response properties are clustered together to form areas. For example, in the monkey and the cat there are about 70 cortical areas, linked by around 1000 connections. Connections between these brain areas may consist of tens of thousands or even millions of nerve fibres. Many of these areas seem specialised to perform different tasks, so, for example, visual area 5 (V5) seems specialised to process information on visual motion and visual area 4 (V4) seems specialised for colour. The number of

**Figure 1.1.** Superolateral view of the left hemisphere of the human cerebral cortex, showing the names of the major gyri and sulci (redrawn from Bindman and Lippold, 1981).
different specialised areas increases with increasing size and complexity of the brain. For example, mice have 15 cortical areas, of which around 5 are visual areas, whereas the cat has 65 cortical areas, of which 22 are visual (Kaas, 1989; Scannell, Blakemore & Young, 1995). It is suggested that the increase in visual areas allows the analysis of an increased number of visual parameters, which in turn allows a more complex and detailed analysis of visual stimuli. There is considerable interaction between neurons dealing with a particular visual parameter, such as colour or motion and, by grouping all such neurons into specialised areas, the amount and the length of connections between neurons are reduced. The arrangement and connections between neurons is largely genetically pre-determined. To have widely interconnected neurons, and to have many different types of neurons with different connections patterns spread throughout the brain, would be extremely difficult to program genetically and would have a greater potential for errors (Kaas, 1989).

Secondly, many of these different areas themselves are subdivided into smaller processing units. For example, in the primary visual area (V1), the cells are organised into columns, within which all the cells have similar response properties. This form of columnar organisation seems to be a common feature within the visual system. Thirdly, a further feature of organisation of the visual system, is lateralisation. On either side of the brain, there is a duplication of visual areas. So there are two V1 areas and two V5 areas, and so on. However, the higher visual areas, such as the inferior temporal cortex in monkeys and the inferior temporal and fusiform gyri in humans, do slightly different tasks on different sides of the brain. So, for example, the recognition of faces is mediated by the right side of the brain. This process of lateralisation allows the brain to carry out a greater variety of tasks with a limited amount of brain tissue.

Humans and Old World primates seem to have a visual system based on a broadly similar organisation. Differences seem to arise between the human and Old world monkey visual systems largely because of the expansion of the cortex in humans, which displaces the higher areas relative to their position in Old World primates. For this reason, during the course of this book I will refer to visual areas by the names originally coined for areas in monkey cortex, but which are now being applied to human visual areas (see Figure 1.2) (Kaas, 1992; Tootell et al., 1995). A problem with coming to grips with the visual system is that different research groups have used different names for the same area. For example, visual area 1 (V1), is also called the primary visual cortex and the striate cortex, and the higher visual areas can be collectively referred to as either the prestriate cortex or the extrastriate cortex. When I come to describe each area, I will use its most common name, but I will also list the other names by which you might encounter the area in other accounts of visual function.
Why is the cerebral cortex a sheet?

It seems that the evolutionary expansion of the cortex may be constrained by the way the cortex is formed during development. Pasko Rakic has put forward a persuasive theory based on the limitations that cell division during development place on the expansion of the cortex (Rakic, 1988, 1995). This model, called the radial-unit hypothesis, proposes that the 1000-fold increase in the expansion of cortical surface area seen in mammalian evolution is the result of changes in cell division that increases the number of cell columns which make up the cortex, without changing the number of cells in each column. Thus the sheet-like structure of the cortex is determined by the constraints of cell division during development. The cortical sheet is folded to produce a series of ridges (gyri), and furrows (sulci). The
simplest explanation for this folding is that you have to fold a large sheet to get it into a small box. But mere folding explains neither the species-specific pattern of sulci and gyri, nor why they provide landmarks to the location of functional areas of cortex, nor why this pattern of folding is altered by lesions of the cortex that cause the brain to ‘re-wire’ (Rakic, 1988). So, what factors control the pattern of folding?

One likely explanation for the placement of cortical folds is to reduce the length of axonal connections (Griffin, 1994; Scannell, 1997; Van Essen, 1997). It is commonly accepted that some, but by no means all, aspects of the organisation of the central nervous system appear to minimise wiring volume (Cowey, 1979; Mitchison, 1991; Cherniak, 1991). Quite simply, an animal that arranges its neurons efficiently, by putting the computationally necessary connections between nearby neurons and leaving ‘non-connections’ between neurons that are far apart, can make do with less white matter and will benefit from a smaller, faster and cheaper brain. Such a brain should also be easier to make with simple developmental and evolutionary processes.

Efficient wiring may be seen in neuronal arbours, cortical maps and in the two-dimensional arrangement of cortical areas (Cowey, 1979; Mitchison, 1991; Cherniak, 1991, 1995; Young, 1992; Scannell et al., 1995). There is also some evidence that the principle applies to the 3-D morphology of cortical folds. Both the cat and macaque appear to fold their cortices in such a way that devotes the available convexities to heavily connected areas and puts the concavities between sparsely connected areas (Scannell, 1997; Van Essen, 1997).

While the importance of efficient wiring is widely accepted, the processes that generate it and its overall importance in explaining major aspects of brain structure have been hotly debated (Cherniak, 1996, Young & Scannell, 1996). Efficient wiring could be produced either by neurons and areas starting in particular locations and then sending projections to neurons in their locality (local wiring) or by neurons and areas starting out with particular connections and then ‘migrating’ to get close to the things with which they connect (component placement optimization, CPO). The fact that wiring is efficient does not distinguish between these possibilities.

Until recently, developmental and evolutionary considerations suggested that local wiring rather than CPO could best account for the observed regularities between connectivity and location. Indeed, the evidence that structures migrated around the brain to minimise wire is questionable (Young & Scannell, 1996). However, when it comes to the 3-D arrangement of cortical areas in relation to sulci and gyri, it does now look as if major brain structures may be positioned in such a way that reduces connection length.
Cortical origami

The cortical sheet is a jigsaw of functionally distinct areas linked by a complex network of cortico-cortical connections. How is the folding coordinated with the wiring? Van Essen has suggested two factors play a key role. The first are intrinsic factors, such as differential growth rates in the grey matter, and second are extrinsic factors, which are based on long-range axonal connections in the underlying white matter. Some of the axonal connections are to subcortical structures and Van Essen hypothesises that the tension generated in these axons produces an inward force to counteract the intraventricular hydrostatic force generated by the CSF. The second type of axonal connections is between different cortical areas. These connections are established at around the time that folding begins, and could generate tension that would lead to folding.

The cortex can fold either outwards or inwards. In an outwards fold, the ridge is directed away from the white matter and the brain interior, and the length of axonal connections between the two banks of the fold is small. Such folds could bring together densely interconnected areas. In an inwards fold, the crease is directed towards the white matter and so the white matter distance between the two banks of the fold is long. Therefore, inwards folds should end up between sparsely connected areas. This suggestion is consistent with results published on connectivity and cerebral folding in the macaque and cat brain (Scannell, 1997). Heavily interconnected areas tend to be separated by gyri and sparsely connected areas seem to be separated by sulci (Figure 1.3).

Thus one has to make a trade-off, reducing the tension in the axonal connections between some cortical areas at the price of increasing the tension in the connections between other areas. The connections between some areas are more extensive than those between other areas, so if one makes an outwards fold at the boundary between two areas that are densely connected and an inwards fold at the boundary between two sparsely connected areas, the overall axonal tension will be reduced. Thus, the eventual shape of the cortical sheet will be determined by the relative density of connections between different areas.

Other aspects of the gross morphology of the brain may follow from the same mechanisms. The link between wiring and folding is supported by evidence from developmental studies. For example, prenatal bilateral eye removal in the macaque alters the pattern of folding in the occipital cortex in the absence of direct mechanical intervention (Rakic, 1988). Thus, even if tension-based factors do not turn out to be the explanation, developmental neuroscientists still need to account for the relationship between wiring and folding, possibly turning their attention to the possibility that growth factors are released by cortico-cortical axons.
While efficient wiring is an attractive principle, it should not blind us to the fact that brains represent a compromise between many competing constraints. As well as saving wire, brains have to produce adaptive behaviour; they have to be made during development, specified by a genome, and based on a design inherited from the parents. It is unlikely that in balancing all these constraints, the brain could be optimal for any one. Indeed, apparent examples of wire-wasting connectivity are widespread; the facts of developmental pruning, the inverted retina, the visual cortex at the wrong end of the brain, and the unconnected thalamic nuclei clustering together and not with the groups of cortical areas with which they exchange axons, all suggest other factors are at work (Scannell, 1997; Young & Scannell, 1996).

Does connectivity predict intelligence?

The way the brain is wired up may play a role in intelligence and conceptual thought in humans, although this remains a controversial area. There seems to be a degree of variation between individuals in the organisation and connectivity of the brain, and this may play a role in some aspects of intelligence and cognition (Witelson et al., 1999).

Albert Einstein died in 1955 at the age of 76. Within 7 hours of his death, his brain was removed and preserved for further study. The gross anatomy of the brain seemed to be normal, but there was something unique in the anatomy of the Sylvian fissure that divides the temporal lobe from the parietal lobe (Witelson, Kigar & Harvey, 1999). The Sylvian fissure is present in the cortex when a child is born, and it has a definite place and pattern. But in Einstein's brain, the Sylvian fissure runs into another major fold in the brain, the so-called post-central sulcus. In fact, it's hard to know where one fold ends and the other begins. That makes a brain region known as the inferior parietal lobule larger. Van Essen hypothesised that a gyrus develops within a region functionally related to cortex to allow for efficient axonal connectivity, between opposite walls of the cortical

![Fig. 1.3 (A). The human brain.](image)

In this and many other mammalian brains, a distinct pattern of folds is the most striking anatomical feature. The pattern is characteristic of species and is related to the mosaic of distinct functional areas that make up the cortex. (B). How folds may influence the length of cortico-cortical connections. In this model, five functional areas (areas 1 to 5) are distributed over 2 gyri. 1 and 2, and 3 and 4, are 'nearest neighbours' (NN), while 1 and 3, and 3 and 5 are 'next door but one' on the cortical sheet. Area 1 is 'nearest neighbour OR next door but one' (ND1) with 2 and 3.

Axons linking areas 1, 2 and 3 would be short, while axons linking 3 and 4 would be long. Thus, given the same axonal diameter, spike rate and axon number, a cortico-cortical connection between 1 and 3 would be more compact, faster and use less energy than a connection between 3 and 4. An efficiently folded cortex might place the folds so that heavily connected areas are together on gyri while sparsely connected areas are separated by sulci (reproduced by courtesy of Dr Jack Scannell).
gyrus; by contrast, sulci separate cortical regions having less functional relatedness. In this context, the compactness of the inferior parietal lobule may reflect an extraordinarily large expanse of highly integrated cortex. The larger region is in the part of the brain that is believed to be important for visual imagery, three-dimensional perception and mathematical intuition (which may be crucial for the thought experiments that led to the formulation of the theory of relativity).

**Analysis techniques: mapping the brain**

Traditional methods of divining the function of brain areas have relied on two lines of approach; the study of human patients who have suffered brain damage or the use of animal models of human brain function. Common causes of head injuries to human patients are strokes, traumatic head injuries such as those suffered in car accidents and carbon monoxide poisoning. The difficulty with this approach is that the damage tends to be widespread, affecting more than one type of visual process. For example, damage that causes visual agnosia (the inability to recognise objects) is often linked to achromatopsia (an impairment of colour perception). The alternative line of investigation has been to use an animal model of human visual function. The advantage of this approach is that artificially induced lesions can be used to remove selectively specific brain areas, to determine their function. Also, the activity of single neurons can be determined using a technique called microelectrode or single-unit recording. In this technique, a glass-insulated, tungsten-wire microelectrode is inserted into an animal’s brain and its position adjusted until it is adjacent to a neuron in a particular brain area. The microelectrode can detect the small electrical changes associated with an action potential, and so the activity of single neurons in response to different visual stimuli can be determined.

Recently, new non-invasive analysis techniques have been developed to examine brain function and these fall into two categories: structural imaging and functional imaging.

**Structural imaging**

*Computerised tomography (CT), or computer assisted tomography (CAT)*, uses X-rays for a non-invasive analysis of the brain. The patient’s head is placed in a large doughnut-shaped ring. The ring contains an X-ray tube and, directly opposite to it on the other side of the patient’s head, an X-ray detector. The X-ray beam passes through the patient’s head, and the radioactivity that is able to pass through it is measured by the detector. The X-ray emitter and detector scan the head from front to back. They are then moved around the ring by a few degrees, and the transmission of radioactivity is measured again.
The process is repeated until the brain has been scanned from all angles. The computer takes the information and plots a two-dimensional picture of a horizontal section of the brain (see Figure 1.4). The patient’s head is then moved up or down through the ring, and the scan is taken of another section of the brain.

A more detailed picture is available from magnetic resonance imaging (MRI). It resembles the CT scanner, but instead of using X-rays it passes an extremely strong magnetic field through the patient’s head. When a person’s head is placed in a strong magnetic field, the nuclei of some molecules in the body spin with a certain orientation. If a radio-frequency wave is then passed through the body, these nuclei emit radio waves of their own. Different molecules emit energy at different frequencies. The MRI scanner is tuned to detect the radiation from hydrogen molecules. Because these molecules are present in different concentrations in different brain tissues, the scanner can use the information to prepare pictures of slices of the brain (see Figure 1.5). Unlike CT scans, which are limited to the horizontal plane, MRI scans can be taken in the sagittal or frontal planes as well.

A new approach to looking at brain structure is a variant of MRI, called water diffusion MRI or dMRI. This specifically allows the wiring of the brain to be explored. It exploits a basic characteristic of biological tissue, which is that water molecules move through and within it, by a process called diffusion. Some materials have the interesting
property that diffusion happens faster in some directions than in others. The name for this phenomenon is anisotropy. The wider the variation in diffusion rate as a function of direction, the more anisotropic a material is. The brain is an interesting system to study because it has a variety of anisotropies. At the surface of the brain, there’s the grey matter (composed primarily of neuronal cell bodies), which is isotropic (i.e. diffusion is at the same rate in all directions). Deeper inside the brain, there’s the white matter (the neuronal axons), which is anisotropic. More specifically, water diffuses more rapidly along an axon than it does across it. So, if one were able to track the movement and speed of water diffusion, it would be possible to infer the position and connections of an axon in the cortex. This is exactly what dMRI does, by tracking the position of hydrogen atoms in water molecules (Le Bihan, 2003). Instead of passing a single radio frequency pulse through the brain, as in standard MRI, two pulses are used, one slightly after the second. From the relative change in position of the water molecules, the rate of diffusion can be determined and the neural connections of the cortex can be inferred.

Functional imaging techniques: PET and fMRI

The above two techniques provide a representation of brain structure, but do not provide any information on how the different parts of the brain function. A method that measures brain function, rather than brain structure, is positron emission tomography (PET). PET measurements depend on the assumption that active areas of the brain
have a higher blood flow than inactive areas. This is because these more active areas use more oxygen and metabolites and produce more waste products. So, an increased blood flow is necessary to supply the former and remove the latter. A PET camera consists of a doughnut-shaped set of radiation detectors that circles the subject’s head. After the subject is positioned within the machine, the experimenter injects a small amount of water labelled with the positron-emitting radioactive isotope Oxygen-15 ($^{15}$O) into a vein in the subject’s arm. During the minute following the injection, the radioactive water accumulates in the brain in direct proportion to the local blood flow. The greater the blood flow, the greater the radiation counts recorded by PET. The measurement of blood flow with $^{15}$O takes about 1 minute. $^{15}$O has a half-life of only 2 minutes, which is important as one does not want to inject long-lasting radioactive material into someone.

Different human brains vary slightly in their relative sizes and shape and, as PET scans do not provide any structural information, they are usually combined with MRI scans to allow the accurate comparison of structural and functional information (e.g. Zeki et al., 1991). Although PET scanning is able to determine roughly which areas are active, its ability accurately to resolve specific regions is limited. A new technique that is now coming into use is functional MRI (fMRI) and this has better resolution. This method is a refinement of the MRI technique and, like PET scanning, it measures regional blood flow (Tank, Ogawa & Urgubil, 1992). Deoxyhaemoglobin (haemoglobin without a bound oxygen molecule) is paramagnetic, and so a blood vessel containing deoxyhaemoglobin placed in a magnetic field alters the magnetic field in its locality, the blood oxygen-level-dependent (BOLD) effect. It is thus possible to map blood flow based on these changes in local magnetic fields.

In recent years, fMRI has largely eclipsed PET, a technique that is now over 30 years old. PET, which uses radioactive tracers to measure blood flow to activated brain regions, is comparatively slow, taking up to a minute to gather enough data for a brain image. As a result, it is necessary to run ‘block trials,’ in which the subject performs a string of similar brief tasks, causing the brain to repeat the same mental process while the data are gathered (Barinaga, 1997). However, a fMRI system can take snapshots of the brain, which take as little as 2 seconds, and so allows the neural response to an individual trial to be imaged (‘an event-related’ recording). fMRI also has much better spatial resolution than PET. A PET scanner can distinguish activated brain areas separated by a centimetre or more. However, fMRI scanners can resolve distances in the order of millimetres. This allows us not only to look at which cortical areas are active during a particular task, but also to look at how different parts of an area function during the task.

It is assumed that a PET or fMRI signal increases in proportion to the amount of blood flow, which is in turn assumed to be proportional to the amount of neural activity. However, neural activity can be due to a number of processes and, to clarify this ambiguity, Niko
Logothetis and his colleagues measured two things simultaneously: an fMRI signal and the electrical activity of neurons in the primary visual cortex of monkeys watching rotating chequerboard patterns using microelectrodes inserted into the cortex (Logothetis et al., 2001). They looked at the relationship between the size of the fMRI signal and three types of electrical activity in neurons: the slowly changing electrical fields produced by input signals to neurons and by their signal-processing activity, the rapid output pulses that individual neurons generate in response and the output signals from collections of neurons. They found that the fMRI signal was most strongly related to the input and local processing of information by neurons, rather than to the output of information by neurons in an area.

These functional imaging techniques have allowed us to match behaviour to the anatomy and function of the brain. So, for example, when we perceive colour, we can now say which brain areas seem to be processing this information to give the sensation of colour. We can also see how different brain areas interact to produce the complex synthesis of different visual sensations that is our everyday experience of the visual world.

What is the relationship between blood flow and neural activity?

At rest, the brain uses about 20% of the oxygen used by the body, although the brain accounts for less than 2% of the body’s mass. The oxygen is used in breaking down glucose to supply the brain with energy. However, when we carry out a visual stimulus presentation in a PET or fMRI experiment, the brief increase in the activity of a brain region (and thus its energy use) is accompanied by increases in blood flow and glucose consumption that far exceed the increase in oxygen consumption (Fox et al., 1989). This is because glucose is being broken down anaerobically in a process called glycolysis to supply energy rapidly to the active neurons. Thus the increase in local blood flow is due to a need for energy in the form of glucose rather than to a need for oxygen. As a result, the supply of oxygen exceeds demand and there is an increase in the amount of oxygen around the active neurons. fMRI is sensitive to changes in the oxygen content of the blood (the BOLD signal), and so it can detect changes in neural activity indirectly (Figure 1.6).

The communication between neurons occurs at synapses and requires the release of a neurotransmitter substance, such as glutamate or acetylcholine, from a presynaptic neuron and their detection by a postsynaptic neuron. To give a crisp, sharp signal it is important that, after a neurotransmitter is released at the synapse, it is removed promptly and recycled, and does not remain active in the synapse. Glutamate, the primary excitatory neurotransmitter in the brain, is taken up by adjacent non-neural cells called astrocytes, where it is converted to glutamine before being returned to the presynaptic
neuron and recycled. The energy needed for the active uptake of glutamate from the synaptic cleft and its processing by the astrocytes is provided by glycolysis. Hence the need for an increased supply of glucose during neural activity, with an absence of a corresponding need for oxygen.

Thus the blood oxygen level rises because of an increase in the processing of glutamate in astrocytes after excitatory neurotransmission. So, the changes in blood flow and oxygen levels measured by functional imaging techniques are linked to neural activity, but this link is indirect, and via astrocyte activity. This finding is also consistent with the experiment by Logothetis (discussed earlier in this chapter), which suggests that the fMRI signal is most strongly related to the input and local processing information by neurons in an area (which requires the release of neurotransmitters by axons synapsing on to neurons in that particular area) rather than the output of information that is mediated by action potentials travelling along these neurons’ axons to other areas of the cortex (and so will not produce a release of neurotransmitters in the original area).

The resolution problem

fMRI has a much better spatial and temporal resolution than PET. However, even an fMRI system has very poor temporal and spatial resolution compared with how fast the brain works and the size of its components. Consider that, as neural activity occurs on a millisecond time scale, a temporal resolution of seconds is still very slow. Moreover, neurons in a cortical area are organised into columns 200–1000 μm in diameter, but standard fMRI has a spatial resolution of only a few millimetres. So, with fMRI you can see that a localised area of cortex is active but you may not be able to tell very much
about what is going on within this area. How can we overcome this problem and improve the spatial resolution of this technique?

Consider what happens when there is cortical activity in a brain area triggered by a visual stimulus. There are two main changes in the relative deoxyhaemoglobin levels. Firstly, there is an increase in oxygen consumption, caused by an increase in the oxidative metabolism of the stimulated neurons, which leads to an increase in the levels of deoxyhaemoglobin. This happens within 100 ms of the neural activity starting (Malonek & Grinvald, 1996; Vanzetta & Grinvald, 1999), and seems to be localised around the neural activity in a cell column (Grinvald, Slovin & Vanzetta, 2000). The higher deoxyhaemoglobin levels and the increased production of metabolites lead to an increased blood flow to the active region. This produces the second change in relative deoxyhaemoglobin levels, which occurs about 0.5 to 1.5 seconds after the onset of electrical activity for reasons discussed above. This second change in deoxyhaemoglobin levels is far larger than the first, because the increased blood flow overcompensates for the reduction in oxyhaemoglobin levels. Most fMRI systems use a magnetic field of around 0.5 to 1.5 teslas, and detect the second larger change in deoxyhaemoglobin levels, which, as we have seen, is less spatially localised in the cortex and is not closely linked in time with the underlying neural activity. However, the introduction of new high-field fMRI systems (4.7 to 9.4 teslas) means that it is starting to become possible to detect the first change in deoxyhaemoglobin levels, which allows much better temporal and spatial resolution and it is possible to directly measure the activity in single cell columns (e.g. Kim, Duoung & Kim, 2000).

Measuring brain activity in real time: MEG and EEG

Measuring blood flow allows us to track changes in brain activity only indirectly. The blood flow changes occur over a period of seconds, whereas the function of neurons is measured in milliseconds. There are two non-invasive ways to try and measure brain activity in real time: Electroencephalography (EEG) and Magnetoencephalography (MEG). EEG uses electrodes placed on the scalp to measure changes in gross electrical activity across the underlying brain. The electrical signal is sampled simultaneously at multiple locations, usually using 32 or 64 electrodes. One problem with this technique is that, as the electrical activity is being measured through the skull and scalp by electrodes on the surface of the head, the signal may be distorted by its passage through the intervening tissue. MEG does not suffer this problem, as it measures the tiny changes in the magnetic field that accompany electrical currents in the brain.

The passage of any electrical current (such as neuronal activity) induces a magnetic field, which changes as the underlying electrical activity changes. MEG measures the magnetic field induced by neural electrical activity, and so allows underlying neural function to be
deduced. The MEG signals are detected by superconducting quantum interference device (SQUID) sensors. As in EEG, the signal is sampled simultaneously at multiple locations. This is accomplished by fitting over 300 SQUID sensors into a helmet-like area that covers the head (Hari, Levanen & Raji, 2000).

The MEG signals that are produced by neural activity are very small, usually in the femto tesla \(10^{-15}\) tesla range. To put this figure into perspective, the magnetic field of the Earth is in the order of \(0.5 \times 10^{-4}\) tesla. So, to avoid contamination of the MEG signal by noise originating from outside sources (such as electrical instruments and power lines), the recordings are carried out typically in a magnetically shielded room.

Although EEG and MEG are able to track changes in neural activity in real time (i.e. with high temporal resolution), they have comparatively poor spatial resolution. As a result, to get the best results, it is becoming common to pair MEG with fMRI, thus allowing good temporal and spatial resolution.

**Transcranial magnetic stimulation (TMS)**

Functional imaging can only establish the association between task performance and a pattern of cortical activity. However, by using transcranial magnetic stimulation (TMS) to inactivate a particular cortical area transiently, it is possible to test the causal link between activity in a region and a particular behaviour.

TMS is based on the principle of electromagnetic induction. A burst of electric current flowing through a coil of wire generates a magnetic field. If the amplitude of this magnetic field changes over time, it will induce a secondary current in any nearby conductor (Pascual-Leone et al., 1999). The size of the induced current will be dependent on how fast the magnetic field changes its size. In TMS, a magnetic coil is held over the subject’s head and, as a brief pulse of current is passed through it, a magnetic field is generated that passes through the subject’s scalp and skull. This time-varying magnetic field induces a current in the subject’s brain, and this stimulates the neuronal tissue. The neurons within the stimulated area fire off in a burst of activity, followed by a period of inactivity while the neurons recover. During the recovery period, it is assumed that the sensory or cognitive functions normally performed by this area will be temporarily deactivated, allowing the experimenter to deduce the function normally undertaken by the area. In many experiments, single pulses of stimulation are used. In others, a series of pulses at rates of up to 50 Hz (this called repetitive TMS or rTMS). This latter procedure can be dangerous and can cause seizures, and so must be used with caution. With this caveat, TMS is an extremely useful technique. It allows investigators to test reversibly whether a particular area undertakes the functions ascribed to it, and has been widely used in exploring visual perception in human observers.
Summary of key points

(1) The human cortex is a strip of neurons, usually divided into six layers, which vary in thickness from about 1.5 to 4.5 mm. To fit it all into the confines of the skull, the cortex is thrown into a series of ridges (gyri), and furrows (sulci).

(2) Why is the cortex a flat sheet? The radial-unit hypothesis suggests that increasing cortical size is based on increases in the number of radial columnar units that make up the cortex, rather than changing the number of cells in each column.

(3) The cortical sheet folds in specific places. The folding is designed to minimise total axon length. In an outwards fold the ridge is directed away from the white matter and the brain interior, and the length of axonal connections between the two banks of the fold is small. Such folds could bring together heavily connected areas. In an inwards fold, the crease is directed towards the white matter and so the white matter distance between the two folds is long. Inwards folds thus should tend to fall between sparsely connected areas.

(4) The cortex is divided into four main lobes: the occipital lobe, the temporal lobe, the parietal lobe and the frontal lobe. These lobes are then subdivided into different functional areas.

(5) There are several rules of organisation for brain organisation. Firstly, neurons with similar patterns of connections and response properties are clustered together to form areas. Secondly, these different areas themselves are subdivided into smaller processing units. Thirdly, corresponding higher visual areas on different sides of the brain do slightly different jobs, a process called lateralisation.

(6) Computerised tomography (CT), or computer assisted tomography (CAT), uses X-rays for a non-invasive analysis of the brain. The X-ray emitter and detector scan the head from front to back, and the process is repeated until the brain has been scanned from all angles. The computer takes the information and plots a two-dimensional picture of a horizontal section of the brain. The patient’s head is then moved up or down, and the scan is taken of another section of the brain.

(7) Magnetic resonance imaging (MRI) passes an extremely strong magnetic field through the patient’s head; this causes the nuclei of some molecules to emit radio waves. Different molecules emit energy at different frequencies. The MRI scanner is tuned to detect the radiation from hydrogen molecules. Because these molecules are present in different concentrations in different brain tissues, the scanner can use the information to prepare pictures of slices of the brain.

(8) Positron emission tomography (PET) measures the flow of radioactively labelled blood to different areas of the brain. It is assumed that an increased flow to a brain area is an indication of increased...
function. A more accurate measure of blood flow is functional magnetic resonance imaging (fMRI), which is a refinement of the MRI technique. This technique takes advantage of the differences in the magnetic profile of oxygenated and deoxygenated haemoglobin to map bloodflow and the metabolic use of oxygen in the brain.

(9) Electroencephalography (EEG) measures the electrical activity of the brain through electrodes placed on the scalp. Magnetoencephalography (MEG) measures the electrical activity of the brain indirectly by measuring the changes in magnetic field induced by the fluctuations in the brain’s electrical activity. These are both functional imaging techniques and, although they have good temporal resolution, they have relatively poor spatial resolution.

(10) In Transcranial magnetic stimulation (TMS) a magnetic coil is held over a volunteer’s head and a brief pulse of current is passed through it; a magnetic field is generated that passes through the subject’s scalp and skull. This time-varying magnetic field induces a current in the subject’s brain, and this stimulates the neuronal tissue and leads to its temporary deactivation.
The eye and forming the image

What is the eye for?

In this chapter we will review the purpose of the eye and how the complex optical and neural machinery within it functions to perform this task. The basic function of the eye is to catch and focus light on to a thin layer of specially adapted sensory receptor cells that line the back of the eye. The eyeball is connected to an elaborate arrangement of muscles that allow it to move to follow target stimuli in the environment. The lens within the eye, which helps focus light, is also connected to muscles that can alter the lens shape and thus its focal length. This allows target stimuli at different distances to be focused on the back of the eye. At the back of the eye, light energy is transformed into a neural signal by specialised receptor cells. This signal is modified in the retina, to emphasise changes and discontinuities in illumination, before the signal travels onto the brain via the optic nerve. In the sections that follow we will examine these procedures in detail.

Light

Light has a dual nature, being considered both an electromagnetic wave, which can vary in frequency and wavelength, and also a series of discrete packets of energy, called photons. Both forms of description are used in explaining how the visual system responds to light. In determining the sensitivity of the visual system to light, such as the minimum threshold of light detection, it is usual to refer to light in terms of photons. However, when discussing colour perception, it is normal to refer to light in terms of its wavelength, measured in nanometres (nm). One nanometre is $10^{-9}$ m. For example, blue light is of comparatively short wavelength (around 430–460 nm), whereas red light is of comparatively long wavelength (around 560–580 nm).

Only electromagnetic radiation with a wavelength between 380 and 760 nm is visible to the human eye (Figure 2.1). The width of the
spectrum is determined primarily by the spectral absorbance of the photopigments in the eye. However, other structures play a role. Light just below the human visible spectrum (300–400 nm) is called ultra-violet (UV). The human lens and cornea absorbs strongly in this region, preventing UV light from reaching the retina (e.g. van den Berg & Tan, 1994). However, the human short wavelength (or blue) photopigment’s absorption spectrum extends into the UV range and, if the lens is removed, such as in cataract surgery, a subject can perceive UV light. A good reason for preventing UV light from reaching the retina is that it is absorbed by many organic molecules, including DNA. Thus, UV light, even of comparatively long wavelengths such as 380 nm, can cause retinal damage and cancer (van Norren & Schelkens, 1990). However, a wide variety of animal species show sensitivity to UV light, ranging from insects to mammals (Tovee, 1995a). Some have developed specific UV sensitive photoreceptors to detect UV light, whereas others have combined a clear ocular media with short-wavelength receptors whose spectral absorbance extends into the UV range. These species use UV light for a wide range of purposes: from navigation using the pattern of UV light in the sky to intra-specific communication using complex UV-reflecting patterns on their bodies.

The structure of the eye

The eyes are suspended in the orbits of the skull, and each is moved by six extra-ocular muscles attached to the tough, fibrous outer coating of the eye (the sclera). Within the orbit, the eye is cushioned by heavy deposits of fat surrounding the eyeball. The eyelids, movable folds of tissue, also protect the eye. Rapid closing of the eyelids (blinking) can occur both voluntarily and involuntarily. Blinks clean and moisten the surface of the eye, and under normal circumstances, we automatically blink about once every 4 seconds. It takes about a third of a second from the beginning of a blink, when the lids first begin to move, until they return to their resting point. For about half of this time, the eyelids are completely closed, reducing the amount of light reaching the retina by around 90%. If an external light is flicked on and off for this length of time, a brief blackout is very noticeable. So, why do we not notice our blinks? One suggestion has
been that visual perception is suppressed during a blink. Evidence for this suggestion comes from an ingenious experiment by Volkman and his colleagues. The eyes lie directly above the roof of the mouth, and a bright light shone on the roof will stimulate the retina, whether or not the eyes are closed. Volkman found that the light intensity required to stimulate the retina during a blink is five times greater than at any other time, strongly suggesting that there is suppression of perception during a blink (Volkman et al., 1980) (Figure 2.2).

Functional imaging has suggested that neural activity mirrors these behavioural studies. Activity in a number of cortical visual areas is suppressed during a blink (Bristow et al., 2005). Particularly affected are those areas that are sensitive to rapid change in visual stimuli, such as visual area 3 or V3 (see Chapter 10). We normally think of these areas as being sensitive to motion, but they are also sensitive to rapid global changes in visual input, such as would be produced by a blink (Burr, 2005).

A suppression of visual sensitivity during blinks explains why darkening is not seen, but it is not sufficient to account for the continuity of visual perception. Functional imaging has shown that, in humans, just after a blink, the posterior parietal cortex is active (Harl et al., 1994). The latency of the parietal activity suggests it is a reaction to the eyeblink, and does not occur in advance as might be expected if it was connected with the generation of a motor command involved in the movement of the eyelids. This parietal activity is not seen if the blinks occur in darkness. The posterior parietal cortex is connected reciprocally with prefrontal cortical areas, which seem to underlie spatial working memory, and it has been suggested that the parietal cortex is continually updated on information about
the nature and structure of objects in a person's surroundings (Goodale & Milner, 1992), and is also kept informed about eyeblinks (Harl et al., 1994). It is believed that this activity in the posterior parietal cortex is important for maintaining the illusion of a continuous image of the environment during each blink, perhaps by filling in the blink with visual sensation from working memory.

A mucous membrane, called the conjunctiva, lines the eyelid and folds back to attach to the eye (Figure 2.3). The eye itself is a roughly spherical ball, about 2.5 cm in diameter. The sclera of the eye is made up of closely interwoven fibres, which appear white in colour. However, at the front of the eye, where the surface bulges out to form the cornea, the fibres of the sclera are arranged in a regular fashion. This part of the sclera is transparent and allows the entry of light. The part of the sclera surrounding the cornea is called the white of the eye. Behind the cornea is a ring of muscles called the iris. In the centre of the ring is an opening called the pupil, and the amount of light entering the eye is controlled by the pupil's diameter. The iris contains two bands of muscles, the dilator (whose contraction enlarges the pupil) and the sphincter (whose contraction reduces it). The sphincter is enervated by the parasympathetic nervous system, which uses the neurotransmitter acetylcholine. When we are interested in something, or someone, there is an unconscious expansion of the pupils. This is an important positive social signal. To mimic this response, and make themselves appear more attractive, women once added drops to their eyes containing the alkaloid atropine. This blocks the action of acetylcholine, causing dilation of the pupil by relaxing the sphincter of the iris. This preparation was made from deadly nightshade and gave this plant its species name of Belladonna, meaning beautiful women.

Figure 2.3. (See also colour plate section.) Cross-sectional diagram of the human eye (reproduced by permission from Burr, 2005. Copyright (2005) Elsevier).
The light then passes through the anterior chamber of the eye to the lens. The anterior chamber is filled with a watery fluid called the aqueous humour. This fluid transports oxygen and nutrients to the structures it bathes and carries away their waste products. This function is normally carried out by blood in other parts of the body, but blood would interfere with the passage of light through the eye. The aqueous humour is being produced constantly by spongy tissue around the edge of the cornea (the ciliary bodies) and, if the drainage is blocked or slowed, then pressure builds up in the eye. This can lead to permanent visual damage (glaucoma), and this is one of the commonest causes of blindness in Western Europe and North America.

The cornea and the lens alter the path of the light such that it will be in focus on the surface of the back of the eye, which is covered by the retina. The lens also inverts the image, so the picture of the world on the retinal surface is upside-down. The inversion is not important as long as the relative spatial positions of the different features of the image are preserved. After passing through the lens, the light passes through the main part of the eye, which contains a clear, gelatinous substance (the vitreous humour), before reaching the retina. Unlike the aqueous humour, the vitreous humour is not being replaced constantly and so debris can accumulate. This debris can impinge on your visual awareness by forming floaters, small opacities which float about in the vitreous (White & Levatin, 1962).

The retina is divided into three main layers: the receptor cell layer, the bipolar layer and the ganglion cell layer (Figure 2.4). The receptor cell layer is at the back of the retina and light has to pass through the transparent, overlying layers to get to it. The photoreceptors

---

**Figure 2.4.** The neural structure of the retina. Light passes through the neural layers before striking the receptors (rods and cones) that contain the photosensitive pigments. The vertical organisation of the retina is from receptor to bipolar cell to retinal ganglion cell. The horizontal organisation is from receptor to bipolar cell to retinal ganglion cell. The horizontal organisation is mediated by horizontal cells at the receptor-bipolar (outer) synaptic layer and by the amacrine cells at the bipolar-retinal ganglion cell (inner) synaptic layer (redrawn from Cornsweet, 1970).
form synapses with bipolar cells, which in turn synapse onto ganglion cells, the axons of which travel through the optic nerve to the brain. These axons come together and pass through the bipolar and receptor cell layer and leave the eye at a point called the optic disc. The optic disc produces a blind spot, as no photoreceptors can be located there. We are not consciously aware of the blindspot, due to a phenomenon called ‘filling in’ (Ramachandran, 1992). Based on the visual stimulus surrounding the blindspot, the visual system fills in the ‘hole’ in the visual image to give the complete picture of the world we are used to. This process seems to be mediated by cortical neurons in areas V2 and V3 (De Weerd et al., 1995). The retina also includes the outer plexiform layer, containing horizontal cells, and the inner plexiform layer, containing amacrine cells. These cells transmit information in a direction parallel to the surface of the retina and so combine or subtract messages from adjacent photoreceptors (Figure 2.4).

Behind the retina is the pigment epithelial layer, and the ends of the photoreceptors are embedded in this layer (Figure 2.5). The specialised photoreceptors are unable to fulfil all their metabolic requirements and many of these functions, including visual pigment regeneration are carried out by the pigment epithelial layer. Behind this layer is the choroid layer, which is rich in blood vessels. Both these layers contain the black pigment melanin, and this light-absorbing pigment prevents the reflection of stray light within the globe of the eyeball. Without this pigment, light rays would be reflected in all directions within the eye and would cause diffuse illumination of the retina, rather than the contrast between light and dark spots required for the formation of precise images. Albinos lack melanin throughout their body, and so have very poor visual acuity. Visual acuity is usually measured using an eye chart, such as the Snellen eye chart, which you are likely to see at any optician’s. The measurement of acuity is scaled to a viewing distance of 20 feet between the observer and the eye chart. Normal visual acuity is defined as 20/20. Someone with worse than normal visual acuity, for example 20/40, must view a display from a distance of 20 feet to see what a person with normal acuity can see at 40 feet. Someone with better than normal visual acuity, for example 20/10, can see from a distance of 20 feet what a normal person must view from 10 feet. Even with the best of optical correction, albinos rarely have better visual acuity than 20/100 or 20/200. For nocturnal or semi-nocturnal animals, like the cat, the opposite strategy is employed. Instead of a light-absorbent coating at the back of the eye, they have a shiny surface called a tapetum. This reflects light back into the eye and, although it degrades the resolution of the image, it increases the probability of a photon being absorbed by a photoreceptor. In low light intensity environments, this increases visual sensitivity and, for a semi-nocturnal hunter, this makes good sense. It also explains why the eye of a cat seems to glow when it catches the beam of a torch or any other light source.
Fragment detached from rod outer segment during normal loss and regeneration of lamellae

Figure 2.5 (See also colour plate section.) A schematic illustrating how the outer segments of rods and cones are embedded in the choroid layer at the back of the eye (redrawn from Young, 1971).
Focusing the image

The ability of the eye to refract or focus light is dependent primarily on two structures: the cornea and the lens. Focusing the image is just what happens when light passes from air into the cornea and from the aqueous humour in the eye to the lens. The relative difference in the density of the two sets of media means that 70% of the eye’s focusing is done by the cornea. However, this focusing is not adjustable, whereas focusing by the lens is adjustable. The lens is situated immediately behind the iris, and its shape can be altered by the ciliary muscles. The lens is usually relatively flat, due to the tension of the elastic fibres that suspend it in the eye. In this flat state the lens focuses distant objects on the retina. When the ciliary muscles contract, tension is taken off these elastic fibres, and the lens becomes more rounded in shape. In this rounded shape condition, the lens focuses nearer objects on the retina. The ciliary muscles thus control whether near or far objects are focused, a process called accommodation. Accommodation is usually integrated with convergence (‘turning together’) of the eyes. When a near object is fixated, the eyes turn inwards, so that the two images of the object are fixated on corresponding portions of the retina.

The refractive or focusing power of the eye is measured in dioptres, the reciprocal of the distance in metres between the eye and an object. For example, an eye with a refractive power of 10 dioptres can bend light sufficiently to focus on an object 10 cm away. In humans with normal vision, the refractive power declines from 14 dioptres at the age of 10 (a focusing distance of only 7 cm, allowing one to focus on the tip of one’s nose) to about 9 dioptres at 20 (a focusing distance of 11 cm), 4 dioptres in the mid-30s (25 cm), 1–2 dioptres in the mid-40s (50–100 cm) and close to zero by the age of 70 (a condition called presbyopia). The change from 4 dioptres to 2 dioptres is the one people notice most as it affects reading. Most people hold books from 30–40 cm from their eyes. The reason for this change in focusing ability is related to changes in the lens’s size, shape and flexibility (Koretz & Handelman, 1988).

The lens consists of three separate parts: an elastic covering (the capsule), an epithelial layer just inside the capsule and the lens itself. The lens is composed of fibre cells produced by the epithelial layer. The most common protein class in the lens is the crystallins. They make up 90% of the water-soluble proteins in the lens of vertebrates. Most of the crystallins are contained in the fibre cells. The unique spatial arrangement of these molecules is thought to be important for the maintenance of the transparency and refractive properties of the lens (Delaye & Tardieu, 1983). The distribution of the crystallins is not uniform throughout the lens. There is a general increase in protein concentration towards the centre of the lens. As a result, the refractive index increases towards the core of the lens, compensating for the changing curvature of the lens.
The fibre cells that make up the lens are being produced constantly, but none is discarded, which leads over time to a thickening of the lens. This cell production continues throughout life and, as a result, the lens gradually increases in diameter and slightly alters in shape. For example, the unaccommodated lens in an infant is 3.3 mm in thickness, whereas by the time a person reaches 70 the unaccommodated lens can be as thick as 5 mm. The old fibre cells in the centre of the lens become more densely packed, producing a hardening (sclerosis) of the lens. This thickening and hardening of the lens reduces its ability to correctly focus light on the retina. Moreover, those fibre cells in the centre eventually lose their nuclei and cell organelles. The crystallin in these cells cannot be replaced and, although it is a very stable protein, over time it does suffer a slight denaturisation (change in structure). This leads to a ‘yellowing’ of the lens, most noticeable in old age. This yellowing acts as a filter, subtly altering our colour perception as we grow older.

Two common problems arise with lens focusing: myopia and hyperopia (Figure 2.6). Myopia (near sightedness) is an inability to see distant objects clearly. This problem can be caused in two ways: (1) refractive myopia, in which the cornea or lens bend the light too much, or (2) axial myopia, in which the eyeball is too long. As a result, the focus point is in front of the retina. Hyperopia (or far sightedness) is an inability to see nearby objects. In the hyperopic eye the focus point is located behind the retina, because the eyeball is too short or because the lens is unable to fully focus the image (as discussed above). The changes in the lens means that hyperopia becomes more and more common with increasing age. In contrast, myopia is more likely to develop in younger people (see below).

The development of myopia

Although myopia and hyperopia are relatively stable conditions in adults, in new-born infants these refractive errors rapidly diminish to produce emmetropia (this is when the length of the eye is correctly matched to the focal length of its optics). The young eye seems to be able to use visual information to determine whether to grow longer (towards myopia) or to reduce its growth and so cause a relative
shortening of the eye (a change towards hyperopia). This process is called *emmetropisation*. In a ground-breaking experiment by Wiesel and Raviola in 1977, it was found that a degraded retinal image can lead to axial eye elongation, a condition called *deprivation myopia* or *form-deprivation myopia*. Further experiments have shown that, if myopia or hyperopia is imposed by the use of spectacles on the young of a variety of species including chicks, tree-shrews and primates, the shape of the developing eye alters to compensate for this change in focal length (Schaeffel, Glasser & Howland, 1988; Hung et al., 1995).

It seems that one factor controlling eye growth is dependent on the local analysis of the retinal image without the necessity of communication with the brain. Severing the optic nerve does not alter the change in eye growth associated with deprivation myopia (Wildsoet & Wallman, 1992). The local retinal mechanism seems to be triggered by retinal image degradation, involving the loss of both contrast and high spatial frequencies (Hess et al., 2006).

Another factor in axial eye growth is the degree of accommodation the eye has to undergo to focus an image. This can be used as a measure of whether the eye is hyperopic (more accommodation) or myopic (less accommodation). However, chicks can still compensate for the addition of spectacle lenses after the ability of the eye to undergo accommodation has been eliminated by brain lesions or drugs (Schaeffel et al., 1990). One reason for linking accommodation to myopia is that atropine (an antagonist for the muscarinic class of acetylcholine receptors), which blocks accommodation, has been said to halt the progression of myopia in children and monkeys (Raviola & Wiesel, 1985). It has been reported that treatment using atropine can produce a reduction in myopia of one dioptre in children, suggesting that atropine reduces the progression of myopia (Wallman, 1994). Similarly, in children with one eye more myopic than the other, the difference can be reduced by treating the more myopic eye.

However, atropine may not act by blocking accommodation because in chicks, which lack muscarinic receptors in their ciliary muscles, it reduces compensation for spectacle lenses (Stone et al., 1988; Wallman, 1994). This suggests that atropine is working at the level of retinal muscarinic receptors, but the levels required to produce myopia inhibition effects are far above the levels required to block muscarinic receptors, implying that non-specific drug effects or even retinal toxicity may be involved. Muscarinic blockers also reduce the synthesis of the sclera in chicks and rabbits, and so these blockers may be interfering with the normal, as well as the myopic, eye growth.

It has been suggested that there may be an inherited susceptibility to develop myopia. For example, children with two myopic parents are more likely to be myopic than children with no myopic parents, and monozygotic (identical) twins are more likely to both be myopic than dizygotic (non-identical) twins (Zadnik et al., 1994; Hammond et al., 2004). However, environment seems to be a stronger factor in
the development of myopia. For example, myopia can be strongly correlated with education. Of Taiwanese students and Hong Kong medical students 70%–80% are myopic, compared to only 20%–30% of the same age group in rural areas. Moreover, as once humans have become myopic they stay myopic, one can compare differences in occurrence at different ages in a population. In Finnish Inuits and in Hong Kong, the young are, on average, myopic whereas the middle-aged are not. This increase in myopia in the young suggests that environmental, rather hereditary factors, are important in the development of myopia (Goldschmidt, 2003).

In conclusion, it seems that visual cues guide the growth of bird and mammal eyes actively towards emmetropia. This would be consistent with the association of education with myopia, as the students’ eyes will grow into focus at the distance of the page, whereas the eyes of a person who largely lives outdoors will grow to focus at infinity.

Clouding of the lens (cataracts)

Another important factor in focusing a clear image on the retina is the transparency of the lens. Clouding of the lens, which is called a cataract, is sometimes present at birth (a congenital cataract), can be caused by eye disease (a secondary cataract), or by injury (a traumatic cataract), but the most common cause of all is old age (a senile cataract). Cataracts develop in roughly 75% of people over 65 and in 95% of people over 85. However, in only 15% of people do the cataracts cause serious visual impairment and in only 5% of cases is surgery necessary. In this case a small opening is made in the eye, through which the lens is removed either by pushing on the lens to force it out and allowing its removal with forceps, or by a method called phacemulsification, which uses ultrasound to remove the lens. To compensate for removal of the lens, the patient may either be given glasses, a contact lens or an intraocular lens (an artificial lens to replace the one removed).

Congenital cataracts may arise from a number of possible causes: aberrant function of the fibre cells, such as alteration of structural proteins or proteins that serve to protect the cell from damage and preserve the clarity of the lens matrix, or the defect may lie in a metabolic pathway, resulting in an accumulation and deposition of insoluble material in the lens. At least two forms of congenital cataracts have been shown to be caused by mutations of the genes for the crystallin protein, which in turn lead to changes in the structure of this protein (Cartier et al., 1994).

Photoreceptors

Once the image has been focused on the retina, this pattern of light must be transformed into a pattern of neural activity that can
accurately represent the image. This transformation or transduction is carried out by the light-sensitive receptor cells (photoreceptors) in the retina. There are two types of photoreceptors: the rods and the cones. The human retina contains around 120 million rods and 6 million cones. The cones are concentrated in a small area of the retina called the fovea (Figure 2.7). They mediate diurnal visual function, and provide high acuity, colour vision. The rods mediate nocturnal vision and provide only low acuity, monochrome vision. There are three types or classes of cones. Between 5% and 10% of the total cone population are blue cones, and they form a ring or annulus around the edge of the fovea. The rest of the cones are red and green in a ratio of 2 : 1. These latter two classes do not seem to be arranged in a regular array, but are randomly mixed together in small patches or clusters (Mollon & Bowmaker, 1992).

Photoreceptors consist of an outer segment connected by a cilium to an inner segment containing the cell nucleus (Figure 2.8). The outer segment contains several hundred thin membrane plates (lamellae), around 750 lamellae in a monkey rod. In rods, the lamellae are free floating discs, whereas in cones they consist of one continuous, folded membrane. Embedded in the lamellae membrane are the photopigment molecules (rhodopsin). A single human rod contains 100 million photopigment molecules. They are so tightly packed together there is only 20 nm between individual molecules, and they make up 80% of the membrane proteins. Each pigment molecule consists of two parts: opsin (a protein) connected by a Schiff-base linkage to retinal (a lipid), which is synthesised from retinol (vitamin A). Retinal is a long chain molecule that can exist in two forms: a straight chain form (all-trans retinal) and a bent form (11-cis retinal). 11-cis retinal is the only form that can bind to the opsin. When the 11-cis
retinal absorbs a photon of light, the long chain straightens to the all-trans form, a process called photo-isomerisation, and the photopigment molecule then eventually breaks into its two constituent parts. When this occurs, it changes colour from a rosy colour to a pale yellow. The photopigment is said to have been bleached.

Transduction

In darkness, the rods and cones have a resting membrane potential of $-40$ mV, considerably different from the usual membrane potential of $-80$ mV found in other neurons. This is because a continuous dark current flows into the outer segment as sodium ($Na^+$) ions move down their electrochemical gradient through open cation channels. The effect of light is to cause hyperpolarisation of the cell membrane by indirectly closing the cation channels in the outer segment membrane. This change in potential is in the opposite direction to the change found in other receptors and neurons, which depolarise when stimulated. The cation channels are normally kept open by cytoplasmic cyclic guanosine $3'\text{-}5'$-monophosphate (cGMP). The photo-isomerisation of rhodopsin precipitates a series of reactions, which result in a rapid reduction in the levels of cGMP. This in turn causes the cation channels to close and the electrical resistance of the outer segment membrane to increase, stopping the dark current. Thus, the cGMP acts as an internal messenger within the cell, transferring news of the detection of light from rhodopsin molecules in the disc membrane to the ion channels in the cell membrane.
As mentioned, when a photon is absorbed by a rhodopsin molecule, the retinal chromophore undergoes photo-isomerisation and changes from the 11-cis to the all-trans configuration as the terminal chain connected to opsin rotates. This transition is very rapid, taking only $10^{-12}$ seconds. The protein then goes through a series of intermediate forms. One of these intermediate forms is metarhodopsin II, which is produced 1 ms after the absorption of a photon. Metarhodopsin II is enzymatically active and binds a globular- or G-protein called transducin to a disc membrane. This protein is composed of three subunit chains ($T_\alpha$, $T_\beta$ and $T_\gamma$). Different isoforms of the three subunits are present in rods and cones and this may explain some of the physiological differences between the two receptor types (Peng et al., 1992). In the inactive state, transducin is bound to a molecule of guanine diphosphate (GDP). Metarhodopsin II catalyses the exchange of a molecule of guanine triphosphate (GTP) for the bound molecule of GDP. The metarhodopsin II–transducin–GTP complex then disassociates into metarhodopsin II, $T_\alpha$-GTP and $T_\beta\gamma$. Metarhodopsin II can catalyse around 500 such exchanges, before it is inactivated by phosphorylation of sequences near the C-terminus. This phosphorylation allows a protein called arrestin to compete with transducin for metarhodopsin II, and so inhibits further catalytic activity.

Each of the released $T_\alpha$-GTP molecules then binds to an enzyme called phosphodiesterase (PDE), which is a complex of four subunits, two inhibitory subunits (PDE$_\gamma$) and two catalytic subunits (PDE$_\alpha$ and PDE$_\beta$). The interaction of $T_\alpha$-GTP with PDE splits off the PDE$_\gamma$ subunits, and $T_\alpha$-GTP–PDE$_{\alpha\beta}$ complex can then catalyse the break-up of cGMP. This break-up reaction produces one molecule of non-cyclic GMP and one $H^+$ ion for every molecule of cGMP hydrolysed. Around 800 molecules of cGMP are hydrolysed before the $T_\alpha$ component of the $T_\alpha$-GTP–PDE$_{\alpha\beta}$ complex becomes deactivated. The deactivation occurs through the conversion of GTP to GDP, leading to the release of PDE$_{\alpha\beta}$, which then reassociates with the PDE$_\gamma$ subunit. The $T_\alpha$ subunit, which is now again bound to GDP, reassociates with the $T_\gamma$ subunits to complete the cycle initiated by the rhodopsin molecule’s absorption of a photon and storage of energy. This two-stage cycle is powered by the $T_\alpha$-induced conversion of GTP to GDP. This system can allow the hydrolysis of 400,000 cGMP molecules within 1 second of the absorption of a single photon. The reduced levels of cGMP cause the sodium channels to close, and the receptors to hyperpolarise. A single photon can close approximately 300 channels, about 3%–5% of the channels that are open in the dark, and internal levels of cGMP fall by about 20% on illumination (Baylor, 1987).

The calcium feedback mechanism

The intracellular concentration of calcium ions ($Ca^{2+}$) also changes over the course of the photo-transduction process. In the dark, $Ca^{2+}$
ions, like Na\(^+\) ions, enter the cell through the open cation channels and are expelled from the cell by an electrogenic calcium-sodium exchanger, which is located in the cell membrane. The transduction process leads to a fall in intracellular cGMP concentration, and the subsequent closure of the cation channels means that Ca\(^{2+}\) ions can no longer enter the cell, but they are still being pumped out of the cell. As a result, the level of intra-cellular calcium falls and does not start to rise again until the cation channels start to reopen as the cell recovers from stimulation. It has been shown that the changing level of Ca\(^{2+}\) ions acts as a feedback mechanism which speeds up a cell’s recovery from light stimulation, and also mediates light adaptation (Koutalos & Yau, 1993). Three mechanisms have been proposed for this action.

(i) It has been suggested that intracellular Ca\(^{2+}\) ions alter the action of guanylate cyclase, the enzyme responsible for the synthesis of cGMP. This is mediated through a Ca\(^{2+}\) binding protein called recoverin. It was suggested that this protein activated guanylate cyclase at low Ca\(^{2+}\) levels. Thus as the levels of intra-cellular calcium fall, the activity of guanylate cyclase should increase. This raises the concentration of cGMP, allowing the cGMP-activated cation channels to reopen and the dark current to flow again.

(ii) Ca\(^{2+}\) ions seem to decrease the cGMP gated cation channel’s affinity for cGMP. Thus, as Ca\(^{2+}\) levels fall, the affinity of the channels for cGMP rises, which helps to offset the fall in cGMP levels.

(iii) The third component of the effect of Ca\(^{2+}\) is through S-modulin, a calcium binding protein homologous to recoverin. This protein lengthens the lifetime of active PDE at high Ca\(^{2+}\) levels, and it inhibits phosphorylation of rhodopsin at the same high Ca\(^{2+}\) levels (Kawamura, 1993). Thus, one of the sites of Ca\(^{2+}\) modulation of the cGMP cascade is at the level of pigment inactivation.

**Signal efficiency**

It is common in textbooks to emphasise the stability of the rhodopsin molecule and imply that the photoreceptors are therefore extremely efficient at signalling the presence of light as there is little or no background noise. Denis Baylor calculated that the spontaneous thermal isomerisation of a single rhodopsin molecule from the 11-cis to the all-trans form should occur about once every 3000 years, or \(10^{23}\) times more slowly than photo-isomerisation (Baylor, 1987). However, the retinal photoreceptors are actually very ‘noisy’. They produce discrete electrical events in the dark that are indistinguishable from those evoked by light. This phenomenon limits visual sensitivity at low light levels, although recently it has been argued...
that, under certain circumstances, noise can play a constructive role in the detection of weak signals via a mechanism known as stochastic resonance (Wiesenfield & Moss, 1995). The random and spontaneous electrical events are strongly temperature dependent, and have therefore been attributed to the thermal isomerisation of retinal, thus initiating the G-protein cascade, which should signal the absorption of a photon. However, the thermal generation of dark events in photoreceptors requires activation energies in the range of 23–27 kcal mol\(^{-1}\), which is significantly less than the energy barrier of 45 kcal mol\(^{-1}\) required for photo-isomerisation of retinal. Recent work has suggested that photoreceptor noise results from the thermal isomerisation of a small proportion of photoreceptor molecules (< 0.01%), in which the Schiff-base linkage between the retinal chromophore and the opsin is unprotonated (Barlow et al., 1993). This deprotonation of the linkage destabilises the photopigment molecule by reducing the energy barrier for isomerisation to 23 kcal mol\(^{-1}\). Interestingly, the horseshoe crab (Limulus) has developed a method for reducing photoreceptor noise during the night, increasing the sensitivity of their eyes and so helping them locate a mate. At night, nerve signals from a Circadian clock in the brain reduces the spontaneous activity in the photoreceptors by reducing the number of photopigment molecules in the unprotonated state. The mechanism for this change seems to be a lowering of the external pH in the vicinity of the photopigment containing membrane (Barlow et al., 1993).

The centre-surround organisation of the retina

There are around 126 million photoreceptors in the retina, each one signalling information on how much light is absorbed at a particular point in the retina. The information from the retina is transmitted to the brain via the axons of the ganglion cells, but there are only one million of these. Therefore, the retina has to condense and reorganise the information from the photoreceptors into a form that can be transmitted through the optic nerve. To consider how it does this, we must ask ourselves what is the basic purpose of a visual system? It is not just to signal the presence or absence of illumination, but instead it is to detect patterns of light from which information relating to the identity of objects and their spatial relationships in the environment can be derived. The first step in this process is the detection of differences in light at adjacent locations, which are likely to signal an edge or border. Such edges can then be used to build up a picture of the environment (see Chapter 8). Regions of uniform illumination are less important as they are unlikely to signal an edge.

The first point at which information relating to regional light differences is extracted is at the level of the retinal ganglion cells. Each ganglion cell is connected to a number of photoreceptors via bipolar cells. Stimulation of the retinal area corresponding to these
photoreceptors alters the activity of the ganglion cell, and this retinal area is called the ganglion cell’s receptive field. The photoreceptors in a particular receptive field do not simply stimulate the ganglion cell, but instead are arranged in what is called a centre-surround organisation. For example, light falling on the centre of the receptive field might result in the corresponding photoreceptors stimulating the ganglion cell (an ON response). Light falling just on the surrounding ring of photoreceptors might inhibit the ganglion cells (an OFF response). This cell is an example of an ON-centre, OFF-surround cell. There are also cells with the opposite arrangement, an OFF-centre and an ON-surround. This opponent interaction is often called lateral inhibition.

If a whole receptive field is illuminated (as in Figure 2.9(a)), both the ON-centre and OFF-surround are stimulated. The ON-centre excites the ganglion cell, but the OFF-surround inhibits it and there is little if any change in the ganglion cell’s firing rate. Now consider what happens if an edge is positioned as shown in Figure 2.9(b). The ON-centre receives an increase in light and will stimulate the ganglion cell, while the OFF-surround receives a reduced level of light and will not inhibit the cell as much as before. The net result is that the ganglion cell is stimulated and will signal the presence of a light/dark boundary. Because the receptive fields of these centre-surround cells are usually concentrically arranged, the cells will respond well
whatever the orientation of the edge. Only by comparing the responses of a number of ganglion cells can the orientation of a stimulus be determined and this comparison is made in the first cortical visual area.

The structure of the centre-surround receptive field is believed to contribute to a well-known optical illusion called the Hermann grid. The Hermann grid consists of a criss-crossing set of white lines on a black background (Figure 2.10). Although all the white lines are of the same luminance, the intersections of the lines appear darker. An explanation for this phenomenon may be based on the relative activity in the centre and surround regions. Consider the receptive field on the left of Figure 2.11. Most of the ON-centre region is strongly stimulated, whereas less than 25% of the OFF-surround field is stimulated. As the response of the cell is based on the difference between the activation in the centre and surround, this cell will be strongly active (as indicated on the firing-rate bar-chart). Now consider the receptive field on the right of the figure on the intersection of the lines. The ON-centre is again strongly stimulated. Up to 50% of the OFF-surround is also stimulated. Thus the resultant activity of this cell is less than that of the first cell. Therefore, the centre of the intersection appears to be darker than the rest of the line.
Light adaptation

Humans can see over a range of illumination levels of about $10^{10} : 1$ (Table 2.1). However, all the information leaving the eye travels in the optic nerve fibres, whose response range is quite limited (perhaps $100 : 1$), and so an enormous range of inputs is mapped on to a very small range of outputs. To cope with this problem, the visual system employs a number of strategies. Firstly, the ganglion cell’s response is dependent on the average illumination of the retina. For example, if one examines the response of an ON-centre/OFF-surround cell, the strength of the response produced by the ON-centre will be dependent on the intensity of the illumination of the surround. The result is the shift of the ganglion cell’s response function as detailed in Figure 2.12. This mobile response function has several important advantages over a fixed relationship. For a fixed relationship, the visual system will be insensitive to all but relatively large changes in input. However, the visual system can detect changes in illumination of less than 1%. Moreover, it is a wasteful system because, at any one time, the light intensities to which the visual system is exposed mostly lie within a small range. So, it is more efficient if the limited response range of the visual pathways are available to handle the whole of the relatively small range of light intensities likely to be encountered at any one time. It is more efficient if there is a flexible

<table>
<thead>
<tr>
<th>Table 2.1. Range of visible light intensities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intensity (candelas/m²)</td>
</tr>
<tr>
<td>The sun at noon</td>
</tr>
<tr>
<td>$10^{10}$</td>
</tr>
<tr>
<td>$10^9$</td>
</tr>
<tr>
<td>$10^8$</td>
</tr>
<tr>
<td>$10^7$</td>
</tr>
<tr>
<td>Damaging</td>
</tr>
<tr>
<td>Filament of a 100 watt light bulb</td>
</tr>
<tr>
<td>$10^6$</td>
</tr>
<tr>
<td>White paper in sunlight</td>
</tr>
<tr>
<td>$10^5$</td>
</tr>
<tr>
<td>$10^4$</td>
</tr>
<tr>
<td>$10^3$</td>
</tr>
<tr>
<td>$10^2$</td>
</tr>
<tr>
<td>Photopic vision</td>
</tr>
<tr>
<td>Comfortable reading</td>
</tr>
<tr>
<td>$10$</td>
</tr>
<tr>
<td>$1$</td>
</tr>
<tr>
<td>$10^{-21}$</td>
</tr>
<tr>
<td>Mesopic vision</td>
</tr>
<tr>
<td>White paper in moonlight</td>
</tr>
<tr>
<td>$10^{-22}$</td>
</tr>
<tr>
<td>$10^{-23}$</td>
</tr>
<tr>
<td>White paper in starlight</td>
</tr>
<tr>
<td>$10^{-24}$</td>
</tr>
<tr>
<td>$10^{-25}$</td>
</tr>
<tr>
<td>Scotopic vision</td>
</tr>
<tr>
<td>Weakest visible light</td>
</tr>
<tr>
<td>$10^{-26}$</td>
</tr>
</tbody>
</table>

Source: Taken from Sekuler & Blake, 1994.
mapping of input on to output, with the operating range of the visual system shifting with the level of ambient light. This shift in the operating range and the accompanying change in sensitivity is called light adaptation. This shift in sensitivity allows a phenomenon called lightness constancy. A particular object or surface will appear equally light, relative to surrounding surfaces, over a range of illumination. Perceptual constancies, such as lightness, colour and object constancies, are a fundamental feature of the visual system. Features such as lightness, colour or shape can be used to identify an object, and then can be used as a cue to recognise it subsequently. These perceptual cues would be useless if they changed with different viewing conditions, such as changes in illumination or viewing angle. As a result, the visual system has developed a number of mechanisms to allow the stimulus features of objects to appear constant under different conditions, and lightness constancy is one of these.

Duplicity theory of vision

Another strategy to expand the visual system’s operating range is specialisation and division of labour. As mentioned, the retina contains two types of photoreceptors: rods and cones, and the range of possible light intensities is divided between them. The rods respond to low intensities and the cones to high intensities (Figure 2.13). There are three cone classes: red, green and blue (see Chapter 3), and the photopic sensitivity shown in Figure 2.13 is based on their combined activity. At intermediate intensity levels there is a degree of overlap between the two photoreceptor systems, with both being

Figure 2.12. The stimulus–response relationship for a cone, showing the amplitude of hyperpolarisation as a function of light intensity. (A). Without light adaptation. Hypothetical relationship shows how the whole range of working intensities is mapped onto the limited response range of a photoreceptor. This results in very poor sensitivity to small changes in illumination. (B). With light adaptation. The stimulus–response relationship is shown for a single cone in the turtle Pseudemys scripta elegans as measured by Normann and Perlman. The resting membrane potential in the dark is indicated by the line at 0 millivolts. Steady background light reduces the sensitivity of the cone, but causes only small changes in the resting potential, permitting the cone to generate relatively large changes in membrane potential for relatively small increments or decrements in illumination about the background level (redrawn and modified from Lennie & D’Zmura, 1988).
active at the same time. Under these conditions, the rod system sums its responses with the responses of the red cone class. The spectral sensitivity of the rods and cones differs (Figure 2.13) and when the rods and cones interact under these conditions, our perception of colour is shifted towards shorter wavelengths. This is called the Purkinje shift.

This division of labour between the two photoreceptor systems is termed the duplicity theory of vision, and was first proposed by Von Kries in 1896. The arrangement can be demonstrated in a number of ways. For example, the dark adaptation curve (the increase in sensitivity which occurs when illumination changes from light to dark) is clearly a two-stage function (Figure 2.14). At high light intensities, the light-adapted eye is at its least sensitive. When the ambient light is turned off, the sensitivity of the eye increases. It does so in two stages. It first increases for about 3–4 minutes, then levels off for about 7–10 minutes, before increasing again for about 20–30 minutes. The first increase in sensitivity is due to the cones, the second to the rods.

This difference is due to the different rates of pigment regeneration in the rods and cones. This can be measured in the following way. A dim measuring beam of constant intensity is projected into
the eye. The beam passes through the retina, hits the back of the eye and is reflected back out. During this procedure much of the light is absorbed by the visual pigment, by other structures in the eye and by the black pigment epithelium. However, some of the light is reflected back out. When the visual pigment is bleached, it absorbs less light. Thus the amount of light that is reflected out of the eye will be a measure of the amount of pigment present. This method is called retinal densitometry, and was used by William Rushton (1961) to measure the concentration of visual pigments during pigment regeneration. There are no rods in the centre of the fovea, and so, if the measuring beam is confined to this region, it is possible to measure the rate of pigment regeneration in cones. His results suggest that the cone pigments take approximately 6 minutes to regenerate fully. To measure the rate of pigment regeneration in rods, Rushton used the technique on the retinas of rod monochromats (subjects who lack cones in their retinas), and found it took 30 minutes for rod pigment to regenerate.

The differences are also evident in the response of the eye to a flickering light. A brief flicker excites the eye for around 0.1–0.2 seconds and, because of the persistence of excitation, rapid successive flashes of light become fused together to give the appearance of being continuous. This phenomenon is used in films and television to give the illusion of motion (the film screen flickers at 24 Hz and that
of the television at 60 Hz). The frequency at which flicker fusion occurs, called the critical frequency for fusion, varies with light intensity. At low light intensity, flicker fusion can occur at 2–6 Hz, but under bright illumination flicker fusion can be as high as 60 Hz. The difference stems largely from the fact that the cones are active under bright illumination, and they have a shorter persistence in response to light. Under poor illumination, the rods mediate vision and they have a longer persistence. For the same reason of differences in persistence, the fusion frequency varies with retinal eccentricity.

Sensitivity, acuity and neural wiring

As mentioned above, there are over 120 million rods and at least 6 million cones, but only 1 million ganglion cells to carry the information from the retina to the brain. This suggests a general convergence factor of 126 : 1, but in reality the convergence factor varies with retinal eccentricity. At the centre of the fovea, the convergence factor may be as low as 1:1, but in the periphery it may be as high as several hundred to one. The degree of convergence determines the spatial resolution of that part of the retina. The resolution (measured as visual acuity) refers to the ability to distinguish differences in the spatial distribution of light in the image. The size of a ganglion cell’s receptive field will be large if the convergence factor is high, but receptive field size will be small if the convergence factor is low. The receptive fields can be thought of as analogous to the grain size in a photograph. The larger the grain size, the poorer the quality of the picture and the poorer the detail that can be discerned. At the centre of the retina (which is dominated by cones), the ganglion cell receptive fields are at their smallest and here spatial resolution is at its best. As one moves out into the periphery (where the rods dominate), the receptive field size rapidly increases and visual acuity rapidly decreases.

An increase in receptive field size does have one advantage: it increases sensitivity. If one considers a single ganglion cell, for it to be sufficiently stimulated to signal the presence of light, a certain amount of excitatory neurotransmitter must be released by the neurons that synapse on to it. If a dim light illuminates the retina, only a small proportion of photoreceptors will be stimulated. If the receptive field is small, with only a few photoreceptors, then perhaps only one of these photoreceptors will be stimulated. The change in level of neurotransmitter released by that cell’s excitation may not be enough to stimulate the ganglion cell. However, if the receptive field is large, the same proportion of photoreceptors will be stimulated, but a larger receptive field may contain several activated photoreceptors. The combined effects of these excited photoreceptors will stimulate the ganglion cell, where a single excited photoreceptor could not.
The rod system is more sensitive than the cone system for two additional reasons. Firstly, on average, the rods have a larger diameter and are longer. The increased diameter increases the probability of a photon passing through an individual rod, and the increased length increases the probability of the photon’s absorption as it passes through the receptor. Secondly, the persistence of the response to the absorption of a photon is longer in rods than in cones. This means that, if a rod is stimulated by the absorption of a photon, but this stimulation is too weak to stimulate the bipolar cell, then the probability of it absorbing a second while the rod is still excited is increased. This second absorption may boost the level of stimulation in the rod such that it may be strong enough to initiate the passage of a signal in the bipolar and ganglion cells.

Summary of key points

(1) For humans, visible light is a narrow range of the electromagnetic spectrum between 360 and 700 nm. Many species can see shorter wavelengths (down to around 300 nm), in a part of the spectrum called the ultra-violet.

(2) Light entering the eye is focused by the cornea and the lens. The cornea is responsible for 70% of the focusing in the eye, but its focal distance is not adjustable. The adjustable, ‘fine-tuning’ of the focusing of the image is carried out by the lens. The focusing power of the eye is measured in dioptres, the reciprocal of the distance in metres between the eye and an object.

(3) Two common problems arise with lens focusing: myopia and hyperopia. Myopia (near sightedness) is an inability to see distant objects clearly. Hyperopia (or far sightedness) is an inability to see nearby objects.

(4) Although myopia and hyperopia are relatively stable conditions in adults, in new-born infants these refractive errors rapidly diminish to produce emmetropia (this is when the length of the eye is correctly matched to the focal length of its optics). The young eye seems to be able to use visual information to determine whether to grow longer (towards myopia) or to reduce its growth and so cause a relative shortening of the eye (a change towards hyperopia).

(5) Once the image has been focused on the retina, this pattern of light must be transformed into a pattern of neural activity that can accurately represent the image. This transformation or transduction of light into neural energy is carried out by the light-sensitive receptor cells (photoreceptors) in the retina. There are two types of photoreceptors: the rods and the cones.

(6) Each photopigment molecule consists of two parts: opsin (a protein) connected by a Schiff-base linkage to retinal (a lipid), which is synthesised from retinol (vitamin A). Retinal is a long-chain molecule that can exist in two forms: a straight chain form
(all-trans retinal) and a bent form (11-cis retinal). 11-cis retinal is the only form which can bind to the opsin. When the 11-cis retinal absorbs a photon of light, the long chain straightens to the all-trans form, a process called photo-isomerisation, and the photopigment molecule then eventually breaks into its two constituent parts.

(7) In darkness, the rods and cones have a resting membrane potential of $-40\,\text{mV}$, because a continuous dark current flows into the outer segment as sodium (Na$^+$) ions move through open sodium channels in the cell membrane. The effect of light is to cause hyperpolarisation of the cell membrane by indirectly closing the cation channels in the outer segment membrane. The cation channels are normally kept open by cytoplasmic cyclic guanosine 3’–5’-monophosphate (cGMP). The photo-isomerisation of rhodopsin precipitates a series of reactions that result in a rapid reduction in the levels of cGMP. This, in turn, causes the cation channels to close and so reducing or stopping the dark current.

(8) The intracellular concentration of calcium ions (Ca$^{2+}$) also changes over the course of the photo-transduction process. The changing level of Ca$^{2+}$ ions acts as a feedback mechanism, which speeds up a cell’s recovery from light stimulation, and also mediates light adaptation.

(9) Each retinal ganglion cell is connected to a number of photoreceptors via bipolar cells. Stimulation of the retinal area corresponding to these photoreceptors alters the activity of the ganglion cell, and this retinal area is called the ganglion cell’s receptive field. The photoreceptors in a particular receptive field do not simply stimulate the ganglion cell, but instead are arranged in what is called a centre-surround organisation.

(10) The response range of the visual system is limited, but the visual system responds to a wide range of light intensities, although we are likely to encounter only a relatively small range of light intensities at any one time. There is a flexible mapping of input onto output, with the operating range of the visual system shifting with the level of ambient light. This shift in the operating range and the accompanying change in sensitivity is called light adaptation. This shift in sensitivity allows a phenomenon called lightness constancy. A particular object or surface will appear equally light, relative to surrounding surfaces, over a range of illumination.

(11) To further expand the visual system’s operating range, the rods and cones function over different ranges of light intensities. The rods respond to low intensities and the cones to high intensities. At intermediate intensity levels there is a degree of overlap between the two photoreceptor systems. This arrangement is termed the duplicity theory of vision.

(12) There are over 120 million rods and at least 6 million cones, but only 1 million ganglion cells to carry the information from the
retina to the brain. At the centre of the fovea, the convergence factor may be as low as 1:1, but in the periphery it may be as high as several hundred to one. The degree of convergence determines the spatial resolution and sensitivity of that part of the retina. A low convergence allows good spatial resolution, but a low light sensitivity. A high convergence means poor spatial resolution, but a higher light sensitivity.
Retinal colour vision

Why do we need more than one cone pigment?

In the vertebrate eye, colour is detected by cone receptors. In the case of humans and other Old World primates, there are three cone classes (Figure 3.1). A blue or short-wavelength pigment absorbing maximally at 420 nm, a green or middle-wavelength pigment absorbing maximally at 530 nm and a red or long-wavelength pigment absorbing at 565 nm (Dartnall et al., 1983). For an animal to be able to discriminate between colours, it must have two or more different classes of cones. This is because a single cone pigment cannot discriminate between changes in wavelength and changes in the intensity of a light. For example, a red cone will respond strongly to a 560-nm light, but weakly to a 500-nm light. However, the same pattern of response can be obtained by a light of fixed wavelength, say 560 nm, and changing intensity, as a single cone class can only signal the number of photons absorbed by its pigment. This pattern of response is called univariance. To make the crucial differentiation between wavelength and intensity, a comparison of signals from two or more cone classes is required. 540-nm and 640-nm lights will produce different patterns of firing in the red and green cones as compared with two 540-nm lights of different intensity. As a general rule of thumb, the more cone classes in an eye, the better will be the wavelength discrimination. Non-primate mammals, which rely heavily on sound and smell, have only two pigments (dichromacy), whereas birds, such as the pigeon, which are highly visually orientated, have five (pentachromacy).

Trichromacy

In 1802, Thomas Young correctly proposed that the human eye detected different colours because it contained three types of receptor, each sensitive to a particular hue. His theory was referred to as the trichromatic (three-colour) theory. It suggested that, for a human
observer, any colour could be reproduced by various quantities of three colours selected from various points in the spectrum, such as red, green and blue. However, it has long been thought that there are four primary colours: red, green, blue and yellow. The trichromatic theory cannot explain why yellow is included in this group. In addition, some colours appear to blend while others do not. For example, one can speak of a bluish-green or a yellowish-green, but one cannot imagine a greenish-red or a bluish-yellow. To account for this, an alternative theory was proposed by Ewald Herring, which regarded mechanisms in the eye sensitive to these colours as undergoing some form of opponent interaction. As we discussed in the preceding paragraph, the responses from the three different cone classes are compared to allow colour discrimination. This is indeed done in an opponent manner. There are three opponent mechanisms (Figure 3.2). The first compares the difference between the red and the green cone classes. The second compares the difference between the blue cones and the sum of the red and green cones (yellow). The final mechanism, is an achromatic (black-white) mechanism detecting differences in luminance. So, the human visual system is trichromatic and also compares the four primary colours in an opponent mechanism.

The basis of the opponent mechanism is the centre-surround opponency described in Chapter 2. The difference in illumination between the centre and the surround of ganglion cell receptive fields can obviously provide the basis of the achromatic mechanism. But this arrangement can also provide colour information. If the centre is composed of cones of one class, and the surround of cones of another class, then centre-surround opponency can produce colour opponency. This idea suggests that the cones should be organised in the fovea in a regular pre-ordained array. However, experimental evidence suggests that the red and green cones are randomly arranged.

![Spectral absorbance curves for the human photopigments.](Figure 3.1)
in the fovea (Mollon & Bowmaker, 1992; Roorda & Williams, 1999). In
the fovea, the receptive fields are very small and composed of a
comparatively small number of cones. Given a random arrangement
of the two cone classes, the probability is that there will be an
unequal number of the two cone classes in the centre and the sur-
round. Thus, there will be colour opponency without the need for a
regular arrangement of cones. This almost haphazard arrangement
emphasises the recent nature of the red and green cones in the Old
World primate eye. All other mammals are dichromats, possessing a
blue cone class and a second cone class that absorbs primarily in the
red–green range. As in primates, the few blue cones form a ring
around the edge of the fovea, and the rest of the fovea is composed
of a single cone class. The centre-surround opponency in this part of
the fovea merely signals achromatic differences. The recent addition
of a second cone class (less than 35 million years ago) has been
overlaid on to this older achromatic opponent system to produce a
new colour opponent system (Mollon, 1989, 1999).

As can be seen from Figure 3.1, the red and green pigments have a
great deal of overlap in their spectral absorbance curves. There is a
separation of only around 35 nm between their maximum spectral
absorbs. Although colour discrimination relies on there being
some overlap between cone pigments to allow comparison of the
responses of two or more cone classes to a light, human colour vision
would be improved by a more equal spacing of the three cone pig-
mets in the spectrum. The reason that the red and the green pig-
mets are not further apart may be linked to the way information
from the cones is processed. As well as using the differences in the
responses of different cone classes to light to determine colour, the

---

**Figure 3.2** A schematic diagram illustrating the relationship of the
three opponent mechanisms. In the
achromatic channel and the
red–green channel, the role of the
blue cones is unclear. They may
under some conditions contribute
to both achromatic and red–green
sensitivity.

---
combined responses of red and green are used in a system that is sensitive purely to changes in luminance and is used to detect fine detail in a scene. For this purpose, the spectral absorbances of the two cone classes must be similar for two reasons. Firstly, the point at which a lens focuses a light is dependent on its wavelength. If the spectral absorbances of the red and green pigments were to be moved further apart in the spectrum, the focal length for the light which primarily stimulates each of the two cone pigments would be significantly different. Thus, when the light for one of the cone classes is in focus, the other will be out of focus (chromatic aberration), and if the signals from the two cone classes are combined, the cumulative image will be degraded. Secondly, if the spectral absorbances of the two cone classes differ too much, then a light of a specific wavelength and intensity may stimulate one class strongly and the other quite weakly. If one is purely looking at luminances, then one class would be signalling a strong luminance and the other would be signalling a weak luminance, which could cause problems when attempting to integrate the signals from both cone classes into a single luminance detection system.

The genetics of visual pigments

When we look at an object or scene, it is easy to assume that what we see is what everyone else sees. However, this need not be so. The different complement of visual pigments in the eyes of ‘colour-blind’ people means that they see a very different picture from the one most of us see. One of the first to wrestle with this problem was the chemist John Dalton. Two hundred years ago he described his own colour blindness in a lecture to the Manchester Literary and Philosophical Society, and the term Daltonism has been subsequently used to characterise this form of colour blindness. Dalton ascribed his colour blindness to a blue tint in the vitreous humour of his eye, which selectively absorbed light in the red–green range (Hunt et al., 1995). A macabre twist in this tale is that he was so convinced he was right, he gave instructions that on his death his eyes should be removed and dissected to confirm his hypothesis. When he died at the ripe old age of 78 in 1844, his physician Joseph Ransome examined Dalton’s eyes and found no discolouration. At this time, the main alternative to Dalton’s theory was that colour blindness arose from a defect in the brain. As a result, Ransome felt bound to report that Dalton had a ‘deficient development’ of the phrenological organ of colour! (Hunt et al. 1995).

We now know that colour blindness arises from a genetic cause, and can be traced through family trees. In fact, Dalton reported that his own brother suffered the same impairment of colour vision as he did. For humans, the different forms of colour blindness are defined in terms of their difference from the ‘normal’ trichromacy (Table 3.1). Someone who has lost the blue pigment is called a tritanope; someone
who has lost the green pigment is called a deuteranope; and someone who has lost the red pigment is called a protanope (see Figure 3.3 for a simulation of the spectrum visible to some of the different forms of colour blind observers, and see Figure 3.4 to see how this affects their perception of a visual scene). Those people with an altered pigment are called anomalous trichromats and these altered photopigments are termed anomalous pigments. For example, someone with an anomalous red pigment would be said to have protanomalous colour vision. Each normal pigment was considered to have two forms of

<table>
<thead>
<tr>
<th>Colour Abnormality</th>
<th>Pigment basis</th>
<th>Effects</th>
<th>Genetic basis</th>
<th>Occurrence (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dichromacies</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protanopia</td>
<td>Red pigment absent</td>
<td>Confuses wavelengths in the range 520–700 nm</td>
<td>(i) Deleted red gene (ii) Hybrid red gene which is either inactive or produces a green pigment</td>
<td>Male: 1.0  Female: 0.02</td>
</tr>
<tr>
<td>Deuteranopia</td>
<td>Green pigment absent</td>
<td>Confuses wavelengths in the range 530–700 nm</td>
<td>Deleted green gene</td>
<td>Male: 1.1  Female: 0.1</td>
</tr>
<tr>
<td>Tritanopia</td>
<td>Blue pigment absent</td>
<td>Confuses wavelengths in the range 445–480 nm</td>
<td>Mutant blue gene</td>
<td>0.001–0.005 (no gender difference)</td>
</tr>
<tr>
<td><strong>Anomalous Trichromacies</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protanomaly</td>
<td>Hybrid red pigment</td>
<td>Abnormal colour matches</td>
<td>Hybrid red gene</td>
<td>Male: 1.0  Female: 0.02</td>
</tr>
<tr>
<td>Deuteranomaly</td>
<td>Hybrid green pigment</td>
<td>Abnormal colour matches</td>
<td>Hybrid green gene</td>
<td>Male: 4.9  Female: 0.04</td>
</tr>
</tbody>
</table>

*Figure 3.3.* (See also colour plate section.) The appearance of the visible spectrum for five types of colour blindness: protanopia, deuteranopia, tritanopia, blue-cone monochromacy and rod monochromacy. Neutral points are indicated by the white region in the spectrum for the three types of dichromat (reproduced by permission from Sharpe et al. 2000 (Copyright 2000)).
anomalous pigment, severe and mild. The terms severe and mild reflect how much the spectral absorbance of the anomalous pigment has been altered with respect to one of the standard photopigments.

In the last few years, the techniques of molecular genetics have been used to try and determine the basis of both normal colour vision and colour blindness. Although initially these studies seemed to confirm the idea of the three standard pigment classes, more detailed studies in the last few years have shown that each of the single pigment classes is actually made up of a number of pigments with slightly different absorption spectra. Moreover, the anomalous pigments seem not to occur at one or two set positions in the spectrum, but form a continuum between the spectral absorbances of the normal pigments. Thus, what has always been called normal colour vision may just be the most commonly occurring forms of a variety of different trichromatic forms that occur naturally in the human population. In addition, the possibility must be considered that some humans may have more than three types of pigment in their eye. For example, some women may have four cone pigments in their retina and can use them in a tetrachromatic colour vision system (Jordan & Mollon, 1993).

There is a striking sex difference in the occurrence of colour blindness: around 8% of men but only 0.5% of women seem to have abnormal colour vision. The most common causes of inherited colour blindness are changes to the red or green pigments; loss of the blue cone pigments is very rare. However, the most frequently occurring form of acquired colour blindness is due to loss or damage to the blue cones, as the blue cones are extremely sensitive to high light intensities or oxygen deprivation. From the study of the inheritance of colour blindness within families, it was possible to deduce that the inheritance of the green and red pigments is sex-linked, but that of the blue and rod pigments is not (autosomal inheritance). However, until very recently the nature of the genes that code for these pigments, and the changes that lead to the loss or modification of the pigments,
could only be speculated upon. The application of the techniques of molecular genetics in the past few years is beginning to reveal the genetic basis of human colour vision and colour blindness.

All visual pigments are composed of retinal (an aldehyde derivative of vitamin A) and a protein called opsin. It is the opsin that varies in different visual pigments, and it is the opsin’s structure that determines where in the spectrum the attached retinal chromophore absorbs light. In 1986, Jeremy Nathans working at Stanford published the first definitive study of the location and nature of the genes that code for the proteins of visual pigments (Nathans et al., 1986a). Nathans found that the gene for the blue pigment is located on chromosome 7 and the gene for the rod pigment is on chromosome 3. The green and red genes are arranged in a head-to-tail array on the X-chromosome. These genes show a 97% homology in their sequence, but only a 40% homology with the blue gene. The regions upstream of the green gene on the X-chromosome are also very similar to each other. As a result it is possible for the sequence just prior to the red gene on one X-chromosome to become paired with the sequences just prior to the green pigment gene on the other X-chromosome during meiosis (see Figure 3.5(a)) (Nathans et al., 1986b). As a result, one of the X-chromosomes carries a red gene but no green gene, and the other X-chromosome carries two green genes and one red gene. A human male with the former X-chromosome lacks a green pigment and will be dichromatic. A human male with the latter X-chromosome is a ‘normal’ trichromat. Additional green genes could be added by similar unequal intergenic recombinations. The number of green genes on a single human X-chromosome has been reported to vary between one and six, with two being the commonest complement. A research group led by Jim Bowmaker in London and John Mollon in Cambridge reported the same pattern of multiple green pigment genes in seven other Old World primate species, including the chimpanzee (Ibbotson et al., 1992; Dulai et al., 1994). However, colour blindness is extremely rare in Old World primate species other than man (e.g. Onishi et al., 1999), suggesting a strong selective pressure against a deviation from this standard trichromacy. It is less clear whether a similar deletion can remove the red gene (Nathans, 1999).

Anomalous trichromats and some dichromats possess a ‘hybrid’ or ‘chimeric’ gene in addition to, or instead of, one of their red or green genes (Nathans et al., 1986b). This hybrid gene is composed of part of a red gene and part of a green gene. So, even if the red gene cannot be deleted, dichromats lacking the red pigment may be produced when the red gene is replaced by a hybrid gene, which is either inactive or codes for a green pigment. Hybrid genes can be formed by unequal intragenic recombination (Figure 3.5B). The green and red genes that code for these opsins are composed of six exons (the sequences of DNA which code for the opsin), separated by five introns (nonsense sequences). The differences between the two genes are confined to
exons 2–5. So, for example, a hybrid gene composed of green exons 1–5 and red exon 6, will produce a green pigment (Merbs & Nathans, 1992a). If a male has such a hybrid gene instead of a red gene, then he will be a dichromat lacking a red pigment. If a hybrid gene has the substitution of exons 2, 3 or 4, then the spectral peak is shifted by 2–5 nm. Substitution of exon 5 shifts the peak absorbance by 15–21 nm. A trichromat with an anomalous pigment has one or more green genes and a hybrid gene with substitutions that seem to include one or more of the exons 2, 3, 4 or 5. A trichromat with an anomalous green pigment instead of the green pigment has a red gene, a hybrid gene and one or more green genes. An anomalous trichromat may then lack a green pigment, but possess a green gene. The reverse situation is also found where the green gene is expressed and the hybrid gene is not. In this case the subject has normal trichromatic colour vision.

This suggests that only genes in specific positions in the array of pigment genes on the X-chromosome will be expressed. The rest

---

**Figure 3.5.** (a) This diagram is a schematic of how inter-genic recombination of the red and green genes can lead to deletion of the green gene. The red gene is shown as a black arrow and the green gene is shown as a white arrow. The wavy lines denote the rest of the X-chromosome. During meiosis, the X-chromosomes line up and exchange genetic material. Correct alignment means that they swap versions of the gene on one X-chromosome for the version on the other. At the end of the process, the chromosomes have the same numbers of genes as before. If the two X-chromosomes do not line up correctly, then unequal exchange of genetic material can occur. In the case of the red and green genes, unequal crossover between misplaced red and green genes can result in the deletion of a green gene from one X-chromosome and the addition of the gene to the other X-chromosome.

(b) Sometimes the misalignment of the X-chromosomes means that the parts of genes, and not whole genes, are exchanged (intra-genic recombination). In this case, crossing over within the misplaced red and green genes can produce hybrid genes. These hybrid genes can either be inactive, and so do not produce a functional cone pigment, or they can produce a pigment with a spectral absorbance different from that of the red and green pigments.
remain inactive. Further evidence for this hypothesis comes from the work of Samir Deeb and his colleagues at Washington University (Winderickx et al., 1992). It is possible to differentiate between some green genes on the basis of a silent substitution in exon 5. A silent substitution means that there is a change in the sequence coding for an amino acid but the new amino acid is homologous with the original and so the absorbance spectrum of the resultant photopigment is unchanged. During the process of the transcription of a protein from a gene, a messenger RNA (mRNA) copy of the gene is produced. Therefore, if a gene is active, one should be able to find these mRNA copies in a cell. If the gene is inactive, the mRNA copies will be absent. Deeb and his colleagues analysed retinal cells for mRNA from subjects who have two different forms of the green gene on an X-chromosome. They found only one type of the mRNA in each subject, suggesting that only one of the two forms of the gene is expressed. The expression of only some of the photopigment genes on an X-chromosome seems to be due to a genetic sequence that controls expression of the green and red genes and allows transcription of genes in only certain positions of the gene array on the X-chromosome. This is called the locus control region (LCR). Each opsin has a promoter region that lies next to exon 1. This promoter region is like an on/off switch for the gene. In each cone cell, the LCR, which lies upstream of the opsin genes on the X-chromosome, bends around to activate the promoter region of one of the opsin genes (Mollon, 1999). The other opsin genes on the X-chromosome remain inactive. So, in each cone cell, only one of the pigment genes is active and only one of the cone pigments is expressed, no matter how many red or green genes are present on the X-chromosome. The probability of the LCR activating a particular gene is linked to the distance between them (Smallwood et al., 2002). The closer the gene and its promoter region are to the LCR, then the more likely they are to be switched on (Figure 3.6). Thus, the first gene in an array is most likely to be activated, the second less likely, and the third and fourth genes (if they are present) are extremely unlikely to be activated. So, in an individual with the
‘normal’ complement of one red gene and one green gene, the red gene is closer and more likely to be activated than the green gene (hence the 2 : 1 ratio of red to green cones in the retina). So, if a hybrid gene is present on the X-chromosome, its position will determine whether it will be expressed or not. If it is one of the first two pigment genes, it is likely to be expressed, if it is further along the gene array it will not.

Failure to express both the red and green pigments is very rare, affecting approximately 1 person in 100 000. Nathans and his colleagues examined this problem in 1989 and they suggest that this deficit could arise in two ways (Nathans et al., 1989). One pathway is a two-step sequence. In the first step unequal intergenic recombination deletes the green. In the second step either a point mutation inactivates the red gene or it is deleted by recombination. The second pathway leads to an immediate loss of green and red gene function by deletion of the LCR.

As a final codicil to Dalton’s story, his eyes were preserved in a glass jar, from which they must have observed the passing years with little favour. However, they have allowed us to solve a final riddle. Based on his description of his colour vision, Dalton has been classified as a protanope. However, more detailed examination of these accounts has shown that deuteranopia was an equally plausible explanation. To finally determine Dalton’s colour vision, several samples of DNA were extracted from the dried-up remains of Dalton’s peripheral retina (Hunt et al., 1995). From these samples, it was possible to partially sequence and determine the presence of the red opsin gene. The green gene seemed to be absent, suggesting that Dalton was a deuteranope. More than 200 years after Dalton first described his condition, science has finally been able to provide an explanation of why his perception differed from normal human colour vision.

The blue cone pigment

Studies on the inheritance of deficits in the blue cone pigment in families have shown that they are inherited autosomally. Whereas red-green colour blindness is inherited recessively, the inheritance of the blue pigment deficit is dominant. Nathans’ study in 1986 showed that the gene for the blue pigment is on chromosome 7, but details of the genetic deficits that caused blue blindness remained obscure. As the blue gene is not adjacent to a similar gene or stretch of genetic material on chromosome 7, it was unlikely to be lost or altered by an unequal crossover event. In 1992, Nathans and his colleagues showed that loss of the blue cone pigment can be caused by mutations at one of two sites in the blue gene (Weitz et al., 1992). The mutations cause the substitution of a positively charged amino acid for a non-polar amino acid in part of the transmembrane section of the opsin. It is believed that the effect of this non-homologous substitution is to disrupt the folding, processing or stability of the protein.
Rhodopsin and retinitis pigmentosa

The rods mediate low intensity, night vision. Other than at the very centre of the retina (the fovea), the rods dominate vision. As we have seen, there are only 6 million cones in the retina, but there are 120 million rods. Thus, changes in rod function have a profound effect on visual perception and lead to problems such as night blindness and retinitis pigmentosa (RP). In night blindness a person’s sensitivity to light is greatly reduced, so in low light conditions that person is largely blind. RP is a far more serious condition, in which progressive degeneration of the rods and the retina eventually leads to complete blindness. It is the most common form of retinopathy (retinal disease) and currently afflicts up to 1.5 million people worldwide. Pathological changes associated with the disease include the progressive death of the rods, and the concomitant development of night blindness (nyctalopia). This form of night blindness differs from that described above, where the sensitivity of the receptors is reduced but they remain alive. In some cases the progression of RP is rapid, and night blindness develops within the first decade of life. In other instances, such symptoms may not manifest until the fifth decade or beyond. The death of the rods is followed by more extensive pathological changes in the retina. The cone cells begin to die off and there is a gradual constriction of the patient’s visual fields. The retina visibly thins and the blood vessels supplying the retinal tissue begin to attenuate. Gradually, black pigmentary deposits build up in the neural retina, the result of damage to the photoreceptor and pigment and epithelium layers, which leads to the migration of pigment-laden cells into the outer and inner retinal layers. Many patients eventually lose all their sight.

Inheritance of these diseases is generally autosomal dominant, and is linked to a mutation of the gene that codes for the opsin component of rhodopsin. Work by Nathans’ group in 1991 and by Daniel Oprian’s group at Brandeis University, Massachusetts in 1994 have shown that a number of mutations can occur in the rhodopsin gene, which lead to changes in the opsin structure (Sung et al., 1991; Rao et al., 1994). Depending upon the site of the mutation, rhodopsin can be altered in one of two possible ways. The structural changes can disrupt the folding of the protein, rendering it inactive. The change in the rhodopsin protein structure also prevents the transport and metabolisation of the inactive rhodopsin, and it accumulates in the rods leading to cell death and to the retinal degeneration associated with retinitis pigmentosa. Alternatively, the mutations can prevent attachment of retinal to the mutant opsin. Under normal circumstances, when a rhodopsin molecule absorbs a photon of light, it splits into retinal and an active form of the opsin, which indirectly alters the cell’s membrane potential and so signals the presence of
light. In the absence of light, most of the opsin molecules are bound to retinal, but there are a few unbound inactive opsin molecules. Oprian suggests that, in retinitis pigmentosa, the mutant opsin cannot bind to retinal and is permanently in the active state. This has the effect of stimulating the cell permanently, which leads to cell death and ultimately to retinal degeneration. In congenital night blindness the mutation is similar: the opsin is permanently in the active state, but it can bind retinal. When the opsin has bound retinal, it does not stimulate the cell, it is only the few unbound opsin molecules that stimulate the cell. The effect of this is to saturate the cell’s response to light partially or fully, but unlike the situation in retinitis pigmentosa this does not kill the cell.

**Better colour vision in women?**

The existence in humans of multiple pigments in the red–green region of the spectrum raises interesting questions about the colour vision of human females. With two X-chromosomes, they have two loci for the green and red genes. If one of the loci is occupied by a hybrid gene, the female will have four pigments in her retina, and be potentially tetrachromatic. If a green locus and a red locus are occupied by different hybrid genes, she is potentially pentachromatic. Three prerequisites must first be met before a woman with an additional pigment can use it as a functional colour vision system. Firstly, the anomalous pigment and its ‘normal’ counterpart must be expressed in a separate cone class and not mixed together. Secondly, the nervous system must be able to discriminate between the signals from the cones with the normal pigments and those with an anomalous pigment or pigments. This may not be easy, the separation between the peak spectral absorbances of the normal pigment and that of its anomalous form can be quite small, a matter of a few nm. Thirdly, the nervous system must be flexible enough to incorporate the signals from an additional cone class into a functional tetrachromatic colour system.

There is a precedence in primates for these three prerequisites. In most New World monkeys there is only a single gene locus on the X-chromosome for a visual pigment in the red–green region of the spectrum (Tovée, 1993). In a given species, there are three possible versions of the gene (*alleles*) which can occur at this locus, and each codes for a cone pigment with a slightly different spectral absorbance. A male, having one X-chromosome, will have one copy of the gene and so will always be dichromatic. However, a female with two X-chromosomes can either have the same allele at both loci, and so have only a single pigment in the middle- to long-wavelength region, or have two different alleles and so have two pigments in this region. Early in embryonic development, one of the two X-chromosomes in a cell is inactivated (a process called *lyonisation*) and this inactivation is preserved in all this cell’s descendants. This
process is random, and so roughly 50% of the cells will have the paternal X-chromosome active and the maternal X-chromosome inactive. In the other 50% of the cells the reverse situation will be true. Lyonisation means that, in a female, the products of the two X-chromosomes will be expressed in different cells. So, in heterozygous females, the two different pigments will be expressed in different cells and can be used in a trichromatic visual system. A study by John Mollon, Jim Bowmaker and myself (Tovee, Bowmaker & Mollon, 1992) at Cambridge has shown that, in the New World monkey, the common marmoset, the separation between the spectral absorbances of two of the pigments in the red–green range is only 6 nm. However, in the marmosets signals from cones containing these pigments can be distinguished and used in a colour opponent mechanism. Moreover, their nervous system is flexible enough to operate either as a di- or a trichromatic system, depending on the number of pigments present in the retina. It is therefore possible that the human nervous system might be sensitive and flexible enough to differentiate between, and use, four cone pigments in a tetrachromatic visual system. The additional pigment would be expressed in a separate cone class due to lyonisation. Experimental evidence for this hypothesis comes from further work at Cambridge, reported by Gabi Jordan and John Mollon on a small population of putative female tetrachromats (Jordan & Mollon, 1993). Behavioural tests of their colour vision suggest that at least one of them is able to use the additional pigment in a functional tetrachromatic system.

Three pigments in normal human colour vision?

Once it was clear that hybrid genes with different proportions of the red and green genes produced pigments with different spectral absorbances, several research groups began trying to find which changes were important. They took as their starting point the hypothesis that the spectral absorbance of cone pigments in the red–green region is based on the net effect of differences in the number and position of hydroxyl groups in the vicinity of the retinal chromophore. Therefore, researchers have concentrated on changes in the gene sequence that would produce these changes. In the early 1990s Gerry Jacobs and Jay and Maureen Neitz at Santa Barbara looked for differences in the gene sequences of cone pigments taken from humans and from two New World monkey species (the squirrel monkey and the saddle-backed tamarin). They interpreted their results as suggesting that changes that caused the substitution of a single non-polar amino acid for a hydroxyl-bearing amino acid at three sites in the opsin (positions 180, 277, 285) account for all the variation in the spectral sensitivity of cone pigments in the red–green region (Figure 3.7) (Neitz et al., 1991). It is suggested that the three substitutions are linearly additive in the number of nanometres by
which they shift the peak sensitivity of the pigment. The substitution at 180 (located in exon 3) has been reported to shift the peak by 5.3 nm, that at 277 (in exon 5) by 9.5 nm and that at 285 (in exon 5) by 15.5 nm. However, Nathans and his colleagues have been able to construct both normal and hybrid genes from complementary DNA clones for the green and red genes and express them in cultured mammalian kidney cells (Merbs & Nathans, 1992a,b). The opsins produced by these genes are then combined with retinal to form photopigments. This technique allows the exon composition of the artificially produced genes to be controlled. For example, a gene can be produced with exon 1 from the green gene and exons 2–6 from the red gene. The production of such hybrid genes allows the effect of individual exons on the spectral absorbance of the resultant pigment to be gauged. The results show that, in addition to exons 3 and 5, altering exons 2 and 4 also change the spectral peak of the resultant pigment. Consistent with this result is the finding in sequencing studies on other Old World primate species by the Bowmaker and Mollon group that suggests that non-homologous substitutions at a minimum of two more amino acid sites may play a part in determining a pigment’s spectral absorbance (Ibbotson et al., 1992). For example, in all the ten primate species examined so far, there is a non-hydroxyl-bearing amino acid at site 233 (exon 4) in pigments with a spectral peak above 560 nm. In the eight species with pigments whose spectral peak is below 540 nm, there is a hydroxyl-bearing amino acid at 233. Moreover, the presence or absence of a hydroxyl-bearing amino acid at 233 alters the effect of non-homologous
substitutions at 180. This suggested that substitutions at certain sites, such as 233, may play a modulating role on the more obvious effects of substitutions at 180, 277 and 285. To investigate this hypothesis, Oprean’s group constructed a total of 28 artificial visual pigments by hybridisation and cloning techniques (Asjeno et al., 1994). The red and green pigments are composed of 364 amino acids, but differ in only 15 of these amino acids. By changing each of these amino acids in turn in the artificial visual pigments, it was possible to determine what effect an individual amino acid had on the absorbance spectra of the pigment. They found that a total of seven amino acids control the spectral position of the visual pigment and not just three as suggested by Jacobs and the Neitzs.

A common test for red–green colour blindness is the Rayleigh match, in which an observer is asked to find the ratio of red to green light that matches an orange light. In the later 1980s Jacobs and the Neitzs published results of Rayleigh matches from populations of Caucasian males, which they claimed showed a bimodality (Neitz & Jacobs, 1986). This suggested the possibility of multiple pigment types within what was conventionally classified as the red pigment. The bimodality is consistent with the results of a microspectrophotometry (MSP) study of human cone pigments by the Bowmaker and Mollon group (Dartnall et al., 1983). In MSP, a beam of monochromatic light is passed through a single, isolated cone outer segment. By varying the wavelength of the light, and measuring how much is absorbed, it is possible to calculate the spectral absorbance of the cone pigment. The MSP study on the cones from seven normal trichromats showed two distinct red cone pigments. Despite this evidence, little notice was taken of this idea. The results were dismissed as experimental artefacts or due to skewed samples. However, in the last few years the idea has been vindicated. Studies of human populations show a common non-homologous substitution in the red pigment at site 180, where the amino acid can be either serine or alanine, and this difference can be correlated with the bimodal Rayleigh distribution (Merbs & Nathans, 1992b; Winderickx et al., 1992). The substitution of serine for alanine at position 180 can shift the peak absorbance of the pigment by 4–6 nm. In addition to two red pigments, it has been reported that there may also be two green pigments, which show the same substitution difference at site 180 (Neitz et al., 1993).

Given that substitutions at several different amino acid sites can alter the peak absorbance of the green and red opsins, it is possible that there are many versions of the green and red pigments naturally occurring in the human population, which may be produced by unequal intragenic crossover or by point mutation. The so-called ‘anomalous’ pigments would then be part of a naturally occurring range of pigments in the green and red region and what has been called normal colour vision may just be the most commonly occurring forms of a variety of different trichromatic forms that occur in the human population.
The evolution of primate colour vision

All Old World primates and the New World Howler monkeys are ‘routine’ trichromats (Surridge, Osorio & Mundy, 2003). All other New World monkeys, like the marmoset discussed above, are polymorphic for colour vision. Within a single species, there are three types of trichromat and three types of dichromat. An obvious question to ask is whether the common ancestors of Old and New World monkeys were trichromatic, and if it was then lost in most of the New World monkey species. DNA analysis of the sequences associated with the M and L genes in Howler monkeys (such as the promoter regions and LCR) suggest that the evolution of trichromacy in Howler monkeys is independent and more recent than that in Old World monkeys (Surridge et al., 2003).

It is believed the second gene site on the X-chromosome in routine trichromacy arose through gene duplication around 35–40 million years ago, which is roughly the time that the Old World and New World primate lineages separated. The evolution of trichromacy may have arisen in one of two ways. Firstly, a polymorphism of the pigment in the red–green region, such as now exists in the New World monkeys, may have arisen through point mutation. Then an unequal recombination event in a heterozygous female could have placed two of the different alleles on the same X-chromosome. This recombination event did not include the LCR for the duplicated gene. So, although there would be two genes on the X-chromosome, there would be only one LCR, the result of which is that only one gene will be activated on each X-chromosome in each cell. The similarity of the red and green genes in the Howler monkeys to the different alleles for red–green pigments in other New World species suggests that this is how trichromacy in the Howler Monkey was produced (Hunt et al., 1998). For Old World primates the sequence of events was slightly different. Gene duplication seems to have arisen in a dichromatic ancestor, with the result that a gene for a single pigment in the red–green range was duplicated. Then, over time, the DNA sequences of the two genes diverged, until the two genes coded for two spectrally separate pigments (Dulai et al., 1999).

What is trichromacy for?

The advantage of trichromacy for primates is usually couched in terms of detecting ripe fruit against a background of leaves (Regan et al., 2001). A possible problem with this hypothesis is that, although Old world primates (including humans) exhibit a uniform trichromacy with the three cone pigments in the same three spectral positions, not all of them eat ripe fruit. The colobines are a major
subgroup in Africa and Asia, and their primary food source is leaves and unripe fruit (Surridge et al., 2003). This has led to the suggestion that primate trichromacy may have developed primarily to detect young nutritious leaves against a background of older, tougher leaves (Lucas et al., 1998). However, as the enhancement of colour discrimination in the red–green range would benefit both ripe fruit detection as well as foraging for leaves, it is not possible to say whether primate trichromacy developed for a fruit or a leaf diet (Sumner & Mollon, 2000). Additionally, it may be that the primary factor in determining the number and spacing of the visual pigments in the spectrum is the colour of background (which is likely to be mature leaves in either case), rather than the colour of the targets (whether it is fruit or young leaves) (Dominy & Lucas, 2001). If the background is the primary determinant, then this could explain why all the Old World primates have the same trichromacy despite differences in diet.

**Summary of key points**

1. The human retina contains three classes of cone: red, green and blue. For this reason, colour vision is said to be trichromatic.
2. The three cone classes are part of three opponent mechanisms. The first takes the difference in the responses of the red and green cones (R/G), the second takes the difference between the responses between the blue cones and the sum of the red and green cones (B/Y) and the third is an achromatic mechanism that detects differences in luminance.
3. Cone pigments are composed of retinal (an aldehyde derivative of vitamin A) and a protein called an opsin. The gene for the blue pigment’s opsin is on chromosome 7, and the red and green pigments’ opsins are on the X-chromosome.
4. Loss or modification of one or more of these pigments leads to an alteration in colour vision, a condition often called colour blindness. Loss or modification of the red and green cone pigments is the most common form of colour blindness, as the genes for the two pigments lie side by side and are extremely susceptible to unequal crossover during the exchange of corresponding genetic material, which occurs during meiosis between the two chromosomes that make up a chromosome pair.
5. Unequal crossover can lead to the deletion of the red or green gene from one X-chromosome and its addition to another. A male with the former X-chromosome will be a dichromat lacking the pigment whose gene has been deleted. Unequal crossover can also cause a mixing or blending of the red and green genes to form hybrid or chimaeric genes. If these genes are non-functional, they lead to dichromacy. If they are functional, they will often produce a pigment with a spectral absorbance intermediate
between that of the red and green pigments. This is called an anomalous pigment.

(6) The gene for the opsin of the rod pigment is on chromosome 3, and mutations of this gene can lead to defects of rod function such as night blindness (nyctolopia) or retinitis pigmentosa.

(7) The red and green pigments are composed of 364 amino acids, but differ in only 15 of these amino acids. A total of seven of these amino acids control the spectral position of the visual pigment in the red–green range.

(8) There may be many versions of the green and red pigments naturally occurring in the human population, and the so-called ‘anomalous’ pigments would then be part of a naturally occurring range of pigments in the green and red region, and what has been called normal colour vision may just be the most commonly occurring forms of a variety of different trichromatic forms that occur in the human population.
The organisation of the visual system

Making a complex process seem simple

Vision is the primary sensory modality in primates such as ourselves, and this is reflected in the complexity of the visual system and the extent of the cerebral cortex used for the analysis of visual information. On the basis of anatomical, physiological and behavioural studies, it is believed that at least 32 separate cortical areas are involved with the processing of visual processing in the macaque monkey (Van Essen et al., 1992). Twenty-five of these areas are primarily visual in function; the remaining seven are also implicated in other functions such as polysensory integration or visually guided motor control. These visual areas occupy about half of the 100 cm² area of each of the monkey’s cortical hemispheres. Two of the areas, V1 and V2, each occupies more than 10 cm² of the cortical surface, but most visual areas occupy less than a tenth of this area. Comparatively little is known of the functional anatomy of the human visual cortex, but it seems to be at least as complex as that of the monkey (Kaas, 1992; Sereno et al., 1995). Fortunately, it is possible to simplify this picture by concentrating on the key visual areas and looking at their functional organisation.

As one moves up the visual system, from the retina to the lateral geniculate nucleus and then on to successive cortical areas, visual neurons become responsive to more and more complex stimuli. For example, in monkeys, in the first cortical visual area (called primary visual cortex or V1) there are neurons responsive to simple lines of different orientations, whereas in one of the higher visual areas (inferior temporal cortex) the neurons respond to complex stimuli, such as faces. However, the visual system is not organised in just a serial, hierarchical pathway. Different aspects of a stimulus (such as its shape, colour and motion) are analysed in separate, parallel pathways. These pathways are usually divided into two broad categories; ‘what’ and ‘where’ pathways. The ‘what’ pathway deals with information about the stimulus features (such as shape and colour) and the
identity of an object, and can be sub-divided into two further pathways: colour and shape. The ‘where’ pathway deals with spatial information about an object, and is usually sub-divided into motion and form derived from motion. This chapter firstly will describe the basic anatomy of the visual system and its connections, and then it will explain how this machinery functions to produce vision.

The retina

As stated above, the visual system can be divided into two or more separate pathways. This separation starts to become evident at the level of the retina. There are several forms of ganglion cell in the primate retina, of which two types (M- and P-cells) constitute about 90% of the cells. The M-class (sometimes also called A- or Pα-cells) account for 10%, and the other 80% are accounted for by the P-class (sometimes called B- or Pβ-cells). The subsequent pathways are often referred to as the M or P pathways. The remaining 10% of the retinal ganglion cells consist of at least eight different types (Rodieck, 1988). The P-cells are selective for wavelength and high spatial frequencies and have slow sustained (tonic) responses, whereas the M-cells are not wavelength sensitive but are sensitive to low spatial frequencies, have transient (phasic) responses and have faster conduction velocities. At any particular eccentricity, the dendritic field of the M-cell is three times larger than that of the P-cell. These differences in the response properties of the neurons shape the functions of the subsequent visual areas.

The lateral geniculate nucleus (LGN)

The axons of all of the ganglion cells come together to form the optic nerve, which passes out of the eye through the optic disc and projects to the dorsal lateral geniculate nucleus (LGN) (Figure 4.1). The nerves from the two eyes join together before they reach the LGN to form the optic chiasm. At this point, axons from ganglion cells from the inner halves of the retina (the nasal sides) cross over and then continue on to the LGN. As the axons from the nasal halves of the retinas cross to the other side of the brain, each hemisphere of the brain receives information from the opposite side of the visual scene. So, if you look straight ahead, the right hemisphere receives information from the left half of the visual field, and the left hemisphere receives information from the right side of the visual field.

The LGN is a folded sheet of neurons, about the size of a credit card, but about three times as thick and found on each side of the brain (Figure 4.2). It consists of six layers. Each layer receives input from only one eye: layers 2, 3 and 5 from the eye on the same side as the LGN, and layers 1, 4 and 6 from the eye on the opposite side. The topographic arrangement of the ganglion-receptive fields is maintained in the LGN,
so that each layer contains a complete map of the retina. The cells in layers 1 and 2 contain larger cell bodies than those in the remaining four layers. The inner two layers are thus called magnocellular or M-layers and the outer four layers are called parvocellular or P-layers. Neurons in the M-layers receive input from M ganglion cells and neurons in the P-layers receive input from type P ganglion cells. The LGN neurons show the same response characteristics as the cells from which they receive input.

**The primary visual cortex (V1)**

The LGN neurons mainly project to the primary visual cortex (also known as the striate cortex or V1). This is the first cortical visual
area and consists of six principal layers (and several sub-layers) arranged in bands parallel to the surface of the cortex. The axons from the LGN terminate on cortical neurons in layer 4 (Figure 4.3). The P-layer neurons send their axons to neurons in the deeper part of this layer (sub-layer 4Cβ), which in turn send their axons to layers 2 and 3 and from there to visual area 2 (V2). The M-layer neurons send their axons to neurons in sub-layer 4Cx, and the information is then relayed to layer 4B and then to V2 and to visual area 5 (V5) (Figure 4.4). Cells in layer 4B are orientation selective and most show selectivity for the direction of movement. Some of these neurons are binocular (require stimulation from both eyes) and show sensitivity to retinal disparity (the difference in the relative position of the stimuli in the visual field of two eyes) (Poggio & Fischer, 1977). But these cells do not unambiguously signal stereoscopic depth, although combining the outputs of many V1 neurons could potentially do so (Cumming & Parker, 1997). The disparity signals may also be important for the rapid involuntary control of vergence eye movements (eye movements...
that bring the images on the two foveae into register) (Masson, Busettini & Miles, 1997).

The P-pathway splits to produce two new pathways in the upper layers of V1. One pathway seems to deal primarily with colour and this is called the P–B pathway. Neurons in the second pathway are sensitive to features such as the orientation of the stimulus and seem to mediate high acuity perception. This pathway is called the P–I pathway. The M-pathway is dominated by a single source, but the P–I and P–B streams receive inputs from a number of different sources. At the top of the cortical hierarchy, the M-pathway leads primarily to the posterior parietal cortex, which processes spatial and motion information. The P–I and P–B pathways project to the inferior temporal cortex, which mediates pattern and object recognition. K or W cells form koniocellular layers between the M and P layers.

Abbreviations: RGC, retinal ganglion cell; SB, small bistratified ganglion cell; LGN, lateral geniculate nucleus; V1 to V5, visual areas 1 to 5; PPC, posterior parietal cortex; IT, inferior temporal cortex (redrawn from Van Essen & Deyoe, 1995).
opponency and respond well to achromatic luminance contrast borders. This suggests that the colour-coded P-cell input is pooled in such a way that colour contrast can be used to identify borders, but that the information about the colours forming the border is lost. The neurons in the inter-blob region are part of the P–I pathway.

The blob cells are not orientation selective, but are either colour or brightness selective. These cells are part of the P–B pathway. The P–B pathway thus seems to carry information complementary to the information carried by the P–I pathway. The colour-opponent blob cells receive input from the colour-opponent P cells in the LGN, although they differ in that their receptive field centres are larger and their colour coding is double opponent (they give opposite responses to different parts of the spectrum in the different parts of their receptive field. For example, the centre might give an on-response to green and an off-response to red, and the opposite set of responses in the surround). A possible means of deriving double opponent cells from opponent cells is shown in Figure 4.5.

The blob and interblob systems thus seem to work in different but complementary ways. The blob cells are colour coded, excited by colours in one region of the spectrum and inhibited by others, and not selective for stimulus orientation. Inter-blob cells are selective for stimulus orientation, but mostly are not colour selective, responding to a line or edge of the correct orientation regardless of its colour.

**Visual area 2 (V2)**

The main target of V1 is V2. Staining for cytochrome oxidase in this area does not reveal a pattern of blobs or inter-blobs, but instead a pattern of stripes running perpendicular to the border between V1 and V2, and which extends over the entire 8- to 10-mm width of V2. There seem to be three types of stripe. There are two darkly staining stripes, one **thick** and one **thin**, separated by more lightly staining **inter-stripes**. (sometimes called **pale stripes**). The neurons in layer 4B of V1 (part of the M pathway) project to the thick stripes (Figure 4.4).
Neurons in the thick stripes show similar response properties to the neurons in layer 4B. They are orientation and movement selective, and many show sensitivity to retinal disparity. The neurons in the blobs project to the thin stripes, and neurons in the thin stripes are not orientation selective, but more than half are colour sensitive (mostly double opponent). The inter-blobs project to the inter-stripes, and neurons in this region are orientation selective, but not direction selective nor do they show colour selectivity. The organisation of V1 is retinotopic, that is the visual field of the retina is mapped on to the surface of the cortex of V1. In V2 there seem to be three separate visual maps (Roe & Ts’o, 1995). Within the thick stripes there is a visual orientation map; within the thin stripes there is a colour map; and within inter-stripes a disparity map. Adjacent stripes are responsive to the same region of visual field. So there are three, interleaved visual maps in V2, each representing a different aspect of the visual stimulus.

Thus, the M pathway projects from layer 4B of V1 to the thick stripes of V2. The P–B pathway projects from the blobs of V1 to the thin stripes of V2, and the P–I pathway projects from the inter-blob region to the inter-stripes.

**Visual area 4 (V4)**

Both sub-divisions of the P pathway, the thin stripes (colour) and inter-stripes (form) project to visual area 4 (V4). V4 and the other visual areas upstream of V2 all stain relatively homogeneously for the cytochrome oxidase enzyme, and no alternative marker is yet known. However, the continued separation of the two sub-divisions of the P-pathway in V4 can be inferred from patterns of connectivity. It is possible to trace connections by using substances, such as peroxidase enzymes or dyes, which will be absorbed by the neurons and transported up or down their axons. This method reveals the neuronal connections of the specific piece of cortex in which the tracer is deposited. When tracer is deposited in different parts of V4, the backfilled neurons in V2 tend to occur in one of two distinct patterns: either they are largely restricted to the thin stripes or they are largely restricted to the interstripes (Shipp & Zeki, 1985; Zeki & Shipp, 1989). This suggests that the separation of the two sub-divisions of the P pathway continues in V4. V4 projects primarily to posterior inferior temporal cortex (PIT). Retrograde tracing from this area also shows a patchy, modular organisation of the cells in V4 (Felleman, Xiao & McClendon, 1997), a finding supported by a combined electrophysiological and optical imaging study (Ghose & Ts’o, 1997). This suggests that the anatomical separation of the two sub-divisions of the P pathway continues in V4.

V4 has received a great deal of attention as the putative site for the development of colour constancy. Primarily this region has been explored functionally in two ways in the macaque: lesions and
electrophysiology. The results of lesion studies suggest a dissociation between hue discrimination and colour constancy. While monkeys with V4 lesions are not impaired significantly on hue discrimination tasks (Wild et al., 1985; Walsh et al., 1993; Heywood, Gadotti & Cowey, 1992), they are impaired on colour constancy tasks (Wild et al., 1985; Walsh et al., 1993). Yet the most striking deficit caused by V4 ablation in monkeys is in pattern discrimination. This fact emphasises that V4 cannot be devoted solely to analysing colour, and also suggests that V4 cannot be the straightforward homologue of the putative colour centre damaged in human cerebral achromatopsia, since such achromatopsics are unimpaired in pattern discrimination but are severely impaired on hue discrimination tasks.

Visual areas 3 (V3) and 5 (V5)

The M pathway projects to visual areas 3 (V3) and 5 (V5), both directly from layer 4B of V1 and through the thick stripes of V2. Most cells in V3 are orientation selective and are believed to be concerned with processing dynamic form. V5 (also known as the middle temporal visual area or MT) is believed to process information on motion and stereoscopic depth. In monkeys, lesions of V5 cause deficits in pursuit eye movements and in discriminating the direction of motion. The M-pathway then projects to the parietal cortex. The parietal cortex seems to be important for the integration of movement and depth into a representation of space. Damage to this region in humans causes a condition called Balint’s syndrome. This has three main symptoms. Firstly, a difficulty in reaching for objects under visual guidance (optic ataxia). Secondly, subjects display a deficit in visual scanning (ocular apraxia). A person may perceive an object normally, but will be unable to maintain fixation. His or her eyes will wander. He or she will not be able to make a systematic scan of a scene and will not be able to perceive the location of the objects seen. Finally, the subject is only able to see one object at a time in a crowded scene (simultagnosia).

The koniocellular pathway

The separation of the P- and M-pathways should not be overemphasised. There is some communication between the two pathways. If the M-layers of the LGN are inactivated by cooling, the visual responses of neurons are reduced both in V4 and in the blobs and interblobs of V1 (Nealey & Maunsell, 1994; Ferrera et al., 1994). This suggests that there is an input from the M pathway into the P pathway. Moreover, there is a third type of cell in the LGN (Figure 4.4). These small neurons are called K or W cells, and are found within the region between the M and P layers, forming what sometimes is called the koniocellular layers. They are believed to receive input from the
small bistratified ganglion cells in the retina, one of the cell types that make up the remaining 10% of retinal ganglion cells (Dacey & Lee, 1994). The koniocellular or K-system is thought to represent the ‘old’ mammalian colour system (Mollon, 1999). It carries signals from the blue–yellow opponent mechanism, the equivalent of the dichromatic mechanisms found in non-primate mammals. The additional red–green opponent system that evolved in primates is carried by the P system output from the retina, along with information on fine spatial detail. In the P system, the information on colour and detail is separated into the blobs and inter-blob regions. The K-system, which carries only colour information, projects directly to the blobs in layers 2 and 3 of V1 (Hendry & Yoshioka, 1994). Thus, the so-called P pathway actually seems to receive input from the P, M and K neurons of the LGN. In contrast, the M pathway does seem to be more segregated. For example, inactivation of the P layers of the LGN has a negligible effect on the responses of neurons in V5 (Maunsell et al., 1990).

The functional organisation

It has been proposed that, in Old World primates, visual information is processed in two broad systems: the ‘what’ system (also called the ventral system), which is concerned with the identification of an object, and the ‘where’ system (also called the dorsal system), which is concerned with the relative spatial position of an object (Mishkin et al., 1983) (Figure 4.6). Damage to the ‘what’ system, through which V1 projects to the temporal lobes, does not impair performance of visuospatial tasks, but does impair performance of object discrimination tasks. Damage to the ‘where’ system, through which the V1 projects to the parietal lobes, produces impairments on visuospatial tasks, but does not impair the performance of object discrimination tasks. The visual cortex of New World monkeys appears to be similarly organised into separate ‘what’ and ‘where’ streams (Weller, 1988). This suggests a common primate plan, which positron emission tomography (PET) scan studies suggest extends to humans (Haxby et al., 1991).

The two streams project to different prefrontal cortical areas (Wilson et al., 1993). The ‘what’ system projects to the cortex of the inferior convexity (IC) ventrolateral to the principal sulcus and the ‘where’ system projects to the dorsolateral prefrontal region (DL). The prefrontal cortex is an important region for working memory. Patricia Goldman-Rakic and her colleagues trained monkeys to carry out visual tasks, and at the same time they recorded from single neurons in either IC or DL. In the first task, the monkeys were trained to stare at a spot on a video screen while an image flashed at one of several locations on the screen and then disappeared. A few seconds later, a cue on the screen signalled the monkeys to move their gaze to where the image had been, indicating that they had remembered the
location of the image. In the second task, the location of the image remained constant but the image itself changed. The monkeys were trained to wait until the image disappeared, and then, after a delay, move their eyes to the right if they saw one image and to the left if they saw another, indicating they remembered information about an object’s features. In the first task, neurons in DL became active during the delay period, while there was no change in activity of neurons in IC. However, during the second task the pattern of activity is reversed: the neurons in IC are active during the delay and the neurons in DL remain quiet. These results suggest that IC mediates working memory for objects and DL mediates spatial working memory.

The M and P pathways have been thought to correspond roughly to these two systems (Livingstone & Hubel, 1988). The P pathway, with information about colour and shape, would seem to be ideal for the ‘what’ system which develops a representation of an object. Similarly, the M pathway, with information about motion, stereopsis and form derived from motion, would seem to be the obvious candidate for the ‘where’ system, which develops a representation of spatial relationships in the visual field. However, there is considerable communication between the two systems at all levels (Harries & Perrett, 1991).

Perception vs. action

An alternative approach has been proposed that takes more account of the output requirements of the system (Goodale & Milner, 1992; Sakata & Taira, 1994). This approach is called ‘what’ vs. ‘how’. In brief, it postulates that the ‘what’ pathway mediates the conscious recognition of objects and scenes, while the ‘how’ pathway provides visuospatial information directly into the motor systems to guide our
actions. This hypothesis has been supported both by clinical evidence and by primate electrophysiology.

Visually guided grasping was studied in a patient with Balint’s syndrome (as mentioned above, in this syndrome bilateral parietal damage causes profound disorders of spatial attention, gaze and visually guided reaching). While this patient had no difficulty in recognising common objects, her ability to pick up the objects remained impaired (Jakobsen et al., 1991). Not only did she fail to show normal scaling of the grasping movement, she also made a large number of adjustments in her grasp as she closed in on an object. These adjustments are rarely observed in normal subjects. Such studies suggest that damage to the parietal lobe can impair the ability of patients to use information about the size, shape and orientation of an object to control the hand and fingers during a grasping movement. By contrast, another patient developed profound visual-form agnosia (an inability to recognise objects) following carbon monoxide poisoning (Goodale et al., 1991). Despite her profound inability to recognise the shape, size and orientation of an object, the patient showed strikingly accurate guidance of hand and finger movements directed at the very same objects. So, despite impaired conscious visual discrimination of the objects, visual information, computed unconsciously, was made available to the action system to direct grasping actions (Milner & Goodale, 1995).

Consistent with the clinical data, single-cell recording in the monkey parietal cortex has found neurons responsive to stimuli that would be used in interacting with the environment. For example, shape-selective neurons have been reported in the intra-parietal sulcus (Taira et al., 1990; Sakata & Kusonoki, 1992; Sakata et al., 1995). Many cells in the lateral bank of the sulcus respond selectively to images of a switch, a lever, a button or a knob in a groove that the monkeys had been trained to manipulate. The hand movements and the shape of the grasp required to manipulate these objects differed (pulling the lever, pushing the button, grasping and pulling the knob, etc.), and the monkey moved its arm and shaped its hand specifically to manipulate each object before the monkey reached out. The responses of neurons in the lateral sulcus started when the monkey saw the object, and the response was reduced if the monkey performed the task in a dark room. This suggests that the neural responses might be stimulated partially by visual inputs. Additionally, there are cells in a more posterior part of the sulcus that respond to more primitive features of stimuli including the 3-D orientation of a pole or a 3-D tilt of a plane (Sakata et al., 1995; Sakata, 1996). The responses were reduced when the degree of binocular disparity is reduced, suggesting that the binocular disparity represents an important cue for these responses. These results suggest that, in the parietal cortex, there is a synthesis of action-orientated visual information with other sensory cues to produce a basis for visually directed movement (Sakata & Taira, 1994; Sakata, 1996).
A key part of the ‘what and how’ hypothesis is that the visuospatial information in the how stream is not available for conscious perception, instead it passes directly to the motor guidance systems in the brain. An intriguing finding in support of this hypothesis is the pattern of hand movements in relation to the Ebbinghaus illusion (see Figure 4.7). In this illusion, two identical central circles appear to differ in size, based on the sizes of the circles surrounding them. However, the Ebbinghaus illusion does not seem to deceive reaching behaviour (Aglioti et al., 1995). So, although the observer’s conscious perception is that the central circles are of different sizes, when they reach out to pick up the central circles (presented as discs), their hand movements are not deceived. However, this finding has been disputed (e.g. Franz, 2001) and the question of whether reaching behaviour resists visual illusions is still to be resolved.

Blindsight

Damage to V1 causes holes (scotomas) in our visual field. People with V1 damage seem to have no conscious perception of the visual stimuli presented in these scotomas. As we have seen in the preceding sections, all the main visual pathways pass through V1, and so damage to this bottleneck might be expected seriously to disrupt vision. However, some patients can respond to visual stimuli presented in their scotomas if they are required to make a forced choice to indicate stimulus parameters. Some patients are able to look towards stimuli presented in their scotomas, to localise them by pointing and to detect and discriminate movement (Cowey & Stoerig, 1991). One patient with complete cortical blindness was able to follow a large moving, striped display with his eyes, despite disclaiming any visual sensation that might explain his visual tracking. Patients can detect and discriminate flicker, orientation and wavelength. Their pupils continue to respond to changes in light level, pattern and contrast and when asked to reach for visual targets, two patients adjusted their grasp so that it matched the size and shape of the object. They could also use the meaning of words flashed in their blindfields in order to select between pairs of words subsequently presented in their intact field. This perception of visual stimuli without conscious knowledge is called blindsight.
The primary projection from the retina is to the LGN, and this exceeds a million fibres per eye. However, the retina also projects to other structures, such as the projection to the superior colliculus (SC) (approximately 100 000–150 000 fibres) (see Figure 4.8). Many of these connections transmit information about the position, size and movement of visual stimuli (Cowey & Stoerig, 1991). It is believed that these connections might mediate the residual vision found in blindsight. This hypothesis is supported by the effects of lesions of these sub-cortical pathways in monkeys. Monkeys also seem to show blindsight (Cowey & Stoerig, 1995), and so provide an animal model for investigating this phenomenon. Damage to V1 has profound effects on the subsequent visual areas. Lesions of either the SC, or the lateral pretectum (with interruption of the accessory optic system) drastically reduce the blindsight capacities of monkeys who had V1 removed (Mohler & Wurtz, 1977; Pasik & Pasik, 1982).

Doubts have been raised as to whether blindsight does actually exist. Alternative suggestions have included the possibility that the
patients' V1 was merely damaged, not destroyed, or that patients responded to light scattered from the stimulus on to the intact retina, or that experimenters employed a lax criterion for detection that was very different from the one used in the normal field. As regards to at least some of the studies, these reservations are groundless. Firstly, blindsight can be demonstrated even in patients in whom V1 or even the entire cerebral hemisphere has been removed surgically. Secondly, when a stimulus that is detected in blindsight is presented on the natural blindspot (the optic disc), it becomes undetectable, despite the fact that the optic disc normally scatters more light than the rest of the retina. Thirdly, a stimulus presented in the blind field and to which the patient is not even asked to respond can influence the response to a companion stimulus presented in the intact field. It is true that, in the case of one patient who was diagnosed as having blindsight, an MRI scan indicated the presence of some remnants of V1 which might have mediated some of the residual vision (Fendrich et al., 1992); however, there is also considerable evidence from patients studied either at autopsy, during surgery or by non-invasive scanning, who had complete loss of V1 (see Cowey & Stoerig, 1993) and an fMRI study of two blindsight patients who showed no activation of V1 (Goebel et al., 2001).

A slightly different criticism of the blindsight concept has come from recent work that suggests there may be conscious perception of movement in human subjects with damage to, or reversible inactivation of, V1 (Barbur et al., 1993; Beckers & Zeki, 1995). It has been suggested that there might be a ‘fast’ pathway that bypasses V1 and connects V5 directly with either the LGN or possibly even with the retina (Beckers & Zeki, 1995). The information conveyed by this fast pathway seems to be sufficient to mediate a conscious, coarse perception of movement, and would enable an individual to react more quickly to moving stimuli than if the information came via V1. This hypothesis is supported by a visual-evoked potential study, which suggests that to some forms of moving stimuli, V5 may become active before, or simultaneously with, V1 (Ffytche et al., 1995). This is not to suggest that V1 is not an extremely important stage in the processing of movement information. The ‘fast’ connection would represent only a few per cent of V5’s connections, and input from connections that come through V1 are required for any form of detailed movement perception and discrimination.

The fast connection hypothesis has limited support in monkey studies. Neurons in V5 continued to respond when transmission through V1 is blocked (Rodman et al., 1989a; Girad et al., 1992), but ablation of the SC largely abolished this activity (Rodman et al., 1989b). This connection is probably related to the need to integrate eye movement (in which the SC plays a role) with external motion as processed by V5. Thus, although there is evidence of connections which bypass V1 on their way to V5, there is little evidence for the suggested fast pathway. The role of V1 seems much more important in the activity of other extrastriate visual areas. Girard and Bullier
(1988) cooled V1 and tested visual responsiveness in V2, where they found that fewer than 2% of neurons responded normally. This is consistent with a similar study in V4 (Schiller & Malpeli, 1978), suggesting these areas are extremely dependent on V1 for visual information.

Summary of key points

(1) The visual system is divided into two or more streams of information, in which different aspects of visual perceptions such as movement, depth, colour and shape are processed separately.

(2) This division is first evident at the level of the retinal ganglion cells. There are two main classes of ganglion cell: the M class (which gives rise to the M pathway) and the P class (which gives rise to the P pathway).

(3) The axons of the ganglion cells form the optic nerve, which projects to the dorsal lateral geniculate nucleus (LGN). The nerves from the two eyes join before they reach the LGN to form the optic chiasm. At this point, axons from ganglion cells from the inner halves of the retina (the nasal sides) cross over and then continue on to the LGN, so each hemisphere of the brain receives information from the opposite side of the visual scene.

(4) The LGN consists of six layers. Each layer receives input from only one eye, layers 2, 3 and 5 from the eye on the same side as the LGN, and layers 1, 4 and 6 from the eye on the opposite side. The cells in layers 1 and 2 contain larger cell bodies than those in the remaining four layers. So cells in layer 1 and 2 are called magnocellular (M) and those in the other layers are called parvo-cellular (P). Neurons in the M layers receive input from M ganglion cells and neurons in the P layers receive input from type P ganglion cells.

(5) The LGN neurons mainly project to the primary visual cortex (V1). This is the first cortical visual area and consists of six principal layers (and several sub-layers) arranged in bands parallel to the surface of the cortex. The axons from the LGN terminate on cortical neurons in layer 4.

(6) The P layer neurons send their axons to neurons in the deeper part of this layer (sub-layer 4Cb), which in turn send their axons to layers 2 and 3 and from there to visual area 2 (V2).

(7) The M layer neurons send their axons to neurons in sub-layer 4Ca, and the information is then relayed to layer 4B and then to V2 and to visual area 5 (V5). Cells in layer 4B are orientation selective and most show selectivity for the direction of movement. Some of these neurons are binocular and show sensitivity to retinal disparity.

(8) The P pathway splits to produce two new pathways in the upper layers of V1. One pathway seems to deal primarily with colour and this is called the P-B pathway. Neurons in this pathway are
found in columns called *blobs*. Neurons in the second pathway are sensitive to features such as the orientation of the stimulus and seem to mediate high acuity perception. This pathway is called the P–I pathway. These neurons are found in the area surrounding the blobs (the *inter-blob region*).

(9) The main target of V1 is V2. The neurons in layer 4B of V1 (part of the M pathway) project to the *thick stripes*, and neurons in this stripe are orientation and movement sensitive. The neurons in the blobs project to the *thin stripes*, and neurons in the thin stripes are not orientation selective, but more than half are colour sensitive (mostly double opponent). The inter-blobs project to the *inter-stripes*, and neurons in this region are orientation selective, but not direction selective nor do they show colour selectivity.

(10) Both sub-divisions of the P pathway: the thick stripes (colour) and inter-stripes (form) project to *visual area 4* (V4), and the two streams remain separate in this area. Some V4 cells respond not to the wavelength of light but to its ‘colour’, a phenomenon known as colour constancy. Damage to the equivalent area in humans also impairs the ability to distinguish colour, a condition called *achromatopsia*. V4 is also important for object discrimination. V4 projects primarily to the *temporal visual cortex*, where there seems to be an integration of form and colour to give a representation of complex objects.

(11) The M pathway projects to *visual areas 3* (V3) and 5 (V5), both directly from layer 4B of V1 and through the thick stripes of V2. Cells in V3 are orientation selective and are believed to be concerned with processing dynamic form. V5 (or MT) is believed to process information on motion and stereoscopic depth. The M pathway then projects to the *parietal cortex*, which is important for the integration of movement and depth into a representation of space. Damage to this region in humans causes a condition called *Balint’s syndrome*. This has three main symptoms: optic ataxia, ocular apraxia and simultagnosia.

(12) Visual information is processed in two broad systems: the ‘*what*’ system (also called the *ventral system*), which is concerned with the identification of an object, and the ‘*where*’ system (also called the *dorsal system*), which is concerned with the relative spatial position of an object. Another approach is to divide the pathways into ‘*what*’ vs. ‘*how*’. In this scheme the parietal cortex is concerned primarily with visual cues important for the conversion of visual cues into spatial information for motor movement and interaction with the environment.

(13) Damage to V1 causes blindness, but under forced choice conditions a patient may display perception of visual stimuli without conscious knowledge. This is called *blindsight*. It is believed that this perception is mediated by sub-cortical pathways that bypass V1 and project directly to the later visual areas.
Primary visual cortex

The visual equivalent of a sorting office?

The primary visual cortex (V1) or striate cortex is an important area in which partially processed information from the retina and LGN is separated and packaged up for more elaborate analysis in the specialised visual areas of the extrastriate cortex. But V1 is more than just a neural version of a post office sorting department. The response properties of most neurons in V1 are very different from those of neurons in the preceding area. New response features, such as sensitivity to lines and bars of different orientations and movements are created, along with a specialisation of some neurons to an existing visual feature such as colour. Moreover, the functional organisation of V1 into repeating columns and modules seems to be a standard pattern in all cortical visual areas, and this pattern of organisation is an efficient way of mapping a multi-dimensional stimulus, such as vision, on to an irregularly shaped piece of two-dimensional cortex.

Visual information passes to the cortex from the LGN through the optic radiation. In the monkey, the first cortical visual area (V1) consists of a folded plate of cells about 2 mm thick, with a surface area of a few square inches. This is a much larger and more complex structure than the LGN, for example, the LGN is composed of 1.5 million cells, whereas V1 is composed of around 200 million. V1 lies posteriorly in the occipital lobe and can be recognised by its characteristic appearance. Incoming bundles of fibres form clear stripes in this area, hence the name striate cortex. Adjacent regions of cortex also are concerned with vision. The area which immediately surrounds V1 is called V2 (sometimes called area 18) and receives input primarily from V1. Each area contains its own representation of the visual field projected in an orderly manner. The topography of the visual field is preserved in the projection to V1 from the LGN, with relatively more cortex being devoted to the fovea (about 300 mm²/degree²) than to peripheral visual field (about 0.1 mm²/degree² at 20 degrees of retinal eccentricity).
Segregation of layer 4 inputs

A general feature of the mammalian cortex is that the cells are arranged in six layers within the grey matter. These layers vary in appearance from area to area in the cortex depending on the density of cell packing and the thickness. In V1, most of the incoming fibres from the LGN terminate in layer 4 (see Figure 4.3). This layer is sub-divided into three sub-layers, A, B and C. Sub-layer C is further subdivided into 4Cα and 4Cβ. Projections from the LGN’s parvocellular layers terminate in sub-layers 4 A and 4Cβ and the upper part of layer 6. The cells in layer 4Cβ then supply cells in layers 2 and 3. The magnocellular layers terminate in layer 4Cα and in the lower part of layer 6. The cells in layer 4Cα supply layer 4B. Both the P and M systems send inputs to the blobs. Thus the separation into P and M streams first observed in the retina is preserved in V1.

Also preserved in layer 4 is the separation of inputs from the two eyes. In cats and monkeys, if the cells in one layer of the LGN receive their input from one eye, the next layer will receive inputs from the other eye. The cells from one LGN layer will project to groups of target cells in layer 4 C, separate from those supplied by the other eye. These groups of cells form alternating stripes or bands in layer 4 C. Above and below this layer, most cells are driven by both eyes, although one eye is usually dominant. Hubel and Wiesel (1977) termed these blocks of cells ocular dominance columns.

Evidence for this alternating projection comes from a number of sources. Firstly, if a small lesion is made in one layer of the LGN, degenerating terminals subsequently appear in layer 4 in a characteristic pattern of alternating stripes (Hubel & Wiesel, 1977). These correspond to areas driven by the eye in whose line of connection the lesion is made. Further, if one injects a radioactively labelled amino acid, such as proline or leucine into the vitreous humour of one eye, it is taken up by the nerve cell bodies of the retina and incorporated into a protein. The labelled protein is then transported from the ganglion cells through their projections to the LGN, and then on through the projections of LGN cells to layer 4 C. The striped arrangement demonstrated in this way is the same as that produced by lesions of the LGN, but now the pattern is in all parts of the primary visual cortex. The colour-selective blobs stained by cytochrome oxidase (described in Chapter 4) lie in the centre of each ocular dominance column.

Cortical receptive fields

There seem to be two broad categories of neurons in V1, which are termed simple and complex neurons. In addition, there is a class of cells found exclusively in layer 4C (where most LGN fibres terminate),
which have a concentric centre-surround receptive field like those of
the LGN and the retinal ganglion cells.

Simple cells are found mostly in layers 4 and 6, and these two
regions receive direct inputs from the LGN (Martínez et al., 2005). One
type of simple cell has a receptive field that consists of an extended
narrow central portion flanked by two antagonistic areas. The centre
may be either excitatory or inhibitory. For such cells, optimal activa-
tion is by a bar of light that is not more than a certain width, that
entirely fills the central area and that is orientated at a certain angle.
The optimal width of the narrow light or dark bar is comparable to
the diameters of the ‘on’ or ‘off’ centre regions of the centre-surround
receptive fields of the LGN and ganglion cells (Figure 5.1). Resolution
has not been lost, but has been incorporated into a more complex
receptive field. The preferred orientation of the light or dark bar
varies with different cells, as does the symmetry of the receptive
field. In some cases the receptive field may just consist of two longi-
tudinal regions facing each other, one excitatory and one inhibitory.
Another type of cell is sensitive to the length of the stimulating bar.
There seems to be an additional antagonistic area at the top or
bottom of the receptive field. As a result, the optimal stimulus is an
appropriately orientated bar or edge that stops in a particular place.
These cells are end-inhibited or end-stopped cells. In spite of the differ-
et proportions of inhibitory and excitatory areas, the two contribu-
tions match exactly and so diffuse illumination of the entire
receptive field produces little or no effect.

Complex cells are found in all layers and are abundant in layers
2, 3 and 5 (Martínez et al., 2005). Complex cells receive their inputs
from simple cells, and their response characteristics are based on

---

**Figure 5.1.** A simple orientation
selective cell (the grey cell in the
diagram) behaves as if it receives
input from several centre-surround
antagonistic cells of the LGN
(shown in white). Flashing a line
with an orientation that stimulates
more of the excitatory centres will
stimulate the cell. Thus a bar
orientated as in (a) will stimulate
the cell, but one oriented as in
(b) will not (redrawn from Hubel &
Wiesel, 1962).
integrating several overlapping simple cell receptive fields (Alonso & Martinez, 1998). In common with simple cells, these cells require a specific orientation of a dark–light boundary, and illumination of the whole receptive field is ineffective (Figure 5.2). However, the position of the stimulus within the receptive field is not important as there are no longer distinct excitatory and inhibitory areas. There are two main classes of complex cells. Both respond best to moving edges or slits of fixed width and precise orientation. One type of cell ‘likes’ a stimulus bar of a particular length, while the other type ‘likes’ end-stopped stimuli (these cells were previously called hypercomplex cells). The best stimulus for these cells requires not only a certain orientation, but also a discontinuity, such as a line that stops, an angle or a corner.

**Spatial frequency**

Although many neurons in V1 do respond to lines and bars, the optimal stimulus is a sine-wave grating (De Valois et al., 1978).
A sine-wave grating looks like a series of fuzzy, unfocused parallel bars. Along any line perpendicular to the long axis of the grating, the brightness varies according to a sine-wave function. Sine-waves can vary in frequency and amplitude. The location of a point along a single cycle of a sine-wave is specified by its phase. The peak is at 0 degrees, the middle of the cycle is at 90 degrees, the trough is at 180 degrees and the middle of the cycle occurs again at 270 degrees. The end of the cycle, 360 degrees, is the same as the beginning of the next one. A sine-wave is classified by its spatial frequency. Because the size of the image of a stimulus on the retina is dependent on how close it is to the eye, visual angle is usually used instead of the physical distance between adjacent cycles. Thus the spatial frequency of a sine-wave grating is measured in cycles per degree of visual angle. Most neurons in the striate cortex respond best when a sine-wave grating of a particular spatial frequency is placed in the appropriate part of the visual field. For orientation-selective neurons the grating must be aligned at the appropriate angle of orientation. In most cases a neuron’s receptive field is large enough to include between 1.5 and 3.5 cycles of the grating (De Valois et al., 1985).

**Texture**

A new class of neurons has been reported in V1 that seem to be responsive to texture (Von der Heydt, Peterhans & Dürrstler, 1992). They are unresponsive to single lines, bars, or edges, but they respond preferentially to a sine-wave or square-wave grating of a particular spatial frequency and orientation. These cells do not seem to be frequency analysers like those described above, but instead respond to the ‘texture’ of the pattern. Natural surfaces have a rough texture, and these cells, sensitive to texture and texture orientation, can not only detect the presence of a surface, but also its orientation and they may contribute to depth perception, for which texture gradients are an important cue.

**Direction selectivity**

Around 10%–20% of complex cells in the upper layers of the striate cortex show a strong selectivity for the direction in which a stimulus is moving. Movement in one particular direction produces a strong response from the cell, but it is unresponsive to movement in other directions. The other complex cells do not show a marked direction preference. How are such direction-selective cells wired up? In 1965, Horace Barlow and William Levick proposed a wiring plan for direction-selective cells in the rabbit retina. This has been modified in order to explain the direction selectivity of the primate complex cells and is illustrated in Figure 5.3. It is suggested that, between simple and complex cells, there are intermediate cells. These cells receive
excitation from one simple cell and inhibition from a second simple cell via a second intermediate cell. The second intermediate cell has a receptive field, which is immediately adjacent to that of the first. If a stimulus moves in the null direction, then the first intermediate cell is excited by one of its inputs just as it is inhibited by the other, whose receptive field has just been crossed. The two effects cancel and the cell does not fire. If the stimulus moves in the opposite direction, then the inhibition arrives too late to prevent the cell from firing.

If an observer views a grating moving in a single direction for several minutes, the observer’s threshold to detect motion in this direction is increased by up to a factor of two. This is an example of selective adaptation, produced by fatiguing the cortical cells that are selective for movement in that direction (Hammond et al., 1985). The arrangement of cortical neurons for the detection of motion suggests a ready explanation for the phenomenon of motion after-effect, sometimes called the waterfall illusion. If you view a stimulus moving in one direction for a period of time, such as a grating or a waterfall, and then fixate a stationary object, it will appear to move in the opposite direction to the moving stimulus. This is because prolonged exposure to downwards motion will fatigue or adapt the cells preferring downward motion. They will have virtually no spontaneous activity, while those preferring upwards motion will have normal levels of spontaneous activity. This biased distribution of spontaneous activity produces a pattern of activity similar to that produced by actual upwards movement and it is believed
that this forms the neural basis of the motion after-effect (Mather & Moulden, 1980).

**Colour**

As described in Chapter 4, staining for the mitochondrial enzyme cytochrome oxidase has shown a matrix of oval patches or **blobs**, approximately $150 \times 200 \ \mu m$ each, and each blob is centred on an ocular dominance column. Within an ocular dominance column, the degree of dominance varies. In the centre of a column, the dominance will be absolute and the neurons will receive input from only one eye. The blobs coincide with these centres of monocularity (Ts’o et al., 1990).

The blobs contain colour-opponent and double colour-opponent cells, but there is a segregation in the different forms of colour opponency. Within a blob, the neurons will be either red/green opponent or blue/yellow opponent, the two forms of opponency are not mixed within a single blob (T’so & Gilbert, 1988; Landisman & Ts’o, 2002). These findings suggest that individual blobs are dedicated to the processing of one colour opponent system. Moreover, different types of blob are not equally represented in V1. There are more red and green cones in the retina than blue cones, and this is reflected in the proportions of the red/green opponent retinal ganglion cells to the blue/yellow ganglion cells (5:2). This difference is also reflected in the proportions of red/green blobs to blue/yellow blobs (3:1), and the blue/yellow blobs seem to be clustered together. This suggests a non-uniform or patchy input of blue/yellow inputs into V1, which would be consistent with the annular organisation of blue cones in the retina. (As discussed in Chapter 3, the blue cones are distributed in a sparse ring around the edge of the fovea.).

The blobs often seem paired. The blob in a particular ocular dominance column is connected by **bridges** to a blob in a neighbouring ocular dominance column. In vertical section, this gives a ladder-like arrangement, with the bridges forming the rungs. Micro-electrode recording and optical imaging studies using voltage-sensitive dyes have shown that these bridges contain colour-selective cells like those in the blobs (T’so & Roe, 1995). Bridges have been found connecting blobs of different colour opponency. In this case, the responses of neurons in the bridges are neither red/green or blue/yellow, but are mixed in spectral selectivity.

**Modular organisation**

Different visual areas seem to be sub-divided into processing **modules** or **super columns**, which receive information from other modules, perform some calculation and then pass it on to other modules. V1 has been suggested to contain 2500 modules each $0.5 \times 0.7 \ mm$ and
each containing 150,000 neurons. The neurons in each module are concerned with the analysis of a particular portion of the visual field. The modules consist of two units, each centred around a blob (Figure 5.4). The neurons in layer 4Cα and 4Cβ are monocular. Each unit receives input from the opposing eye. The two units, however, exchange information and 80% of the neurons in the other layers are binocular. As mentioned, neurons in the blobs are sensitive to wavelength, and neurons in the inter-blob region are sensitive to orientation, and in the case of the complex cells, movement as well. Thus, each module contains neurons sensitive to wavelength, movement and lines or edges of particular orientations within a specific portion of the visual field. Within a module, the response characteristics of a cell are arranged systematically. If a neuron in one position responds best to a line orientated at 45 degrees, then a neuron a short distance away will respond best to a line orientated at 50 degrees. Each 25 μm of lateral movement encounters a neuron that responds to a line rotated by 10 degrees (Figure 5.5). Travelling across both halves of the module there are two 180 degree rotations in the orientations of lines that best stimulate the neurons.

The orderly progression of stimulus orientation in columns across the cortex, as shown in the traditional ‘ice cube’ model, seems to be an idealised version of cortical organisation. The micro-electrode experiments described above can only be used to map small areas of V1, but a new technique for simultaneously mapping activity across the cortex has been developed by Gary Blasdel (Blasdel & Salma, 1986) (Figure 5.6). Blasdel removed that part of the monkey’s skull covering the cortex of V1, and replaced it with a glass window. He then injected a voltage-sensitive dye on to the surface of the cortex.
The colour of this dye changes with the strength of the electrical current passing through it, and so can be used to analyse patterns of electrical activity in the striate cortex. If the visual stimuli used are bars of different orientations, it is possible to map the distribution of the orientation columns, and, if the stimuli are presented monocularly, it is possible to map the ocular dominance columns. This technique has shown that, although there are regions of cortex where a smooth gradation of orientation occurs, there are also discontinuities and distortions in the mapping, where columns with markedly different preferences are sometimes neighbours (Figure 5.7). This makes sense if you think about it. The stimulus can vary in orientation, ocular dominance and two dimensions of...
retinal position (up or down and from side to side) and to map this on to an irregular-shaped piece of two-dimensional cortex will require the occasional discontinuity to allow most of the retinotopic relationships to be preserved (Young, 1993a).

Summary of key points

(1) In cats and monkeys the left and right eyes supply alternate layers of the LGN. The cells from one LGN layer project to groups of target cells in layer 4 C, separate from those supplied by the other eye. These groups of cells form alternating stripes or bands in layer 4 C. Above and below this layer, most cells are driven by both eyes, although one eye is usually dominant. Hubel and Wiesel termed these blocks of cells **ocular dominance columns**.

(2) There seem to be two broad categories of neurons in V1, which are termed **simple** and **complex** neurons. In addition, there is a class of cells found exclusively in layer 4 C (where most LGN fibres...
terminate), which have a concentric centre-surround receptive field like those of the LGN and the retinal ganglion cells.

(3) Around 10%–20% of complex cells in the upper layers of the striate cortex show a strong selectivity for the direction in which a stimulus is moving. Movement in one particular direction produces a strong response from the cell, but it is unresponsive to movement in other directions.

(4) Staining for the mitochondrial enzyme cytochrome oxidase has shown a matrix of blobs on the surface of V1, approximately 150 x 200 μm each, and each blob is centred on an ocular dominance column. The blobs contain colour-opponent and double colour-opponent cells, but there is a segregation in the different forms of colour opponency. Within a blob, the neurons will be either red/green opponent or blue/yellow opponent. The two types of blob are present in unequal proportions (three red/green blobs to one blue/yellow blob), and the blue/yellow blobs seem to be clustered together. The blobs often seem paired. The blob in a particular ocular dominance column is connected by bridges to a blob in a neighbouring ocular dominance column.

(5) Different visual areas seem to be sub-divided into processing modules or super columns. V1 has been suggested to contain 2500 modules each 0.5 x 0.7 mm and each containing 150 000 neurons. Each module contains neurons sensitive to wavelength, movement and lines or edges of particular orientations within a specific portion of the visual field. Within a module, the response characteristics of a cell were proposed to be arranged systematically, such as in the ice cube model. However, optical imaging studies have shown that, although there are regions of cortex where a smooth gradation of orientation occurs, there are also discontinuities and distortions in the mapping, where columns with markedly different preferences are sometimes neighbours.
Plate 1. (Figure 1.6) The neural basis of functional magnetic resonance imaging (fMRI). (a) Viewing a stimulus such as a checkerboard produces marked changes in the areas of the brain that respond to visual stimuli, as seen in these positron emission tomographic (PET) images. These changes include increases in glucose use and blood flow that are much greater than those in oxygen consumption. As a result, there is an increase in the oxygen level in those areas (supply exceeds demand). PET is generally used to monitor blood flow. fMRI detects the changes in oxygen availability as a local change in the magnetic field. The resulting fMRI signal is a ‘blood-oxygen-level-dependent’ (BOLD) signal. (b) These metabolic and circulatory changes are driven by electrical potentials arising from the input to, and information processing within, the dendrites of neurons. (c) An attractive explanation for the BOLD signal invokes the preferential use of glycolysis in nearby non-neuronal cells (astrocytes) to handle an increase in the release of the neurotransmitter glutamate (Glu), which must be converted to glutamine (Gln) before it is returned to the neuron. Glycolysis consumes glucose to produce energy, but does not require oxygen (reproduced with permission from Raichle, 2001. Copyright (2001) Macmillan Publishers Ltd (Nature)).
Plate 2. **(Figure 2.1)** Light is a narrow band in the spectrum of electromagnetic radiation. Only electromagnetic radiation with a wavelength between 380 and 760 nm (one nanometre is $10^{-9}$ m) is visible to the human eye. The spectral sensitivity curves of the human eye are dependent on the nature of the cone pigments. Other species have different cone pigments and can detect different ranges of electromagnetic radiation. For example, some birds seem to have five cone pigments, including one that absorbs in the ultraviolet. The brightly coloured plumage of birds visible to us is only a fraction of the patterns and colours birds can see. All non-primate mammals, including cats, dogs and horses, have only two cone pigments in their retinas and have poorer colour vision than humans (reproduced with permission from Carlson et al., 2006. Copyright (2006) Pearson).

<table>
<thead>
<tr>
<th>Gamma ray</th>
<th>X rays</th>
<th>Ultraviolet rays</th>
<th>Infrared rays</th>
<th>Radar</th>
<th>Television and radio broadcast bands</th>
<th>AC circuits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength in nanometres</td>
<td>400</td>
<td>500</td>
<td>600</td>
<td>700</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Plate 3. **(Figure 2.2)** Volkman’s system for bypassing the eyelids and stimulating the eye through the roof of the mouth (reproduced by permission from Burr, 2005. Copyright (2005) Elsevier).
Fragment detached from rod outer segment during normal loss and regeneration of lamellae

Plate 4. (Figure 2.5) A schematic illustrating how the outer segments of rods and cones are embedded in the choroid layer at the back of the eye (redrawn from Young, 1971).
Plate 5. (Figure 2.11) A possible explanation of the Hermann Grid illusion. Two examples of centre-surround receptive fields are superimposed onto the grid. Above them, a bar chart shows the firing rate of a cell in response to the different patterns of stimulation at the two positions (reproduced by courtesy of Dr Jenny Read).

Plate 6. (Figure 3.1) Spectral absorbance curves for the human photopigments. There is a blue or short-wavelength pigment absorbing maximally at 420 nm, a green or middle-wavelength pigment absorbing maximally at 530 nm and a red or long-wavelength pigment absorbing at 565 nm (redrawn from Toveé 1995b).

Plate 7. (Figure 3.3) The appearance of the visible spectrum for five types of colour blindness: protanopia, deuteranopia, tritanopia, blue-cone monochromacy and rod monochromacy. Neutral points are indicated by the white region in the spectrum for the three types of dichromat reproduced by permission from Sharpe et al. 2000 (Copyright 2000).
Plate 8. (Figure 3.4) A scene from a fruit market as perceived by a normal trichromat, a protanope, deuteranope and a tritanope (reproduced with permission from Sharpe et al., 2000 (copyright 2000)).

Plate 9. (Figure 3.7) The photopigment molecule is composed of an opsin (a protein which is located in the cell membrane) and 11-cis retinal (an aldehyde derivative of vitamin A). The opsins form seven α-helical, hydrophobic regions within the cell membrane, linked by straight chain extra-membrane hydrophilic loops. The membrane regions form a bundle or palisade within which retinal is bound by a Schiff’s base to a lysine residue located in the centre of the seventh helix. The spectral absorbance of cone pigments in the middle- to long-wave region is thought to be based on the net effect of differences in the number and position of hydroxyl groups in the vicinity of the retinal chromophore. Five potentially important sites are indicated. Changes in the amino acids at positions 180, 277 and 285 seem to directly alter the spectral absorbance of the pigment. Substitutions at sites 233 and 230, seem to play a modulating role on the effects of changes at the other three sites (redrawn from Toveé, 1995b).
Plate 10  (Figure 4.7) (a). An example of the Ebbinghaus illusion. The two central circles are of equal diameter, but the ring of smaller circles makes the central circle on the left appear larger than the central circle on the right. The annulus of large circles on the right has the opposite effect. (b). Volunteers in the Aglioti study reached to grasp discs presented in the context of the small and large circle annuli. Grip aperture was recorded using an optoelectronic recording device (reproduced with permission from Plodowski & Jackson, 2001. Copyright (2001) Elsevier).

Plate 11  (Figure 5.6) The technique of optical imaging. A sensitive video camera is used to record changes in light absorption that occur as the animal views various stimuli presented on a video monitor. Images are digitised and stored in a computer in order to construct maps that compare patterns of activity associated with different stimuli (reproduced with permission from Purves et al. (2004). Copyright (2004) Sinauer Associates).
Plate 12. (Figure 5.7) This figure shows, at the top, a schematic representation of the surface of part of V1, showing the activation when the monkey viewed contours of different orientations (indicated on the right of the picture). Areas that were most active during the presentation of a particular orientation are indicated by the colour that represents that orientation (the bars are shown on the right). Thus, red and green show maximal activities in response to horizontal and vertical, while blue and yellow show greatest activation by left and right oblique. On the bottom left of the figure is an enlarged section of the cortical surface, showing the distinctive 'pin-wheel' arrangement with activation patterns spiralling outwards from the centre. On the bottom right of the figure is a 3-D diagram of the cortex under the pinwheel (redrawn with permission from Kandel, Schwartz & Jessel, 2000. Copyright (2000) McGraw-Hill).

Plate 13. (Figure 7.1) Estimates of the relative spectral power distribution of daylight phases across the visible spectrum, normalized to equal power at 560 nm (reproduced with kind permission from Bruce McEvoy from the website http://www.handprint.com).
Plate 14. (Figure 8.3) A schematic representation of the object recognition pathway. Through a hierarchy of cortical areas, from V1 to the inferior temporal cortex, complex and invariant object representations are built progressively by integrating convergent inputs from lower levels. Examples of elements for which neurons respond selectively are represented inside receptive fields (RFs; represented by circles) of different sizes. Feedback and horizontal connections are not shown but are often essential to build up object representations. The first column of bottom-up arrows on the right indicates the progressive increase in the 'complexity' of the neuronal representations. In the second column, figures are an estimate of the response latencies and in the third column are estimates of the RF sizes (modified with permission from Rousselet, Thorpe & Fabre-Thorpe (2004). Copyright (2004) Elsevier).
Plate 15. **(Figure 8.7)** This figure shows activity patterns on the surface of the monkey IT in response to different stimuli. Each activity patch corresponds to the top of a column of cells extending down through the cortex. (a) Distributions of active spots elicited by three different objects. (b) An example in which simplified stimuli elicited only a subset of the spots evoked by the more complex stimuli. (c, d) Examples in which new activity appeared when the original stimulus was simplified (reproduced with permission from Wang et al., 2000. Copyright (2000) Macmillan Publisher Ltd (Nature Neuroscience)).

Plate 16. **(Figure 8.10)** A figure illustrating the pattern of activation on the surface of the cortex to successive presentations of a head viewed at different angles. The colour of the strip above the image of the head indicates which activation pattern corresponds to which head (reproduced from Wang et al., 1996. Reprinted by permission of the AAAS).
Plate 17  (Figure 8.11) An illustration of the different patterns of activity seen in sparse and distributed coding. The blue and yellow pixel plots represent a hypothetical neural population. Each pixel represents a neuron with low (blue) or high (yellow) activity. In distributed coding schemes (left column), many neurons are active in response to each stimulus. In sparse coding schemes (right column), few neurons are active. If the neural representation is invariant (i.e. responsive to the same face independent of viewing position) (top row), then different views of the same person or object evoke identical activity patterns. If the neural representation is not invariant (bottom row), different views evoke different activity patterns. The results for face cells suggest that neural representation is extremely sparse and invariant (reproduced with permission from Connor, 2005. Copyright (2005) Macmillan Publisher Ltd (Nature)).
Plate 18. (Figure 8.12) Examples of some of the faces and non-face stimuli which have been used to stimulate face cells (reproduced with permission from Foldiak et al., 2004. Copyright (2004) Elsevier).
Plate 19 (Figure 8.15) An illustration of the experiments carried out by Tomita et al. (1999). (a) The bottom-up condition in which visual stimuli (cue and choice pictures) were presented in the hemifield contralateral to the recording site (‘electrode’) in the inferior temporal cortex. The monkey had to choose the correct choice specified by the cue. The bottom-up sensory signals (black arrow) would be detected in this condition. (b) The top-down condition. As in the bottom-up condition, but the cue was presented in the hemifield ipsilateral to the recording site, whereas the choice was presented contralaterally. In posterior-split-brain monkeys, sensory signal cannot reach visual areas in the opposite hemisphere, so only top-down signals (pale arrow) could activate inferior temporal neurons through feedback connections from the prefrontal cortex (reproduced with permission from Tomita et al., 1999. Copyright (1999) MacMillan Publisher Ltd (Nature)).

Plate 20 (Figure 9.1) Functional imaging maps showing the face-selective regions in the fusiform gyrus and the superior temporal sulcus. Regions shown in red to yellow responded more to faces than to houses. Regions shown in blue responded more to houses than to faces. The upper figures are lateral views of the folded cortical surface. The next row of images shows the cortical surfaces of each hemisphere tilted back 45° to show both the lateral and ventral surfaces of the temporal lobe. In the next set of images, the cortical surfaces are inflated to show the cortex in the sulci, indicated by a darker shade of grey. The lower images show the entire cortical surface of each hemisphere flattened into a two-dimensional sheet (reproduced with permission from Haxby et al., 2003. Copyright Elsevier (2003)).
Plate 21. (Figure 9.4) The pattern of eye-movements when judging facial expression for a control observer (top images) and for SM (bottom images). The circles show fixations and the red lines show saccades (reproduced with permission from Adolphs et al., 2005. Copyright (2005) Nature).

Plate 22. (Figure 10.2) Detail from Portraits in an Office by Edward Degas. The red line indicates a series of saccades, broken by six fixations. The set of small snapshots below the picture labelled a to f, shows the content of these fixations (reproduced with permission from Miller & Bockisch, 1997. Copyright (1997) Nature).
Plate 23 (Figure 10.4) An example of A.I.'s head movements while viewing one of Yarbus's pictures. A.I.'s head movements are characteristic of normal eye movement during viewing such a picture (reproduced with permission from Gilchrist et al., 1997. Copyright (1997) Nature).

Plate 24 (Figure 10.5) A lateral view of a monkey brain, which illustrates the pathway for corollary discharge to interact with visual perception. There is a pathway that runs from the SC in the midbrain to the MDN of the thalamus and then on to the FEF. This pathway is believed to carry the corollary discharge from the SC to the FEF (redrawn from Munoz, 2006).

Plate 25 (Figure 10.14) The illusion called Leviant's Enigma. Fixation of the centre will result in the perception of rotatory motion in the circles (reproduced with permission from Zeki, 1994. Copyright (1994) Royal Society).

Plate 27. (Figure 11.9) The anatomical correlates of hemi-neglect. Most studies of brain damaged patients suggest that sites of the lesions involve the right inferior parietal lobule (Brodmann Areas (BA) 39 and 40, coloured in red) and the supramarginal gyrus (at the temporo-parietal junction, black–grey area). Neglect can also be found with damage to the more dorsal and medial regions of the frontal premotor cortex, and to the superior temporal gyrus (shown in the lighter blue areas). Neglect after right frontal damage is less frequent and usually associated with lesions to the frontal premotor cortex, particularly to the more ventral parts (BA 44 and ventral BA 6, dark blue area) (reproduced with permission from Halligan et al., 2003. Copyright (2003) Elsevier).
Plate 28 (Figure 12.6) An illustration of the Necker cube (left). It is perceived as a three-dimensional object, and either the ‘back’ face (centre) or the ‘front’ face can be seen as being in front. The visual system alternates between these two alternative perceptions.

Plate 29 (Figure 12.7) The brain areas that contained cells whose activity correlated with the monkey’s subjective perception. The percentage of percept-related cells increases in the ‘higher’ visual centres (reproduced with permission from Leopold & Logothetis, 1999 Copyright (1999) Elsevier).

Plate 30 (Figure 12.8) An example of the stimuli used for binocular rivalry. In this case there are two experimental conditions. In condition (a), the observer is shown two sets of stimuli; faces to one eye and pictures of places to the other. This is the rivalry condition; the images are too different to be fused into a single percept. The observer sees either the face or the place at any one time, alternating between each. In condition (b), a face is shown to one eye and nothing to the other, and then a blank to the eye which had seen a face, and a place image to the eye which had previously seen the blank (reproduced with permission from Tong et al., (1999). Copyright (1998) Elsevier).
Visual development: an activity-dependent process

Variations on a theme

The development of the visual system is under the control of both genetic and environmental factors. The connections are refined and cut to fit on the basis of neural activity that is constantly flickering through the visual system from the retina. Following birth, it is environmental stimulation that elicits neural activity in the visual system. Cells in the retina, LGN and V1 of newborn, visually naive monkeys and kittens have receptive field and response properties very much like those of the adults. However, there are differences in their visual systems, such as in layer 4 of V1 where the projections from the LGN terminate. At birth, the cells in layer 4 are driven by both eyes, as projections from the LGN spread over a wide region of layer 4, whereas in the adult a layer 4 cell is driven by either eye but not by both. The adult pattern of ocular dominance columns in layer 4 is established over the first 6 weeks of life, when the LGN axons retract to establish separate, alternating zones in layer 4 that are supplied exclusively by one eye or the other (Figure 6.1).

In early life, the connections of neurons in the visual system are susceptible to change and can be affected irreversibly by unbalanced neural activity passing through them. For example, closure of the lids of one eye during the first 3 months of life leads to blindness in that eye. This is not because the eye no longer functions properly, but because the neurons in the visual cortex no longer respond to the signals the eye sends to them. Lid closure in adult animals has no such effect. It seems that, for the visual system to be correctly wired up, it must receive stimulation from the eyes to guide its development and allow connections to be strengthened or weakened, depending on the activity in the system. The most favoured theory for the mechanism underlying neural plasticity in adult animals was proposed back in the 1940s by Hebb. He suggested a coincidence detection rule such that, when two cells are simultaneously active, the synapse connecting them is strengthened (Hebb, 1949) (see...
Figure 6.2). The discovery of a putative cellular substrate for learning long-term potentiation (LTP) by Lomo in 1966 has resulted in a veritable deluge of studies. This work has been centred very largely on the hippocampus, an important area for learning and memory. In the hippocampus, LTP is characterised by an abrupt and sustained increase in the efficiency of synaptic transmission brought on by a brief high frequency stimulus. It may persist in the in vitro hippocampal slice for hours and in a freely moving animal for days (Bliss & Collingridge, 1993).

Although LTP does seem to be the most likely candidate for the mechanism of activity-dependent synaptic plasticity, it continues to be extraordinarily difficult to determine exactly how this synapse strengthening comes about. One generally agreed feature is that N-methyl-D-aspartate (NMDA) receptors, a sub-type of glutamate receptor, mediate the entry of Ca$^{2+}$ in CA1 of the hippocampus and thus induce LTP, although NMDA receptors are not necessarily involved in LTP at other sites. The NMDA receptors open in the presence of L-glutamate when the post-synaptic membrane is depolarised.
sufficiently to expel the channel blocker Mg$^{2+}$. Much of the current debate concerns the site that controls LTP expression: is it presynaptic or postsynaptic? Is control dependent on the specific experimental condition? In spite of this continuing tussle, the evidence for LTP as a general model of synaptic plasticity in the adult brain is increasing. But, what of the plasticity involved in the developing brain – is there a common mechanism? This chapter will examine the evidence for changes in the visual system with changes in visual input, and the possible mechanisms that might mediate these changes.

**Monocular or binocular deprivation**

The segregation of the LGN axons to form ocular dominance columns seems to be dependent on balanced activity from the two eyes. If this activity is interrupted and the balance between the two eyes is altered, then the result is a series of changes in the organisation of the visual system. One rather drastic way of altering the balance of activity is to close one eye in a developing animal. Rearing kittens with one eye sutured (monocular deprivation) causes a series of changes throughout the visual system and drastically reduces the perceptual capabilities of the eye that has been sutured during the early development. In the LGN, neurons connected to the deprived eye were reduced in size by 40% relative to the neurons connected to the other eye (Wiesel & Hubel, 1963). Further studies on the terminal fields of the LGN cells in layer 4 showed that LGN axons connected to the deprived eye occupied less than 20% of the cortical area, and the other non-deprived eye had expanded its representation to cover more than 80% of the thalamic recipient zone (LeVay, Stryker & Shatz, 1978). Single unit recording studies have shown that stimuli presented through the formerly deprived eye failed to influence the majority of cells in the striate visual cortex (Figure 6.3). The undeprived eye becomes the primary effective route for visual stimuli.

Under conditions of dark rearing (binocular deprivation), the organisation of the visual system and the selectivity of the cells initially continue to develop, despite the lack of visual stimuli (Buisseret & Imbert, 1976). When both eyes are closed in newborn monkeys for 17 days or longer, most cortical cells (such as the simple and complex cells) respond largely as normal to visual stimuli (Daw et al., 1983). The organisation of layer 4 seems to be normal and in other layers most cortical cells are stimulated by both eyes. The major difference is that a large proportion of cells could not be driven at all, while others were less tightly tuned to stimulus orientation. Binocular deprivation in kittens leads to similar results except that more cortical cells continue to be binocularly driven (Wiesel & Hubel, 1965). Longer visual deprivation (3 months or more) leads to a more marked effect. The visual cells become weakly responsive or totally unresponsive to visual stimuli, and the weakly responding cells lack
orientation, direction and stereo selectivity (Sherk & Stryker, 1976; Pettigrew, 1974). It seems that some of the results of monocular deprivation can be prevented or reduced by binocular deprivation. It may be that the two eyes are competing for representation in the cortex and, with one eye closed, the contest becomes unequal.

What, then, is the physiological basis for this ocular dominance shift associated with monocular deprivation? Such a shift can be prevented by modifying neuromodulator and neurotransmitter functions in the cortex (e.g. Shaw & Cynader, 1984; Bear & Singer, 1986; Reiter & Stryker, 1988), for example, by the infusion of glutamate into the cortex for a 2-week period during monocular deprivation. Control recordings during the infusion period show that cortical neurons in general fail to respond well to visual stimuli from either eye during the infusion period. The lack of ocular dominance modification seems to be the result of the reduced ability of the cortical cells to respond to the unbalanced LGN afferent input. Effective inputs representing the two eyes are greater than that of the deprived

![Figure 6.3](image-url)

Ocular dominance histograms in cells recorded from V1 in cats. (a) Recordings for 223 cells of adult cats. Cells in groups 1 and 7 of the histogram are driven by one eye only (ipsilateral or contralateral). All the other cells have inputs from both eyes. In groups 2, 3, 5 and 6, the input from one eye is dominant. In group 4, both eyes have a roughly equal influence. (b) Recordings from 25 cells of a kitten that was reared with its right eye occluded until the time of the experiment. The dashed bar on the right indicates that five cells did not respond to the stimulation of either eye. The solid bar indicates that all 20 cells that were responsive to stimulation responded only to the eye that was opened during rearing (redrawn from Wiesel & Hubel, 1963).
eye. It seems that, although changes associated with monocular deprivation have been found at the level of retino-geniculate terminals, in the LGN cell bodies, in the LGN terminal and in cortical cells’ responses, the primary site of binocular competition is cortical, and other changes in the visual system are secondary to the primary cortical competition.

This change in ocular dominance, as in all major rewiring in the visual system, occurs during a limited period following birth, often called the critical or sensitive period. It seems that the visual system is only capable of rewiring itself in this small temporal window and can do very little more once this opportunity has elapsed. For kittens, deprivation for only 3 days between the fourth and fifth week causes a large change in the pattern of ocular dominance (Hubel & Wiesel, 1970). If deprivation was started later than the eighth week, similar effects were observed, until even long periods of deprivation at four months caused no effect (Figure 6.4). Hubel and Wiesel concluded that the critical period for susceptibility to monocular deprivation begins in the fourth week and extends to about 4 months of age. Deprivation does not have to be long to cause large effects if it occurs during the critical period. Occluding one of the eyes of a 4-week-old kitten for a single day causes a large effect on the pattern of ocular dominance (Olson & Freeman, 1975) (Figure 6.5). However, it seems that the critical period is not fixed (Cynader, 1983). If cats are reared in the dark until long after the end of the chronologically defined critical period and only then brought into the light for monocular deprivation, this deprivation can still produce marked effects on cortical ocular dominance. Dark-reared cats seem to undergo a new critical period in the first few weeks after they are brought into the light. So, not only is the rewiring activity dependent but so is its initiation.

A similar situation to that described above in non-human mammals for monocular deprivation may also occur in human subjects. There is increasing evidence that amblyopia (a large reduction in the visual acuity in one eye) may sometimes occur in humans who, as young children, had the reduced use of one eye because of patching following an eye operation. This evidence has been provided by investigating the histories of 19 patients with amblyopia and finding that they all had their amblyopic (low visual activity) eye closed in early life, following an eye operation, with most of the closures occurring within the first year of life (Awaya et al., 1973). This type of amblyopia is called stimulus-deprivation amblyopia.

Image misalignment and binocularity

The changes in the ocular dominance columns are merely the most obvious effect of changes in the balance of neural input into the visual cortex. The majority of cells in the normal visual cortex are binocular, and during post-natal development, when the ocular dominance columns are being established, the connections to individual
cells from both eyes are also being refined. Unsurprisingly, monocular deprivation leads to most cells in the visual cortex being monocular. Under normal conditions, the input to a cell from the two red eyes is from corresponding areas of the retina. Misalignment of the images in the two eyes can be accomplished either by cutting the eye muscles or by fitting the animal with a helmet that contains small optical prisms. This disruption does not alter the absolute magnitude of activity. Under these conditions, most cells can only be driven monocularly, rather than binocularly as in normal animals, and the ocular dominance columns seem more sharply delineated (Lowel & Singer, 1993). Whereas 80% of cortical cells in normal cats are binocular, only 20% of the cells in cats with cut eye muscles respond to the stimulation of both eyes (Hubel & Wiesel, 1965). Similarly, 70% of cortical cells in monkeys are binocular, but less than 10% of cells are binocular in a monkey that has worn a prism-helmet for 60 days (Crawford & von Noorden, 1980). These neurological changes translate into striking behavioural effects. For example, prism-reared monkeys are unable to detect depth in random-dot stereograms, suggesting that they have lost the ability to use binocular disparity to perceive depth (Crawford et al., 1984).

It appears that it is not just the magnitude or balance of neural activity that is important, but also the temporal pattern of this activity. This hypothesis is supported by experiments in which the retina was deactivated with tetrodotoxin, and the optic nerve was stimulated directly. This allowed the temporal relationship of the neural activity from the two eyes to be controlled directly. Many more cortical cells were found to be monocular under a regimen of separate stimulation through the two optic nerves than were found with simultaneous stimulation (Stryker & Strickland, 1984). This synchronised activity of the two inputs to the same cell could be used in a Hebbian process for strengthening synapses from both inputs. The mechanism of this strengthening could be a form of LTP called ‘associative LTP’, in which the paired activity of two inputs to a cell results in the strengthening of both inputs. Uncorrelated activity from the two eyes seems to lead to a weakening and possible elimination of synapses in the visual cortex. This is another form of neural plasticity, long-term depression (LTD). A similar result can be demonstrated in the development of orientation selectivity by V1 neurons. If electrical stimulation is used to introduce artificially correlated activity into the visual system, the development of orientation selectivity is disrupted, emphasising once again how important the relative temporal pattern of activity is for developing neural connectivity (Weliky & Katz, 1997; Weliky, 2000).

### Image misalignment in humans

Some people have an imbalance in the eye muscles that upsets the co-ordination between their two eyes. This condition is called strabismus.
The misaligned eye can either turn inwards (esotropia) or outwards (exotropia). Just as cutting the eye muscles in experimental animals causes a loss of cortical cells that respond to stimulation of both eyes, there seems to be a similar lack of binocularly driven cells in people who had strabismus as young children. Strabismus can be corrected by a muscle operation that restores the balance between the two eyes. However, if this operation is not performed until the child is 4–5 years of age, a loss of binocularly driven cells seems to occur. This can be measured by the tilt after-effect (Figure 6.6), because of the phenomenon of interocular transfer. If an observer looks at the adapting lines with one eye and then looks at the test lines with the other eye, the after-effect will transfer between the eyes. This transfer, which is about 60%–70% as strong as the effect that occurs if the adaptation and test lines are viewed with the same eye, indicates that information from one eye must be shared with the other. The degree of transfer can be used to assess the state of binocularly driven cells. When surgery is carried out early in life, interocular transfer is high, indicating good binocular function, but if the surgery is delayed, interocular transfer is poor, indicating poor binocular function. The critical period for binocular development in humans seems to begin during the first year of life, reaches a peak during the second year, and decreases by 4 to 8 years (Banks, Aslin & Letson, 1975) (Figure 6.7).

The reduction in binocular neurons in people with strabismus reduces their ability to see depth, as much of our depth perception comes from comparing the differences in visual input between the eyes (see Chapter 11). However, in some cases this reduced depth perception might actually be an advantage. An artist has to translate the complexity of the 3-D world into a 2-D picture. It might be an advantage if you already see the world as a flat, 2-D image. A possible artistic candidate for this phenomenon is the seventeenth-century Dutch painter, Rembrandt van Rijn (Figure 6.8). An analysis of a set of his self-portraits (24 oil paintings and 12 etchings) showed that, in all but one painting, the eye on the right of the painting looked straight ahead, and the one on the left looked outwards (Livingstone & Conway, 2004). This eye alignment suggests exotropic strabismus.
As the authors of this study point out, art students are often advised to close one eye to flatten their perception, and so, for an artist, this impaired depth perception might be an advantage, rather than a handicap.

Selective rearing: manipulating the environment

Another way of altering the visual input is to raise animals in a tightly controlled visual environment, dominated by a certain visual stimulus and deficient in others. This does not alter the balance of activity between the eyes, but does alter the pattern of activity produced by each eye. These experiments have usually been carried out either by placing infant animals in an environment containing stripes of only one orientation (e.g. Blakemore & Cooper, 1970) or by fixing infant animals with goggles that present vertical stripes to one eye and horizontal stripes to the other (e.g. Hirsch & Spinelli, 1970).
Blakemore and Cooper kept kittens in the dark from birth to 2 weeks of age and then placed them in a large vertical tube for 5 hours every day. For the rest of the day, they remained in the dark. The inner surface of the tube was covered with either horizontal or vertical stripes. The kittens sat on a plexi-glass floor and the tube extended above and below them to ensure that there were no visible corners or edges in their environment other than the stripes on the side of the tube. The kittens wore neck ruffs to prevent them altering the orientation of the stripes by turning their heads. After 5 months, the selective rearing was stopped and the kittens remained in the dark except for brief periods when their vision was tested. The kittens displayed a number of defects in their visual behaviour. Their head movements were jerky when following moving objects, they tried to touch distant objects and often bumped into things. Most important of all, they seemed to be blind to stripes orthogonal to the orientation of the environment in which they were reared.

Following these behavioural tests, Blakemore and Cooper recorded from cells in the visual cortex to determine the optimum stimulus orientation for different cells. Most of the cells of the ‘horizontally reared’ cats responded primarily to horizontal stimuli and none at all responded to vertical stimuli. The opposite is true of the ‘vertically reared’ cats. These results have been confirmed by subsequent experiments (Muir & Mitchell, 1975). The results of Hirsch and Spinelli’s experiments (1970) using goggles showed the same pattern of effects. In single-cell recording experiments, they have found few cells in the visual cortex where the preferred orientation deviated from the orientation of the environmental stimulus by more than 5–10 degrees.

A single hour of exposure in a striped tube can drastically alter the preferred orientation of cells in the visual cortex. Blakemore and Mitchell (1973) kept a kitten in the dark until they recorded from its visual cortex at 42 days of age. As in kittens exposed to vertical stripes for much longer periods of time, most cells responded best to vertical or near-vertical orientations.

A number of different types of environment have been used to explore cortical plasticity further, such as moving white spots, random arrays of point sources of light and stripes moving in one particular direction (e.g. Van Sluyters & Blakemore, 1973; Pettigrew & Freeman, 1973). In each case the majority of cortical cells responded to the stimuli that were present in their environment and responded very weakly to anything else. An interesting example of selective rearing is shown in cats reared under conditions of stroboscopic illumination, where continuous retinal movement is prevented. This results in a deficit in the direction selectivity of the cortical cells, which is expressed behaviourally as a deficit in motion perception (Cynader & Cherneneko, 1976). The development of other cortical cell properties, such as orientation and stereo-selectivity, is unaffected.

It seems that, during the critical period, a number of changes are made to the wiring of the visual system, and this fine tuning is activity dependent. Without neural activity to stimulate and alter
the strength of synaptic connections, the normal response properties of visual cells will not develop.

**Impoverished visual input in humans**

The selective rearing experiments have been used as a model of condition in humans called *astigmatism*. This is caused by a distortion in the cornea, which results in an image that is out of focus either in the horizontal or the vertical orientation. A person who has an astigmatism at an early age is exposed to an environment in which lines in one orientation are imaged sharply on the retina, but lines 90 degrees from this orientation are out of focus. Freeman and Pettigrew (1973) showed that cats reared with an artificial astigmatism, created by wearing a mask containing astigmatic lenses, develop cortical cells that favour whichever orientation is in a sharp focus during rearing. This result in cat vision resembles a condition known as *meridional amblyopia* in humans. People whose vision is not corrected very soon after birth seem to show the same perceptual changes as animals reared in a selective environment or with goggles. As a result, even if the optical errors are corrected subsequently, the subject’s vision will still be poor, as he or she will not have the cortical machinery to process the new information available to the visual system.

**The critical period**

LTP and LTD are linked closely to the function of NMDA receptors. These receptors are found in the visual cortex of both cats and kittens (Fox, Sato & Daw, 1989), and the blocking of these receptors prevents the ocular dominance shift that occurs after the monocular deprivation (Bear *et al.*, 1990). These studies provide strong circumstantial evidence for a role for LTP and LTD in activity-dependent refinement of the visual cortex. Whilst most experiments on the development of the visual cortex have used monkeys and cats, the same effects can also be found in rat visual cortex (Fagiolini *et al.*, 1994), and in the last few years the hapless rat has formed the basis of brain slice preparations to investigate neural plasticity. The occurrence of LTP in the adult rat visual cortex was first reported some years ago (Artola & Singer, 1987), but recently, both LTP and LTD were also found to occur in the white matter of layer 3 of visual cortex in post-natal rats (Kirkwood, Lee & Bear, 1995). Most interestingly of all, a form of LTP was reported that only occurs during the critical period. Moreover, when the critical period is shifted by binocular deprivation, the occurrence of this form of LTP shifts with it, so the two are always in register.

Such a coincidence between LTP occurrence and a critical period is not confined to the visual system. In the rat somatosensory cortex,
connections can be altered radically if the input from the sensory vibrissae around the rat’s nose and mouth is manipulated. In this case, the critical period has no overlap with the regulating input to the visual cortex but it is much earlier and seems to be confined to the first post-natal week (Schlagger, Fox & O’Leary, 1993). Crair and Malenka (1995) have found a form of LTP in the somatosensory cortex that can only be induced during this first week. In addition, they present evidence for the involvement of NMDA receptors in this critical period, which adds to the likelihood that the LTP found in the adult brain is the same, or very similar, to that involved in the major construction that is carried out in the post-natal developing brain. Unlike the situation in the adult brain, however, these NMDA receptors must undergo some chemical or structural change correlated with the decline in LTP with the end of the critical period. The molecular composition of NMDA receptors can change during development, although the trigger for this is, as yet, unknown (Williams et al., 1993).

What we see, shapes how we see it

The development of the visual system combines features of both a hard-wired network and a self-organising neural net. The basic structure is pre-determined and is largely unaffected by the neural activity passing through it. However, for all the complex connections to be specified precisely in advance would be an epic task, and the opportunity for error during development would be immense. Therefore, the fine tuning of the connections, including the target cell for a particular LGN afferent as well as the balance and weighting of the synapses, is an activity-dependent process mediated by specific forms of LTP and LTD. As a result, our visual experience in the period immediately following birth is vitally important in shaping the functional organisation of the visual system, and an imbalance in visual stimulus will be mirrored by an imbalance in the visual system’s organisation.

Summary of key points

(1) The development of the visual system is dependent upon balanced neural activity from the eyes during a critical period early in post-natal development. Disruption of this activity through monocular deprivation or a controlled environment disrupts the organisation of the visual system.

(2) Monocular deprivation in young mammals leads to the visual cortex becoming unresponsive to the covered eye. A similar situation can be found in children who have had one eye patched.

(3) Binocular deprivation has a less dramatic effect initially, although longer visual deprivation (three months or more) leads to visual
cells becoming weakly responsive or totally unresponsive to visual stimuli; the weakly responding cells lack orientation, direction and stereo-selectivity.

(4) Some of the results of monocular deprivation can be prevented or reduced by binocular deprivation or by the infusion neurotransmitter blockers; it seems that the two eyes are competing for representation in the cortex; with one eye closed, the contest becomes unequal.

(5) If there is a misalignment of the images in the two eyes during early development, the proportion of visual neurons that are binocular is drastically reduced. These neurological changes translate into striking behavioural effects, suggesting that such animals have lost the ability to use binocular disparity to perceive depth.

(6) A similar lack of binocularly driven cells can also be found in human subjects who, during early childhood, have had an imbalance in the eye muscles that upsets the co-ordination between the two eyes. This condition is called strabismus.

(7) If an animal is raised in a controlled environment where it only sees certain stimuli, such as horizontal lines, it will be behaviourally and neurophysiologically unresponsive to the lines of other orientations.

(8) The selective-rearing experiments have been used as a model of a condition in humans called astigmatism. This is caused by a distortion in the cornea, which results in an image that is out of focus either in the horizontal or the vertical orientation.

(9) The neural basis of cortical plasticity is made up of two mechanisms called long-term potentiation (LTP) and long-term depression (LTD). Both forms are found in the visual cortex and one type of LTP is only found during the critical period.
Colour constancy

The colour constancy problem

One of the most important functions of the visual system is to be able to recognise an object under a variety of different viewing conditions. For this to be achieved, the stimulus features that make up that object must appear constant under these conditions. If stimulus parameters do not form a reliable ‘label’ for an object under different conditions, they are considerably devalued in their use to the visual system. For example, if we perceive a square shape on a video screen and the area it covers increases or decreases, we experience a sense of movement. The square seems to get closer or further away. The visual system assumes that the size of the square will not change, so that changes in its apparent size will signal changes in its relative distance from us. This is called object constancy. This is a sensible assumption, as under normal conditions, objects seldom change in size. Another example is lightness constancy. Over the course of a normal day, light levels change significantly, but the apparent lightness of an object will change very little. The visual system scales its measure of lightness to the rest of the environment, so that the apparent lightness of an object will appear constant relative to its surroundings. A similar problem exists with the perception of colour. Over the space of a day, the spectral content of daylight changes significantly (Figure 7.1). This means that the spectral content of light reflected from an object changes too. One might expect that objects and surfaces acquire their colour due to the dominant wavelength of the light reflected from them, thus a red object looks red because it reflects more long-wave (red) light. However, surfaces and objects retain their colour in spite of wide-ranging changes in the wavelength and energy composition of the light reflected from them. This is called colour constancy, and is not only displayed by humans and primates, but by a wide range of species from goldfish to honeybees. So it seems there is no pre-specified wavelength composition that leads to a colour and to that colour alone. If colours did change with every change in illumination,
then they would lose their significance as a biological signalling mechanism since that object could no longer be reliably identified by its colour.

The Land Mondrian experiments

Some of the most important and influential studies on colour constancy were made by Edwin Herbert Land (1909–1991). Land was a Harvard University drop-out, who went on to become one of the most successful entrepreneurs in America. He developed a method for producing large sheets of artificial polariser, and in 1937 founded the Polaroid Corporation to market his invention (Mollon, 1991). Polaroid filters, for visible and infra-red light, were soon being used in cameras and sunglasses, and in wartime for range-finders and night adaptation goggles. This development was followed up in 1948 with an instant camera, which could produce a picture in 60 seconds, and Land and his company became very rich. However, for the last 35 years of his life, Land’s chief obsession was with colour and colour constancy. As part of his experiments, he had observers view a multicoloured display made of patches of paper of different colours pasted together (Land, 1964). This display was called a Colour Mondrian, from the resemblance it bore to the paintings of the Dutch artist Piet Mondrian. The rectangles and squares composing the screen were of different shapes and sizes, thus creating an abstract scene with no recognisable objects to control for factors such as learning and memory. No patch was surrounded by another of a single colour and the patches surrounding another patch differed in
colour. This was to control for factors such as induced colours and colour contrast. The patches were made of matt papers which reflected a constant amount of light in all directions. As a result, the display could be viewed from any angle without affecting the outcome of the experiment.

The display was illuminated by three projectors, each equipped with a rheostat that allowed the intensity of the light coming from the projector to be changed. The first projector had a filter so that it only passed red light, the second projector only passed green light and the third projector only passed blue light. The intensity of light produced by each projector was measured using a telephotometer, so the relative amounts of the three wavelengths in the illumination could be calculated.

In one experiment, the intensity of light reflected from a green patch was set so that it reflected 60 units of red light, 30 units of green light and 10 units of blue light. Test subjects reported the green patch as being green in colour even though it reflected twice as much red as green light, and more red light than green and blue light put together. So, this is a clear example of the perceived colour of the patch not corresponding with the colour of the predominant wavelength reflected from it.

This experiment was repeated but under slightly different conditions. The subject still observed the same patch, illuminated by the same light, but this time the patch was viewed in isolation. The surrounding colour patches were not visible. This is called the void viewing condition. In this case the perceived colour of the patch corresponded to the wavelength composition of the light reflected from it. If the surround was then slowly brought into view, the colour of the patch was immediately reported to be green. This suggests that the perceived colour of the patch was determined not only by the wavelength composition of the light reflected from it, but also by the wavelength composition of the light reflected from the surrounding surfaces. If the position of the green patch was changed within the Mondrian, so that the surrounding patches were different, the perceived colour remained the same. This suggested that the relationship between the perceived colour and the wavelength composition of the patch and its surrounding patch or patches was not a simple one.

Reflectance and lightness: the search for constancy in a changing world

To construct a representation of colour that is constant with changes in the spectral illumination of a surface, the visual system must find some aspect of the stimulus which does not change. One physical constant of a surface that does not change is its reflectance. For example, a red surface will have a high reflectance for red light, and a low reflectance for green and blue light. If the intensity of the light
incident upon the object changes, the proportions of red, green and blue light reflected from the object will not (Figure 7.2). Therefore, the visual system must ignore the information related to light intensities and concentrate purely on relative reflectance.

One way of doing this is to compare the reflectance of different surfaces for light of the same wavelength. So, for example, consider two surfaces, a red and a green one. The red surface will have a high reflectance for long-wave light and so reflect a high proportion of red light. The green surface will have a low reflectance for red light, and therefore only a small proportion of red light will be reflected from it. So, if the patches are illuminated by a red light, the red patch will always appear lighter, regardless of the intensity of the red light. Thus, the biological correlate of reflectance is lightness (Zeki, 1993). By determining the efficiency of different surfaces in a scene for reflecting light of a given wavelength, the brain builds a lightness record of the scene for that particular wavelength.

When an entire scene is viewed, each surface will have a different lightness at every wavelength depending upon its efficiency for reflecting light of that wavelength. The record of that scene in terms of areas that are lighter or darker, is called its lightness record (Zeki, 1993). In ordinary daylight, as in most light sources, there is a mixture of wavelengths, and each set of wavelengths will produce a separate lightness record. Land’s Retinex theory (the name is derived
from retina and cortex) proposes that, in the visual system, the lightness records obtained simultaneously at three different wavelengths are compared in order to construct the colour of a surface (Land, 1964, 1983). This comparison will be unrelated to the wavelength composition of the illuminating light, and therefore will not be affected by the relative intensity of the lights of different wavelengths.

The colour that we perceive is thus the end product of two comparisons: the comparison of the reflectance of different surfaces for light of the same wavelength (generating the lightness record of the scene for that wavelength), and the comparison of the three lightness records of the scene for the different wavelengths (generating the colour). Colour therefore, is a comparison of comparisons (Zeki, 1993). When the wavelength composition of the light illuminating a surface changes, the intensities of light reflected from all the surfaces in the display will change, but the comparisons will remain the same because the reflectances do not themselves change.

Land has suggested an algorithm for generating these comparisons (Land, 1983). In it, the logarithm of the ratio of the light of a given wavelength reflected from a surface (the numerator), and the average of light of the same wavelength reflected from its surround (the denominator) is taken. This constitutes a designator at that wavelength. The process is done independently three times for the three wavelengths.

The biological basis of colour constancy

Colour constancy requires the comparison between the light from an object and the light reflected from other objects and surfaces to compensate for the spectral composition of the illuminating light. Until recently, it was thought that neurons capable of making this comparison did not occur until V4, where the receptive fields were sufficiently large (Schein & Desimone, 1990). Consistent with this theory, Semir Zeki found cells in V4 which appeared to show colour constancy (so, for example, cells responsive to green would continue to signal green, despite changes in the spectral composition of the illuminating light, as long as a surface continued to be perceived as green) (Zeki, 1983). He called these cells colour-only. Cells in V1 seemed to alter their responses with changes in the spectral composition of the illuminating light regardless of the perceived colour, and he called these cells wavelength-only. However, recent studies on the responses of visual neurons and their receptive fields have suggested that a large receptive field may not be necessary. Visual cells respond to stimuli within their receptive field. Stimuli presented outside the receptive field do not elicit a direct response from the cell. However, stimuli presented in the region surrounding the receptive field can modulate the cell’s response to a stimulus presented within its receptive field (Lennie, 2003). As a result, the region corresponding to the traditional receptive field is often called the classical receptive field, and the surrounding region which
modulates the cell’s response is called the non-classical or extra-classical receptive field.

This modulation may form the basis for the initial calculations necessary for colour constancy. Consider the simplest example of the background altering colour perceptions. If one sees a green patch on a green background, it appears to be less green than a green patch that is observed on a grey background. The difference, or contrast, between the colour of the patch and the background alters our perception of the colour of the patch. It seems that colour contrast plays an important role in building up a colour constant perception of the world, as factoring out the colour of the background is likely to also factor out the colour of the illuminant (Hurlbert, 2003). Recent studies have found V1 neurons that respond to colour contrast (Wachtler et al., 2003; Hurlbert et al., 2001). When presented with a patch of colour that completely covered the classical receptive field against a neutral grey background, each cell will have a preferred colour. Additionally, a background of a cell’s preferred colour will inhibit its response to the preferred colour. Thus the cell generates a measure of contrast, which seems to be based on interactions between the classical and extra-classical receptive fields. These measures can form the basis for the lightness record needed by the retinex theory to generate colour constancy. Individual cells cannot represent colour contrast accurately, but the activity of a whole population of such cells could.

This is not to say that colour constancy is computed in V1. It is probably a gradual process, in which it is calculated by successive computations in V1, V2 and then finally in V4, where full colour constancy is finally realised. This would be consistent with lesion studies, which have shown that the removal or damage of V4 in monkeys leaves them able to discriminate wavelength, but impaired on colour constancy (e.g. Wild et al., 1985).

**Colour constancy and the human brain**

The perception of colour in humans was initially associated with activation of a ventromedial occipital area (in the collateral sulcus or lingual gyrus, see Figure 7.3) in three separate PET studies (Corbetta
et al., 1991; Zeki et al., 1991; Gulyas & Roland, 1991). Because V4 contains colour selective cells, it has been speculated that this area is the homologue of V4. The location of this area agreed well with the location of lesions associated with achromatopsia, which is close, but medial to the posterior fusiform area activated by faces. That the colour and face-selective areas are close to each other would be consistent with evoked potential studies from chronically implanted electrodes in epilepsy patients (Allison et al., 1993, 1994). The proximity of these two areas would explain the frequent association of achromatopsia with prosopagnosia (the inability to recognise faces).

However, the situation seems to be more complicated than this. The neurons in monkey V4 are selective for features relevant to object recognition, including shape and colour (Zeki, 1983; Desimone & Schein, 1987), and therefore one would predict that the human homologue of V4 would show the same feature selectivity. However, of the two PET studies that examined colour and shape, one found that shape perception also activated the venteromedial occipitotemporal region (Corbetta et al., 1991), but the other did not (Gulyas & Roland, 1991). Moreover, lesions of monkey V4 produce significant impairments in form perception (Schiller & Lee, 1991), but form perception is usually spared in patients with achromatopsia. Also, the monkey V4 lesions do not seem to produce the profound and permanent colour impairment that is seen in patients with achromatopsia (Schiller & Lee, 1991; Heywood et al., 1992). Thus, although an area in human cerebral cortex has been located that is selective for colour, it may not be the homologue of monkey V4. An alternative candidate has been suggested in a study by Hadjikhani et al. (1998). They used fMRI to map brain activity in response to colour, and found a new area that is distinct anatomically from the putative human V4. This area (which they called Visual area 8 or V8) is located in front of human ‘V4’, and responds more strongly to colour than the surrounding areas and, unlike human ‘V4’, is activated by the induction of colour after-effects. They suggest that, for humans, V8 may be the neural basis for colour constancy and the conscious perception of colour (Hadjikhani et al., 1998; Heywood & Cowey,
1998). However, Semir Zeki has proposed that ‘V8’ should actually be lumped together with the putative human ‘V4’ into the ‘V4 complex’, and that V8 should be more properly named V4\(\alpha\) (Bartels & Zeki, 2000). This latter approach stresses the strong connections between the putative human ‘V4’ and V8, and sees V8 as functionally part of a single colour processing unit along with human ‘V4’ (Figure 7.4).

### Summary of key points

1. **Surfaces and objects retain their colour in spite of wide-ranging changes in the wavelength and energy composition of the light reflected from them.** This is called *colour constancy*.
2. Edwin Land investigated colour constancy by using a multi-coloured display made of patches of paper of different colour pasted together (a *Colour Mondrian*).
3. When the spectral composition of the light illuminating the Mondrian was altered, the perceived colours of the patches remained the same. However, if a patch was viewed in isolation (the *void viewing condition*), the perceived colour of the patch corresponded to the wavelength composition of the light reflected from it. This suggests that the perceived colour of a patch was determined not only by the wavelength composition of the light reflected from it, but also by the wavelength composition of the light reflected from the surrounding surfaces.
4. **One physical constant of a surface that does not change with changes in the spectrum illumination is its *reflectance*. The biological correlate of reflectance is the perceived *lightness* of a surface.**
5. The record of a scene in terms of areas which are lighter or darker, is called its *lightness record*. Land’s *Retinex theory* proposes that, in the visual system, the lightness records obtained simultaneously at three different wavelengths are compared to construct the colour of a surface.
6. Some neurons in monkey V1 and V2 are sensitive to the wavelength composition of light, but do not show colour constancy. However, the responses of some cells in monkey V4 show the same colour constancy characteristics as those of a human observer viewing the same stimuli.
7. **The neural basis of human colour constancy is unclear. A putative V4 area has been identified, but an additional area, called V8 or V4\(\alpha\), may also play an important role in the development of colour constancy.**
Object perception and recognition

From retinal image to cortical representation

In the primary stages of the visual system, such as V1, objects are coded in terms of retinotopic co-ordinates, and lesions of V1 cause defects in retinal space, which move with eye movements, maintaining a constant retinal location. Several stages later in the visual system, at the inferior temporal cortex (IT) in non-human primates, the receptive fields are relatively independent of retinal location, and neurons can be activated by a specific stimulus, such as a face, over a wide range of retinal locations. Deficits that result from lesions of IT are based on the co-ordinate system properties of the object, independent of retinal location. Thus, at some point in the visual system, the pattern of excitation that reaches the eye must be transposed from a retinotopic co-ordinate system to a co-ordinate system centred on the object itself (Marr, 1982). An outline of such a transformation can be seen in Table 8.1.

At the same time that co-ordinates become object centred, the system becomes independent of the precise metric regarding the object itself within its own co-ordinate system, that is to say the system remains responsive to an object despite changes in its size, orientation, texture and completeness. Single-cell recording studies in the macaque suggest that, for face processing, these transformations occur in the anterior IT. The response of the majority of cells in the superior temporal sulcus (STS) is view-selective and their outputs could be combined in a hierarchical manner to produce view-independent cells in the inferior temporal cortex. As a result, selective deficits to higher visual areas, such as IT, cause the inability to recognise an object or classes of object. This defect in humans is called an agnosia.

Early visual processing

Visual recognition can be described as the matching of the retinal image of an object to a representation of the object stored in memory.
For this to happen, the pattern of different intensity points produced at the level of the retinal ganglion cells must be transformed into a three-dimensional representation of the object, which will enable it to be recognised from any viewing angle. The cortical processing of visual information begins in V1, where cells seem to be selective for the orientation of edges or boundaries. Boundaries can be defined not just by simple changes in luminance, but also by texture, colour and other changes that occur at the boundaries between objects. So, what principles guide the visual system in the construction of the edges and boundaries that form the basis of the object representation?

The answer may lie, at least partially, with the traditional gestalt school of vision, which provides a set of rules for defining boundaries (see Table 8.2). For example, under the gestalt principle of good continuity, a boundary is seen as continuous if the elements from which it is composed can be linked by a straight or curved continuous line. Figure 8.1(a) illustrates an illusory vertical contour that is formed by the terminations of the horizontal grating elements. There is no overall change in luminance between the left and right halves of

<p>| Table 8.1. | A summary of Marr’s model of object recognition. Marr viewed the problem of vision as a multi-stage process in which the pattern of light intensities signalled by the retina is processed to form a three-dimensional representation of the objects in one’s surroundings |</p>
<table>
<thead>
<tr>
<th>Level</th>
<th>Description achieved</th>
</tr>
</thead>
<tbody>
<tr>
<td>The raw primal sketch</td>
<td>Description of the edges and borders, including their location and orientation</td>
</tr>
<tr>
<td>The full primal sketch</td>
<td>Where larger structures, such as boundaries and regions, are represented</td>
</tr>
<tr>
<td>The 2½-dimensional sketch</td>
<td>A fuller representation of objects, but only in viewer-centred co-ordinates; this is achieved by an analysis of depth, motion and shading as well as from the structures assembled in the primal sketch</td>
</tr>
<tr>
<td>The three-dimensional model</td>
<td>A representation centred upon the object rather than on the viewer</td>
</tr>
</tbody>
</table>

| Table 8.2. | The gestalt principles of organisation |
| Rule | Boundaries defined |
| Pragnanz | Every stimulus pattern is seen in such a way that the resulting structure is as simple as possible |
| Proximity | The tendency of objects near one another to be grouped together into a perceptual unit |
| Similarity | If several stimuli are presented together, there is a tendency to see the form in such a way that the similar items are grouped together |
| Closure | The tendency to unite contours that are very close to each other |
| Good continuation | Neighbouring elements are grouped together when they are potentially connected by straight or smoothly curving lines |
| Common fate | Elements that are moving in the same direction seem to be grouped together |
| Familiarity | Elements are more likely to form groups if the groups appear familiar or meaningful |

(Perrett & Oram, 1993).
the figure, yet a strong perceptual border exists. The operation of continuity can also be seen in Figure 8.1(b), where an illusionary bar seems to extend between the notches in the two dark discs. The illusory light bar is inferred by the visual system to join the upper and lower notches and the break in the central circle. In Figure 8.1(c), the illusory light bar perceptually is absent. Here, the notches are closed by a thin boundary and each notch is therefore seen as a perceptual entity in its own right in accordance with the gestalt principle of closure. Psychologists have speculated that contours defined by good continuity were constructed centrally, rather than extracted automatically by neural feature detectors working at some stage of visual processing (Gregory, 1972). The illusory contours have therefore been given various labels including cognitive, subjective or anomalous. However, recent neurophysiological and behavioural results have disproved this idea, and suggest that these illusory contours are extracted very early in the visual system.

Physiological studies have shown that specific populations of cells in early visual areas (V1 and V2) do respond selectively to the orientation of contours defined by good continuity (Peterhans & von der Heydt, 1989; Grosof et al., 1993). Cells in V1 and V2 respond to illusory contours defined by the co-linearity of line terminations and signal the orientation of this illusory contour. Moreover, about one-third of the cells tested in V2 responded to illusionary contours extending across gaps as well as they did to normal luminance contours, and the cells seem to exhibit equivalent orientation selectivity for real and illusory edges. This neurophysiological evidence is supported by the findings of Davis and Driver (1994), who used a visual search task to distinguish between early and late stages in the processing of visual information. For example, among many jumbled white letters, a single red one is discerned instantly (a phenomenon called ‘pop out’), but a single L among many Ts needs more time to be detected. This result is taken to suggest that colour differences are extracted early in the visual system, but differentiation of similar letters is the result of more complex processing at a higher level. This procedure can be quantified by measuring the time it takes for a single odd
feature to be detected among a number of background features. A rapid reaction time, which is largely independent of the number of background features, is taken to be indicative of processing at an early stage in the visual system. Davis and Driver used figures outlined by illusory contours based on the Kanizsa triangles (Figure 8.2), and their results were consistent with the processing of these features occurring early in the visual system.

Thus, the early cortical visual areas contain the neural machinery that is involved in the definition of boundaries in different regions of the retinal images. While many of these boundaries and contours are defined by luminance changes, analysis of subjective contours provides powerful supplementary cues to object boundaries.

A visual alphabet?

As we move up the object-processing pathway in monkeys (V1–V2–V4–posterior IT–anterior IT) (see Figure 8.3), the response properties of the neurons change. The receptive field of a cell gets significantly larger. For example, the average receptive field size in V4 is 4 degree$^2$, which increases to 16 degree$^2$ in posterior IT, and to 150 degree$^2$ in anterior IT. Most cells along the V4, posterior IT and anterior IT pathway also have receptive fields close to, or including, the fovea (75% of anterior IT cells included the fovea). The increase in receptive field allows the development of a visual response that is unaffected by the size and position of a stimulus within the visual field. The cells also respond to more and complex stimuli. In V4 and
in posterior IT, the majority of cells have been found to be sensitive to the ‘primary’ qualities of a stimulus, such as colour, size or orientation, whereas cells in anterior IT seem to be sensitive to complex shapes and patterns (Figure 8.4).

How cells in IT encode a representation of objects is a knotty problem. An interesting approach has been taken by Keji Tanaka. He has tried to determine the minimum features necessary to excite a cell in anterior IT (Tanaka et al., 1992; Tanaka, 1997). This method begins by presenting a large number of patterns or objects while recording from a neuron, to find which objects excite that cell. Then, the component features of the effective stimulus are segregated and presented singly or in combination (see Figure 8.5), while assessing the strength of the cell’s response for each of the simplified stimuli. The aim is to find the simplest combination of stimulus features to which the cell responds maximally. However, even the simplest ‘real world’ stimulus will possess a wide variety of elementary features, such as depth, colour, shape, orientation, curvature and texture and may show specular reflections and shading (Young, 1995). It is therefore not possible to present all the possible feature
combinations systematically, and the simplified stimuli that are actually presented in the cell’s receptive field typically are a subset of the possible combinations. Hence, it is not possible to conclude that the best simplified stimulus is optimal for the cell, only that it was the best of those presented (Young, 1995).

Tanaka found a population of neurons in IT, called *elaborate cells*, which seemed to be responsive to simple shapes (Tanaka et al., 1991; Fujita et al., 1992). Cells in IT responsive to such simple stimuli seem to be invariant with respect to the size and position of a stimulus and of the visual cues that define it (Sary, Vogels, & Orban, 1993). Moreover, Tanaka found that closely adjacent cells usually responded to very similar feature configurations. In vertical penetrations through the cortex, he consistently recorded cells that responded to the same ‘optimal’ stimulus as for the first test cell tested, indicating that cells with similar preferences extend through most cortical layers. In tangential penetrations, cells with similar preferences were found
in patches of approximately 0.5 mm$^2$. These results suggested to Tanaka that the cells in IT are organised into functional columns or modules, each module specifying a different type of shape (Figure 8.6). This hypothesis has been supported by studies that combine intrinsic optical recording and single cell recording. This intrinsic optical recording measures the local changes in blood flow and blood oxygenation on the surface of the brain. It can show which patches of cortex are active in response to a particular visual stimulus. Combining it with single cell recording allows an experimenter not only to see which parts of the cortex are active in response to a stimulus (and presumably to processing information about the

**Figure 8.5.** An example of the procedures used by Tanaka and his colleagues in determining which features are critical for the activation of individual elaborate cells in IT. Among an initial set of three-dimension object stimuli, a dorsal view of the head of an imitation tiger was the most effective for the activation of a cell. The image was simplified while the responses of the cell were measured, the final result being that a combination of a pair of black triangles with a white square was sufficient to activate the cell. Further simplification of the stimulus abolished the responses of the cell (redrawn from Tanaka, 1992).

**Figure 8.6.** Schematic diagram of the columnar organisation of inferior temporal cortex. The average size of columns across the cortical surface is 0.5 mm. Cells in one column have similar but slightly different selectivities (redrawn from Tanaka, 1992).
stimulus), but also to what individual cells in the active patches are responding. These studies show a patchy distribution of activity on the surface of IT, roughly 0.5 mm in diameter, which would be consistent with a columnar organisation (Wang et al., 1996, 1998; Tsunoda et al., 2001). Within each ‘patch’, cells seem to be responding to a similar simple shape. If these modules are 0.5 mm² in width, then there could be up to 2000 within IT. However, allowing for the fact that many may analyse the same type of shapes, and many may analyse more complex patterns such as faces, the number of different simple shapes is probably only around 600 (Perrett & Oram, 1993).

This gave rise to the idea that these simple shapes form a ‘visual alphabet’ from which a representation of an object can be constructed (Stryker, 1992; Tanaka, 1996). The number of these simple shapes is very small by comparison with the number of possible visual patterns, in the same way as the number of words that can be constructed from an alphabet is very large. Each cell would signal the presence of a particular simple shape if it were present in a complex pattern or object. Consistent with this hypothesis, an intrinsic recording study has shown that visual stimuli activated patches on the cortical surface (presumably columns) distributed across the surface of IT (Tsunoda et al., 2001). When specific visual features in these objects were removed, some of the patches became inactive. This suggests that these inactive patches correspond to functional columns containing cells responsive to the visual feature that has been removed from the stimulus, a conclusion supported by subsequent single recording in that part of IT cortex corresponding to the activation patch (Tsunoda et al., 2001).

On some occasions when visual stimuli were simplified, although some patches became inactive, other new patches became active. These new patches were not active previously to the more complex (unsimplified) stimulus and are active in addition to a subset of the previously active patches (see Figure 8.7). This suggests that objects are represented not just by the simple sum of the cells which are active in different columns, but also by the combination of active and inactive cells. This increases the number of possible activation patterns, and so helps to differentiate different objects precisely with different arrangements of features. Such combinations may allow the representation of changes in our viewpoint of an object, such as when it is rotated, or occluded, or when it changes in size (Figure 8.8).

The shape selectivity of the elaborate cells is greater than that anticipated by many theories of shape recognition. For example, Irving Biederman (1987) described a theory of shape recognition that deconstructed complex objects into an arrangement of simple component shapes. Biederman’s scheme envisaged a restricted set of basic 3-D shapes, such as wedges and cylinders, which he called geons (geometrical icons). Examples of these figures are shown in Figure 8.8. These geons are defined only qualitatively. One example
is thin at one end, fat in the middle and thin at the other. Such qualitative descriptions may be sufficient for distinguishing different classes of objects, but they are insufficient for distinguishing within a class of objects possessing the same basic components (Perrett & Oram, 1993). Biederman’s model is also inadequate for differentiating between perceptually dissimilar shapes (Figure 8.9(b) and (c)) (Saund, 1992). Perceptually similar items (Figure 8.9(a) and (b)) would be classified as dissimilar by Biederman’s model. The single cell studies provide direct evidence that shape and curvature are
coded within the nervous system more precisely than would be expected from Biederman’s recognition by components model.

**Complex objects in 3-D: face cells**

There is evidence that the cellular coding of at least some complex patterns and objects does not remain as a collection of separate codes for its component shapes. The most studied example is the face cell. For nearly 30 years it has been known that there are neurons in the monkey visual system that are sensitive to faces. These face cells have been studied in most detail in the anterior inferior temporal (IT) cortex and in the upper bank of the superior temporal sulcus (STS), but they also occur in other areas such as the amygdala and the inferior convexity of the prefrontal cortex. Characteristically, the optimal stimuli of face cells cannot be deconstructed into simpler component shapes (Wang *et al.*, 1996). In general, these cells show virtually no response to any other stimulus tested (such as textures, gratings, bars and the edges of various colours) but respond strongly to a variety of faces, including real ones, plastic models and video display unit images of human and monkey faces. The responses of many face cells are size and position invariant; the cell’s response is maintained when there is a change in the size of the face, or if the position of the face within the cell’s receptive field is altered (e.g. Rolls & Baylis, 1986; Tovee *et al.*, 1994). Face cells do not respond well to images of faces that have had the components rearranged, even though all the components are still present and the outline is unchanged (e.g. Perrett *et al.*, 1982, 1992). Face cells are even sensitive to the relative position of features within the face; particularly important is inter-eye distance, distance from eyes to mouth and the amount and style of hair on the forehead (e.g. Yamane *et al.*, 1988; Young & Yamane, 1992). Moreover, presentation of a single facial component elicits only a fraction of the response generated by the whole face, and removal of a single component of a face reduces, but does not eliminate, the response of a cell to a face. This suggests that the face cells encode holistic information about the face, because the entire configuration of a face appears to be critical to a cell’s response (Gauthier & Logothetis, 2000).

Most face cells in the anterior IT and STS are selective for the viewing angle, such as the right profile of a face in preference to any
other viewing angle. These cells are described as view-dependent or viewer-centred. A small proportion of the cells are responsive to an object, irrespective of its viewing angle. These view-independent or object-centred cells, may be formed by combining the outputs of several view-dependent cells. For example, view-independence could be produced by combining the responses of the view-dependent cells found in the STS. This hierarchical scheme would suggest that the response latency of such view-independent cells would be longer than that of the view-dependent cells, which proves to be the case. The mean latency of view-invariant cells (130 ms) was significantly greater than that for view-dependent cells (119 ms) (Perrett et al., 1992).

Studies that have combined optical imaging with single-cell recording have revealed a patchy distribution of cellular activity on the cortical surface in response to faces, consistent with face cells being organised into functional columns (Wang et al., 1996, 1998) (Figure 8.10). However, the imaging also showed that, rather than discrete columns with little overlap, there was significant overlap in activity to different face orientations. This may mean that stimuli are mapped as a continuum of changing features (Tanaka, 1997). Such a continuous map could produce a broad tuning of cortical cells for certain directions of feature space, which would allow the association of different, but related images, such as the same object from different viewpoints or under different illumination. This would
obviously be an important mechanism in the development of a stimulus-invariant response. However, feature space is a vast multi-dimensional area in which even the simplest ‘real world’ stimulus will possess a wide variety of elementary features, such as depth, colour, shape, orientation, curvature and texture, as well as specular reflections and shading (Young, 1995). Thus, a continuous representation would have to be reduced in some way to fit the limited dimensions possible in the cortex. Ultimately, a columnar organisation is more likely, with cells in several columns responsive to stimuli that have features in common, and becoming jointly active as appropriate, a scheme that can also give rise to stimulus invariance.

**Functional divisions of face cells: identity, expression and direction of gaze**

Faces can vary in a number of ‘dimensions’, such as identity, expression, direction of gaze and viewing angle. Different populations of face cells seem to be sensitive to specific facial dimension, and insensitive to others. For example, Hasselmo et al. (1989) studied face cells in the STS and anterior IT with a set of nine stimuli consisting of three different monkey faces each displaying three different expressions. Neurons were found to respond to either dimension independently of the other. Cells that responded to expressions tended to cluster in the STS, whereas cells that responded to identity clustered in anterior IT. Further investigation has shown that there are also face cells in the STS that are responsive to gaze direction and orientation of the head (both of which are cues to the direction of attention) rather than expression (Perrett et al., 1992; Hasselmo et al., 1989). There seem to be five ‘classes’ of face cell in the STS, each class tuned to a separate view of the head (full face, profile, back of the head, head-up and head-down) (Perrett et al., 1992). There are an additional two subclasses, one responding to the left profile and one to the right profile.

Consistent with this finding of an anatomically segregated, functional specialisation in processing different dimensions of facial information, removal of the cortex in the banks and floor of the STS of monkeys results in deficits in the perception of gaze direction and facial expression, but not in face identification (Heywood & Cowey, 1992). Perrett et al. (1992) has suggested that the STS face cells may signal social attention, or the direction of another individuals gaze, information clearly crucial in the social interactions of primates.

Other face populations also seem to be responsive to a specific dimension. The face cells in the amygdala seem to be sensitive to a range of facially conveyed information, including identity, emotion and gaze (Gothard et al., 2007; Hoffman et al., 2007). The neurons responsive to different aspects of facially conveyed information are located in anatomically separate regions of the amygdala. These
different neurons may play a role in influencing eye movements in assessing faces and the information they signal, and may help orientate the observer’s behaviour and cognition towards important social signals (Calder & Nummenmaa, 2007). The face cells in the prefrontal cortex are sensitive to facial identity and seem to play a role in working memory (O’Scalaidhe et al., 1997). The functional organisation of the different face cell populations suggests the existence of a neural network containing processing units that are highly selective to the complex configuration of features that make up a face, and which respond to different facial dimensions (Gauthier & Logothetis, 2000).

There seem to be some homologies between the human and monkey face processing systems. An area of the fusiform gyrus in humans has been implicated in face identification and may be the homologue of the face area in anterior IT. There is also a region in the STS of both humans and monkeys that appears to be important for the processing of eye gaze and other facial expressions. Additionally, the human amygdala seems to play an important role in directing eye movements in the process of recognising facially expressed emotion (Adolphs et al., 2005), and this is consistent with the finding of face cells responsive to expression and gaze in the monkey amygdala (Hoffman et al., 2007).

The grandmother cell?

Temporal lobe face cells appear superficially to resemble the gnostic units proposed by Konorski (1967) or the cardinal cells proposed by Barlow (1972). These units were described as being at the top of a processing pyramid that began with line and edge detectors in the striate cortex and continued with detectors of increasing complexity until a unit was reached that represented one specific object or person, such as your grandmother, leading to the name by which this theory derisively became known. This idea had two serious problems. Firstly, the number of objects you meet in the course of your lifetime is immense, much larger than the number of neurons available to encode them on a one-to-one basis. Secondly, such a method of encoding is extremely inefficient as it would mean that there would need to be a vast number of uncommitted cells kept in reserve to code for the new objects one would be likely to meet in the future.

Although individual cells respond differently to different faces, there is no evidence for a face cell that responds exclusively to one individual face (Young & Yamane, 1992; Rolls & Tovée, 1995; Foldiak, 2004). Face cells seem to comprise a distributed network for the encoding of faces, just as other cells in IT cortex probably comprise a distributed network for the coding of general object features. Faces are thus encoded by the combined activity of populations or ensembles of cells. The representation of a face would depend on the emergent
spatial and temporal distribution of activity within the ensemble (Rolls & Toveé, 1995; Rolls, Treves & Toveé, 1997). Representation of specific faces or objects in a population code overcomes the two disadvantages of the grandmother cell concept. The number of faces encoded by a population of cells can be much larger than the number of cells that make up that population. So, it is unnecessary to have a one-to-one relationship between stimulus and cell. Secondly, no large pool of uncommitted cells is necessary. Single cell experiments have shown that the responses of individual neurons within a population alter to incorporate the representation of novel stimuli within the responses of existing populations (Rolls et al., 1989; Toveé, Rolls & Ramachandran, 1996).

The size of the cell population encoding a face is dependent on the ‘tuning’ of individual cells. That is to say, how many or how few faces do they respond to? If they respond to a large number of faces, then the cell population of which they are a part must be large in order to signal accurately the presence of a particular face. A large cell population containing cells responsive to a large number of faces is termed distributed encoding. If a cell responds only to a small number of specific faces, then only a small number of cells in the population is necessary to distinguish a specific face. This is termed sparse encoding (see Figure 8.11). Single cell recording experiments in monkey IT cortex have found that the face-selective neurons are quite tightly tuned and show characteristics consistent with sparse encoding (Young & Yamane, 1992; Abbott, Rolls & Toveé, 1996). Several studies have shown large sets of visual stimuli (including faces, objects and natural scenes) to face cells (Rolls & Toveé, 1995; Foldiak et al., 2004). Examples of some of these images are in Figure 8.12. The responses of the neurons were tuned tightly to a sub-group of the faces shown, with very little response to the rest of the faces and to the non-face stimuli. These results suggest that the cell populations or ensembles may be as small as 100 neurons.

Are face cells special?

There seem to be two levels of representation of different classes or categories of visual stimuli in the brain, which are shaped by how much information you need to derive from a particular image class. If you only have to have make comparatively coarse discriminations, such as between different categories of objects (i.e. cat vs. dog), then this may be mediated by a distributed code across populations of elaborate cells. However, if you have to make fine, within-category discriminations, such as between faces, then a population of cells may become specialised for this specific purpose.

Evidence for this approach comes from experiments in which monkeys were trained to become experts in recognising and discriminating within a category of objects sharing a number of common features. Logothetis and Pauls trained monkeys to discriminate
within two categories of computer generated 3-D shapes: wire-frames or spheroidal ‘amoeba-like’ objects (Figure 8.13). The animals were trained to recognise these objects presented from one view and then were tested on their ability to generalise this recognition. Single-cell recording from anterior IT during this recognition task revealed a number of cells that were highly selective to familiar views of these recently learned objects (Logothetis et al., 1995; Logothetis & Pauls, 1995). These cells exhibited a selectivity for objects and viewpoints that was similar to that found in face cells. They were largely size and

Figure 8.11. (See also colour plate section.) An illustration of the different patterns of activity seen in sparse and distributed coding. The blue and yellow pixel plots represent a hypothetical neural population. Each pixel represents a neuron with low (blue) or high (yellow) activity. In distributed coding schemes (left column), many neurons are active in response to each stimulus. In sparse coding schemes (right column), few neurons are active. If the neural representation is invariant (i.e. responsive to the same face independent of viewing position) (top row), then different views of the same person or object evoke identical activity patterns. If the neural representation is not invariant (bottom row), different views evoke different activity patterns. The results for face cells suggest that neural representation is extremely sparse and invariant (reproduced with permission from Connor, 2005. Copyright (2005) MacMillan Publishers Ltd (Nature)).
translation invariant, and some cells were very sensitive to the configuration of the stimuli. In short, these cells showed the same response properties as face cells, but to computer-generated object categories. In a further set of experiments, Logothetis has shown that IT cells in monkeys trained to make discriminations between different categories of related objects become sensitive to those specific diagnostic cues that allow the categorisation to be made (Sigala & Logothetis, 2002) (Figure 8.14).

These results suggest that the properties displayed by face cells can be duplicated for other object categories that require fine within-category discrimination over a sustained period of time. Face cells may only be ‘special’ because the difficulty of the task in discriminating and interpreting facially conveyed information requires a dedicated neural network. Equally difficult tasks also can produce a similar neural substrate to mediate this discrimination.

This is not to say that there are not specific regions of cortex responsive to faces. fMRI has been used to identify regions in the monkey cortex which are active in response to faces (Tsao et al., 2003; Pinsk et al., 2005). As might be expected, these are in IT and STS. The
activity patterns are not spread throughout IT and STS, but are found in discrete ‘clumps’. When the researchers then used micro-electrodes to record from neurons in these clumps, over 97% of the cells are face-selective (Tsao et al., 2006). It makes sense to ‘clump’ cells with similar response properties together, as they are likely to be communicating with each other the most. Previous studies have shown that the stimulus preferences of IT neurons are shaped by local interactions with the surrounding neurons. Wang and his colleagues (2000) recorded neural responses to a set of complex stimuli before, during and after applying bicuculline methiodide. This chemical blocked local inhibitory input to the cells from which they were recording. This blocking was to broaden the range of stimuli to which a neuron responded. The study suggests that inhibitory inputs from cells within a feature column, and surrounding feature columns, act to ‘sharpen’ the stimulus preferences of cells in IT cortex. To keep the connections short and to improve the efficiency of the brain, it thus makes sense to keep these neurons close together.
Visual attention and working memory

Despite the vast number of neurons that comprise the visual system, its ability to process fully and store in memory distinct, independent objects is strictly limited. Robert Desimone has suggested that objects must compete for attention and processing ‘space’ in the visual system, and that this competition is influenced both by automatic and cognitive factors (Desimone et al., 1995). The automatic factors are usually described at pre-attentive (or bottom-up) processes and attentive (or top-down) processes. Pre-attentive processes rely on the intrinsic properties of a stimulus in a scene, so that stimuli that tend to differ from their background will have a competitive advantage in engaging the visual systems attention and acquiring processing space. So, for example, a ripe red apple will stand out against the green leaves of the tree. The separation of a stimulus from the
background is called **figure-ground segregation**. Attentive processes are shaped by the task being undertaken, and can override preattentive processes. So, for example, it is possible to ignore a red apple and concentrate on the surrounding leaves. This mechanism seems to function at the single cell level. When monkeys attend to a stimulus at one location and ignore a stimulus at another, micro-electrode recording shows that IT cell responses to the ignored stimulus are suppressed (Moran & Desimone, 1985). The cell’s receptive field seems to shrink around the attended stimulus.

Analogous processes seem to occur within short-term visual memory. The effect of prior presentation of visual stimuli can be in either of two ways: suppression of neural response or by enhancement of neural response. Repeated presentation of a particular stimulus reduces the responses of IT neurons to it, but not to other stimuli. This selective suppression of neural responses to familiar stimuli may function as a way of making new or unexpected stimuli stand out in the visual field. This selective suppression can be found in IT cortex in monkeys passively viewing stimuli and even in anaesthetised animals (e.g. Miller, Gochin & Gross, 1991), suggesting it is an automatic process that acts as a form of temporal figure-ground mechanism for novel stimuli, and is independent of cognitive factors (Desimone et al., 1995).

Enhancement of neural activity has been reported to occur when a monkey is carrying out a short-term memory task actively, such as **delayed matching to sample** (DMS). In the basic form of this task, a sample stimulus is presented, followed by a delay (the retention interval), and then by a test stimulus. The monkey has to indicate whether the test stimulus matches or differs from sample stimulus. Some neurons in monkey IT maintain a high firing rate during the retention interval, as though they are actively maintaining a memory of the sample stimulus for comparison with the test stimulus (Miyashita & Chang, 1988). However, if a new stimulus is presented during the retention interval, the maintained neural activity is abolished (Baylis & Rolls, 1987). This neural activity seems to represent a form of visual rehearsal, which can be disrupted easily (just as rehearsing a new telephone number can be disrupted easily by hearing new numbers), but this still may be an aid to short-term memory formation (Desimone et al., 1995).

In another form of DMS task, a sample stimulus was presented followed by a sequence of test stimuli and the monkey had to indicate which of these matched the sample. Under these conditions, a proportion of IT neurons gave an enhanced response to the test stimulus that matched the sample stimulus (Miller & Desimone, 1994). Desimone has suggested that the basis of this enhanced response lies in signals coming in a top-down direction from the ventral prefrontal cortex, an area which has been implicated in short-term visual memory (Wilson et al., 1993). Like IT neurons, some neurons in ventral prefrontal cortex show a maintained firing rate during the retention interval. This maintained firing is interrupted temporarily
by additional stimuli shown during the retention interval, but the activity rapidly recovers. Desimone speculates that this maintained information about the sample stimulus may be fed back from the prefrontal cortex to the IT neurons so that they give an enhanced response to the correct test stimulus (Desimone et al., 1995). This hypothesis is supported by a recent split-brain study (Tomita et al., 1999). The appearance of an object cued a monkey to recall a specific visual image and then choose another object that was associated with the cue during training. In the intact brain, information is shared between the two cerebral hemispheres. Tomita et al. severed the connecting fibres between the two hemispheres, so that the IT in each hemisphere could only receive bottom-up from one-half of the visual field. The fibres connecting the prefrontal cortices in the two hemispheres were left intact. The cue object was shown in one-half of the visual field. The activity of IT neurons in the hemisphere that did not receive bottom-up input (i.e. received input from the hemi-field in which the cue object was not shown) nevertheless reflected the recalled object, although with a long latency. This suggests that visual information travelled from IT in the opposite hemisphere to the prefrontal cortices and then down to the ‘blind’ IT (Miller, 1999). Severing the connections between the prefrontal cortices abolished this activity in the ‘blind’ IT (Tomita et al., 1999) (Figure 8.15).

Preattentive memory processes are sensitive to stimulus repetition and automatically bias visual processes towards novel or infrequent
Attentive processes are important when we search for a particular stimulus in a temporal sequence of different stimuli. Together, these two types of process determine which stimulus in a crowded scene will capture our attention.

A similar feedback seems to function in the dorsal stream. Neurons in the posterior parietal (PP) cortex and the DL region are sensitive to the spatial relationships in the environment. There seems to be co-activation of these areas during spatial memory tasks (Friedman & Goldman-Rakic, 1994), and the reversible inactivation of either area through cooling leads to deficits in such tasks (Quintana & Fuster, 1993). Neurons in both areas show a maintained response during the delay interval, like those in the IT and IC regions, and the maintained activity in the PP cortex can be disrupted by cooling of the DL region (Goldman-Rakic & Chafee, 1994). This suggests that feedback from prefrontal areas is important for the maintenance of the neural activity in the higher visual association areas that is associated with visual working memory.

**Fine-tuning memory**

It is well known that there is extra-thalamic modulation of the cortical visual system at all levels, and that this includes the prefrontal cortex and the higher association areas (Foote & Morrison, 1986). Recent studies have concentrated on the dopaminergic innervation of the prefrontal cortex, and it has been shown that changes in dopamine levels are associated with working memory deficits in monkeys (Robbins et al., 1994). These studies have an immediate clinical relevance, as changes in the dopamine innervation of the prefrontal cortex have been implicated in working memory deficits in both Parkinson’s disease and in schizophrenia. Williams and Goldman-Rakic (1993) have established that the pre-frontal cortex is a major target of the brainstem dopamine afferents that synapse onto the
spines of pyramidal neurons. The same spines often also have excitatory synapses from the sensory inputs arriving at the prefrontal cortex, and this arrangement has the potential to allow direct dopamine modulation of local spinal responses to excitatory input (Goldman-Rakic, 1995).

Dopamine receptors of a particular sub-type (D1) are concentrated in the prefrontal cortex, primarily on the spines of pyramidal cells (Smiley et al., 1994), and iontophoresis of a D1 antagonist enhances the activity of neurons in the DL region during the intertrial periods of spatial memory tasks. These DL neurons seem to display spatially tuned ‘memory fields’ (Williams & Goldman-Rakic, 1995). The neurons respond maximally during the delay period to targets that had appeared in one or a few adjacent locations (the memory field), but they do not respond to targets in any other location. Different neurons seem to encode different spatial locations, so it is possible that a precise spatial location could be encoded by a population of neurons.

The D1 antagonist causes an enhancement of the delay activity for stimuli in a cell’s memory field, but not for any target locations outside this memory field. This effect is dose-dependent: higher levels of D1 antagonists inhibited cell firing at all stages of the spatial memory task, it did not matter whether the target stimulus was shown in the memory field or outside it (Williams & Goldman-Rakic, 1995). These results suggest that intensive D1 receptor blockade may render prefrontal cells unresponsive to their normal function inputs, and Williams and Goldman-Rakic (1995) suggest that this may be through indirect mechanisms involving inhibitory local circuits. This possibly explains the reports that deficits in working memory are produced by injection of D1 antagonists (Arnsten et al., 1994), and delay period activity is inhibited by non-selective dopamine antagonists (Sawaguchi, Matsumura & Kubota, 1990).

A clinical application?

Visual working memory thus seems to be dependent on interactions between the prefrontal cortex and the higher association areas. This activity is modulated by dopamine through the D1 receptors. D1 is merely one of a number of dopamine receptor sub-types found within the pre-frontal cortex; these have different distributions and seem to have different functions. Moreover, it seems that other neurotransmitter systems, such as the cholinergic system, may also play a modulatory role in pre-frontal memory function. This is underlined by the facts that most drugs that have been used to alleviate the symptoms of schizophrenia act through D2 dopamine receptors, and that the new wave of neuroleptic drugs used in psychiatric treatment act through serotonin (5-HT) receptors. Nevertheless, Williams and Goldman-Rakic’s results show that specific doses of
selective D1 antagonists can alter the ability of primates to carry out a memory task, and suggest that the use of antagonists or agonists selective for specific receptor subtypes, combined with electrophysiology in awake, behaving monkeys, may be the way to cut through the Gordian knot of neurotransmitter interactions in prefrontal cortex.

**Visual imagery and long-term visual memory**

Visual areas in the brain may also have a role to play in long-term memory and visual imagery. If we close our eyes and summon up the image of a particular person, object or scene, it seems that at least some of our visual areas become active. Although long-term memory is thought to be mediated primarily by the hippocampus and its associated areas, these areas all have extensive back projections, both directly and indirectly, to the visual system. Functional imaging studies (such as PET and fMRI), have shown that, in recall of objects, the higher visual areas are active, and that damage to these areas impairs recall (Roland & Gulyas, 1994; Kosslyn & Oschner, 1994; Le Bihan et al., 1993). However, there has been considerable debate about the extent of the re-activation of the visual system and whether it involves the early visual areas, such as V1 and V2. Kosslyn and Oschner (1994) have argued that mental imagery requires the activation of all the cortical visual areas to generate an image, whereas Roland and Gulyas (1994) have pointed out that, if the brain has already produced a representation of a particular stimulus in the temporal or parietal cortex, why should it need to do it all over again? The evidence from functional imaging studies for either argument has been inconclusive. Using PET, Roland reported that early visual areas do not become active (Roland & Gulyas, 1994), but many other PET studies and fMRI studies have shown activation of these areas (Kosslyn & Oschner, 1994; Le Bihan et al., 1993). Studies from brain damaged subjects are equally contradictory (see Roland & Gulyas, 1994; Kosslyn & Oschner, 1994; Moscovitch, Berhmann & Winocur, 1994). However, a transcranial magnetic stimulation (TMS) study suggests that early visual areas, such as V1, do need to be active for image recall (Kosslyn et al., 1999). TMS focuses a magnetic field on a targeted brain area, inducing electrical currents in the targeted area which transiently inactivate it (see Chapter 1). Kosslyn asked eight volunteers to compare the lengths of pictured bars, either while looking at the picture or while holding its image in memory. TMS impaired the volunteers' performance at both perception and imagery, when compared to a control condition that focused the magnetic field outside the brain, creating the same scalp sensations as TMS without affecting any brain areas. This study, taken in conjunction with the functional imaging and clinical evidence, suggests that all the cortical visual areas are active during visual imagery and recall from long-term visual memory.
Summary of key points

(1) The pattern of different luminance intensity points produced at the level of the retinal ganglion cells must be transformed into a three-dimensional representation of the object, which will enable it to be recognised from any viewing angle.

(2) Some aspects of the traditional gestalt school of perception may guide the visual system in the construction of the edges and boundaries that form the basis of the object representation. However, these seem to be automatic, rather than cognitive processes, and are implemented in early visual areas (such as in V1 and V2).

(3) The response properties of visual neurons become more complex as one moves up the visual system, and neurons in monkey inferior temporal cortex (IT), called elaborate cells, seem to be responsive to simple shapes. The elaborate cells seem to be organised into functional columns or modules, each module specifying a different type of shape.

(4) It has been suggested that the simple shapes coded for by the elaborate cells can form a ‘visual alphabet’ from which a representation of an object can be constructed.

(5) Some neurons seem to be responsive to more complex shapes than the elaborate cells; some of these neurons are the face cells, which may represent the neural substrate of face processing. These neurons also seem to have a columnar organisation.

(6) Neurons seem to comprise a distributed network for the encoding of stimuli, just as other cells in IT cortex probably comprise a distributed network for the coding of general object features. Stimuli are thus encoded by the combined activity of populations or ensembles of cells.

(7) The activity of visual neurons in monkey IT seems to be important in the maintenance of short-term visual memory. This activity is dependent at least partially on feedback projections from areas in the frontal cortex, which have been implicated in visual working memory.

(8) In visual imagery, when we close our eyes and summon up the image of a particular person, object or scene, it seems that the visual system becomes active. This activation is believed to be mediated by feedback projections from higher areas, such as the hippocampus.
Face recognition and interpretation

What are faces for?

The recognition and interpretation of faces and facially conveyed information are complex, multi-stage processes. A face is capable of signalling a wide range of information. It not only identifies the individual, but also provides information about a person’s gender, age, health, mood, feelings, intentions and attentiveness. This information, together with eye contact, facial expression and gestures, is important in the regulation of social interactions. It seems that the recognition of faces and facially conveyed information are separate from the interpretation of this information.

Face recognition

The accurate localisation in humans of the area, or areas, important in the recognition of faces and how it is organised has plagued psychologists and neuroscientists for some years. The loss of the ability to recognise faces (prosopagnosia) has been reported in subjects with damage in the region of the occipito-temporal cortex, but the damage, whether through stroke or head injury, is usually diffuse. The subjects suffer not only from prosopagnosia, but usually from other forms of agnosias too, and often from impaired colour perception (achromatopsia). However, functional imaging has allowed more accurate localisation (see Figure 9.1), and these studies have suggested that the human face recognition system in many ways mirrors that of the non-human primates discussed in the previous chapter.

The superior temporal sulcus (STS) in humans (as in monkeys) seems sensitive to the direction of gaze and head angle (cues to the direction of attention) and to movement of the mouth (important for lip reading), as well as to movement of the hands and body (Allison, Puce & McCarthy, 2000). The activation of the STS in response to these latter stimuli suggests that it is involved in the analysis of biological motion, but taken overall the response pattern of the STS suggests
that it is sensitive to the intentions and actions of other individuals (i.e. it is processing socially relevant information). The identity of the face seems to be processed in part of a separate brain area called the fusiform gyrus (e.g. Kanwisher, McDermott & Chun, 1997; Grill-Spector, Knouf & Kanwisher, 2004). This seems to be the equivalent of the face-selective area reported in the monkey’s anterior inferior temporal (IT) cortex (Kanwisher, 2006).

A study by Truett Allison and his colleagues recorded field potentials from strips of stainless steel electrodes resting on the surface of extrastriate cortex in epileptic patients being evaluated for surgery. The electrodes were magnetic resonance imaged to allow precise localisation in relation to the sulci and gyri of occipito-temporal cortex. They recorded a large amplitude negative potential (N200) generated by faces and not by the other categories of stimuli they used (Allison et al., 1994). This potential was generated bilaterally in regions of the mid-fusiform and inferior temporal gyri. Electrical stimulation of this area caused transient prosopagnosia. To confirm this result, Allison then used fMRI techniques to study blood flow during the same face recognition task and found activation of the same areas of the brain as indicated by field potential recording (Puce et al., 1995).

Additional evidence that this area is responsible for face processing comes from an elegant study that used a morphing programme to create three sets of images (Rotshtein et al., 2005). In the first set, the
two morphed images were identical (this set served as a control), the second used two different pictures of the same person (so the physical arrangement of features altered across the image sequence, but the identity of the face did not) and in the third they morphed between two pictures of different people: Marilyn Monroe and Margaret Thatcher. Although the physical features gradually altered across the sequence of images in this set, the perception of identity does not show this gradual shift, as face recognition is a categorical judgement. The face was seen as either Marilyn Monroe or Margaret Thatcher (Figure 9.2).

Pictures from these image series were shown to volunteers while they were in an fMRI scanner. The different stimulus series allowed them to disassociate the areas of the brain that dealt with the physical features of a face, and those which dealt with the identity of the face. Changes in the physical features of the faces are linked to activity in the occipital cortex, which includes the human homologues of the early visual areas, such as V1 and V2. Changes in identity are linked to activity in the fusiform gyrus, and lateralised to the right side. This suggests a specialised region in the right fusiform gyrus sensitive to facial identity.

A number of behavioural features also have been taken to suggest that face processing is unique, separate from object processing. For example, if faces are shown upside-down, then the speed and accuracy of identification by observers is reduced, relative to faces shown the right way up. A similar pattern generally is not found in object recognition. This result is interpreted as showing that inverted faces are processed and recognised on the basis of the components that make up a face, rather than as a unique pattern, and has been a ‘diagnostic feature’ of the unique nature of face recognition (Moscovitch et al., 1997). A patient (C.K.) with severe object agnosia, but unimpaired face recognition shows that there is at least some degree of separation of face processing from other object processing areas. C.K. could perform as well as controls as long as the face was upright, but if it was inverted, he was severely impaired as compared to controls. These results are neatly mirrored by those from a patient called L.H, who suffered from a selective impairment of face recognition. L.H. was severely impaired
in the recognition of upright faces, but significantly better at inverted faces (Farah et al., 1995). Moscovitch and his colleagues concluded that the face recognition was based on two mechanisms: the first recognised a face as a specific pattern under limited viewing conditions but, under conditions where this holistic system is unable to recognise a face, it is processed by a general object processing system. It is this latter mechanism that is impaired in C.K. but spared in L.H.

The area mediating face recognition seems to be in close proximity to the area mediating higher-order colour vision as prosopagnosia is frequently associated with achromatopsia. Allison and his colleagues recorded potentials evoked by red and blue coloured checkerboards (Allison et al., 1993). These potentials were localised to the posterior portion of the fusiform gyrus and extended into the lateral portion of the lingual gyrus. Electrical stimulation of this area caused significant colour effects in the patient’s visual perception, such as coloured phosphenes and less commonly colour desaturation (Allison et al., 1993). This finding is consistent with the position of lesions causing achromatopsia (Zeki, 1990), post-mortem anatomical studies of the human cortex (Clarke, 1994) and PET scan studies (Corbetta et al., 1991; Watson, Frackowiak & Zeki, 1993), and this region may be the human homologue of monkey V4.

Laterality and face recognition

There is considerable evidence from psychophysical experiments and brain damaged subjects that the left and right hemispheres process face information differently, and that right hemisphere damage may be sufficient to cause prosopagnosia. Presentation of faces to the left visual field (and therefore initially to the right hemisphere) of normal subjects leads to faster recognition than presentation to the right visual field (left hemisphere), and to greater accuracy in recognition. The right-hemisphere advantage disappears when faces are presented upside-down, and right-side damage disrupts recognition of upright faces, but not inverted faces (Yin, 1969; 1970). It seems that, in the right hemisphere, upright faces are processed in terms of their feature configuration, whereas inverted faces are processed in a piecemeal manner, feature by feature (Carey & Diamond, 1977; Yin, 1970). In the left hemisphere both upright and inverted faces seem to be processed in a piecemeal manner (Carey & Diamond, 1977). Allison and his colleagues reported that normal and inverted faces produce the same N200 pattern in the left hemisphere, but in the right hemisphere the N200 potential was delayed and much smaller in amplitude in response to the inverted face.

These findings are consistent with the clinical and neuropsychological studies, which suggest that patients with brain damage in the right hemisphere show a greater impairment on face processing tasks than patients with the equivalent damage in the left hemisphere (De Renzi et al., 1994). Although the complete loss of face recognition
capacities seems to be associated with bilateral damage (Damasio et al., 1990), there are suggestions that unilateral right-hemisphere damage might be sufficient (De Renzi et al., 1994; Sergent & Signoret, 1992). One of the most common causes of prosopagnosia is cerebrovascular disease. The infero-medial part of the occipito-temporal cortex (including the fusiform gyrus, lingual gyrus and the posterior part of the parahippocampal gyrus) is supplied by branches of the posterior cerebral arteries, which originate from a common trunk, the basilar artery. It is therefore common to find bilateral lesions when the basilar artery is affected. Moreover, when a unilateral posterior cerebral artery stroke does occur, it is common for further ischaemic attacks to occur in the cortical area served by the other posterior cerebral artery (Grusser & Landis, 1991). It is therefore not surprising that prosopagnosic patients are commonly found with bilateral lesions of the occipito-temporal cortex. However, Landis and his colleagues (Landis et al., 1988) report the case of a patient who had become prosopagnosic after a right posterior artery stroke, and who died 10 days later from a pulmonary embolism. The autopsy revealed a recent, large infero-medial lesion in the right hemisphere and two older clinically silent lesions, a micro-infarct in the lateral left occipito-parietal area and a right frontal infarct. The short delay between symptom and autopsy suggests that a right medial posterior lesion is sufficient for at least transient prosopagnosia. Although it might be argued that, in this case, some recovery of face processing ability might have occurred with time, there is also evidence of unilateral right-hemisphere damage producing long-lasting prosopagnosia. Grusser and Landis (1991) cite more than 20 cases of prosopagnosic patients who are believed to have unilateral, right-hemisphere brain damage on the basis of intra-operative and/or neuroimaging techniques. In many of these patients prosopagnosia has existed for years. Although intra-operative findings and neuroimaging techniques are less precise than autopsy results, and small lesions may go undetected in the left hemisphere, the lesion data in humans does suggest that face processing is primarily, if not exclusively, a right-hemisphere task.

Further evidence for a face-specific recognition system lateralised to the right-side comes from a functional imaging study by Truett Allison. He reasoned that, if faces were processed separately by the visual system, then if a face were seen at a time while the object recognition system is already occupied with processing a number of other objects, an additional cortical area should be activated, and in principle this additional activation should be detectable using current imaging techniques. But, if faces were processed by the general object recognition system, then no such additional activation should ‘pop out’. To test this hypothesis, Allison and his colleagues used fMRI to measure the activation evoked by faces, compared with flowers, presented in a continuously changing montage of either common objects or ‘scrambled’ objects (McCarthy et al., 1997).

This experiment was really making two comparisons. The first was between activation induced by the faces under the two montage
conditions. It was assumed that the scrambled objects would not stimulate the higher object recognition areas, but would act as controls for stimulus features such as luminance and spatial frequency. So, seen amongst the scrambled object montage, the faces should activate areas that process them both as a unique pattern (processed by the putative face-recognition area) and as a collection of shapes that make up a face (processed by part of the general object processing system). Presented amongst the object montage, the faces should stimulate the face processing area, which should not have been activated by the object montage alone. The part of the general object processing system that is stimulated by faces as a collection of objects should already be activated by the montage of objects, so only the new area activation should be that specific to faces.

The second comparison was between the patterns of activation in response to faces vs. that evoked by flowers. It was assumed that flowers would be processed solely by the general object processing system, rather than by a ‘flower recognition’ area. So, showing flowers amongst the scrambled montage should produce no difference in activation, as the general object processing system should already be fully activated.

The results seem to be consistent with this set of predictions. Bilateral regions of the fusiform gyrus were activated by faces viewed amongst the scrambled object montage but, when viewed amongst the object montage, the faces were activated differentially by a focal region in the right fusiform region. Flowers amongst scrambled objects also caused bilateral activation, but did not cause any additional activation when presented amongst the object montage. This suggests not only that face recognition involves a specialised region, but also that recognition of a face as a unique pattern is mediated by the right side of the brain. Recognition by the left side of the brain seems to occur by a piecemeal processing of the components that make up the face, rather than a processing of the whole image as a single coherent pattern.

How specialised is the neural substrate of face recognition?

It is possible to train observers to recognise and discriminate between artificial patterns called greebles, and the discrimination of these greebles shows the same inversion effect as seen in faces (Gauthier, 1999). The inversion effect is also seen in other ‘expert’ discriminations, such as for dogs amongst dog breeders. Thus, the possibility has been raised that face recognition and discrimination is mediated by an ‘expert’ discrimination system that also mediates other ‘expert’ judgements. This has been supported by some functional imaging studies. For example, functional imaging showed increased activity in the face-sensitive regions of the fusiform gyrus, as subjects became expert in discriminating ‘greebles’ (Gauthier et al., 1999,
2000a). But Nancy Kanwisher has argued that the greebles have face-like attributes and the reported activity in ‘face-specific’ regions could be due to face-selective mechanisms being recruited for expert within-category discrimination of these stimuli that share properties in common with faces (Kanwisher, 2000). However, bird experts and car experts were scanned with fMRI while viewing birds, cars, faces and objects (Gauthier et al., 2000b). The activity in a face-selective region of the fusiform gyrus is weakest during viewing of assorted objects, stronger for the non-expert category (birds for car experts and vice versa), strongest for the expert category (cars for car experts and birds for bird experts) and strongest for faces. Gauthier has argued that this illustrates that the ‘face-specific’ area is not face specific, but is part of the neural substrate that mediates any fine within category discrimination (Tarr & Gauthier, 2000).

However, as Kanwisher points out, the degree of cortical activation to non-face stimuli for the experts is comparatively small compared to faces, and several pieces of evidence suggest an anatomical separation of face processing from other object processing systems. Firstly, if the same area mediates the recognition of faces and other expert categories, then damage to the face recognition system will also impair other expert discriminations. However, a man with severe prosopagnosia was able to learn to discriminate greebles, suggesting a separate system mediates this process (Duchaine et al., 2004). Secondly, the degree of activity in the putative face-selective area in the fusiform gyrus can be correlated on a trial-by-trial basis with detecting and identifying faces, whereas the equivalent tasks for expert non-face object discrimination (such as detecting and identifying cars by car experts) activated other adjacent regions, but not the face-selective area (Grill-Spector, Knouf & Kanwisher, 2004). Thirdly, and finally, functional imaging has located specific regions in monkey IT and STS that are active in response to faces, and single cell recording in these regions shows that at least 97% of the cells are face selective (Tsao et al., 2006). These results suggest that there is a specific, anatomically discrete region of the fusiform gyrus specialised for detecting and discriminating faces.

The amygdala and fear

Although the recognition of facial identity and the configuration of facial features that signal expression seem to occur in the fusiform gyrus in humans, other brain structures may also play a role in decoding facially signalled information. The amygdala (so-called for its resemblance to an almond in its size and shape) is an area which has received a great deal of attention in this regard. It is directly linked with sensory regions on the input side and with motor, endocrine and autonomic effector systems on the output side (Amaral et al., 1992). In monkeys, bilateral removal of the amygdala produces a permanent disruption of social and emotional behaviour (part of
the Klver–Bucy syndrome). This evidence suggested that the amygdala is an important route through which external stimuli could influence and activate emotions. This hypothesis was supported by models of the functional connectivity of the primate cortex, which show the amygdala to be a focal point in the passage of sensory information to the effector areas (Young & Scannell, 1993). Neurons in the monkey STS (an area which projects strongly to the amygdala) are sensitive to the facial expression, direction of gaze and orientation of faces (Hasselmo et al., 1989; Perrett et al., 1992) and neurons in the amygdala also show selectivity to faces and features such as the direction of gaze (Brothers & Ring, 1993) (Figure 9.3).

In humans, the location of the amygdala, buried deep in the temporal lobe, means that selective damage to the amygdala is very rare. However, an example of this condition was reported by Damasio and his colleagues. They studied a woman (S.M.), of normal intelligence, who suffers from Urbach–Wiethe disease. This is a rare, congenital condition, which leads in around 50% of cases to the deposition of calcium in the amygdala during development. In the case of S.M. computed tomography (CT) and magnetic resonance imaging (MRI) scans have shown that this condition has caused a nearly complete bilateral destruction of the amygdala, while sparing the hippocampus and other neocortical structures (Tranel & Hyman, 1990; Nahm et al., 1993). S.M.’s face recognition capabilities seem to be normal. She could recognise familiar faces and learn to recognise new faces (Adolphs et al., 1994). However, when tested with faces showing six basic emotions (happiness, surprise, fear, anger, disgust and sadness) and asked to rate the strength of those emotions, she displayed a severe impairment in rating the intensity of fear relative to the ratings of normal subjects and brain-damaged controls. S.M. was then asked to rate the perceived similarity of different facial expressions (Adolphs et al., 1994). The results from normal subjects suggested that facial expressions have graded membership in
categories of emotion, and that an expression can be a member of more than one emotion category. For example, happy and surprised expressions were rated as very similar, and elements of one expression may be present in the other. The results from S.M. did not show this graded categorisation, suggesting that S.M. cannot interpret a blend of emotions expressed by a face. Instead, she categorises the expression on the basis of the prototypical emotion expressed.

These results were interpreted as showing that the amygdala was a key element in making a link between the perception of facial expression with a particular emotion in the observer. For example, an expression of fear may induce a shared feeling of fear in the observer. The amygdala would thus help to give 'meaning' to an expression and so aid in its interpretation (Adolphs et al., 1994; Hamann et al., 1996). However, further experiments have shown that, when confronted with a face, S.M. fails to look at the eyes (Adolphs et al., 2005). Instead, she looks at the nose and the mouth.

By comparison, normal observers will usually look at the eyes first, as these are very expressive (Yarbus, 1967). The fact that S.M. is impaired only in the recognition of fear is an intriguing finding. Firstly, it shows how important the eyes are in expressing this feeling, and secondly, it shows how much of the other expressions is signalled by other features such as the mouth (Vuilleumier, 2005). When S.M. is directly instructed to look at the eyes, she is able to perform the task normally (i.e. she can recognise fear). This suggests that, instead of a perceptual deficit in the recognition of fear, S.M. is impaired in her pattern of eye movements and in her allocation of visual attention (Figure 9.4).

So, what role does the amygdala normally play in face recognition? Well, neurons in the amygdala are sensitive to the direction of gaze (Kawashima et al., 1999) and experiments that have used black and
white circles to simulate eyes suggest that the amygdala responds strongly to a shape configuration simulating wide, fearful eyes (Sander, Grafman & Zeller, 1999). Patric Vuilleumier suggests that the amygdala is part of a fear detection system that targets the eyes in the face, to rapidly detect both fear in an expression and in the direction of gaze (and thus the cause of the fearful expression) and so allow a rapid reaction to a potentially dangerous situation (Vuilleumier, 2005).

Other regions in the brain also have been implicated in the processing of other facial expressions. For example, the response to expressions of disgust is an adjacent area, the anterior insular which shows preferential activation (Philips et al., 1997). The insular has been identified as the gustatory cortex in primates and plays an important role in the appreciation of both tastes and odours. This suggests that the appreciation of visual stimuli depicting other’s disgust may be closely linked to the perception of potentially aversive stimuli. This apparent separation of the neural mechanisms for the perception of fear and disgust can be explained in terms of their different evolutionary functions: fear can be seen as part of a mechanism for the detection and appraisal of danger and threat, whereas disgust plays a role in the determination of the risk of contamination and disease (Sprengelmeyer et al., 1997) (Figure 9.5).

---

**Figure 9.5.** Faces from a standard set (14) were computer-transformed (15) to create two levels of intensity of expressed fear and disgust. Examples of faces depicting 100% neutral, 75 and 150% disgust, and 75 and 150% fear are demonstrated, together with an example of a stimulus depicting a mildly happy expression (75% neutral and 25% happy) that was used as the neutral baseline (Philips et al., 1997). Copyright (1997) MacMillan Publishers Ltd (Nature).
The frontal cortex and social interaction

The frontal cortex seems to play an important role in the interpretation of visual information, such as that conveyed by faces. It has been known for many years that damage to the frontal lobes leads to a disruption of emotional and social behaviour. A particularly well-known example is the case of Phineas P. Gage. Phineas was a 25-year old construction foreman for the Rutland and Burlington Railroad Company in New England. In 1843, in order to lay new railway tracks across Vermont, the railroad company carried out controlled blasting to level the uneven ground. Phineas was in charge of this blasting. His job was to drill holes in the rock, partially fill them with explosive powder, cover the powder with sand, and then tamp it down with an iron rod before laying a fuse. On the 13th of September, Phineas made a mistake. In a moment of abstraction, he began tamping directly over the explosive powder, before his assistant had covered it with sand. The resultant explosion threw the fine-pointed, 3 cm thick, 109 cm long tamping rod up into the air like a javelin and it landed several yards away. Unfortunately, Phineas had been bending over the tamping rod and during its trajectory it passed through his face, skull and brain. Although he was initially stunned, he rapidly regained consciousness and was able to talk, and even walk away from the site of his injury, if a little unsteadily. However, subsequently he was a changed man. Prior to his injury he had been an intelligent, responsible, socially well-adapted individual. Subsequent to his injury, although his intelligence and other faculties were unimpaired, he became unreliable and offensive, with no respect for social conventions. He lost his job and led a wandering, transient existence until his death 12 years later. No autopsy was performed, but his skull (and the fateful tamping rod) were recovered for science, and have been on display at the Warren Anatomical Medical Museum at Harvard University (Figure 9.6).

A team led by Damasio has tried to calculate the brain areas damaged in the accident, based on the skull and the contemporary account of the attending physician, Dr John Harlow (Damasio et al., 1994). They conclude that damage was limited to the ventromedial region of both frontal lobes (an area called the orbitofrontal cortex or OFC), while the dorsolateral region was spared (Figure 9.5). This pattern of damage has been found in modern patients who have an inability to make rational decisions in personal and social matters, and are impaired in processing emotion. This effect is particularly marked in people who have suffered damage to this region early in life. In addition to the impairments in social interaction shown by patients who suffered OFC damage as adults, the behaviour of early-onset patients includes stealing and violent assaults (Anderson et al., 1999). In many ways the early-onset patients resemble psychopathic
individuals, who are characterised by high levels of aggression and antisocial behaviour performed without guilt or empathy for their victims (Dolan, 1999). Damasio has suggested that emotion, and its underlying neural substrate participate in decision making within the social domain, and that an important site for this process is the OFC. This region has reciprocal connections to subcortical structures, such as the amygdala and hypothalamus, which control basic biological regulation, the processing of information on emotion and social cognition. As the case of the early-onset OFC patients demonstrates, early damage to the OFC disrupts this integration and thus the acquisition of the specific forms of knowledge necessary for regulating inter-personal and social behaviour.

Faces as a social semaphore

The primate face has undergone a remarkable transformation. Its neural innervation, musculature and flexibility have increased extensively, from the almost rigid mask of some New World monkeys to the flexible, highly mobile face of the great apes, which reaches its height of sophistication and elaboration in humans. What is the point of it all? It is not simply for identification purposes: a pattern of facial features unique to an individual need not be mobile to signal his or her identity. It seems that, as primates have developed more complex social groups, the primate face has developed into a kind of semaphore system, capable of signaling a wide range of complex social information. The recognition and interpretation of these social
cues are extremely important for the smooth functioning of a social group and for an individual’s place within the hierarchy of this group. The increasing complexity of facial musculature and innervation seems to have been paralleled by an increasing sophistication of the neural representation of facially signaled information. However, there is strong evidence for a dissociation between recognition of facially conveyed information, such as identity and emotion, and the interpretation of this information. In both humans and monkeys, the recognition component seems to occur in the temporal visual cortex, whereas interpretation seems to occur in subsequent structures.

The recognition of faces in the human cerebral cortex seems to be mediated by a discrete region of the fusiform gyrus, but the function of this area is dependent on interactions with other cortical areas. Damage to the fusiform gyrus and to the area mediating face recognition can lead to prosopagnosia, and damage to the amygdala can lead to problems in social integration as characterised by the Kluver–Bucy syndrome. The role of the perception of disgust is less well known and less understood. However, people suffering from obsessive-compulsive disorder (OCD) also seem to suffer from an impairment of the perception of disgust (Sprengelmeyer et al., 1997). OCD seems to have its genesis in childhood, and it is possible that the failure to pick up social cues relating to items or activities that normally generate disgust may lead to the development of aberrant social behaviours, such as the washing or checking rituals observed in OCD. These results emphasise the importance of being able to process facially conveyed information correctly for a very social primate like humans, and the difficulties that can arise when things go wrong.

Summary of key points

(1) Damage to certain areas of the brain can cause a specific form of agnosia called prosopagnosia. This is the inability properly to process visual information on facial identity or facially conveyed information, such as expression or direction of gaze.

(2) In humans, the area of the brain which mediates face-recognition is the fusiform gyrus, and the region which mediates direction of attention, lip reading and biological motion is the superior temporal sulcus (STS).

(3) Face processing seems to show lateralisation. The right side of the human cortex seems to be specialised to process faces as a single pattern. The left side processes faces in a ‘piecemeal’ manner, that is, feature by feature rather than as a single pattern.

(4) Although recognition of faces or facial expression seems to occur in the fusiform gyrus and related areas, subsequent structures have an important influence on these processes. Of particular importance in the recognition of fear is the amygdala, whose normal functioning is necessary for the pattern of eye movement necessary for the detection of fear in a face.
The neural response to expressions of disgust is in the anterior insular. The insular has been identified as the gustatory cortex in primates and plays an important role in the appreciation of both tastes and odours. This suggests that the appreciation of visual stimuli depicting other’s disgust may be closely linked to the perception of potentially aversive stimuli.

Another important area for the interpretation of emotion in the context of social interactions is a region of the ventromedial frontal cortex called the orbitofrontal cortex (OFC). Damage to this area, such as in the famous case of Phineas Gage, disrupts the interpretation of information relating to emotion and a subject’s ability to function socially.
Motion perception

The illusion of continuity

In determining the nature of the movement of an object or scene across the retina, the visual system has to determine if the eyes are moving, the head or body is moving or the object itself is moving. To determine whether the eyes are moving, it seems that the cortical motor areas that control eye movement simultaneously send a signal to the visual system (the corollary discharge theory). For example, if the eye muscles of volunteers are temporarily paralysed, and they are asked to try and move their eyes, the volunteers report that the scene seems to jump to a new position, even though their eyes do not move and the scene does not change (Stevens et al., 1976; Matin et al., 1982).

It is important for the visual system to know about eye movements and to be able to compensate for their effects, as under normal circumstances our eyes are constantly moving. The reason for this constant movement can be found in the organisation of the retina. High acuity, colour vision is limited to the central 2 degrees of the visual field subserved by the fovea. Outside this small window, the spatial sampling of the retinal image declines sharply with increasing distance from the fovea (Perry & Cowey, 1985). Similarly, the packing of colour-sensitive cones declines by a factor of about 30 as one moves from central vision to 10 degrees of eccentricity (Curcio et al., 1991). Indeed, as you read this page, only about 16 letters are fully processed, and the rest of the text can be turned into Xs without it being noticed or without reading performance being impaired (Underwood & McConkie, 1985). This concentration on the central visual field is continued in the cortex. For example, in V1, three to six times as much space is devoted to the representation of central vision as is devoted to the periphery (Azzopardi & Cowey, 1993).

So how, from this tiny visual window (see Figure 10.1), is the high acuity, colour image of the world we believe we see constructed? It seems that, despite our impression of a stable visual image, our eyes are always moving even when we are looking at a single object in a scene. This allows all or most of the features of a scene to be brought into the high acuity centre of the visual field. Our visual image seems
to be constructed by the repeated foveation of different objects, or parts of objects, the use of short-term working memory of each snapshot and the predictive properties of the visual system (such as in ‘filling in’) (Young, 1993b). Our perception of the world is merely a ‘best guess’ of what is really out there, generated by our cortex.

**Saccades**

There are two forms of involuntary eye movement, which are called *micro-saccades* and *saccades*. When we fixate a scene, our eyes are not absolutely still but make constant tiny movements (called micro-saccades or tremors). These occur several times per second and are random in direction, and are about 1–2 minutes of arc in amplitude. If an image is stabilised artificially on the retina, eliminating any movement relative to the retina, vision fades away after about a second and the image disappears. Artificially moving the image on the retina causes the perception of the image to reappear. The neurons of the visual system rapidly adapt to a stationary stimulus and become insensitive to its continued presence. Therefore, micro-saccades are necessary to allow the perception of stationary objects (Martinez-Conde, Macknik & Hubel, 2000).

When we visually explore our environment, our eyes do not move in smooth continuous movements. Instead, our eyes fixate an object for a brief period (around 500 ms) before jumping to a new position in the visual field (see Figure 10.2). These rapid eye movements are called saccades. Saccades can reach very high velocities, approaching 800 degrees per second at their maximum. The size of the saccade is typically 12–15 degrees, but with significant numbers of both larger and smaller amplitudes (Figure 10.3).

Although a saccade may be used to foveate a moving stimulus, the eye must somehow track the stimulus subsequently as it moves through the visual field. This is done using *pursuit eye movements*, sometimes called *smooth eye movements*. Unlike the jerky saccades, pursuit movements are smooth. They are not ballistic. The neural
signals being sent to the extraocular muscles, which mediate pursuit movement, are being constantly updated and revised, allowing the speed and direction of the pursuit movements to alter with changes in the speed and direction of the target. These movements have a maximum target velocity of around 30 degrees per second. When the whole visual scene moves, then a characteristic pattern of eye movements occurs called an opto-kinetic nystagmus (OKN). An example of this situation is when you look out of the window of a moving vehicle, and for this reason OKN was once known as railway nystagmus. OKN has two components, called the fast and slow phases. In the slow phase there is a smooth pursuit of the moving field, which stabilises the image on the retina. If the velocity of field movement increases...
above 30 degrees per second, the eyes lag progressively behind, and the stabilisation is less effective. The slow pursuit phase alternates with fast, saccadic eye movements that return the eyes to the straight ahead position. The OKN seems to be a primitive form of eye movement control, designed to prevent displacement of the retinal image during locomotion.

Suppression of perception during saccades

During rapid eye movements, like those made while reading this book, you will not be conscious of a visual ‘smear’ caused by the movement of the image across the retina. During each saccade the retinal image is displaced at a speed of several hundred degrees per second, and such displacements of the image are perceived as movement if they occur when the eyes are stationary. So, why is our vision not always being interrupted by the smearing effects of the constant saccades our eyes are making, even when we observe a single stationary object? The obvious answer is that some form of suppression of the signal from the eye occurs when it makes a saccade. One can easily see this effect by looking in a mirror and changing the fixation from the image of the pupil, to the edge of the eye. The movement your eye makes is invisible in the mirror. This is not because the movements are too small or too fast to be seen, as they are easily observed when looking at another person’s eyes (Morgan, 1994). However, this is not to say that perception is entirely suppressed during a saccade. If you look at a rail track from a fast moving train, the sleepers only become visible when you make a saccade against the direction of the train, and thus briefly stabilise the image of the track on the retina.

This suppression of perception seems to be confined largely to the M-pathway (Burr et al., 1994). As the M-pathway is primarily sensitive to motion and the P-pathway is primarily sensitive to colour and high acuity, it is possible to tease out the influence on visual perception of the two systems by using the right stimuli. If one uses stimulus gratings of high spatial frequency, and which are equiluminant, but differ in colour, then their perception is unaffected by saccades (Burr et al., 1994). But, if one uses coarse gratings, containing no colour, then their perception is suppressed by saccades. Moreover, if the spectral sensitivity of a subject is measured for brief stimulus presentations either during, or before, a saccade, then the spectral sensitivity function during a saccade shows the sensitivity to wavelength expected of the P-pathway, and the spectral sensitivity function outside the saccade is that expected of the M-pathway (Uchikawa & Sato, 1995). Thus, it seems that the M-pathway is suppressed during saccades, but the P-pathway is not.

This hypothesis has been supported by an fMRI study, which measured the cortical activity evoked by a stimulus presented either
before, during or after a saccade (Kleiser et al., 2004). They found saccadic suppression of activity in several cortical areas, including V5 and V4. V5 is part of the M-pathway, and V4, although an important part of the P-pathway, also receives significant input from the M-pathway. So reduced neural activity in V4 would be consistent with suppression of the M-pathway (Burr, 2005).

So, why suppress the M-pathway and not the P-pathway? The image during a saccade is moving very rapidly, and would probably primarily stimulate the M-pathway. It is therefore important to suppress its action to allow the creation of a stable image of the external world. The P-pathway is probably not stimulated very much by the moving image during the saccade, so there has been no selective pressure to develop a mechanism to suppress its activity.

**What happens if you don’t have saccades?**

Saccades seem to be intrinsic to our perception of the world, so what would happen if we didn’t have them? The chances of someone having such a deficit are extremely unlikely, but just such an observer was reported by Ian Gilchrist and John Findlay (Gilchrist et al., 1997). A.I. was a female psychology undergraduate at Durham University where, by chance, her unusual vision was discovered by Gilchrist and Findlay, who were members of the Psychology faculty. As a result of a congenital degeneration of optic muscles, A.I. has had no eye movements since birth. However, she has no problems with reading or with any other visual task. The process of reading potentially provides a good example of how visual processes and eye movements are co-ordinated to sample the available visual information. Text is scanned in the typical saccadic manner, the eyes alternating between short rapid movements and brief fixations, during which time the gaze is stable (Gilchrist et al., 1997). Most saccades are to the right of the page, as the position of fixation jumps along the line of text. These saccades are usually seven to nine characters long, and the fixations between saccades last between 200 and 250 ms. It is during each fixation that information is gathered from the text. At the end of each line, the reader makes a return saccade to the left that places the eyes at the beginning of the next line.

So, without eye movements, how does A.I. manage to read? The answer seems to be that she uses movements of the head to compensate for the absence of eye movements (Gilchrist et al., 1997). Her head movements during reading show the same pattern of rapid ‘saccadic’ movements interspersed with ‘fixations’ (Figure 10.4). The ‘saccadic’ head movements tend to shift fixation position by approximately six characters, and the fixations themselves have an approximate duration of 200 ms. In addition, A.I. makes a characteristic large return movement at the end of each line. Her ‘head-saccades’ are not restricted to reading, but appear to be used in all viewing situations. When she views a picture, she uses ‘saccadic'
head movements to scan the scene broken by periods of fixation, suggesting that her saccadic head movements are how she usually explores a visual scene. All the movements used to move gaze position seem to have been transferred from the non-functional eye muscles to her neck muscles. Instead of her eyes being constantly on the move, her whole head is moving constantly to shift her gaze and allow her to sample the visual world.

How to stabilise the visual world

So how to stabilise your perception of the visual world when your eyes are constantly moving? Suppression of the motion sensitive M-pathway is only part of the story. Even if we don’t perceive the movement of the visual scene produced by our saccades, we still have to cope with the change in the visual scene produced by our eyes moving from one fixation point to another and integrate these snapshots of the world into a coherent percept. The basis of this stabilisation of our visual perception seems to be based at least partially on monitoring eye movement commands. As mentioned earlier in this chapter, in parallel with the motor signal generating a saccade, a second copy is sent to the visual system to be used to compensate for the eye movement. This is called the corollary discharge.

The superior colliculus (SC), a multi-layered structure located at the top of the brainstem (Figure 10.5), plays an important role in the
generation of eye movements (Carpenter, 2000). The upper layers of the SC contain a map of the visual world. The deeper layers contain motor neurons, which if electrically stimulated produce eye movements that return the position of the eye to specific positions in the visual world. The SC also generates a corollary signal that is relayed to the frontal eye field or FEF (an area known to be involved in the planning of voluntary saccades), by neurons in the medio-dorsal nucleus (MDN) of the thalamus. This signal allows visual neurons in the FEF to spatially shift their receptive fields in anticipation of the change of visual scene produced by a saccade (Munoz, 2006). So the FEF neurons compensate for the saccade before it happens, and this helps stabilise perception. It is possible to interrupt this pathway using a microinjection of muscimol (a neurotransmitter blocker) into the MDN (Sommer & Wurtz, 2002). This prevents the corollary signal from reaching the FEF. This, in turn, blocks the normal anticipatory shift in the FEF receptive fields in response to a saccade (Sommer & Wurtz, 2006) and appears to impair the accuracy with which a monkey is able to integrate the spatial position of previous fixations to form a stable perception of the visual world (Sommer & Wurtz, 2002).

Navigating through the world: go with the flow?

So far, we have discussed the motion produced by our own eye movements. An additional source of global motion is derived from our own movement through the world. This movement produces a sensation called optic flow. Like eye movements, optic flow produces visual changes that cover the entire visual field, whereas the movement of an object in the environment covers only a part of it. Optic flow has long been thought to play a role in our ability to navigate through the environment. Oncoming objects can provide cues as to where you are heading as their images seem to expand from a central point in the visual field (the focus of expansion or FoE) (Figure 10.6). The FoE indicates the target towards which you are heading. It can also provide cues about the relative position of objects in the environment as nearer objects will appear to move faster than more distant ones. Eye movements add another dimension of difficulty to
this calculation, because the FoE will not remain in a fixed position on the retina. As we have seen above, our eyes are constantly moving, drawn to some interesting scene, or object or person. Thus, to use optic flow to calculate the direction in which you are heading, your brain has to compensate for these movements.

For example, when observers sit facing an expanding image on a computer monitor, they experience the sensation of moving through space. However, if the observers are asked to move their eyes to track the movement of a marker across the screen, the observers can still identify the heading direction as signalled by the FoE on the screen (Warren & Hannon, 1988, 1990). Thus, although the FoE on the screen is shifted across their retina by eye movements, the brain is still able to compensate and calculate the correct heading. In a variant of this experiment, observers keep their eyes fixed on the same spot on the computer screen and the computer 'shifts' the screen image to simulate eye movements. This artificially alters the position of the expanding image and the FoE on the observer's retina. As there are no eye movements, the brain has to rely on purely visual cues to compensate for the movement of image. Observers see themselves travelling on a false path that curves away from the actual heading (Royden, Banks & Crowell, 1992; Banks et al., 1996). This suggests that, under normal conditions, information from the corollary discharge allows the brain to compensate for eye movements and calculate heading from the remaining optic flow information.

The neural basis of this behaviour seems to be located in a subdivision of the medial superior temporal (MSTd) part of the parietal cortex. Single-cell recording from MSTd neurons, while monkeys watched flow fields on a computer screen, shows that each neuron fires most strongly when the FoE is in a particular part of the visual field and it reduces its firing rate as the FoE is shifted away from that area (Andersen, 1997). If the FoE is shifted in the visual field by the movement of the monkey’s eyes, nearly half the MSTd neurons tested seem to compensate for the shift and their firing rate remains unchanged. However, if the movement is produced by the computer and the eyes remain still, the neurons’ firing rate changes, suggesting that the neurons have problems compensating for the movement and maintaining an accurate representation of the true heading. It seems that the neurons need the corollary discharge signal to tell
them when the eye is moving. If they fail to receive this signal, then they can no longer compensate for the shift in the FOE and calculate the correct heading.

However, these neuronal characteristics are only correlational: they show that MSTd neurons could carry out the computations required for optic flow, not whether they actually do (Wurtz, 1998). Britten and van Wezel (1998) addressed this problem by training monkeys to indicate whether the direction of heading was to the left or right. The experimenters then found that, after recording from neurons using a micro-electrode, if they then passed a small electrical current through the same microelectrode at frequencies roughly equivalent to the highest sustained discharge rate of these MSTd neurons, they could artificially introduce a signal that mimicked that produced by the neurons. This microstimulation produced an alteration in the monkey’s indication of the heading direction. When the artificially activated neurons had a preference for the left heading, the shift in the monkey’s discrimination was to the left of where it should be. This experiment provides direct evidence that the MSTd neurons contribute to the determination of heading from FOE patterns (Wurtz, 1998).

**Going against the flow?**

An alternative strategy to using optic flow to navigate our way through the world is to use visual direction (i.e. you keep looking at the place you want to go to!). This strategy requires the observer to monitor the angular direction of the target relative to his or her body. The simplest way of doing this is to keep the target straight ahead. If the target drifts off to one side, the observer merely turns his or her body to realign it towards the target. In the real world, optic flow and visual direction strategies are mutually redundant, since either source of information will allow you to reach your target. However, when displacing prisms are placed over the eyes during locomotion, it is possible to dissociate the information provided by optic flow and direction-based strategies. When a displacing prism is placed before the eye, the image of the world is shifted on the retina by an amount determined by the power of the prism (Figure 10.7(a)). The result is that objects that appear to be straight ahead, are actually located to one side. If observers use the visual direction strategy while wearing prisms, when they attempt to walk directly towards the target, they will actually transcribe a curved path (Harris & Rogers, 1999). This is because each step towards the perceived (but incorrect) position of the target will need to be followed by a re-alignment of the body towards the perceived (but still incorrect) target position prior to the next step. The displacing prisms do not change the properties of the flow filed such as the FOE, so the FOE will still coincide with the perceived target direction. So, a flow-based navigation strategy should be unaffected by prisms.

**Figure 10.7.** (a) With a prism held in front of the eye, the retinal image of an object located straight ahead will be deflected to one side (solid line indicates light path). The apparent location of the object will thus be displaced (dotted line and dotted square). (b) If visual direction guides locomotion, then the observer will walk in a curved trajectory (solid line) when viewing through prisms rather than the straight path predicted by use of optic flow (dotted line) (redrawn from Harris & Rogers, 1999. Copyright (1999) Elsevier).
When the above experiment was carried out, the results supported the use of a visual direction strategy (Rushton et al., 1998). Observers wearing prisms that deflected by 14 or 16 degrees transcribe a curved trajectory when walking towards a target. Harris and Rogers suggest that this means that, when both visual direction and optic flow information are available, the visual direction strategy is dominant (Harris & Rogers, 1999). However, they concede that optic flow is still a potentially useful source of information for the essential task of assessing the speed of locomotion, which in turn is important for the regulation of gait and perhaps more cognitive aspects of navigation such as memorising and learning routes.

The neural basis of motion detection

The perception of movement of an external stimulus can be produced in a number of different ways (see Table 10.1). Movement perception seems to be mediated by the M-pathway. The M-pathway projects to areas V3 and V5 (otherwise called middle temporal area or MT), both directly and through the thick stripes of V2. The highest visual association area in the M-pathway is the parietal cortex, where information about motion, velocity and depth integrated to form a spatial representation of the environment seems to be encoded.

Most cells in V3 are orientation selective and are believed to be concerned with processing dynamic form and 3-dimensional structure from motion (3-D SFM) (Zeki, 1993). An example of 3-D SFM can be demonstrated easily. If a piece of wire is bent into a complex, 3-D shape, and then illuminated such that it casts a shadow on a screen, an observer will not be able to determine the wire’s shape from the shadow. However, if the wire is rotated, its 3-D shape is immediately apparent (Wallach & O’Connell, 1953). Another example is the perception of optic flow and motion in depth. As we have discussed above, optic flow fields provide useful information on the direction and speed of movement through an environment.
PET scan studies have shown that the human equivalent of V3 is differentially activated by these optic flow fields (de Jong et al., 1994). V3 is part of the processing stream that allows the calculation of self-movement from optic flow fields, and V3 projects to a number of higher visual areas including MST.

V5 is an important area in the processing of visual information. It is perhaps analogous to area V4 in the P-pathway, and a considerable amount of research has been focused on its function and organisation. In monkeys, lesions of V5 cause deficits in discriminating the direction of motion, and single-cell recording techniques have shown that all V5 neurons responded better to moving stimuli than to stationary stimuli, and that most of them give the same response, regardless of the colour or shape of the test stimulus (Albright, 1984). Each V5 neuron responds preferentially to a particular speed and direction of motion. Some of these neurons show a more complex
analysis of motion. If a stimulus is made up of two gratings drifting across a VDU screen in different directions, a human observer will perceive a single plaid pattern moving across the screen in a direction that is intermediate between the directions of the two gratings which make up the pattern (Figure 10.9). We do not see the movement of two separate gratings (component motion), but the global motion of the pattern. Single-unit recording has shown that neurons in layer 4B of V1 respond to component motion, but many neurons in V5 respond to global motion (Movshon et al., 1985). That is to say, the responses of neurons in V5 correspond to our perception of motion. Figure 10.10 displays another example of component and global motion. The edges of the square, viewed in isolation, seem to move in different directions, but the global motion of the square is in a single direction. To perceive the motion of objects in our environment, it is necessary to make the jump in processing from component to global motion sensitivity.

Using single-cell recording techniques, Thomas Albright (1984) mapped the characteristics of movement-sensitive neurons in area V5. Neurons with similar preferences seem to be grouped together in columns running perpendicular to the surface of the cortex. Consistent with the organisation of many, if not all, visual areas, V5 is believed to be divided into modules or super-columns (Figure 10.11).
Each module consists of a pair of rectangles, arranged side by side. Moving along the long axis, one encounters neurons with directional sensitivities that vary systematically, in a clockwise or counterclockwise fashion. The neurons in adjacent portions of each rectangle have motion sensitivities orientated in the opposite direction.

Bill Newsome and his colleagues trained a monkey to discriminate the direction of motion of a group of moving dots. The difficulty of the task could be varied by altering the proportion of dots moving coherently to the proportion of dots moving in a random direction (Figure 10.12). Macaque monkeys were able to detect the direction of global motion when only 2%–3% of the dots were moving in the same direction (Newsome & Paré, 1988). Newsome then carried out an ingenious experiment on V5 using chemical lesions. He destroyed V5 on the left side of the brain, and left it intact on the other side. As a result, the intact V5 was able to act as a control for the lesioned V5.
The crossover of the axons from the retinal ganglion cells at the optic chiasm means that the right side of your visual field is processed by the left hemisphere and the left visual field by the right hemisphere. In the right visual field of a monkey with a left hemisphere lesion, the detection threshold for global motion was raised by a factor of ten, relative to the unlesioned side (Newsome & Paré, 1988). The monkeys’ ability to perceive stationary objects and stimuli was unaffected. Lesions to V5 do not abolish all motion sensitivity, as there are other parallel pathways processing motion information, such as through V3.

The behaviour of single, directionally-selective, motion cells in V5 have been used to predict the behaviour of the whole animal on Newsome’s motion discrimination task (Britten et al., 1992). In each session, a V5 neuron was isolated with a microelectrode and the cluster of dots was positioned on its receptive field, and the speed of the dots was matched to the cell’s preferred speed. The dots either moved in the neuron’s preferred direction, or in the opposite direction, and the monkey was required to indicate in which direction the dots moved. The monkey’s performance improved as the proportion of dots moving coherently increased. This improvement in performance was paralleled with an increase in strength of the neuron’s response if the dots were moving in its preferred direction. For roughly half the neurons tested, the proportion of dots showing coherence at which the neuron reliably signalled the correct direction closely matched the monkey’s behavioural threshold. This is not to suggest that the monkey’s decision is based on the response of a single cell, rather that the V5 neurons contribute to a population or ensemble of neurons. The activity of these populations forms the basis from which the monkey’s behaviour was computed.

Furthermore, microstimulation of small groups of neurons with the same, or similar, direction preferences alters the performance of the monkeys on the discrimination task (Salzman & Newsome, 1994). For example, if an electric current is passed through the microelectrode, which had previously been used to record from a cell, that cell is stimulated to respond as if its preferred visual stimulus had appeared in its receptive field. On the discrimination task, stimulating neurons with a preference for downward motion increased the probability of the monkey signalling that he had seen a downward movement, even when the dots were moving at random. The effect increased for currents of up to 40 micro-amps, but tended to reverse for larger currents (Murasagi et al., 1993). This has been interpreted as suggesting that larger currents activate neurons outside the targeted column, which have different direction preferences.

V5 seems to be divided into two subdivisions that analyse different aspects of motion, which seem to be related to two broad areas of function. These are the motion of an object through the environment and the motion effects caused by our own movement through the environment (Van Essen & Gallant, 1994). These subdivisions project to separate visual areas within the parietal lobe: medial superior
temporal (MSTl and MSTd). Neurons in MSTl seem to be responsive to the motion of an object through the environment, whereas as we saw earlier in the chapter, neurons in MSTd seem sensitive to motion caused by movement of our eyes or of ourselves. Neurons in this latter area are responsive to changes in certain parameters of a stimulus, such as an increase or decrease in its size (such as might be produced by moving toward or away from it), its rotation (such as might be produced when tilting our heads) and shear (such as might be produced when moving past objects at different distances) (Saito et al., 1986; Duffy & Wurtz, 1991; Orban et al., 1992). In addition, some cells are sensitive to mixtures of these stimulus parameters, such as spiral motion patterns that have components of both rotation and expansion (Graziano et al., 1994). Thus cells in MSTd seem ideally suited to encode the visual changes that occur when we move and allow us to interact with our environment.

Human V5

The position of V5 in humans has been mapped using non-invasive imaging techniques. Semir Zeki combined the techniques of PET and MRI to analyse the position of V5 in 12 normal subjects (Watson et al., 1993). The PET scanning was used to determine the position of areas of increased cerebral blood flow produced when subjects viewed a moving checkerboard pattern, compared to viewing the same pattern when it was stationary. The position of V5 based on PET was then compared with an image from the same brain obtained by MRI, allowing the position of V5 to be related to the gyral configuration of individual brains. The exact size and shape of the brain vary from individual to individual, and this is reflected in the position of V5, which can vary by 18–27 mm. However, there is a consistent relationship between the position of V5 and the sulcal pattern of the occipital lobe. V5 is situated ventrolaterally, just posterior to the meeting point of the ascending limb of the inferior temporal sulcus and the lateral occipital sulcus (see Figure 10.13). This position has been confirmed by fMRI studies and by histological studies of post-mortem human brains (Tootell et al., 1995a; Clarke, 1994).
Damage to human V5 seems to produce a similar set of deficits as lesions of V5 in monkeys. In 1983, a 43-year-old female patient (L. M.) was reported with brain damage, who shows a severe impairment in the perception of motion (akinetopsia), although otherwise her visual perception seems normal (Zihl et al., 1983). She has normal acuity, stereo and colour vision, and has no impairment of visual space perception or of the visual identification of shapes, objects or faces. A high resolution MRI scan has shown that L. M. has bilateral damage to V5 (Shipp et al., 1994). Although L. M. does not have complete loss of motion perception, and can detect the presence of slowly moving objects, her perception of motion is generally impaired and she describes moving objects as undergoing episodic shifts in location. A graphic example is her perception of water being poured into a glass from a glass jug. She can see the glass and the jug, but cannot see the water and the change in water level in the glass until water has stopped being poured. A similar effect can be produced by temporarily inactivating V5 in normal human subjects using transcranial magnetic stimulation (TMS). In this process a magnetic field is used to induce an electric current in a specific area of the brain. This temporarily inactivates the brain area and, when this technique is applied to human V5, the result seems to be complete akinetopsia (Beckers & Zeki, 1995).

Activity in human V5, like that of monkey V5, can be correlated with perception of motion effects by the individual. The M-pathway is largely colour blind. If a subject views a moving grating made up of alternating red and green stripes, as long as the stripes differ in luminance, motion will be perceived. However, if the two colours are equiluminant, and the only cues to the existence of moving stripes are colour differences, then the perception of motion is strongly reduced. These changes correlate with activity in V5. Single-unit recording in monkeys and fMRI studies in humans have shown a reduction in the activity of V5 when the two colours

Figure 10.14. (See also colour plate section.) The illusion called Leviant's Enigma. Fixation of the centre will result in the perception of rotatory motion in the circles (reproduced with permission from Zeki, 1994. Copyright (1994) Royal Society).
become equiluminant (Tootell et al., 1995b). The activity of V5 also can be correlated with the perception of motion in visual illusions. In illusions such as the enigma illusion (Figure 10.14) or the waterfall illusion (see Chapter 5), there is no actual motion in the stimulus, but a human subject perceives movement as occurring. When the subject perceives motion, whether by actual movement or through illusion, V5 is active (Zeki, Watson & Frackowiak, 1993; Tootell et al., 1995b).

Summary of key points

(1) The structure and organisation of the retina means that only the central 2 degrees of the visual field are fully elaborated. To produce our perception of the world, the eyes have to be moving constantly. There are two types of involuntary eye movement: micro-saccades and saccades.

(2) Micro-saccades are small movements of the eye that destabilise the image on the retina and prevent the photoreceptors in the retina from adapting to a continuous stimulus.

(3) Saccades are short, jerky movements of much larger size than micro-saccades and are used by the eye to explore the visual environment. During saccadic movement there seems to be a suppression of activity in the M-pathway, which is normally sensitive to motion.

(4) A third form of eye movement is called pursuit eye movement. This is under voluntary control and allows us to track moving objects.

(5) Perception of external motion is analysed mainly within the M-pathway. An important stage in this pathway is Visual Area 5 (V5), whose neuronal responses to moving stimuli can be correlated with our own perception of these stimuli. The neurons in this area are arranged in a columnar and modular fashion, as seems to be the case in most, if not all, visual areas.

(6) Damage to, or temporary inactivation of, V5 in humans causes deficits in the perception of motion (akinetopsia). Human V5 is active when we perceive motion, whether this perception is caused by real movement or by illusion.
Brain and space

The final frontier

The perception of depth is essential to the generation of a three-dimensional representation of the spatial relationships in our surroundings; a representation which is essential if we are to be able to interact with our environment in any meaningful way. The visual system has two sets of depth cues: oculomotor and visual (Figure 11.1). They are termed cues because they must be learnt through association with non-visual aspects of experience. Oculomotor cues are based on the degree of convergence (a measure of the angle of alignment) of the eyes and the degree of accommodation (change in shape) of the lens. The visual cues can be both monocular and binocular. The monocular cues include interposition, relative size, perspective and motion parallax. Binocular cues are based on the disparity between the different views of the world from the two eyes. From this disparity, a three-dimensional or stereoscopic representation can be generated. The information on depth, together with information about movement and velocity, seem to be integrated with information from other sensory modalities to produce a map of perceptual space which is common to all our senses. This integration seems to occur in the posterior parietal cortex. Damage to this area causes profound impairments in our perception of space, including that occupied by our own bodies.

Oculomotor cues

When you fixate an object, your eyes are *accommodated* and *converged* by an amount dependent on the distance between you and that object (Figure 11.2). To be seen clearly, close objects need more accommodation and convergence than do objects further away. By monitoring the degree of muscle strain, it is possible to determine values for the angle of convergence and the amount of accommodation. When you fixate an object more than a few metres from you, the muscle controlling accommodation is in its most relaxed state. So, as a potential
depth cue, accommodation would be useful only within the region immediately in front of you. Even within this region it is very inaccurate. This is true of most binocular vertebrates. The most striking exception is the chameleon (*Chameleo jackson*). The state of focus of the eye’s lens acts as the principal distance cue, and not only is the focusing mechanism of the chameleon lens very fast and its range unusually extensive, it is also accurate enough to supply the muscles of the projectile tongue with all the information they need to shoot the sticky tip exactly the right distance to catch insect prey (Ott & Schaeffel, 1995). Part of the reason for this reliance on monocular cues may come from the way the chameleon’s eyes seem to move independently and so limit the use of binocular cues.

The use of convergence can operate only over a limited range. The convergence angle formed by your two eyes vanishes to zero (because your eyes are looking straight ahead) when you are looking at objects more than 6 metres away. Below this distance, convergence can be used as a reliable depth cue.

### Interposition

When one figure occludes part of another (*interposition*), the partially occluded object is perceived as the more distant. The use of these cues develops, such that, by 7 months of age, human infants can judge relative distance solely on the basis of interposition. Young children often use interposition in their simple drawings, even though they are unable to reproduce any other pictorial depth cues. There is evidence that children as young as 7 months old determine distance solely on the basis of interposition (Yonas, 1984). Brain damage can remove the ability to use this cue, while leaving the ability to use other depth cues intact (Stevens, 1983).

A striking example of interposition is the *Kanizsa triangle*, where distinct edges seem to be found where none exist (Figure 8.2). The visual system seems to employ a kind of interpolation process where
separate edges and contours in the same spatial neighbourhood are perceived to be connected if a connection can be formed by a simple line or curve and if the operation is consistent with the principle of interposition. This interposition is accomplished by neurons in V2 (Peterhans & von der Heydt, 1991). Although neurons in both V1 and V2 can respond to illusory contours defined by the co-linearity of line terminations, only neurons in V2 responded to illusionary contours extending across gaps. Additionally, even though we are unable to see the portions occluded, we assume the occluded objects are complete and do not have the occluded part missing. This process, which is called amodal completion, is also mediated by interpolation.

Relative size

As the distance between yourself and an object varies, the size of the image of that object varies on your retina. This is true of any object viewed from any angle. So, if you are familiar with the size of an object, you can judge from the size of its retinal image how far away the object is. Also, if a second object appears in the vicinity of the familiar object, one can judge the size of the second object by reference to the size of the familiar object. This scaling effect is used by building companies in show homes. The furniture is built specially to be 10% smaller than the standard size and, as you scale the proportions of the unfamiliar rooms with reference to the relative size of the furniture, you will over-estimate the size of the rooms. It helps give an impression of size and spaciousness seldom found in modern buildings, and more importantly as far as the builders are concerned, helps sell the houses.

Perspective

Perspective refers to changes in the appearance of surfaces of objects or surfaces as they recede into the distance. There are four forms of perspective cue. Linear convergence refers to the way parallel lines seem to converge with distance, and is a way of giving the impression of depth in pictures and illustrations (Figures 11.3 and 11.4). Another form is called texture gradients. Most surfaces have a texture, and the density of this texture will appear to increase with viewing distance. Texture gradients can therefore provide information on the distance and slope of surfaces, as well as information on the size of objects located on those surfaces. Moreover, rapid changes in texture gradients can signal the presence of edges or corners. The third form of perspective cue is called aerial perspective. This refers to the way objects further away seem less clear than those close up. This is due to the scatter of light as it travels through the atmosphere, which has the result of reducing the contrast of the image (O’Shea et al., 1993). The degree of scatter is dependent on two factors; the distance between the object and the observer and the medium through
which the light passes. For example, if the air contains a lot of dust or a lot of moisture droplets such as in a fog, more light will be scattered. The final perspective cue is shading. In the natural environment, light always comes from above, and so the pattern of shading can be used to derive depth (Ramachandran, 1988) (Figure 11.5).
Motion parallax

As you move in the environment, the objects around you are constantly altering their position within your visual field. If you are travelling in a car or train, and you fixate a particular object within the passing scene, the pattern of relative movement around this object will demonstrate a phenomenon called motion parallax. Objects closer than the one you have fixated will appear to move in the opposite direction to yourself, whereas more distant objects will appear to move more slowly but in the same direction as yourself. The relative apparent motion of objects within your field of view as you move (motion parallax) provides a strong cue to the relative distance of objects from the observer.

Stereopsis

At close range, animals with overlapping visual fields have stereoscopic information available to them from the disparate images obtained at the two eyes. Each of the eyes sees a slightly different view of the world due to the horizontal separation of the two eyes. This binocular disparity is generated in the following manner. If the two eyes are focused on an object or point (B), the images in each eye are said to lie on corresponding points on the two retinas (Figure 11.6). The images cast by a nearer or more distant object or point (A and C) will fall on disparate points on the two retinas, and the amount of disparity will depend upon the distance between A and B, and between B and C. If the brain can compute this disparity, it will give precise information about the relative position of objects in the world. Stereopsis (whose literal meaning is solid appearance) requires binocular retinal disparity of some elements of a visual stimulus. The disparity effect can be demonstrated easily using a stereoscope, which was first developed by Sir Charles Wheatstone in 1838. This device presents two drawings (which together constitute a stereogram) separately to the two eyes. The two images are seen as one, and appear in 3-D. Wheatstone’s results demonstrate that re-introduction of retinal disparity gives the appearance of 3-D.

Neurons responsive to binocular disparity can be found in a number of visual areas in primates, including V1, V2, V3, V5 and IT neurons, but it is in V1 that disparity selectivity first occurs, where signals from the two eyes first converge on single neurons (Poggio & Fischer, 1977; Poggio et al., 1988). There seems to be four basic types of neuron. Tuned-excitatory neurons respond optimally to objects at or very near the horopter (i.e. zero disparity), whereas tuned-inhibitory neurons respond at all disparities except those near the horopter. Near cells respond best to objects that lie in front of the horopter, whereas far cells have a preference for objects that lie behind the horopter. However, individually these neurons do not signal depth.
unambiguously, although combining the outputs of many V1 disparity neurons may potentially do so (Cumming & Parker, 1997). In addition to contributing towards the perception of depth, the disparity signals may also be important for the rapid involuntary control of vergence eye movements (i.e. eye movements in which the images on the two foveae are brought into register) (Masson, Busettini & Miles, 1997).

The neural basis of three-dimensional space representation

Depth and motion cues can be used, in conjunction with other sensory cues, to produce a three-dimensional representation of our environment. The posterior parietal cortex (PPC) is the most likely site for such spatial relationships to be represented in the brain. PPC cells receive visual, auditory, somatosensory and vestibular sensory inputs, and the integration of these inputs could be used to form a map of perceptual
space. This is a map of the location of objects around us with respect to ourselves which is common to all our senses.

The PPC is composed of the superior and inferior parietal lobules. In humans damage to the superior parietal lobule causes deficits in complex somaesthetic judgements. These deficits can manifest themselves in several ways. A subject may be unable to recognise the shape of objects by touch alone (astereognosis). This symptom is often linked to a more general deficit of body image, where a subject is unable to assimilate spatial impressions of the positions of their limbs and body to build up an accurate body image (amorphosynthesis). A related disorder is asomatognosia, which is the denial of the existence of a part of the body. A patient suffering from this condition may deny that one of his limbs belongs to him or her.

The representation of space in the PPC seems to be divided into three categories: personal, peripersonal and extrapersonal space (Halligan et al., 2003). Personal space is the space occupied by our own body. Its co-ordinates are based on the orientation of the head, signalled by the vestibular system, together with information about the position of the neck and limbs. This information is primarily represented in the superior parietal lobule. Peripersonal, or near space is the space surrounding us, within which we can reach out and touch objects. Its neural representation requires the integration of retinal foveal signals with oculomotor and limb movement information, which occurs within a sub-division of the inferior parietal lobule. Extrapersonal, or far space, is the space beyond peripersonal space. Its representation requires the integration of visual, auditory, oculomotor cues and whole body signals. This again seems to be represented in a sub-division of the inferior parietal lobule. The three forms of spatial representation seem to be functionally and anatomically separate. For example, neglect of personal space can occur without neglect of peripersonal space and vice versa (Guariglia & Antonucci, 1992; Cocchini et al., 2001). A similar double dissociation has been reported between peripersonal space and extrapersonal space (Halligan & Marshall, 1991; Vuilleumier et al., 1998).

The problem of visual neglect

Damage to early visual areas, such as V1, cause scotomas (holes in our visual perception). Damage to higher association areas in the ‘what’ pathway, such as IT in monkeys or the fusiform gyrus in humans, impairs the ability to recognise objects (agnosias). Equally, damage to the higher regions of the ‘where’ pathway does not produce scotomas in our visual field, but has a more subtle effect. Small lesions on one side of the cortex reduce the accuracy of localising objects on the contralateral (opposite) side. The effect is most marked with lesions of the right side. As the size of the lesion increases, so does the magnitude of effect, until at a certain point the whole of the contralateral side is ignored or neglected. For example, patients may often
dress or shave only one side of the body, draw one side of a picture and attend to only one half of space, both near and far (Andersen, 1989) (see Figure 11.7). Unilateral neglect can also extend to imagery and memory. For example, Milanese patients with right cortical lesions were asked to imagine they were standing at one end of the Piazza del Duomo facing the cathedral and describe from memory the buildings along the sides of the piazza (Figure 11.8). The patients

**Figure 11.7.** An example of picture copying by a hemi-neglect patient. The pictures on the right were copied by the patient based on the pictures on the left (reproduced with permission from Bloom & Lazerson 1988 Copyright (1988) W. H. Freeman & Co.).

**Figure 11.8.** (See also colour plate section.) The Piazza del Duomo in Milan (redrawn with permission from Kandel, Schwartz & Jessel, 2000. Copyright (2000) McGraw-Hill).
described only the buildings on the side contralateral to the intact cortical lobe (Bisiach & Luzzatti, 1978). However, when asked to imagine they were standing on the other side of the piazza with their backs to the cathedral, they then described the opposite set of buildings. Similarly, patients suffering from neglect of the left side of space were unable to spell the beginnings of short words, as if spelling involved reading from an imaginary screen, the left side of which was faded (Baxter & Warrington, 1983).

Bisiach and his colleagues (1985) tested whether the boundary of neglect moved with the retinal field or whether it was anchored to head or body. Their conclusion was that their patients used at least two co-ordinate systems: one relating to the body axis and the other relating to line of sight (oculomotor). Hence, neglect is seldom purely retinotopic. The space they ignore does not move each time they move their eyes, but tends to be centred on a point passing through the centre of the body or head (the egocentre).

There is a suggestion that patients with unilateral neglect have a degree of ‘unconscious’ perception of objects within their neglected field, although they may deny this. For example, a patient with right cortical damage (and therefore left field neglect) performed a forced-choice task in which he was asked to decide whether there were differences between two houses, one with flames coming out of the left hand window. The patient’s ability to discriminate between the two images was not above chance in this task, but if asked which house he preferred, he consistently chose the house without flames, suggesting an unconscious processing of the differences between the houses (Marshall & Halligan, 1988; Bisiach & Rusconi, 1990).

The neural basis of neglect

Studies of visual function in patients with cortical damage have suggested that the neural basis of visual neglect is centred on damage to the inferior parietal lobule (Stein, 1992; Halligan et al., 2003). Damage to this region has been associated with disruption of spatial vision and spatial orientation as well as with spatial neglect. Symptoms include deficits in reaching and pointing to visual targets, avoiding obstacles, learning and remembering routes, judging distance and size, recognising spatial relations, fixating a target and following a moving stimulus.

In humans, left neglect is much more common than right. Hemispheric specialisation has led to a concentration of visuospatial functions in the right PPC. Since lesions of the left PPC seldom give rise to neglect, it seems likely that the right PPC duplicates partially the spatial functions of the left PPC for the right hemifield. This inference is supported by the fact that patients with right PPC damage and left neglect often show some degree of inattention to targets even up to 10 degrees in their right (ipsilateral) field. This result is supported by the finding that, in monkeys, many parietal
neurons have receptive fields that extend well into the ipsilateral field, suggesting that more neurons in the right PPC have bilateral receptive fields than neurons in the left PPC. John Stein has suggested that the right PPC has become specialised for representing three-dimensional space and the direction of attention within this space, and the left PPC has become specialised to directing attention to temporal order (Stein, 1992). Human attributes such as speaking, logic and calculation require the skill of being able to sequence events in time accurately, and these attributes are impaired most by lesions of the left side.

However, patients with lesions in this area tend to have damage that extends beyond PPC and also affects adjacent regions, such as the superior temporal gyrus, and it has been suggested that the neglect is not actually due to damage in PPC. Karnath et al. (2001) studied patients with damage clustering in the right PPC and patients with damage to right superior temporal gyrus. Karnath interpreted his results as suggesting that patients with hemi-neglect, but no blindness, had lesions centred on the superior temporal gyrus, whereas patients with just PPC lesions showed blindness in parts of the visual field, but no visual neglect. However, this interpretation has been challenged by Peter Halligan and John Marshall, who insist that the PPC patients also showed neglect, and that a more accurate conclusion is that damage to the superior temporal gyrus in addition to other areas can cause neglect (Halligan et al., 2003). It is likely that the neural basis of neglect comprises a number of interconnected cortical and subcortical areas, and that damage to this network can cause visual impairments including neglect (Figure 11.9).
Summary of key points

(1) The perception of visual depth is based both on oculomotor and visual cues. Oculomotor cues are based on changes in accommodation and convergence when the eye focuses on an object.

(2) There are both monocular and binocular visual cues. The monocular cues include:
   (i) Interposition, which is when one object partially occludes another, and the occluded object is perceived as being more distant. The occluded objects are assumed to be complete, a process called amodal completion.
   (ii) Relative size, which is where smaller objects are assumed to be more distant than larger objects.
   (iii) Perspective, which is the change in the appearance of surfaces or objects as they recede into the distance.
   (iv) Motion parallax, which uses the pattern of relative movement around an object to derive a measure of depth.

(3) Binocular cues are based on the slightly different views of the world which the two eyes see due to their horizontal separation. The difference in the two views is called disparity and it is used to calculate a three-dimensional or stereoscopic representation of an object.

(4) Sensory information on depth and space is integrated in the posterior parietal cortex (PPC), an area that is important for matching sensory input to motor output to allow us to interact with the world.

(5) Damage to the superior parietal lobule produces an impairment of body image.

(6) Damage to the inferior parietal lobule leads to deficits in visuospatial perception.

(7) The representation of space is divided into three categories: personal, peripersonal and extra personal space. Personal space is the space occupied by your body, peripersonal space is the space within reach and extrapersonal space is the space beyond peripersonal space.

(8) People with damage to the brain in the region of the PPC show a phenomenon called visual-neglect (‘ignoring’ part of the visual field). This has been ascribed primarily to damage to the inferior parietal lobule.
What is perception?

Putting it all together

As we have seen in previous chapters, visual information is broken down into its components and processed, in parallel, in specialised areas, so that cells in different areas show a preference for different combinations of, for example, colour, motion, orientation, texture, shape and depth. This is all carried out in a complex network of 32 visual areas connected by at least 305 connections (Van Essen, Anderson & Felleman, 1992). These connections can run in three ‘directions’. Firstly, from lower areas (such as V1) to higher areas (such as V2). These are called feed-forward connections. Secondly, all these feed-forward connections have reciprocal feedback connections running from higher to lower areas. Thirdly, there are also lateral connections running from areas of equivalent processing complexity. In addition to these problems, there are the differences in how fast different visual parameters are processed (Zeki, 2003). For example, location is perceived before colour, and colour is perceived before motion and orientation (Zeki & Moutoussis, 1997; Pisella et al., 1998). It seems a far from trivial task to re-integrate all of this information from this complex spatial and temporal network into the seamless, coherent perception of the world we all share.

There are two obvious problems. Firstly, we have to put all the different visual features of an object back together in the right spatial and temporal relationship to one another. Secondly, we seldom see a single object in isolation. How, when we are dealing with the perception of two or more objects simultaneously, do we differentiate which features belong to which object? Failure to carry out this task correctly is called the superposition catastrophe.

Neuronal oscillations

In 1981, Rudolf von der Malsburg noticed a problem similar to those outlined above with neural network models. These simulations had
the serious drawback that they could only represent the presence of two or more features in their input, having no means by which to segregate and bind the different features of two objects that were presented at the same time. The different features of the two objects therefore would be confused, producing non-existent feature conjunctions. Von der Malsburg suggested that a binding mechanism was required, by which the different features of different objects could be associated separately. He suggested that the timing of the neural action potentials or spikes could be used in a temporal code to show which features were associated with which object. The idea of a temporal code seemed like a good idea: temporal relations between the firing of sensory neurons might provide a solution to the problem of binding the activity distributed both between and within cortical areas into coherent representations. The case for a temporal code in feature binding received considerable support when experimental results on neuronal oscillations in cat visual cortex were reported.

Two research groups in Germany made micro-electrode recordings from the visual cortex of cats and found that neurons in several visual areas of the cat exhibited short-duration periods when their activity was oscillatory (Eckhorn et al., 1988; Gray & Singer, 1989) (Figure 12.1). These oscillations were said to be ‘stimulus-related’ because the frequency spectrum during stimulation showed peaks in-between 30 and 70 Hz, in contrast to the low frequency peaks (1 to 30 Hz) of background activity. It was suggested that these phenomena related to von der Malsburg’s idea of a temporal code: neurons responding to different features of the same object could oscillate in phase, while cells responding to the features of different objects would oscillate out of phase. Both groups demonstrated that synchronisation of oscillatory activity could occur (Eckhorn et al., 1988; Gray et al., 1989; Gray & Singer, 1989), and showed also that the likelihood of phase coherence declined with the separation of the two electrodes, and with the dissimilarity in the orientation preferences of the recorded neurons (Gray et al., 1989). Since then, however, attempts to replicate and extend these findings have had mixed results, and the empirical evidence remains controversial and incomplete (e.g. Lamme & Spekreijse, 1998; Golledge et al., 2003; Palanca & DeAngelis, 2005).

Additionally, a number of questions have been raised about whether it is theoretically possible for synchronisation to act as a temporal binding mechanism. As Shadlen and Movshon (1999) point out, the synchrony theory is not a theory of how binding is achieved, but only a theory of how binding is signalled. Synchrony could be used to label the cells in an ensemble, but how is the visual system to decide which cells need to be labelled as processing parts of a single object? Shadlen and Movshon suggest that this would probably require information from higher visual areas that could take a more ‘global’ view of the available information, a suggestion supported by the perceptual deficits found in observers with damage to posterior parietal cortex (PPC) (see Chapter 11). Studies on a subject
with bilateral PPC damage (see Figure 12.2), suggest that PPC damage significantly impairs both the ability to see more than one object at a time (simultagnosia) and the ability to bind the elements of an object into a coherent percept (Friedman-Hill et al., 1995, 2003). This disruption of the ability to bind more than one set of features into a coherent percept of an object is suggestive of an impairment in a putative binding mechanism. By contrast, damage to lower visual areas produces a range of problems (as we have seen in earlier

**Figure 12.1.** An example of how the proposed synchronisation of neural responses are influenced by visual stimuli. Multi-unit activity was recorded from two sites in cat V1 that were separated by 7 mm. The cells responded most strongly when the stimulus bar was at a vertical orientation. (A), (B), (C) show plots of the receptive fields (shown as hatched rectangles) under three different stimulus paradigms. (A) A long continuous light bar moving across both fields. (B) Two independent light bars moving in the same direction. (C) The same two bars moving in opposite directions. The circle represents the centre of the cat’s visual field, and the black line drawn across each receptive field indicates the preferred orientation of those cells. (D), (E), (F) show correlograms obtained with each stimulus paradigm. Using the long light bar, the two oscillatory responses were synchronised as indicated by the strong modulation of the crosscorrelogram with alternating peaks and troughs (D). If the continuity of the stimulus was interrupted, the synchronisation became weaker (E), and it totally disappeared if the motion of the stimulus was not coherent (F). The change in the stimulus configuration affected neither the strength or oscillatory nature of the two responses. The graph superimposed on each of the correlograms represents a Gabor function that was fitted to the data to assess the strength of the modulation. The number in the upper right corner indicates what Singer and his colleagues call the ‘relative modulation amplitude’, a measure of the correlation strength that was determined by calculating the ratio of the amplitude of the Gabor function to its offset. Abbreviation: ns, not significant. Scale bars indicate the number of spikes (reproduced with permission from Engel et al., 1992. Copyright (1992) Elsevier Trends Journals).
chapters), but none seems to show characteristics suggesting a failure of feature binding (Shadlen & Movshon, 1999). This pattern of deficit might suggest that synchronous activity recorded from cells in the early visual areas like V1 would be based on feedback from these higher visual association areas. But the question then is why would the visual system want to label the V1 cells like this when it has already worked out the answer? It would seem a redundant exercise.

How else to solve the problem

The need for a temporal binding mechanism is based on the assumption that the visual scene we perceive comes from a uniform input of information across the whole 180 degrees of visual field. However, in primates high acuity is supported by only the central 2 degrees of the retina, based on the fovea. The packing of the cone cells that mediate both high acuity and colour vision declines rapidly with increasing eccentricity. The cone density decreases by a factor of 30 between the centre of the fovea and an eccentricity of 10 degrees (Curcio et al., 1991). The number of photoreceptors synapsing onto a single ganglion cell also increases with increasing eccentricity (Perry & Cowey, 1985). As a result, the ‘grain size’ of the picture increases with increasing retinal eccentricity. This is also true of colour, where colour discrimination thresholds rise as an exponential function of eccentricity (Nagy & Wolf, 1993). These responses are reflected in the relative proportions of neural space given over to processing the input from central and peripheral retina. The representation of the central visual field in the primary visual cortex is up to six times larger than that for the periphery (Azzopardi & Cowey, 1993). The cumulative result of this organisation is to create almost a tunnel vision effect, where only the visual information from the centre of the visual field is fully sampled and analysed. This is accentuated by the effects of attention. If a monkey’s attention is focused on a stimulus, then the response of a visual cell to that stimulus is not enhanced, but the cell’s response to other stimuli within its receptive
field is reduced (Moran & Desimone, 1985). The net effect of attention seems to be to reduce the functional size of a cell’s receptive field.

So, how are the different features of the visual input, limited though they are by the physical constraints of the visual system, coherently integrated within the cortex? Anatomical and functional modelling studies have shown that there is considerable reconvergence of processed visual information in the frontal lobe, the rostral STS and the limbic system (Van Essen et al., 1992; Young, 1992; Young et al., 1995). We have already seen how the response properties of visual neurons have become more and more complex. Thus, integration could be brought about by hierarchical convergence, rather than by the synchronisation of populations of neurons spread widely in the brain (Golledge, Hilgetag & Tovee, 1996). This integration would be aided by attentional mechanisms including those mediated by the parietal cortex. As we have seen, damage to the PPC disrupts the ability to view more than one object at a time, and a normal functioning PPC seems necessary for the ability to bind the elements of an object into a stable percept (Friedman-Hill, Robertson & Triesman, 1995). This includes the phenomenon of perceptual learning. People can reconstruct meaningful objects accurately out of fragmentary or ambiguous evidence. However, naive subjects often experience a long latency before the object is recognised, unless they have been previously exposed to an unambiguous version of the object in the appropriate context. Once they have recognised the unambiguous version of the object, even the ambiguous version of the stimulus can be recognised almost instantly (Tovee, Rolls & Ramachandran, 1996). This represents a striking example of rapid perceptual learning, which seems to be implemented by modifying the responses of neurons in the visual system. If one records from IT neurons in a monkey performing the task outlined above, one finds that face-selective neurons do not respond initially to an ambiguous face but, once they have seen the face in an unambiguous situation, they give an enhanced response to subsequent presentations of the ambiguous face (Tovee, Rolls & Ramachandran, 1996). The responses of the neurons mirror the responses of the observer (Figure 12.3).

A study of human subjects performing this task while undergoing PET neuroimaging also finds enhanced activity in the anterior fusiform gyrus to the ambiguous image following the cueing presentation of an unambiguous image (Dolan et al., 1997). This area corresponds to the face area, and the enhanced activity would be consistent with changes in the response properties of neurons in this area corresponding to perceptual learning. Also active are the PPC and, during the learning phase of the experiment, the medial parietal cortex (Dolan et al., 1997). As we have seen, damage to the lateral parietal cortex causes severe impairment of spatial attention and feature binding, and the medial parietal cortex has been implicated in memory-related imagery (Fletcher, 1995). It seems logical that these areas should be involved in the reconstruction of a face from fragmentary sensory information. The PPC seems to be vital for
assembling a recognisable image from the disjointed parts of the ambiguous image, and the medial parietal cortex vital for providing a template or plan on which to base this construction. These processes result in a long-term, possibly permanent, change in the response properties of the face-selective neurons in the fusiform gyrus, facilitating subsequent recognition of the ambiguous image.

What is perception?

This may seem an easy question to answer. It is an internal representation of our external environment. But is it really so simple? Do we actually see the external environment with any accuracy? And, as perceptual learning shows, the nature of our representation may change even when there has been no change in the external stimulus. Perception seems to be less a representation of our environment, but rather an interpretation, and our interpretation may change based on cognitive rather than perceptual factors. The visual stimulus does not have to change for our perception to be transformed. Two phenomena that illustrate the artificial nature of our internal perception are change blindness and perceptual rivalry.

Change blindness

As we have seen in Chapter 10, we can only perceive a very limited part of a visual scene in detail and colour at any one time. Our eyes are moving constantly to sample different parts of the scene. The logical conclusion of this is that it should be possible to change parts of the scene that are not being attended to, and we should not notice. A phenomenon called change blindness. In the real world, there are some clues to potential changes as, although the rod-dominated
periphery of the retina is not able to support detailed colour perception, it is sensitive to motion. Any change in the scene would usually be accompanied by motion, such as a ball bouncing in to view, and this would guide a saccade to fixate this change in the visual scene. However, such a change will not be detected during a saccade (when the motion pathway is suppressed) or the change can be produced artificially when viewing a scene on a computer screen. This effect is particularly marked if there is a break or some brief interruption to our viewing of a scene. A common paradigm is to show a scene, followed by a blank screen for around 100 ms, and then a return to a modified version of the original scene (Rensink et al., 1997). The visual disruption provided by the blank screen masks the change in the visual scene. By flickering continuously between the images, it is possible to measure how long the observer takes to notice the change (McConkie & Loschky, 2003). Under such conditions, even quite extensive changes to a visual scene are difficult to detect. In case you feel that this situation is too artificial, and that change blindness has very little relevance in the real world, a number of more dramatic examples are also available. When an experimenter asked a passer-by for directions, their view of each other was interrupted by two men walking between them carrying a door (Figure 12.4). During this interruption, the experimenter swapped places with a colleague of different appearance and dress. Surprisingly, only half the passers-by tested noticed this quite dramatic change in the person to whom they were talking (reproduced with permission from Simons & Levin, 1998. Copyright (1998) Psychonomic Society).

So what does change blindness actually tell us about the nature of our perception? It suggests that we seem to retain remarkably little
detail about a visual scene. For example, a scene may contain several objects and, although basic information about the objects may be remembered such as an object’s identity or category (e.g. it is a phone), very little detail about that specific object is recalled. The mental representation built up by repeated fixations of a visual scene seems to be quite sparse, but then again, more detail may not be necessary. After all, under normal conditions, additional information can be accessed easily by making a simple eye movement (McConkie & Loschky, 2003).

Additionally, another well-known example of change blindness is the ‘gorillas in our midst’ experiment (Simons & Chabris, 1999). Observers were shown film clips of seven people throwing basket balls to each other. The observers were asked to count how many times the ball was passed. In the middle of the clip, someone walked through the scene either carrying an umbrella or dressed in a gorilla suit (Figure 12.5). However, because the observers were concentrating their gaze on following the ball, nearly half of them failed to notice that anything out of the ordinary had occurred. This is a graphic illustration of how little of the visual world we are actually taking in at any one time, despite our perception that we are sensitive simultaneously to all the elements in a visual scene.

Perceptual rivalry

Consider the Necker cube in Figure 12.6. It is possible for us to see either face of the cube as being at the front of a 3-D box. Although the stimulus remains the same, our perceptual interpretation fluctuates over time between the two alternative views. This kind of perception is called multistable perception or perceptual rivalry, as the two alternative interpretations are thought to be perceptual rivals. It is believed that this unstable perception occurs because the brain is receiving ambiguous information about the nature of the object. Faced with this ambiguity, the brain alternates between different perceptual states over time (Blake & Logothetis, 2002). The Necker cube is an example of monocular rivalry. Another form of
perceptual rivalry is binocular rivalry. In this form of rivalry, a different image is shown to each eye. Normally, our eyes see two slightly different images due to their horizontal displacement, and they fuse these images to produce a single image. The slight differences in the position of objects in the two images are used to calculate a 3-D representation of the world, a process called stereopsis (see Chapter 11). However, if the images presented to the eyes are too different, then fusion cannot occur. Instead, an observer sees first one image and not the other, and then perception is reversed, with the second image becoming visible and the first disappearing. The input from one eye seems to be temporarily suppressed (Blake & Logothetis, 2002). This phenomenon has been used extensively to explore the basis of perception.

In a set of experiments carried out in Niko Logothetis’s laboratory, monkeys were presented with a binocular rivalry task (Leopold & Logothetis, 1996; Sheinberg & Logothetis, 1997). For example, a right-orientated line grating was presented to one eye, and a left-orientated grating to the other. The monkey had to pull a lever on his right-hand side when he saw one percept, and a lever on his left when he saw the other. This allowed the experimenters to know what the monkey was ‘seeing’ at any one time. Simultaneously, they recorded from neurons in different visual areas in the monkey’s cortex, including V1, V2, V4, V5, the inferior temporal cortex (IT), the superior temporal sulcus (STS) and medial superior temporal sulcus (MST).

The activity of some of the neurons remained constant throughout the experiment, suggesting they were responding to either of the rivalrous stimuli whether the monkey ‘perceived’ it or not. However, some neurons altered their responses over time, the strength of the response being correlated with whether the monkey ‘perceived’ one stimulus or the other (i.e. the behaviour of these neurons reflected the perception of binocular rivalry in the monkey). The proportion of cells that reflected the monkey’s perception increased in the higher visual areas (Figure 12.7). So, only a few cells in earliest cortical visual areas (V1 and V2) show percept-related activity, with a considerably increased proportion in V4, MT and MST, and finally with nearly all the cells in IT and STS closely matching the animal’s perceptual state (Leopold & Logothetis, 1999). Additionally, another study has shown that, in the LGN (a pre-cortical visual area), neurons show no modulation of activity during binocular rivalry (Lehky & Maunsell, 1996).

**Figure 12.6.** (See also colour plate section.) An illustration of the Necker cube (left). It is perceived as a three-dimensional object, and either the ‘back’ face (centre) or the ‘front’ face can be seen as being in front. The visual system alternates between these two alternative perceptions.
Functional imaging studies in humans have shown a similar pattern of activity, with the higher visual areas and parts of the prefrontal cortex showing a modulation of activity with binocular rivalry (Lumer, Friston & Rees, 1998; Lumer & Rees, 1999). A good example is a study that used faces and places as stimuli. Faces stimulate the fusiform gyrus (see Chapter 9), whereas place images of places, such as houses or streets, preferentially stimulate the parahippocampal area. Tong et al. (1999) used fMRI to record their observers’ brain activity while presenting images of faces to one eye and images of places to the other eye to produce binocular rivalry (see Figure 12.8). When the observers reported that they saw a face, the fusiform gyrus was active, but activity in the parahippocampal area was suppressed. When observers reported seeing a scene, the pattern of activation was reversed. These results suggest that the early visual areas primarily are processing the actual physical attributes of the stimulus, whereas the processing of the higher visual areas reflects our interpretation of the visual stimulus.
The illusion of perception

Phenomena like perceptual learning, change blindness and perceptual rivalry illustrate that what we think we see is exactly that: what we think we see and not necessarily an accurate representation of the real world. The limitations of our visual system mean that we can only sample and process a fraction of the available information, and that our perception of the world is more of an educated guess, based on this limited sample and past experience, rather than a detailed and precise picture of the world.

Summary of key points

(1) Different features of a visual stimulus are analysed in different specialised visual areas. This raises the question of how all this distributed information is re-integrated, and how to prevent features from different objects being mixed together (the superposition catastrophe).

(2) It has been suggested that a binding mechanism is required, by which the different features of different objects could be associated separately. The timing of the neural action potentials or spikes could be used in a temporal code to show which features are associated with which object.

(3) Neurons in cat visual areas exhibit short duration periods when their activity is oscillatory (30–70 Hz). It has been suggested that these periods are part of a temporal code: neurons responding to different features of the same object would oscillate in phase, while cells responding to the features of different objects would oscillate out of phase.

(4) Although there is general agreement that oscillatory activity does occur under some circumstances, the evidence for their synchronisation under conditions where they could function as a temporal binding mechanism is inconclusive.

(5) An alternative scheme for the integration of different processing streams is by hierarchical convergence on cells in a subsequent higher integrative cortical area, a hypothesis supported by anatomical, lesion and single-unit recording studies.

(6) Observers can reconstruct meaningful objects accurately out of fragmentary or ambiguous evidence. However, naïve subjects often take some time to recognise such an object, unless they have previously seen the unambiguous version. Once they have recognised the unambiguous version of the object, even the ambiguous version of the stimulus can be recognised almost instantly. This is called perceptual learning.

(7) As we can only perceive a very limited part of a visual scene in detail and colour at any one time, it is possible to change parts of
the scene that are not being attended to, and we do not notice. This phenomenon is called change blindness.

(8) Although the visual properties of an ambiguous stimulus (such as the Necker cube) remain the same, our perceptual interpretation fluctuates over time between the two alternative views. This kind of perception is called multistable perception or perceptual rivalry.

(9) The Necker cube is an example of monocular rivalry.

(10) In binocular rivalry, a different image is shown to each eye. As the images are too different to fuse, our perception alternates between the two images.

(11) Binocular rivalry can be used to explore the separation between neural activity and actual perception of a stimulus. As we move up the visual system towards the higher processing areas, our neural activity increasingly mirrors our perception, rather than being the simple processing of the visual attributes of a stimulus.


REFERENCES


REFERENCES


REFERENCES


REFERENCES


Watson, J.D., Myers, R., Frackowiak, R.S.J. et al. (1993b). Area V5 of the human brain: evidence from a combined study using positron emission tomography and and magnetic resonance imaging. Cerebral Cortex, 3, 79–94.


Williams, G.V. & Goldman-Rakic, P.S. (1993). Characterization of the dopamine innervation of the primate frontal cortex using a dopamine specific antibody. Cerebral Cortex, 3, 199–222.


A-Cells, see M ganglion cells
accommodation 25, 27, 164–5
acetylcholine 21
achromatopsia 69, 106–7, 133, 136
acuity 23, 40
adaptation 37
aerial perspective 166
agnosia 109
albinos, visual acuity in 23
alleles 55
akinetopsia 161–3
amblyopia
meridional 98
stimulus-deprivation 93
amodal completion 166
amorphosynthesis 170
amygdala 120–1, 139–42, 144
anomalous trichromacy 48–9
anterior chamber of the eye 22
aqeous humour 22
asomatognosia 170
astereognosis 170
astigmatism 97–8
atropine 21, 27
attention 126–9
autosomal inheritance 49
B-cells, see P ganglion cells
Balint’s Syndrome 69
binocular disparity 65
binocular rivalry 182–3
binocularity 93–4
bipolar cells 23, 33
blindsight 73–6
blindspot 23
blinking 19–21
blobs 66–7, 70, 79, 84
blue cone pigment 53
blue cone pigment genes 50, 53
blue cones 29, 44, 45, 46
bottom-up processes 126
bridges 84
cardinal cells 121
cataracts 28
centre-surround organisation 34–5, 36
chameleon 165
change blindness 180–2
chimeric gene 50
choroid layer 23
chromatic aberration 47
chromosome 3 50
chromosome 7 50
ciliary bodies 22
classical receptive field 105–6
colour constancy 101–7
colour Mondrian 102–3
complex cells 79–81
counter-assisted tomography 8–9
counterbalanced tomography 8–9
cones 29, 40
(see also blue cones, green cones, red cones)
conjectiva 21
convergence 25, 164–5
cornea 21
corollary discharge theory 147, 152–3
cortical areas 2–4
cortical folding 4, 5, 6–7
critical period 93, 95, 98–9
crystallins 25–6, 28
cyclic guanosine 3’-5’-monophosphate (cGMP) 30–2
cytochrome oxidase 66, 67, 68, 84
Dalton, John 47, 53
Daltonism 47
dark adaptation curve 39
dark current 30
delayed matching to sample (DMS) 127
deprivation monocular deprivation 91, 93
binocular deprivation 91
deuteranope 48
dichromacy 44
dioptries 25
direction selectivity 82–4
disparity 168–9
distributed encoding 122
dMRI, see water diffusion MRI
doamine 129–31
dorsal system 70
dorsolateral prefrontal region (DL) 70–1
double-opponent cells 67, 84
duplicity theory 38
egocentre 172
Einstein, Albert 7–8
elaborate cells
columnar organisation of 114–16
shape selectivity of 114
electroencephalography (EEG) 14
emmetropia 26, 28
emmetropisation 27
end-inhibited cells 81
end-stopped cells 81
enigma illusion 162
ensemble encoding 121–2
exons 50, 57
extra-classical receptive field 105–6
extra-personal space 170
extrastriate cortex 3
face cells 118–26
far cells 168
far space 170
fibre cells 25–6
figure-ground segregation 127
filling in 23
floaters 22
flicker fusion frequency 40
focus of expansion (FOE) 153–5
focusing the eye 25–8
frontal eye field (FEF) 153
functional magnetic resonance imaging (fMRI) 11–14
fusiform gyrus 106, 134–6, 138–9, 145
Gage, Phineas 143
ganglion cells
bistratified ganglion cells 70
M ganglion cells 63
P ganglion cells 63
gceans 116–17
gestalt principles 110–11
glaucoma 22
globular (G) proteins 121
gnostic units 121
good continuity 110
gorillas in our midst 182
grandmother cells 121
greebles 138–9
green cones 29, 44, 45, 46
pigment 50, 53
pigment genes 50, 51, 52
guanosine diphosphate (GDP) 31
guanosine triphosphate (GTP) 31
gyri 2
INDEX

Hebb, Donald O. 89
Hering, Ewald 45
Horopter 168
hybrid genes 50, 51, 52, 55
hypercomplex cells 81
hyperopia 26, 27
hyperpolarisation 30
ice-cube model 85
illusory contours 110–11
inferior temporal cortex (IT) 109, 112–16, 120, 121, 122–5, 127–9, 179
inter-blob region 66, 85
interocular transfer 95
interposition 165–6
interstripes 67–8
introns 50
inverted faces 135–6
iris 21
isomers 29–30
IT, see inferior temporal cortex
K cells 69
Kanizsa triangle 112, 165
Kluver-Bucy syndrome 140
koniocellular layers 69–70
lamellae 29
Land, Edwin Herbert 102
lateral geniculate nucleus (LGN) 63
lateral inhibition 34
lateral pretectum 74
lateralisation 3, 136–8, 174
lens 22, 25–6, 28
yellowing of 26
LGN, see lateral geniculate nucleus
light adaptation, see adaptation
lightness
constancy 37, 101
record 104
linear convergence 166
lingual gyrus 106, 136
locus control region (LCR) 52–3, 59
long-term depression (LDP) 94, 98
long-term potentiation (LTP) 90–1, 94, 98–9
lyonisation 55
M ganglion cells 63, 64
M pathway 67, 69–70, 71
magnetic resonance imaging (MRI) 9–10
magnetoencephalography (MEG) 14–15
magnocellular (M) layers 64, 65
medial superior temporal (MST) 154–5, 157, 160–1
medio-dorsal nucleus (MDN) 153
melanin 23
memory fields 130
metarhodopsin II 31
microelectrode recording 160
microsaccades 148
microspectrophotometry (MSP) 58
microstimulation 160
motion aftereffect 83–4
motion detection
component motion 158
global motion 158
illusion 162–3
neural basis in V1 82–4
neural basis in V3 156–7
neural basis in V5 157–60
motion parallax 168
MT, see visual area 5
Muller-Lyer illusion 167
multistable perception 182
myopia 26–8
N-methyl-D-aspartate (NMDA) 90
near space 170
near cells 168
neglect 170–4
night blindness 54
non-classical receptive field 105–6
nyctalopia 54
nystagmus 149
object constancy 101
ocular apraxia 69
ocular dominance column 79, 91–3
oculomotor cues 164–5
OFF-surround 34–5, 36
ON-centre 34–5, 36
opponent colour mechanisms 45, 84
opsin 29
optic ataxia 69
optic chiasm 63
optic disc 23
optic flow 153–5
optic nerve 23
opto-kinetic nystagmus (OKN) 149–50
orbitofrontal cortex (OFC) 143–4
oscillations, neuronal 176–8
P-B pathway 66, 67, 69
P-I pathway 66, 67
P ganglion cells 63, 64
P pathway 66, 68, 69–70, 71
Pz cells, see M ganglion cells
Pβ cells, see P ganglion cells
pale stripes 67
parvocellular (P) layers 64, 65
pentachromacy 44
perceptual learning 179–80
peripersonal 170
personal space 170
perspective 166
phosphodiesterase (PDE) 31
photoisomerisation 29–30
photons 18
Piazza del Duomo, Milan 171–2
pigment epithelial layer 23
Ponzo illusion 166
pop out 111
population encoding 121–2
positron emission tomography (PET) 10–11
posterior parietal cortex (PPC) 20–1, 129, 169–70, 173–4, 176–7, 178
inferior lobule 170
superior lobule 170
pre-attentive processes 126–9
presbyopia 25
prestriate 3
primary visual cortex, see visual area 1
promotor region 52–3, 59
prosopagnosia 107, 133
protanope 48
pupil 21
Purkinje shift 38
pursuit eye movements 148
Rayleigh match 58
receptive field 34–5
red cones 29, 44, 45, 46
pigment 50, 53
pigment genes 50, 51, 52
reflectance 103–4
refraction of light 25
relative size 166
Rembrandt 96
retina 22–3, 29, 33, 63
retinal 29–30
retinal densitometry 39
retinex theory 104
retinitis pigmentosa (RP) 53–5
retinopathy 54
rhodopsin 29, 30, 31
rhodopsin gene 50, 54
rods 29, 30, 40–1
saccades 148, 149–53
sclera 19
scotoma 73
selective rearing
sensitive period, see critical period
shading 167
signal efficiency
simple cell 79–81
simultagnosia 69
single unit recording, see microelectrode recording
smooth eye movements, see pursuit eye movements
sparse encoding 122
spatial frequency
stereopsis 65, 168–9
stereoscope 168
stochastic resonance
striate cortex, see visual area 1
stripes, see interstripes, pale stripes, thick stripes, thin stripes
STS, see superior temporal sulcus sulci 2
superior colliculus (SC) 74–5, 152–3
superior temporal gyrus 173–4
superior temporal sulcus (STS) 109, 118–19, 120–1, 124–5, 133
suppression of perception during blinking 19–21
during saccades 150–1
tapetum 23
temporal binding mechanism 175–80
tetrachromatic colour vision 56
texture 82
texture gradients 166
thick stripes 67–8
thin stripes 67–8
three-dimensional structure from motion (3D-SFM) 156
tilt aftereffect 95
top-down processes 126
transcranial magnetic stimulation (TMS) 15, 131, 162
transduction 29, 30–3
tremors, see microsaccades
trichromacy 44–5, 58, 59
tritanope 47
tuned-excitatory neurons 168
tuned-inhibitory neurons 168
ultra-violet light 19
univariance 44
Urbach-Wiethe disease 140
V4 complex 107
V4z 107
ventral system 70
visual alphabet 116
visual area 1 (V1) 3, 62, 64, 73, 74, 75, 111, 131
visual area 2 (V2) 62, 65, 67–8, 73, 74, 75, 111, 131
visual area 3 (V3) 20, 69, 156–7
visual area 4 (V4) 68–9, 105–7, 112
visual area 5 (V5) 3, 65, 69
visual area 8 (V8) 107
visual imagery 131
visual memory
long-term 131
short-term 126–9
vitamin A 50
vitreous humour 22
void viewing condition 103
voltage-sensitive dyes 85–6
water diffusion MRI (dMRI) 9, 10
waterfall illusion 83
wavelength 18–19
what versus how pathways 71–3
what versus where pathways 62, 70–1
Wheatstone, Sir Charles 168
Working memory 126–9
Young, Thomas 44
X-chromosome 50, 51, 52, 55, 59