1

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 cm^2 per hemisphere in small-brained rodents to 1100 cm^2 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

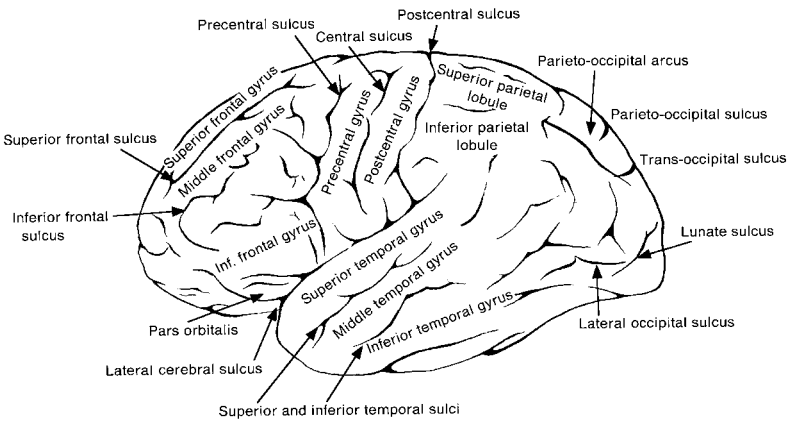

Figure 1.1. Superolateral view of the left hemisphere of the human cerebral cortex, showing the names of the major gyri and sulci (redrawn from Bindman and Lippold, 1981).

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.

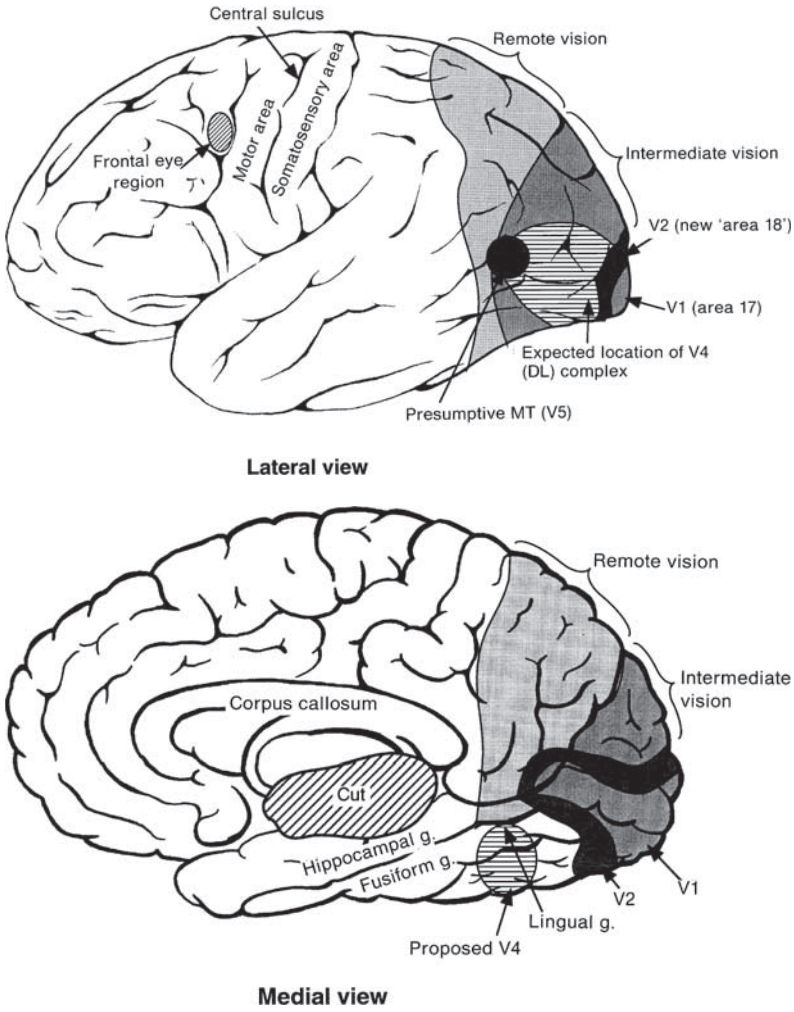

Figure 1.2. The putative location of some of the important visual functions in human visual cortex, shown both in lateral view and in medial cross-section. Abbreviations: V1, the primary visual cortex, also called the striate cortex ; V2, visual area 2; V4, visual area 4, also called the dorsolateral (DL) complex in New World primates; MT, middle temporal, also called visual area (V5) (redrawn and modified from Kaas, 1992).

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.

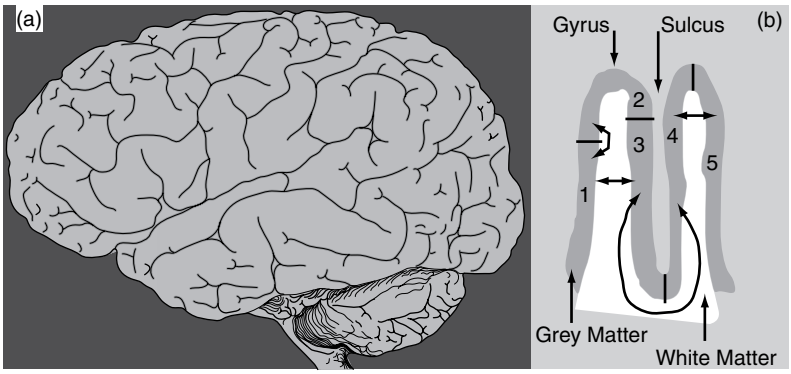

Figure 1.3 (A). The human brain. In this and many other mammalian brains, a distinct pattern of folds is the most striking anatomical feature. The pattern is characteristic of species and is related to the mosaic of distinct functional areas that make up the cortex. (B). How folds may influence the length 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' (NDB1) 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).

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

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.

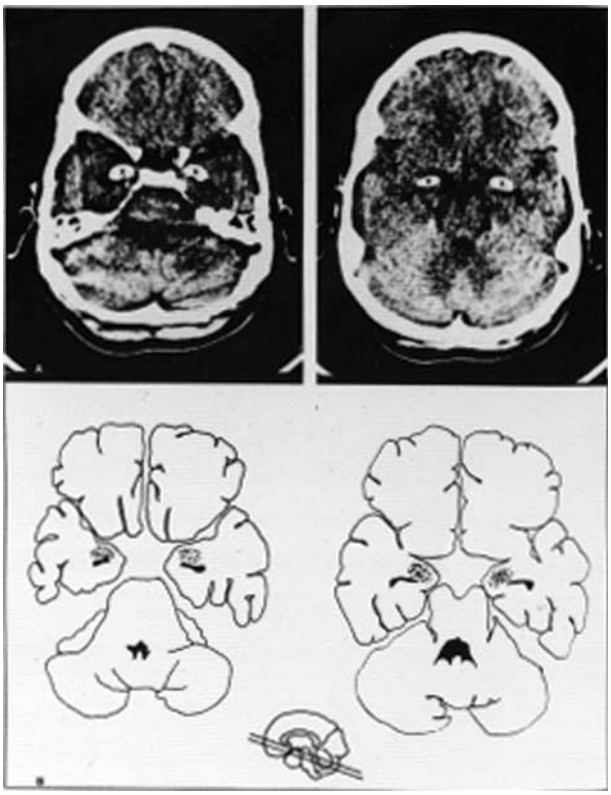

Figure 1.4. Transverse CT scans of a female patient (S.M.) with Urbach–Wiethe's disease. In this condition deposits of calcium are laid down in a brain area called the amygdala (indicated by X marks on the figure). The destruction of the amygdala disrupts the interpretation of facial expression (see Chapter 9) (reproduced with permission from Tranel & Hyman, 1990. Copyright (1990) American Medical Association).

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

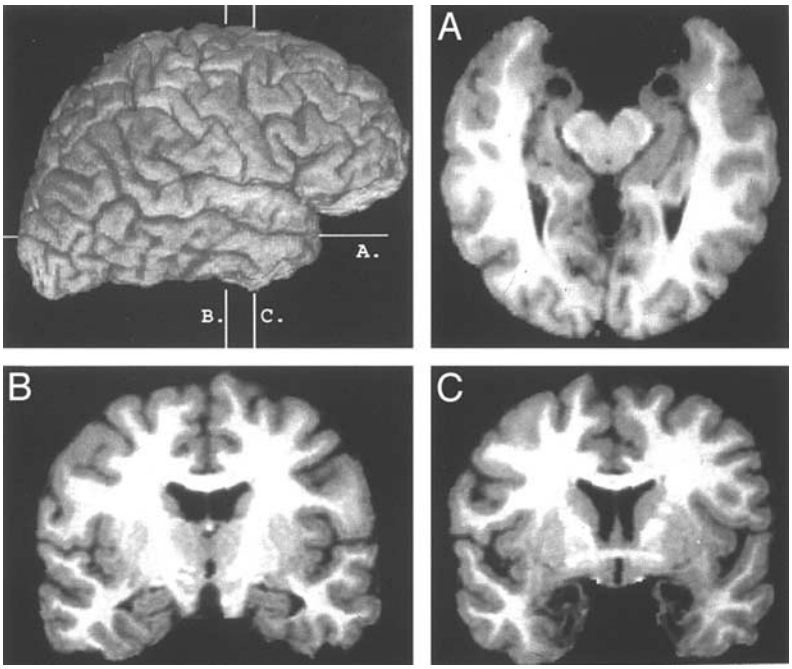

Figure 1.5. An MRI scan of the same patient's (S.M.) brain. The Axial and coronal slices (labelled as *A* and *C*) show a lack of signal at the amygdala (reproduced with permission from Heberlein & Adolphs, 2004. Copyright (2004) National Academy of Sciences, USA).

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