

# Perceptions, Sensations and Cortical Function: Helmholtz to Singer

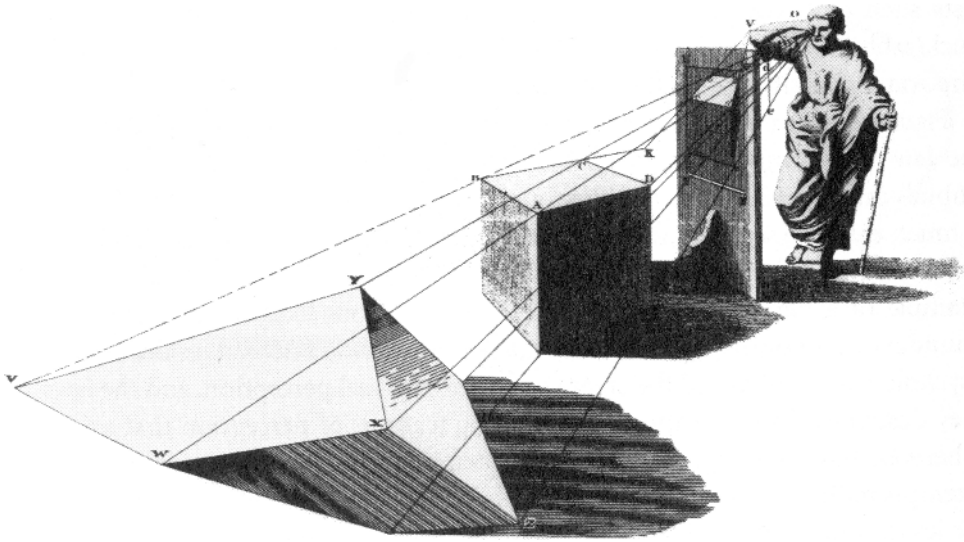
## 1.1 Visual Illusions and their Interpretation by Cognitive Scientists

Helmholtz (fig. 1.1), in his *Treatise on Physiological Optics*, suggested that the formation of a perception involves the development of an unconscious hypothesis based on inductive inferences gained from sensations. For him perceptions are conclusions of unconscious inferences the premisses of which are unconscious and (more or less) indescribable sensations and (unconscious) generalizations about the correlation between past sensations and objects perceived. The viewer shown in fig. 1.2 takes the strangely shaped object in the foreground, looked at with one eye, to be a cube because it has all the identifiable features along the line of sight that a cube has. On Helmholtz's hypothesis inductive inferences are made by the person in fig. 1.2 on the basis of the sensations due to the rays of light from the object, and these support the most likely hypothesis: namely, the perception of a cube.

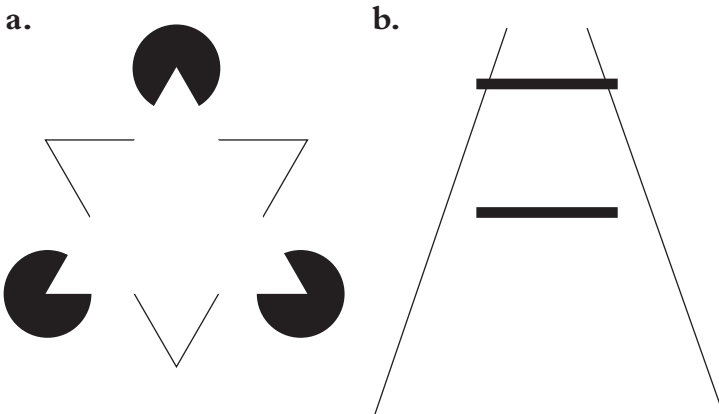
A variety of illusions (e.g. The Ponzo illusion, Kanizsa's illusion, the Ames Room illusion) have been taken as explicable in terms of Helmholtz's theory (Glynn, 1999). That is, these illusions can be explained by reference to the brain's drawing inferences from its past experience to form hypotheses about the objects of its present experience. In the Kanizsa illusion (fig. 1.3a) a ghostly white triangle emerges as a consequence of our inferring that this is the obvious way of interpreting the missing sectors in the three black discs and the edges of the black triangle. In the Ponzo illusion (fig. 1.3b) the upper horizontal bar looks longer than the lower one because the near vertical converging lines are interpreted as railway tracks with parallel lines receding into the distance. Another example of this alleged process of inductive inference is provided by the Adelbert Ames distorting room which produces the experience of extraordinary variations in size of people placed at different positions in the room (fig. 1.4). This room is constructed so that when it is viewed through an eyehole with one eye, an image is produced on the retina identical with that of a rectangular room of uniform height, whereas actually the far wall recedes and both the floor and the ceiling slope, as shown in the small diagram. When people are placed in the far corners of the room, their size is judged in relation to the dimensions of the room on the assumption that this is rectangular. Yet another example which is taken to support Helmholtz's hypothesis



**Fig. 1.1.** Helmholtz. Sketch by Franz von Lenbach (1894). Courtesy of the Siemens-Forum, Munich.

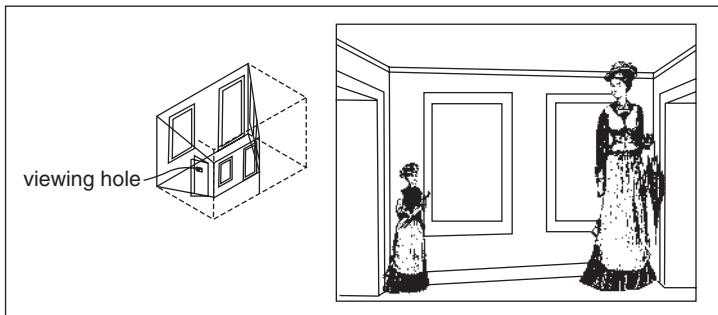


**Fig. 1.2.** Drawing to illustrate Helmholtz's argument on how a perception is formed. (Glynn, 1999, p. 197.)

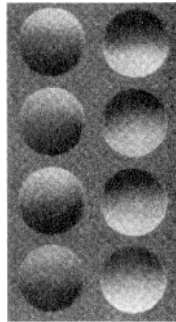


**Fig. 1.3.** a: the Kanizsa illusion. b: the Ponzio illusion. (Glynn, 1999, p. 196.)

is provided by the Ramachandran photograph of fig. 1.5. This shows bumps and hollows that reverse on inversion of the photo. Whether it is interpreted as bumps or hollows is a function of the shading which is ambiguous, depending on the direction of the light. One interpretation of this is that we assume that the light comes from above rather than that the objects face one direction and the different shadings result from different light sources. In the Maurits Escher repeated pattern of fishes and birds (Plate 1.1), the same outline is



**Fig. 1.4.** The Adelbert Ames distorting room. (Glynn, 1999, p. 196.)



**Fig. 1.5.** Ramachandran's bumps and hollows. (Glynn, 1999, p. 196.)

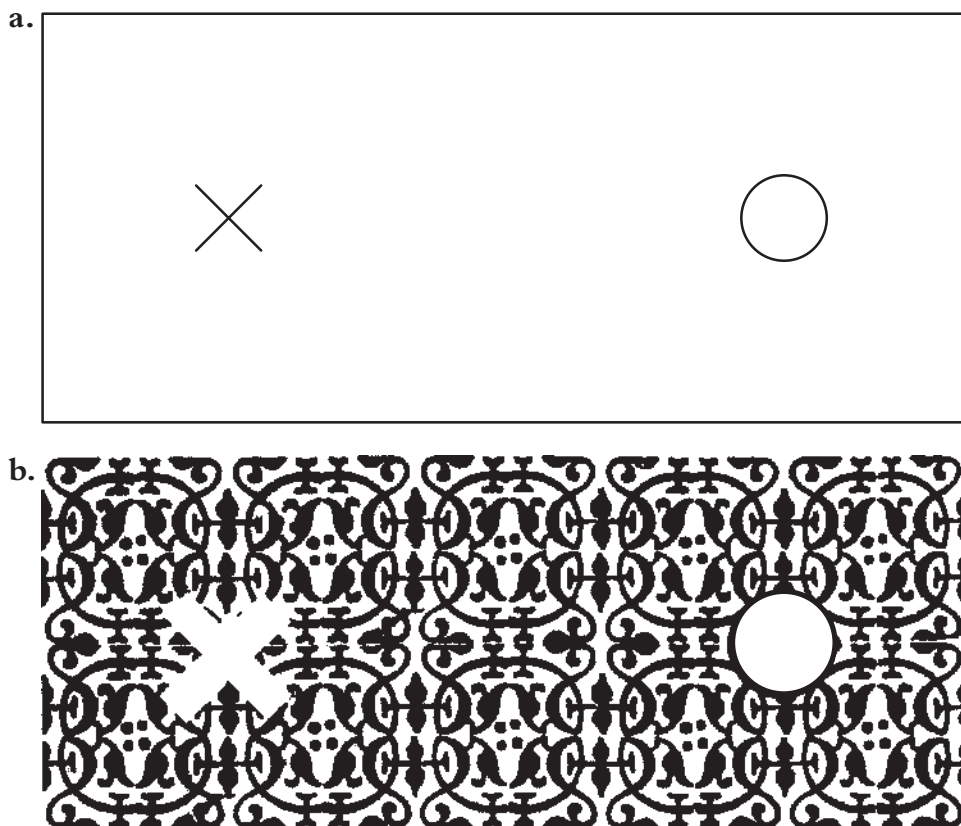
shared by the two different figures. As contours normally only outline an object against its background, Helmholtz's theory holds that the visual system selects either the fishes or the birds for attention, with the other becoming background.

Contemporary neuroscientists support Helmholtz's theory. Thus Glynn comments that 'explanations of this kind do not tell us how the brain manages to make the inferences though they provide a clue to the kind of information processing that may be involved' (Glynn, 1999, p. 195), and Kandel and his colleagues suggest that 'Illusions illustrate that perception is a creative construction based on unconscious conjecture about many of the assumptions the brain makes in interpreting visual data' (Kandel et al., 1991, p. 433). Furthermore, Damasio emphasizes that

When you and I look at an object outside ourselves, we form comparable images in our respective brains. . . . But that does not mean that the image we see is the copy of whatever the object outside is like. Whatever it is like, in absolute terms, we do not know. The image we see is based on changes which occurred in our organisms . . . when the physical structure of the object interacts with the body. . . . The object is real, the interactions are real, and the images are as

real as anything can be. And yet, the structure and properties in the image we end up seeing are brain constructions prompted by the object. . . . There is . . . a set of correspondences between physical characteristics of the object and modes of reaction of the organism according to which an internally generated image is constructed. (Damasio, 1999, p. 320)

A phenomenon that is often used to provide what is taken to be a rather dramatic example of the extent to which the cortex makes inferences based on visual sensations to arrive at what we perceive is given by the phenomenon of ‘filling in’. Fig. 1.6a presents a black cross and a circle on a white background, which should be viewed about 25 cm away with the left eye closed. Focus on the cross and slowly bring the figure towards your right eye; the circle will eventually disappear from your vision, as the image falls on the part of the retina where the optic nerve begins and there are no photoreceptors present. This is your ‘blind spot’. Contemporary neuroscientists, following Helmholtz, suggest that the visual cortex fills in the gap in the blind spot to make it the same as the white background or surroundings. A more dramatic example of this ‘filling in’ is



**Fig. 1.6.** The phenomenon of ‘filling in’. (Glynn, 1999, p. 199.)

given by repeating this kind of experiment but using instead fig. 1.6b, which instead has a white cross and a white disc on a patterned background. This time focus on the white cross and slowly bring the figure towards the right eye with the left eye closed as before. Again the disc disappears, but this time the pattern is continuous across the region previously occupied by the disc. Thus, it would seem, the blind spot does not normally give rise to a black area in one's visual field but is continually 'filled in' by an unconscious inference during normal vision.

### 1.1.1 Misdescription of visual illusions by cognitive scientists

Helmholtz and his contemporary followers interpret visual illusions in terms of the particular theory of perception, mentioned above: namely, that physical stimuli to the retina are transmitted to the brain, where they become sensations, which are conceived to be the raw material from which perceptions are synthesized by the unconscious mind. However, this theory is incorrect. For there are no visual 'sensations' in the brain, although pressure on the brain may produce a sensation: namely, a headache. There is no such thing as combining sensations to form a perception. Furthermore, perceptions cannot be conclusions of unconscious inferences the premisses of which are unconscious and more or less indescribable sensations and (unconscious) generalizations about the correlations between past sensations and objects perceived. So illusions, such as the Ponzo illusion, Kanizsa's illusion, and the Ames room illusions, are not explicable in Helmholtz's terms.

To perceive something is not to form a hypothesis. A hypothesis is an unconfirmed proposition or principle put forward as a provisional basis for reasoning or argument, a supposition or conjecture advanced to account for relevant facts. Only human beings, not their brains, can form hypotheses and make inferences. There is no such thing as a brain's putting forward a proposition as a basis for reasoning or argument or acting on a supposition. Hypotheses are formed on the basis of data which consist of information that is thought to provide evidential support for the hypothesis. However, the brain does not, and could not have, information in this sense. Only thinking creatures *with brains* can form hypotheses or conjectures on the basis of the information available to them.

A perception – i.e. a person's perceiving something – is not a hypothesis, but an event or occurrence, and so cannot be the conclusion of an inference, which is a proposition, not an event or occurrence. Finally, inferences are neither conscious nor unconscious mental processes. For inferences are not processes at all. Rather, inferences are transformations of propositions in accordance with a rule, derivations of propositions from premisses in conformity with a pattern of derivation. But perceiving something does not involve transformations of propositions by a perceiver (or his brain).

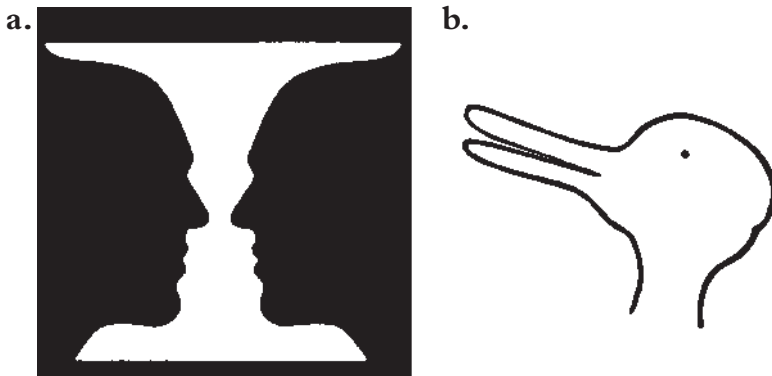
Locke, and the British empiricist tradition that he originated, conceived of ideas and impressions as the result of the impact of the material world on our nerve endings. This misconception is the source of the thought that perceiving always involves sensations which are, on Helmholtzian theory, the premisses of unconscious inferences. For Helmholtzian 'sensations' are, in effect, empiricist 'ideas' or 'impressions'. It is also the source of the equally misguided and far more widespread thought that what is seen (or heard, etc.) when we see (or hear, etc.) something is a picture or image (visual or auditory). This representationalist



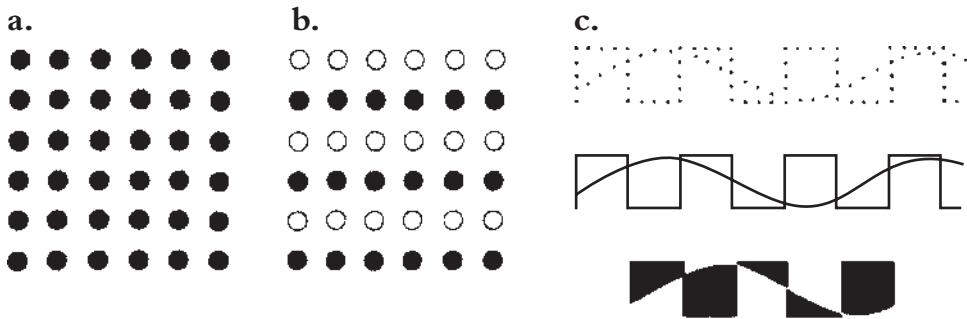
view is defended by Damasio above, but is confused. For what one perceives by the use of one's perceptual organs is an object or array of objects, sounds and smells, and the properties and relations of items in one's environment. It is a mistake to suppose that what we perceive is always, or even commonly, an image, or that to perceive an object is to have an image of the object perceived. One does not *perceive* images or representations of objects unless one perceives paintings or photographs of objects. (Of course, one may *have* after-images or conjure up mental images, but one does not *perceive* them.)

## 1.2 Gestalt Laws of Vision

After the First World War, Max Wertheimer (1924), Kurt Koffka (1935) and Wolfgang Köhler (1929), following Helmholtz, determined to find the laws that relate what we perceive to what we are actually looking at. In particular, they were concerned with how the overall configuration of a scene, rather than particular elements in it, informed the interpretation of the scene. The way in which the *Gestalt* or configuration of the scene provides us with an interpretation is dramatically illustrated by means of Edgar Rubin's vase or two faces (fig. 1.7a), depending on what is assumed to be background or figure, or by Jastrow's duck-rabbit (fig. 1.7b). The laws formulated by the Gestalt psychologists are illustrated by means of fig. 1.8. The 'law of proximity' is shown by the fact that the circles in fig. 1.8a are seen as arranged in horizontal or vertical lines rather than oblique lines, because they are further apart in the oblique lines. The 'law of similarity' is illustrated by fig. 1.8b, in which the dots are now seen as horizontal lines because those forming horizontal lines are more similar than those forming vertical lines. An example of the 'law of good continuation', which states that we perceive the organization that interrupts the fewest lines, is shown in fig. 1.8c, in which the small dots are seen to form a wavy line superimposed on a profile of battlements rather than as the succession of shapes shown at the bottom of the figure. Cognitive scientists believe that Gestalt psychology has shifted the fundamental empiricist's question 'What are the basic components of this



**Fig. 1.7.** a: Edgar Rubin's vase or two faces. b: Jastrow's duck-rabbit. (Glynn, 1999, p. 199.)



**Fig. 1.8.** Illustrations of the Gestalt 'Laws' of (a) 'proximity'; (b) 'similarity'; (c) 'good continuation'. (Glynn, 1999, p. 200.)

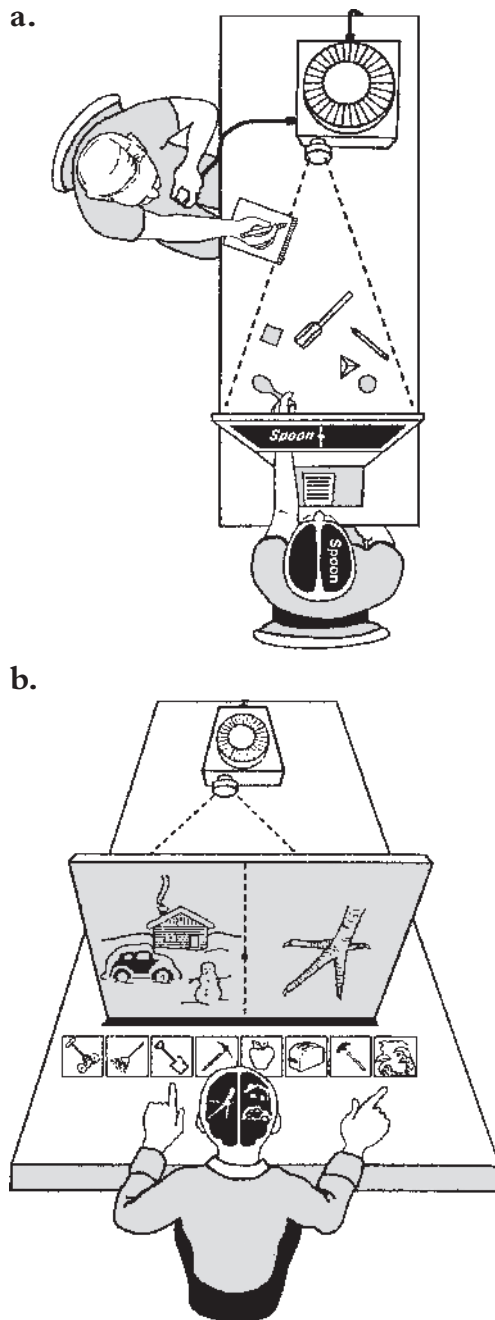
perception?' to 'What neural transformation produces this perception?', thus offering a common scheme for merging psychological and neurobiological investigations into the process of vision.

### 1.3 Split-Brain Commissurotomy; the Two Hemispheres may Operate Independently

Patients with intractable epilepsy often undergo surgery to relieve the condition. This involves cutting the corpus callosum that connects the two halves of the brain. In the early 1960s Michael Gazzaniga and Roger Sperry showed that the two hemispheres of such patients possess their own specializations, with the left hemisphere dominant for language and speech and the right being largely causally responsible for visual motor tasks (Gazzaniga and Sperry, 1967). If a visible object (or picture of an object) occurs in the right visual field (so that concomitant neuronal activity occurs only in the left hemisphere) of such split-brain patients, then they can describe what they see; but if the visible object occurs in the left visual field (with concomitant neuronal activity restricted to the right hemisphere), they cannot. In this case, the patients could point at a similar object to that presented in their left visual field if asked to, but they were not able to say what it was (Gazzaniga et al., 1965).

Similar results were found for the other sensory modalities of touch, sound and smell. In addition, the right hemisphere was shown to be causally responsible for the processes involved in controlling the left hand, whereas the left hemisphere was causally implicated in the control of the right hand. Gazzaniga and Sperry concluded that each hemisphere in humans is causally implicated in different aspects of thought and action. Fig. 1.9a shows the kind of experimental set-up used by Gazzaniga and Sperry to collect data in early split-brain studies (see Gazzaniga, 1995; Baynes and Gazzaniga, 2000). Presenting a written name for an object to the left visual field (involving the right hemisphere) of a





**Fig. 1.9.** a, b: experimental techniques for observing the behaviour of split-brain patients. (Baynes and Gazzaniga, 2000, p. 126.)

patient provides the condition for the patient to use his left hand to select the correct object by touch (in this case 'spoon'; Gazzaniga, 1983). In fig. 1.9b, when presented with *bilateral* picture displays, the patient confabulates about the choices he previously made with his left hand. As Gazzaniga describes the experiments, the right hand selects a rooster to match the claw seen when the left hemisphere is involved, but the patient states that the shovel selected by the left hand was needed to clean out the chicken house. According to Gazzaniga and Le Doux (1978), 'the left hemisphere Interpreter has no knowledge of the snow scene seen by the right hemisphere' in this test. These and other experiments led Gazzaniga (1997, p. 1392) to suggest that 'mind left dealt with the world differently than mind right seemed to be the major conclusion of studies during the era'.

Sidtis (1981) suggested that axons of the corpus callosum which carry information relating to the sensory dimensions of a stimulus are located in a different part of the callosum to those that carry semantic information. His experiment is illustrated in fig. 1.10. The patient underwent a staged callosal section in which the posterior half of the callosum alone was sectioned. This was taken by Sidtis to show that the sectioning prevented the transfer of information relating to the sensory but not the semantic dimensions of the stimulus.

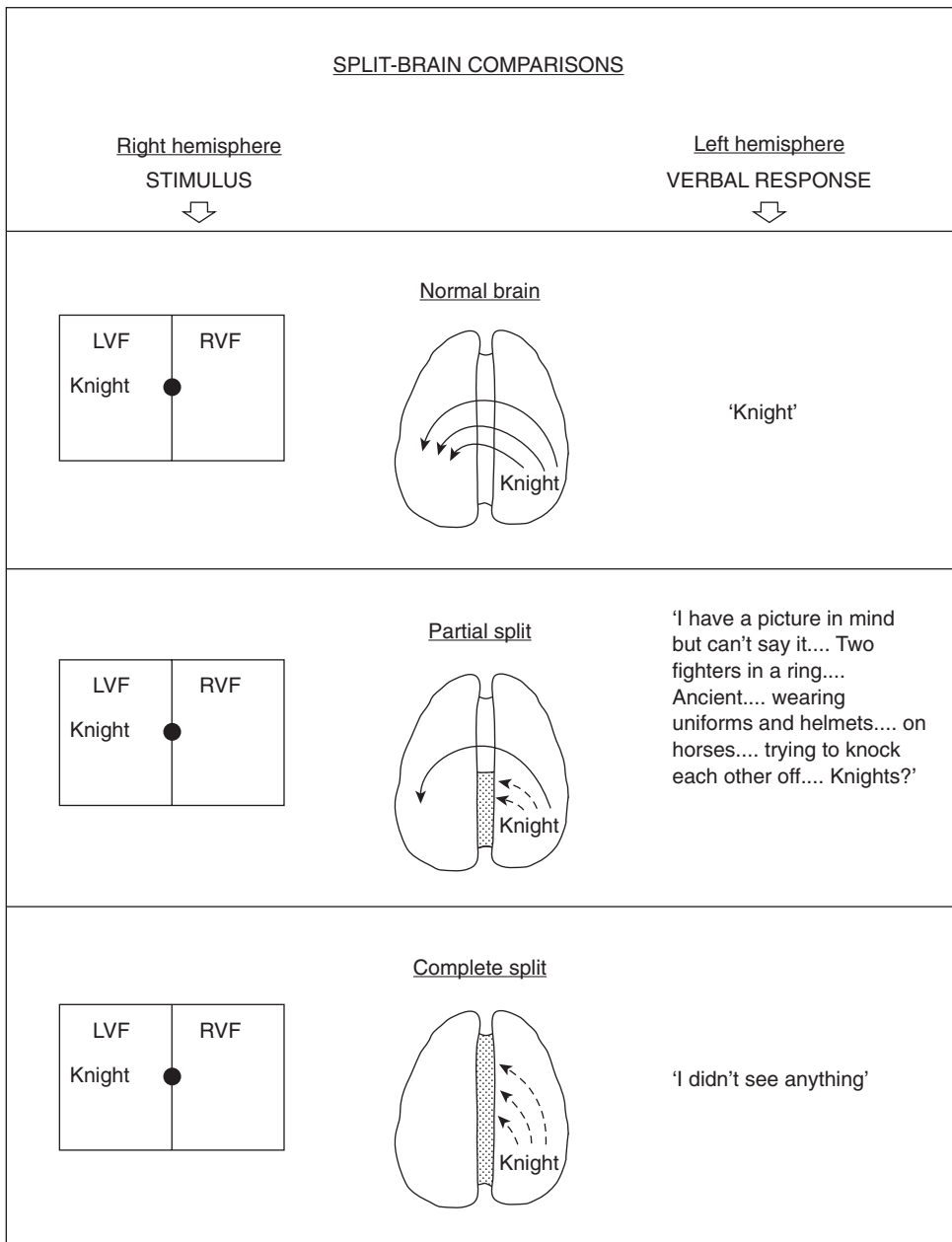
### 1.3.1 Misdescription of the results of commissurotomy

According to the above interpretation of the results of commissurotomy by Gazzaniga and Le Doux, the hemispheres of the brain may possess knowledge and can perceive. However, only human beings can know and perceive, not their brains – which can neither see nor hear, neither write nor speak, nor interpret anything or make inferences from information. The hemispheres of the brain cannot be said to be aware or unaware of anything; they cannot intelligibly be said to recognize or misrecognize anything. They do not make choices or judgements of grammaticality, and they are neither knowledgeable nor ignorant.

The additional claim that the forms of functional dissociation consequent upon commissurotomy produce two minds, one belonging to the left hemisphere, the other to the right, is also awry. The brain does not have a mind, and neither do the two hemispheres of the brain. It is human beings, not their brains, that are said to have minds. To ascribe a mind to a creature is to say that it is a creature with a distinctive range of capacities: in particular, capacities for concept-exercising thought, self-consciousness, memory and will.

### 1.3.2 Explaining the discoveries derived from commissurotomies

The general form of the explanation for the observations of Gazzaniga and his colleagues on the dissociation of functions following commissurotomy is that severing the corpus callosum deprives human beings of their capacity to exercise normally co-ordinated functions. And that in turn is to be explained in terms of the disconnection of neural groups that are causally implicated in the exercise of the relevant capacities. The transmission of



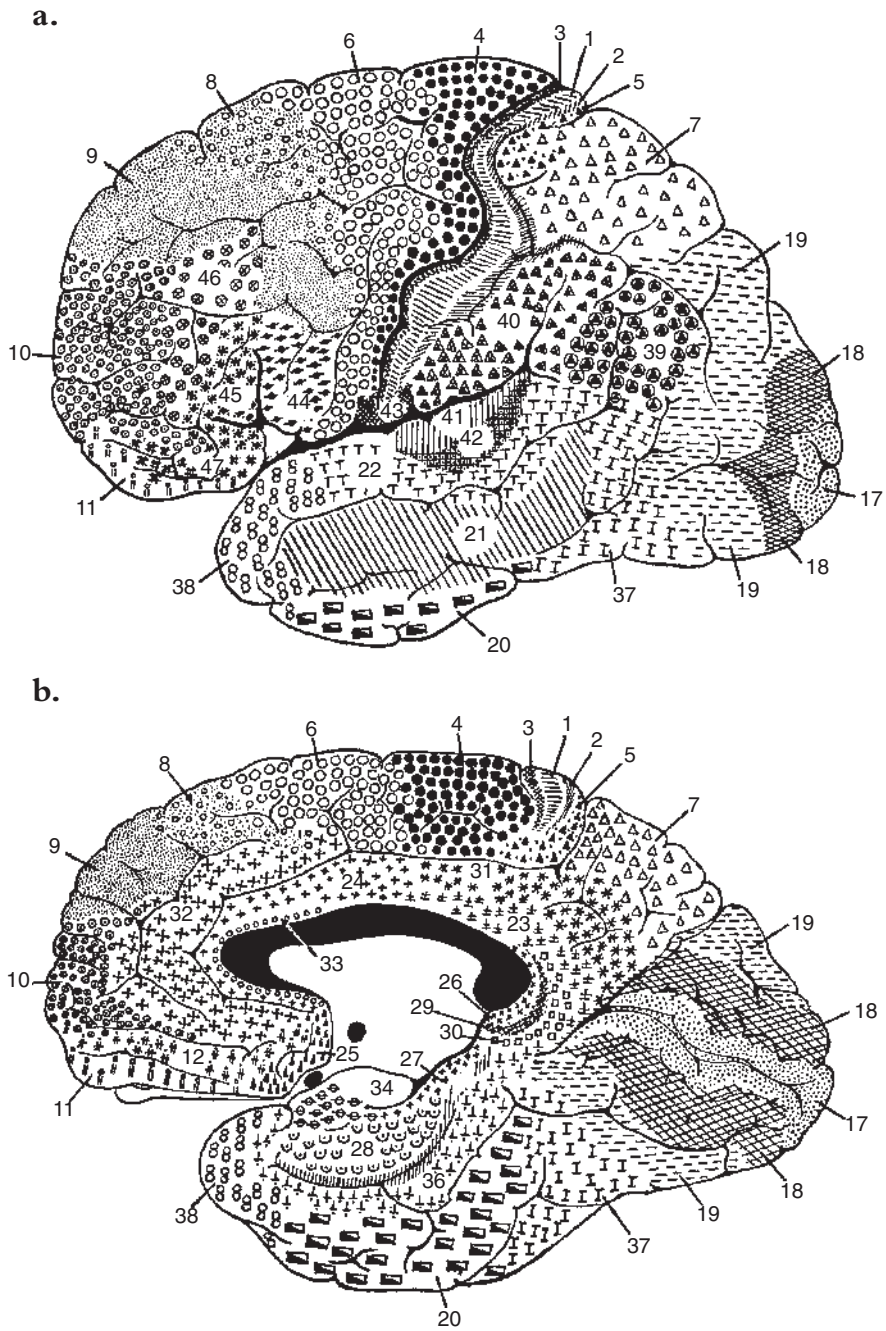
**Fig. 1.10.** Evidence that axons have different functions in different parts of the corpus callosum. (Gazzaniga, 1995, p. 222.)

neural signals across the corpus callosum is a necessary condition of a person knowing, and being able to say, what is visually presented to him (under the experimental conditions in question). It is this which is prevented by commissurotomy. Nevertheless, the patient is still able to respond to what is visually presented to him by making choices with his hands, even though he does not know why he thus points, and confabulates a tale to explain it.

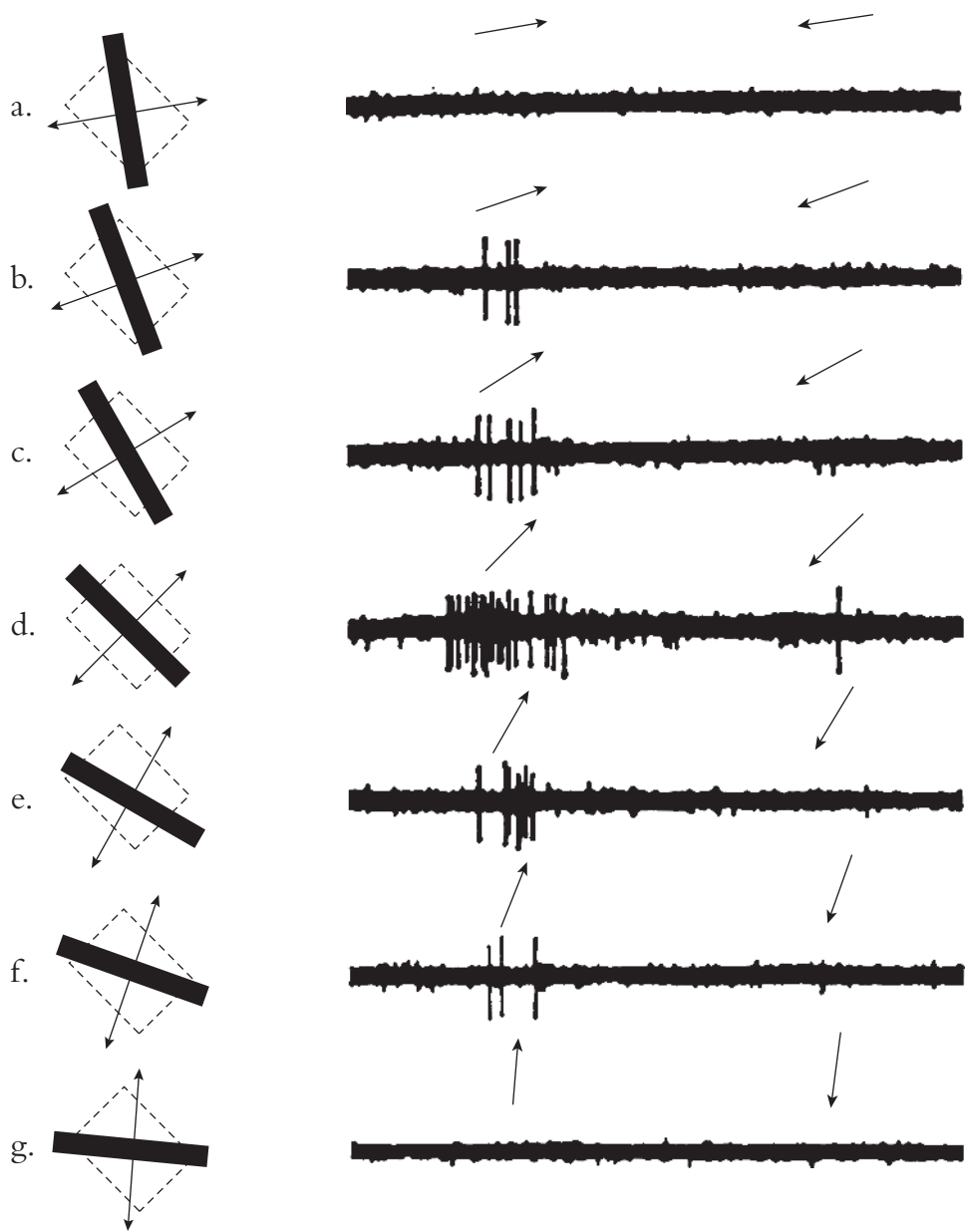
## 1.4 Specificity of Cortical Neurons

David Hubel and Torsten Wiesel as well as Vernon Mountcastle discovered that the cortex contains neurons which are excited by very specific stimuli. These discoveries were founded on the notion of the receptive field of sensory neurons, such as the ganglion neurons in the retina which connect it to the brain. In this case 'the receptive field' refers to that region of the photoreceptor sheet which, when stimulated by a spot of light, changes the frequency of impulse firing in the ganglion neuron being studied. The receptive fields of retinal ganglion neurons consist of two concentric circles; when a spot of light shines on the inner circle, impulse firing in the neuron is increased for some ganglion neurons but decreased for others; if a spot of light shines in the surrounding region outside this, then a reciprocal change in firing is found to that in the centre of the field. There are, therefore, ganglion neurons with a central excitatory region and surround inhibitory region, and others with a central inhibitory region with surround excitation. Hubel and Wiesel, in the late 1950s and early 1960s, examined the receptive fields of neurons in the first relay-station between the retina and the cortex in cats: namely, the lateral geniculate nucleus, and found that they were similar to those of retinal ganglion neurons. Next they examined the receptive fields of neurons in a region of the cortex in cats and monkeys to which the principal neurons in the lateral geniculate nucleus project, called the primary visual cortex or area 17 (Hubel and Wiesel, 1959, 1962, 1968). The referring expression 'area 17' is a designation that dates back to the work of Brodmann (1909), who divided the cortex up into a number of distinctive areas based on differences in the arrangement and types of neurons as well as in the pattern of the myelinated fibers. Such a 'cytoarchitectural' map of the human cortex is shown in fig. 1.11 for both its convex surface (a) and its medial surface (b). There is evidence for a 'modular' organization of brain function that is more fine-grained than that suggested by the cytoarchitectural maps of Brodmann. In the case of the visual system many of these modules have been given 'V' numbers, such as V1 (which encompasses area 17) as well as V2, V3 and V4 (each of which encompass part of area 18). It is therefore more appropriate to adopt this nomenclature when referring to pathways in the visual cortex (see Plate 1.4b and c).

The receptive fields discovered by Hubel and Wiesel in V1 were often much more complex than those of retinal ganglion neurons. Some neurons could be stimulated by lines, bars, squares and rectangles of light, rather than by the spots of light used to characterize the receptive fields of retinal ganglion neurons and neurons in the lateral geniculate nucleus. Fig. 1.12 shows recordings from the receptive field of a neuron in the visual cortex made by Hubel and Wiesel in 1968; the field is indicated by the broken rectangles in the left



**Fig. 1.11.** Brodmann's chart of cortical areas: **a**, convex surface; **b**, medial surface. (After Brodmann, 1909; from Carpenter, 1976.)



**Fig. 1.12.** Neuronal orientation selectivity in primary visual cortex (area 17 or V1). (Hubel and Wiesel, 1968, p. 219.)

column. In this case, the visual stimulus that was viewed by the monkey consisted of a bar of light that was moved back and forth through the receptive field of the cell in each of seven different orientations, indicated in rows a to g. The different directions of motion used for each orientation are indicated by the small arrows. Recorded traces of neuronal activity are shown at the right, in which the horizontal axis represents time, and each vertical line represents an action potential. This neuron responded most strongly to a bar of light oriented along the diagonal (stimulus d), particularly when the bar was moved through the receptive field from lower left to upper right. Neurons of this type were found by Hubel and Wiesel to be common in the visual cortex, and Hubel commented that

The *map* of the receptive field of a cell is a powerful and convenient shorthand description of the cell's behaviour, and thus of its output. Understanding it can help us to understand why the cells in the intermediate stages are wired up as they are, and will help explain the purpose of the direct and indirect paths. If we know what ganglion cells are telling the brain, we will have gone far toward understanding the entire retina. (Hubel, 1988, p. 39)

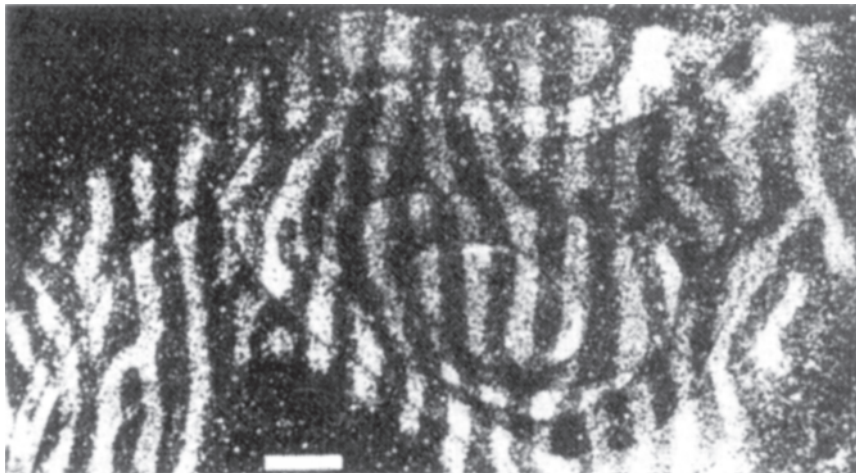
The research of Hubel and Wiesel culminated in the discovery that neurons in the primary visual cortex with particular receptive field properties are organized in narrow 100–200  $\mu\text{m}$ -wide vertical columns passing from the surface of the cortex to the white matter (Hubel and Wiesel, 1977). Such a columnar organization had previously been discovered for neurons with particular receptive field properties in the somatosensory cortex by Vernon Mountcastle (1957). In the case of the visual cortex there is also an arrangement by which the input from each eye is segregated into columns. Fig. 1.13 shows a beautiful anatomical representation of these ocular dominance columns. In order to display such columns, the right eye of a monkey was injected with radio-labelled proline and fucose, which is transported transneuronally to the cortex. Fig. 1.13a shows a dark field autoradiograph of a tangential section of area V1 of the right hemisphere obtained after 10 days' exposure. Radioactivity can be seen in the form of white stripes, which correspond to thalamic axon terminals in layer 4 of the cortex that relay input from the injected eye. The alternating dark stripes depict the position of the afferents from the geniculate axons subserving the uninjected eye (Hubel et al., 1977). Fig. 1.13b shows a reconstruction of the ocular dominance columns in area V1 of the right hemisphere, showing the regular layout of the columns. These discoveries for both the visual cortex and the somatosensory cortex were then taken to indicate that sensory cortex in general is organized along columnar lines in which neurons with similar receptive field properties are found in a particular column.

#### 1.4.1 Cardinal cells

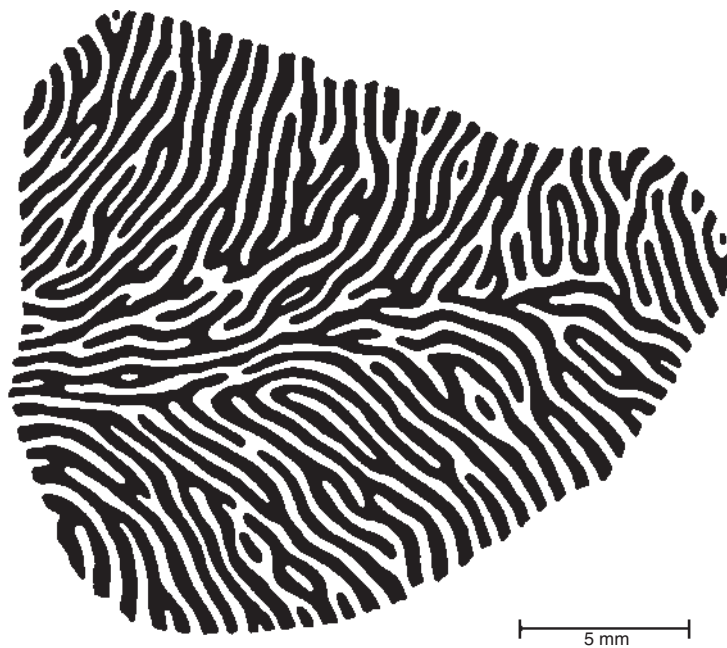
Hubel and Wiesel suggested that the complex properties of neurons in the primary visual cortex (i.e. area V1, see Plate 1.4b) could be thought of as arising from combinations of the 'simple' centre-surround receptive fields possessed by retinal ganglion neurons and neurons in the lateral geniculate nucleus. This idea indicated that there might be a hierarchical increase in complexity of neuronal receptive fields for neurons at progressively higher levels of the cortex – that is, at levels progressively further removed from the retinal input.



**a.** Normal



**b.** Reconstruction: normal ocular dominance columns



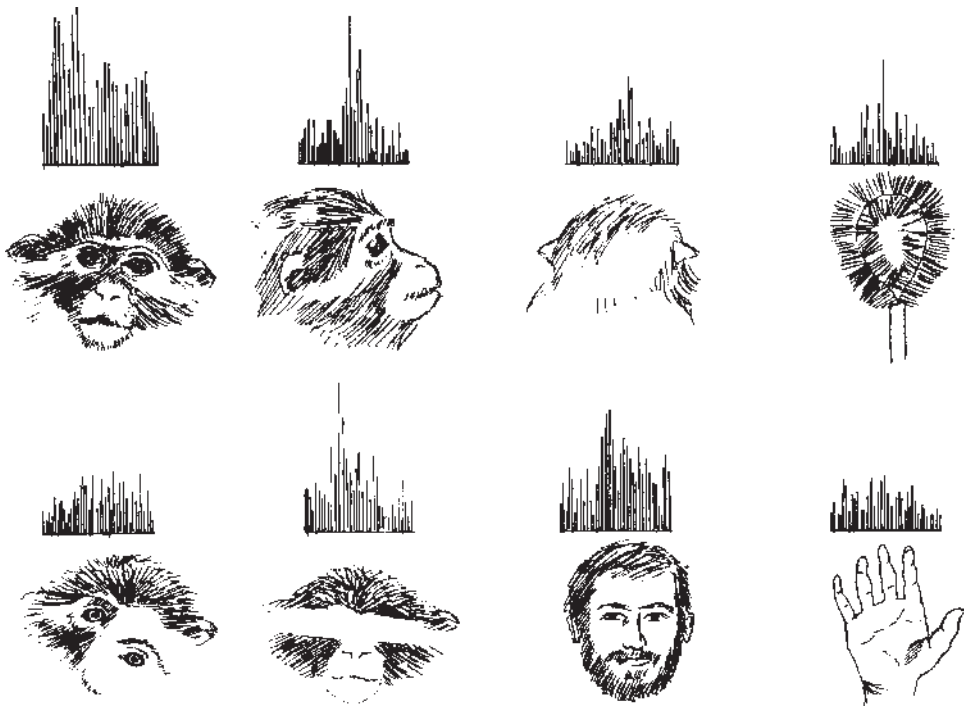
**Fig. 1.13.** Anatomical representation of ocular dominance columns in primate visual cortex. (a: Hubel et al., 1975, p. 584; b: Hubel and Wiesel, 1977, p. 35, reproduced from LeVay, Hubel and Wiesel, 1980, reprinted with permission of Wiley-Liss, Inc., a subsidiary of John Wiley & Sons, Inc.)

Barlow (1972) gave these neurons with very complex receptive field properties the title 'cardinal neurons'. Such neurons might, for example, fire maximally when faces were presented in their receptive field, which could be quite large – that is, a large proportion of the entire visual field. The main function of cardinal neurons was not to respond to some specific characteristic of the retinal illumination (as in the primary visual cortex) but 'to continue responding invariantly to the same external pattern' (in this case to faces). The neuron doctrine in perception fostered the belief that the pathways that are active for a given sensory scene converge and produce activity in a single cell (named 'a cardinal cell' or sometimes 'a grandmother cell') or a group of cells (cardinal cells) 'whose role is to represent the scene'. This idea was motivated, at least in part, by the thought that if the animal is to see, the brain must combine the information derived from the retina to produce a representation of the visual scene. Horace Barlow suggested in 1972 that the hypothesized cardinal cells 'do not represent arbitrary or capricious features in the environment, but features useful for their representative role' – that is, 'their role as correlates of features of the object perceived', and that they can be active in combinations, 'thus having something of the descriptive power of words' (Barlow, 1997, pp. 421–2).

Gross and his colleagues (1969) discovered neurons in the inferior temporal cortex of monkeys (IT, see Plate 1.4b) which possessed just the properties which Barlow had predicted for a class of cardinal neurons. In the monkey's temporal lobe there are neurons that fire impulses at maximal rates when a monkey is viewing a specific object. Some of these neurons respond specifically to the presentation of faces; indeed, these neurons discharge specifically, depending on whether the faces are presented in profile or face on. Fig. 1.14 shows the relationship between a series of images presented and the rate of firing of a temporal lobe neuron. When the monkey is looking directly at the image of another monkey face on, there is maximal firing; but when the image of the head gradually turns around so that it appears only in profile, then the firing occurs at a much lower level. And although somewhat 'monkey like', if the picture of a toilet brush is presented, then the rate of firing of the neurons is much less than when the image of another monkey face is presented. In addition, if an image is presented consisting of the juxtaposition of different elements of the face in a bizarre geometry, again the firing rate is not nearly as high as it is when those elements are put together to make up a proper monkey face. Furthermore, it can be shown that if different elements of the monkey's face such as the mouth or the eyes are taken away, then the firing rate will drop. Finally, the pattern corresponding to the presentation of a monkey's face is different from that resulting from presentation of a human face. The work of Gross and his colleagues has therefore shown that there are neurons in the temporal lobe that are of the 'cardinal' type as postulated by Barlow, inasmuch as they fire impulses vigorously only when a particular kind of object – in the case illustrated in fig. 1.14 a particular kind of face – is presented.

#### 1.4.2 Misdescription of experiments leading to the conception of cardinal cells

The claim made by Barlow that cardinal neurons have a representative role and 'thus have something of the descriptive power of words' cannot be sustained. For the sense in which



**Fig. 1.14.** Firing rate of neurons in inferior temporal cortex of a monkey in response to different images of a monkey face, brush, human face and a hand. (Bennett, 1997, fig. 1.5.)

the excitation of a group of cells represents a certain feature in the visual field is the sense in which a wide ring in a tree trunk represents a year with ample rainfall. That has nothing whatsoever to do with the lexical or semantic sense in which a sentence represents the state of affairs it describes, or with the iconic sense in which a picture represents what it depicts. Furthermore, neither in the iconic nor in the lexical sense could there be any representations of the external world in the brain. Representations, in the lexical sense, presuppose a rule-governed system of symbols and a practice of their employment in the lives of language-using, symbol-employing creatures. It is not intelligible to suppose that the brain has or uses *a language*, any more than it is intelligible that the brain should *draw pictures* and look at them or *make maps* and consult them. And even if we were to suppose that there is nevertheless some as yet unexplained sense in which the brain can be said to contain *maps* (as is commonly argued today) or *pictures* (on the pineal gland, as Descartes supposed), this will contribute nothing to an explanation of how animal vision occurs or is rendered possible. For one cannot explain what the neural processes of seeing items in the visual field consist in by referring us to the brain's *seeing* a picture or map of the objects in the visual field. For, first, we do not know what it would be for the brain to *see* anything (after all, it has no eyes!). Secondly, even if we could give sense to the brain's seeing things,

how would the process of human sight have been explained? To claim that for a human being to see something, his brain must see a representation of the same thing explains the puzzling by reference to the unintelligible. What human beings (and other animals) see are, among other things, objects in their environment. How could representations in the brain, which they cannot see or read, help them to see what is in their environment? Certainly not by means of anything *symbolically* or *iconically* represented. For in whatever legitimate sense (if any) there is to the supposition that there is a representation of what is seen in the brain, that representation is neither what the owner of the brain sees nor anything that is communicated to him.

### 1.5 Multiple Pathways Connecting Visual Cortical Modules

The conception of the visual pathway construed in terms of different modules was advanced by David Marr in the 1970s as part of what he called the ‘computational view’ of sensory processing (Marr, 1982). According to this theory, the process of sensory perception can best be analysed by assigning specific processes to different modules. Marr pointed out that evidence for such modules is apparent in the random-dot stereograms produced by Bela Julesz in 1960. An example of such a stereogram is shown in Fig. 1.15. The left and right images in this figure are identical except for a central square region that is displaced slightly in one image. When fused binocularly, the images yield the impression of the central square floating in front of the background. Marr held that such percepts are caused solely by the stereo disparity between matching elements in the images presented to each eye, so that the analysis of stereoscopic information can proceed independently in the absence of other information. Marr put forward the idea that the study of perception can be subdivided into

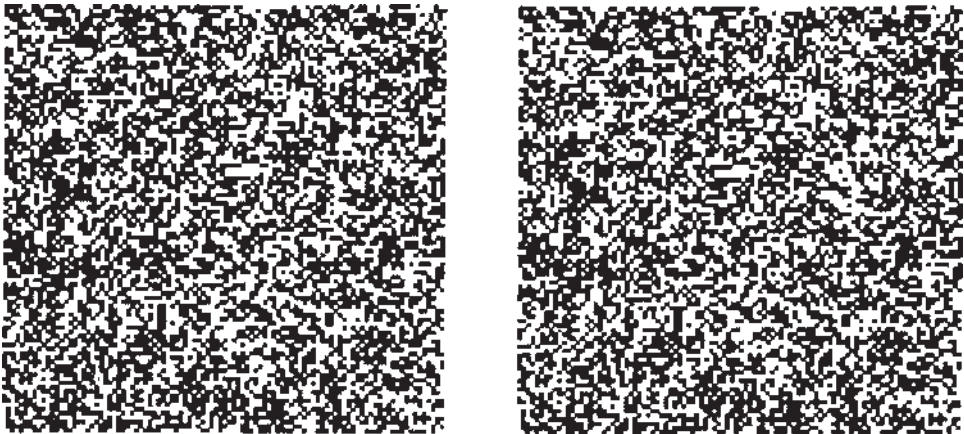
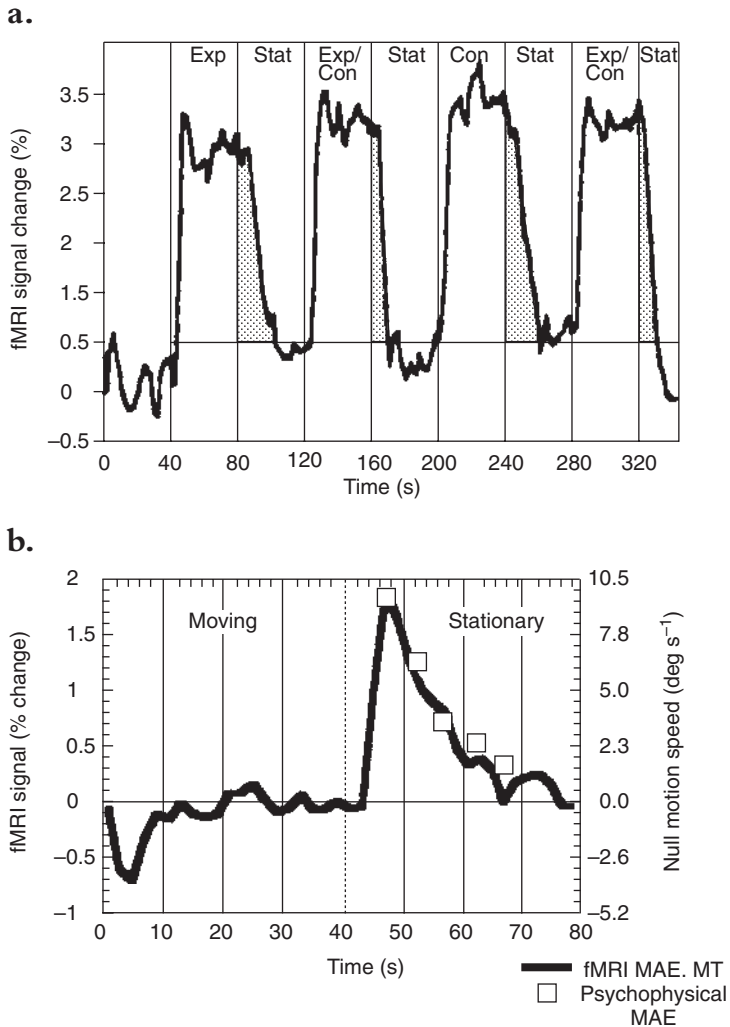


Fig. 1.15. Random-dot stereogram. (Marr, 1982, fig. 1.1.)

specialized parts, each of which can be treated separately. Such parts are 'independent modules of perception' (1982, p. 10).

Damasio and others have presented evidence that the human visual cortex possesses specialized modules like those found in other primates. Lesions in one or more of these modules due to disease or injury lead to the expected behavioural deficits corresponding to the visual experience of the patient. For instance, damage to the occipital and subcalcarine portions of the left and right lingual gyri (containing areas V2 and V4) gives rise to the condition of achromatopsia, i.e. loss of colour vision, so that such patients see the world only in shades of grey (Rizzo et al., 1993). Damasio and colleagues (1979, 1983) have also reported that patients with lesions in the temporal segment of the left lingual gyrus are afflicted with a condition called 'colour anomia'. In this case, they experience colours in the normal way, and are able to rank hues of different saturation, but they use colour names incorrectly. So, for example, they use the word 'blue' or 'red' when shown green or yellow; and, given a colour name, they will point to the wrong colour. Patients with lesions in the left posterior temporal and inferior parietal cortex lose the ability to produce appropriate word morphology. For example, they produce distorted colour names like 'buh' for 'blue'.

Following the discoveries of Hubel and Wiesel in the 1960s, Semir Zeki and others in the 1970s examined the receptive field properties of single neurons outside the primary visual cortex involving higher centres concerned with visual perception in primates. They showed that the properties of these receptive fields were very complex. Using the properties of these fields as the criterion, the visual cortex outside the primary area was shown to be compartmentalized into different regions or modules. Zeki's research has also been taken as supporting this idea of modules, particularly in showing that one area of visual cortex, designated the middle temporal area (MT or area V5; see Plate 1.4b and c) in primates, possesses neurons that are responsive to motion and to the particular direction of movement of an object. A spectacular example of the functioning of the middle temporal area in humans is shown in the phenomenon called the 'waterfall effect'. This name derives from the fact that if a subject looks for some time at water streaming downward in a waterfall and then turns away and looks, for example, at the opposite bank of the river, then the trees on the bank will appear to be momentarily moving in the opposite direction to the flow of the water in the waterfall. Thus the stationary objects (the trees) appear to move. The time course of this illusion can be described quantitatively using psychophysical experiments in conjunction with brain imaging. These show that the middle temporal area is involved (Tootell et al., 1995). Plate 1.2a shows a visual stimulus that consists of a series of concentric rings, either expanding or stationary. Plate 1.2b indicates that the middle temporal area is excited by the moving rings as determined by the non-invasive visual imaging technique, functional magnetic resonance imaging (fMRI; note that the brain is shown in normal and inflated format). The same increase in activity in the middle temporal area is observed when stationary concentric rings are viewed immediately after observing the rings expanding, but not when observing just stationary rings without prior exposure to moving rings, or after prior exposure to rings moving in one direction and then moving in the opposite direction. It is known that the subject experiences an after-effect of apparent visible motion in the first case but not in the latter two cases; that is, the stationary rings appear to move in the



**Fig. 1.16.** **a:** the changes in the fMRI signal during real and illusory visual motion. **b:** the quantitative relationship between the visual after-effect and activity in the middle temporal area as measured by magnetic resonance imaging. (Tootell et al., 1995)

opposite direction to that used in the conditioning period (the period of contracting and expanding stimuli) for some time thereafter.

Fig. 1.16a depicts the changes in the magnetic resonance imaging signal during real and illusory visual motion. The strength of the signal in the middle temporal area is shown for the case of the moving concentric rings either continuously expanding (Exp), continuously contracting (Con), reversing direction (expanding then contracting, Exp/Con) or stationary

(Stat). Following the periods of continuous unidirectional local motion (the expanding or contracting stimuli), a visual motion after-effect is seen by the subject in the physically stationary rings. Following the period of reversing the direction, no motion after-effect was reported by the subject. The magnetic resonance imaging response during the period of expansion of the rings, contraction of the rings, or reverse direction expansion/contraction is about the same (between 3.0 and 3.5 per cent). The important point to notice is that the magnetic resonance imaging response immediately after the single-direction stimulus (i.e. when the motion after-effect was being experienced) remained high for a considerable period of about 20 seconds after the stimulus offset, much longer than that of the case when there is no after-effect, such as the case of reverse expansion/contraction in the stimulus, for which the magnetic resonance imaging signal returns to the no stimulus baseline in about 5 seconds. Thus the middle temporal area remains active during the period when there is no stimulus but the subject experiences the visual after-effect.

Fig. 1.16b shows the quantitative relationship between the visual after-effect and activity in the middle temporal area as measured by magnetic resonance imaging (compare with fig. 1.16a). The line gives the magnetic resonance imaging amplitudes during and after single-direction expansion of the concentric rings minus the amplitudes during and after reversing-direction conditions. During the first 40 seconds there was no difference between the activation produced by a single-direction versus reverse-direction stimulus (see fig. 1.16a). During the next 40 seconds, the magnetic resonance imaging results show that as the subjects observed stationary stimuli, the magnetic resonance imaging was elevated for about 20 seconds following the single-direction stimulus. These results may be compared with the psychophysical data indicated by the open squares. These measured the period as reported by the subject of the time course of the visual illusion which was experienced while observing the stationary rings in a related series of psychophysical experiments. There was very good agreement between the time course of the decline in the visual illusion given by the open squares and that of the activity in the middle temporal area as indicated by the magnetic resonance imaging. The results show that illusions are accompanied by neural activity in the brain, and in the case of the waterfall effect this is to be located in the middle temporal area.

Another area, designated V4 in the primate cortex, possesses neurons that are excited by particular wavelengths of light as well as, on occasion, responding to the orientation of lines. This indicates that they are concerned with colour and form. Still other adjacent areas, designated V3 and V3a, have been shown to be selective for form alone (Zeki, 1993). All this work led Zeki to speculate that the visual cortex of primates possesses modules: namely, separate areas that are functionally specialized for various properties of objects in the visual world such as motion, form and colour. Zeki (1999) expresses this modularity of the visual pathway in vivid terms, contrasting the modular system concerned with colour with that concerned with faces: 'Assuming that, through the operation of its logic, the brain makes an inference about certain physical properties of surfaces, interpreted as colour, it seems difficult not to believe that it uses the same inferential method to deduce, for example, the expression on a face or the appearance of an object' (p. 2058).

More than thirty of these modules have now being identified by Van Essen and his colleagues (Felleman and Van Essen, 1991). Plate 1.3a shows many different cortical areas on

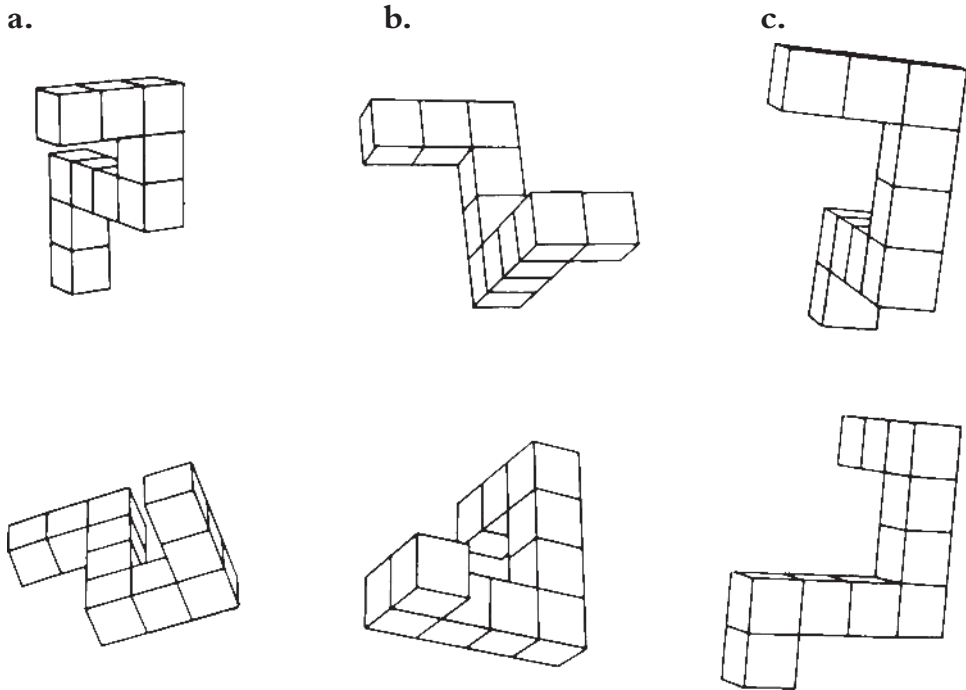


the right-hand side of the macaque's brain, with those colour-coded indicating the position of different cortical modules devoted to vision. The two smaller figures on the smaller scale on the left show the view of the right-hand side of the brain viewed from the outside (the upper one) and from the inside as if the brain were cut in half (the lower one). The folded macaque cortex has been unfolded using sections of the brain, so that the spatial relationships between the different cortical modules can be appreciated. The different names of these modules have in most cases been abbreviated to initials. The interconnections of these modules are shown in Plate 1.3b, with each line representing many axons (from a few hundred thousand to millions) passing in both directions (see also Plate 1.4c). The retinal ganglion cells at the bottom of the figure project to the lateral geniculate nucleus (LGN) in the thalamus, which in turn then projects to the first visual area in the cortex, V2 (which has four parts). At the very top, HC is for hippocampus, and ER for entorhinal cortex. Note that at the highest levels, module 46 receives a very large number of inputs. These can best be appreciated by reference to Plate 1.5, which shows the topological organization of the macaque cortical visual system. Reciprocal connections are coloured red; one-way projections going from left to right are coloured blue; and one-way projections going from right to left are green. A total of 301 connections is represented, of which 62 are one-way. This non-arbitrary structure is a best-fit representation in two dimensions of the connectional topology of this system, in which the positions of areas are specified by their positions, being ones that minimize the distance between connected areas and maximize the distance between areas that are not connected. The analysis represents in a spatial framework the organizational structure of the network of cortico-cortical connections between modules of the visual cortex.

## 1.6 Mental Images and Representations

Helmholtz's conception of perception as a matter of unconscious hypothesis formation was further developed in the late twentieth-century computational theory of visual perception. In its most sophisticated form, this conception was elaborated by David Marr, who 'adopted a point of view that regards visual perception as a problem in information processing' (Marr, 1980, p. 203). The informational input is conceived to be the light array (which he referred to as an 'image') falling upon the retina, and the output is held to be the construction of efficient and useful symbolic descriptions of objects in view. According to Marr, 'vision is the *process* of discovering from images what is present in the world and where it is' (Marr, 1980, p. 3). It is the process of transforming the information implicit in an image into an explicit description of what is seen. Marr conceived of the brain as operating a system of symbols that represent features of an image in order to construct descriptions. By a series of computational operations on the symbolism, the brain can, in the final stage of the visual process, produce a description of shapes of objects, their distance, orientation and identity. Marr suggested that 'if we are capable of knowing what is where in the world, our brains must somehow be capable of representing this information' (Marr, 1980, p. 3).

Marr suggested that a strong argument in favour of this representational process is provided by the experiments of Shepard and Metzler (1971), illustrated in fig. 1.17. This shows



**Fig. 1.17.** Drawings similar to those used by Shepard and Metzler in their experiments on ‘mental rotation’. (Marr, 1982, fig. 1.2.)

line drawings of simple objects that differ from one another either by a three-dimensional rotation or by a rotation plus a reflection. Those in (a) are identical as a clockwise rotation of the page by  $80^\circ$  shows; those in (b) are also identical, and again the relative angle between the two is  $80^\circ$ , but a rotation in depth is required to make the first coincide with the second; those in (c) are not identical, as no rotation brings them into congruence. Shepard and Metzler showed that the time taken to decide whether a pair are congruent varied linearly with the angle through which one figure had to be rotated in order to be brought into correspondence with the other. This led them to the idea that mental representations of the shapes of the pair existed, that one of them is being rotated at constant velocity relative to the other in mental space until they match. Marr (1982) considered this kind of experiment to offer good evidence for the existence of mental representations involved in the normal recognition of objects in the visual field.

In the late 1970s Anne Treisman and her colleagues performed psychological experiments in order to develop models of feature perception and integration in a visual search task that could be tested by neurobiologists. Treisman and Gelade (1980) established that the search time for a unique item is faster when all items differ by one attribute than if all items differ by two or more attributes. If a subject is instructed to identify whether an item is present or not in an image such as that in Plate 1.6a, the unique stimulus ‘pops

out', and subjects take about the same time to find the stimulus regardless of how many items are present in the display, as shown in the graph of Plate 1.6c. Treisman took this to be consistent with a pre-attentive process in the display in which all attributes are scanned for the particular feature at once. If, however, the unique item differs by two attributes, as in Plate 1.6b, then it does not pop out. In this case of a conjunctive search, the more items present, the longer the search takes, as shown in the graph of Plate 1.6c. Treisman regarded this as being consistent with a serial search and successive shifts of attention (see Treisman et al., 1977).

In order to account for the differences in feature and conjunction searches Treisman (1986) suggested a hypothetical model of feature perception and integration, shown in Plate 1.7, taken to indicate how different types of visual information are processed separately and then combined into a coherent image. The elementary properties of objects in the visual field (such as colour, orientation, size and distance) are taken to be processed in separate parallel pathways, each of which generates a map that is tuned to a specific feature. Thus object perception is thought to begin with the parallel analysis of its component features, which does not involve attentional mechanisms. Stimuli automatically activate cells tuned to their features on these maps, which are representational structures that indicate the presence or absence of features. In the colour search of Plate 1.6a the subject would simply have to determine whether the feature blue was active in the colour maps. Conjunction searches, as in Plate 1.6b, are much more complicated, involving in this case the checking of activation on two different maps and their association with a particular location. Cognitive psychologists now take it that targets that give flat search functions, like that in Plate 1.6c, are dealing with visual primitives which are taken as the basic building blocks of perception, and are found by neurobiologists in the early parts of the visual pathway. As to the later parts of the visual pathway, as Crick comments:

We can see how the visual parts of the brain take the picture (the visual field) apart, but we do not yet know how the brain puts it all together to provide our highly organized view of the world – that is, what we see. It seems as if the brain needs to impose some global unity on certain activities in its different parts so that the attributes of a single object – its shape, color, movement, location, and so on – are in some way brought together without at the same time confusing them with the attributes of other objects in the visual field. (Crick, 1994, p. 22)

### 1.6.1 Misconceptions about images and representations

Marr's suggestion that there can be symbolic descriptions of objects in view in the brain does not make sense. A description is a form of words or symbols, a sentence expressing a proposition that specifies an array of features of an object, event or state of affairs. It can be true or false, accurate or inaccurate, detailed or rough-and-ready. A symbolic description may be written down or spoken; it may be encoded for concealment or for transmission. But there is no such thing as a description in the brain. For something to be a (semantic) symbol, it must have a rule-governed use. There must be a correct and an incorrect way of using it. It must have a grammar determining its intelligible combinatorial possibilities

with other symbols, which is elucidated by explanations of meaning that are used and accepted among a community of speakers. There can be no symbols in the brain; the brain cannot use symbols and cannot mean anything by a symbol. A symbol is used only if the user means something by it – but brains cannot mean anything. To mean something by a symbol is roughly to intend the symbol to signify such-and-such a thing – but brains can have no intentions.

Marr's idea that the output of the computational process is the production of a description of visible objects that is embodied in an internal representation which is made available as a basis for decisions is confused. The 'output' of the neuro-visual process, in so far as it can be said to have an 'output', is that the creature sees. But to see something is no more to construct or produce a description than it is to construct Helmholtzian hypotheses.

Shepard and Metzler's rotation experiments have also been misinterpreted. The fact that the time taken in all of these experiments is proportional to the angle of rotation of the figures visualized does not suggest that it takes longer to perform a greater rotation at constant velocity 'in mental space' than to perform a lesser rotation, since there is no such thing as rotating a mental image at constant (or variable) velocity – only such a thing as imagining an object rotating at constant (or variable) velocity. One can imagine a rotating object. But to imagine an object moving quickly does not mean that anything moved quickly in the imagination. The idea that it must take longer to imagine rotating a figure by 90° than to imagine rotating it by 45° is as misconceived as the thought that it must take longer to paint a slow-moving figure than it takes to paint a fast-moving one. There is no obvious reason why it should take longer to imagine a figure rotating through 90° than to imagine it rotating by 45° – for one is at liberty to imagine the first figure rotating fast and the second more slowly. To assume that it takes longer to match a figure rotated by 90° than to match one rotated by 45° because the figure is being rotated at constant velocity adds a further incoherent hypothesis to the misconception.

Treisman's interpretations of her experiments in terms of representational structures that indicate the presence or absence of features is also awry. One does not perceive representations of objects, unless one perceives paintings or photographs of the objects. To see a red apple is not to see an image or representation of a red apple. Nor is it to have an image in one's mind or brain, although one can conjure up images in one's mind, and sometimes images cross one's mind independently of one's wish or will. But the mental images we thus conjure up are not visible, either to others or to ourselves – they are 'had', but not seen.

Crick's assertion that 'the visual parts of the brain take the picture . . . apart' is likewise a misdescription. The visual scene is not a picture, although it may contain a picture if one is in an art gallery. The electrochemical reactions of the rods and cones of the retina to the light falling on them cause a multitude of responses in different parts of the 'visual' cortex, but that is not correctly characterized as 'taking the picture apart'. Nor does the brain have to 'put it all together' again in order to provide our view of the world. For our 'view of the world' is not a picture of the world (or of the visible scene), and the attributes of the visibilia in front of us do not have to, and cannot, be 'brought together'. For the colour, shape, location and movement of the red geraniums swaying in the wind cannot be taken apart (there is no such thing as separating these attributes from the objects of which they

are attributes), and the colour, shape, location and movement of the geraniums cannot be brought together in the brain, since these attributes are not to be found in the brain, either together or separately.

The pattern of neural firing that is a causal response to a stimulus in the visual field can be described without resort to Marr's idea of 'symbolic descriptions'. There are no 'symbols in the brain', but there are neural events that are causally correlated with certain other phenomena, such as seeing. We use our sense-faculties, such as vision, for apprehending how things are in our environment. The sense-organs are not information-transmitters, although we acquire information by their use. The neural correlates of features in a visual scene are neither 'representations' nor 'symbols'.

The experiments of Shepard and Metzler suggest that it may take longer to work out how a certain figure will appear when rotated thus than to work out how the same or another figure will appear when rotated otherwise. For one needs to exercise one's imagination – that is, one's powers to think of possibilities, to work out where this part of the figure will lie in relation to that part if the whole figure is rotated by 90°. One needs to think about the rotation of a figure, not to rotate an imaginary figure (since there is no such thing). In so thinking, one may, but need not, imagine a rotating figure. (And it is important to remember that thinking about something does not imply saying anything to oneself.)

We have seen that Crick's concern that the activities of the differently located cells which respond severally to colour, shape, location and movement when one is viewing an object need to be united somehow in order to form an image is misplaced. For no image is or needs to be formed in order to see the object which is visible. It may well be, however, that the firing of certain neurons in the brain is a causal condition for being able to see an object in the visual field, and that some of these are required to fire in response to shape, others in response to colour, and yet others in response to motion, etc. It is plausible to suppose that these functionally related groups of cells must fire more or less simultaneously if the animal is to enjoy normal visual perception. What is not plausible, because not intelligible, is that these functionally related groups of cells must form an image of anything or enable the brain to form an image. Their normal functioning is what makes it possible for *an animal* (not *the brain*) to see (not to form images).

## 1.7 What and Where Pathways in Object Recognition and Maps

The modules that compose the visual pathway from the retina to higher visual centres follow two diverging streams in the cortex (see Plate 1.4a): one pathway extends dorsally to terminate within the parietal lobe, including the motion detection area MT and the visual areas of the posterior parietal cortex; the other pathway extends ventrally to terminate in the temporal lobe (including V4 and the inferior temporal cortex). The work of Mortimer Mishkin and his colleagues in the early 1980s suggested that these two pathways serve different functions: the dorsal pathway is concerned with *where* an object is in visual space (motion, distance); the ventral pathway is concerned with *what* an object is (form, colour, texture, all of which are involved in object recognition) (Ungerleider and Mishkin, 1982).

Plate 1.4a illustrates these pathways, showing lateral views of the rhesus monkey brain, indicating the two major pathways both originating from V1, with arrows indicating the dorsal ‘where’ cortical stream, which takes a dorsal route to the parietal cortex, as well as a ventral ‘what’ cortical stream, which takes a ventral route to the temporal cortex. A simplified version is given in Plate 1.4c of the modules that participate in these two streams, together with their interconnections, the lines indicating both forward- and backward-projecting axons. The ventral stream (the ‘magnocellular’ (M) for the large lateral geniculate nucleus (LGN) neuron stream) consists of modules in the striate cortex (V1) which project from there to the middle temporal (MT) modules that are concerned with movement, as mentioned above. Projections proceeding from these to modules that include the medial superior temporal cortex (MST), fundus superior temporal cortex (FST), ventral intraparietal cortex (VIP), and finally to the posterior parietal cortex (PP) and superior temporal polysensory cortex (STP). The dorsal stream (the ‘parvocellular’ (P) for the relatively small lateral geniculate nucleus neuron stream) consists of modules in the striate cortex (V1) which project from there to extrastriate cortex (V2) and to V4 and finally to modules in the inferior temporal cortex (IT). It is clear that the ‘cardinal’ cells of the inferior temporal cortex that are involved in face recognition, mentioned above, fit neatly into the idea that the inferior temporal cortex module is involved with the ‘what’, i.e. with identification of objects.

Gerald Edelman has sought explanations for why the modules in the lateral geniculate nucleus and cortex, illustrated in Plate 1.4c, receive at least as many connections in the backward direction as in the forward direction to the ‘what’ and ‘where’ final pathways. He refers to these backward projections as ‘reentrant’: that is, nerves which make connections in the reverse direction to those along the principal pathway. An example of a reentrant pathway is that made by nerves which project back from the primary visual cortex (V1) to the lateral geniculate nucleus (LGN) in the thalamus, which is the reverse of the forward pathway from the retina to the thalamus to the primary visual cortex. Referring to the modules in the visual pathway as maps, he comments that:

The visual system of the monkey, for example, has over thirty different maps, each with a certain degree of functional segregation (for orientation, color, movement, and so forth), and linked to the others by parallel and reciprocal connections. Reentrant signalling occurs along these connections. This means that, as groups of neurons are selected in a map, other groups in reentrantly connected but different maps may also be selected at the same time. Correlation and coordination of such selection events are achieved by reentrant signaling and by the strengthening of interconnections between maps within a segment of time. (Edelman, 1992, p. 85)

## 1.8 Misuse of the Term ‘Maps’

When certain features of the visual field can be mapped on to the firings of groups of cells in the cortex, then the idea of maps in the brain arises. Neuroscientists often take these ‘maps’ as playing ‘an essential part in the representation and interpretation of the world by

the brain, *just as the maps of an atlas do for the reader of them*' (Blakemore, 1990, p. 265; our italics). It is by no means evident what could be meant by the claim that the topographical relations between groups of cells that are systematically related to features of the perceptual field play an essential role in the brain's interpreting something. To interpret, literally speaking, is to explain the meaning of something, or to take something that is ambiguous to have one meaning rather than another. But it makes no sense to suppose that the brain explains anything or that it apprehends something as meaning one thing rather than another. The claim that 'brain maps' (which are not actually maps) play an essential part in the brain's 'representation and interpretation of the world' cannot be 'just as the maps of an atlas do for the reader of them'. For a map is a pictorial representation, made in accordance with conventions of mapping and rules of projection. Someone who reads an atlas must know and understand these conventions in order to read anything off from the maps at all. The brain is not akin to the reader of a map, since it cannot be said to know any conventions of representations or methods of projection or how to read anything off the topographical arrangement of firing of cells in accordance with a set of conventions. For the brain cannot follow rules or conventions – since that presupposes knowledge of the rules and an intention to comply with them. Furthermore, the cells are not arranged in accordance with conventions at all, and the correlation between their firing and features of the perceptual field is not a conventional but a causal one.

## 1.9 The Binding Problem and 40 Hz Oscillations

The binding problem has been taken to arise when considering how the neurons responsive to the distance, textures, colours, different orientations of lines and edges of say, a house are interrelated so that the animal will be able to perceive the house as a unified object. More generally, neuroscientists have been puzzled about how the modularization of cellular function in the neural processes involved in perception enables the perceiving animal to apprehend a unified object in the visual field. Eric Kandel and Robert Wurtz, in a discussion interestingly entitled 'Constructing the Visual Image' (Kandel et al., 2000, ch. 25), explain that 'information about' (i.e. presumably, electrochemical responses to) form, motion and colour is carried by parallel pathways. This, according to Kandel and Wurtz, creates the 'binding problem':

How is information carried by separate pathways brought together into a coherent visual image? . . . How does the brain construct a perceived world from sensory information and how does it bring it into consciousness? . . . what the visual system really does [is] to create a three-dimensional perception of the world which is different from the two-dimensional image projected onto the retina. (Kandel and Wurtz, 2000)

How is information about color, motion, depth and form, which are (*sic*) carried by separate neural pathways, organized into cohesive perceptions? When we see a square purple box we combine into one perception the properties of colour (purple), form (square), and dimensions in depth (box). We can equally well combine purple with a round box, a hat or a coat ...



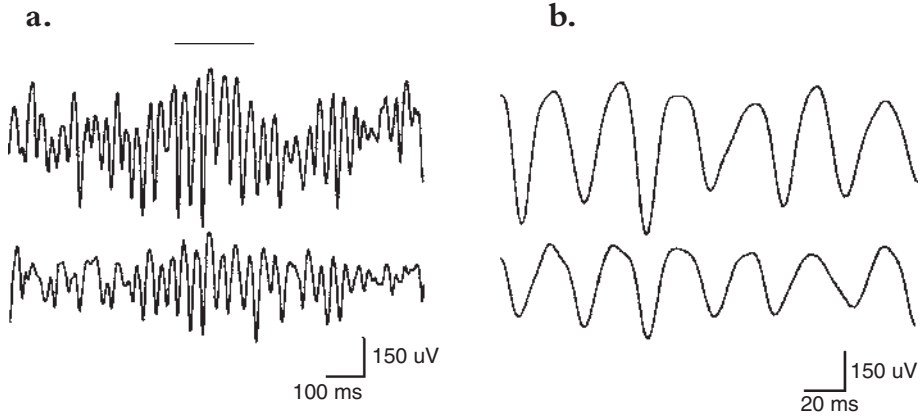
. . . visual images are typically built up from the inputs of parallel pathways that process different features – movement, depth, form and color. To express the specific combination of properties in the visual field at any given moment, independent groups of cells must temporarily be brought into *association*. As a result, there must be a mechanism by which the brain momentarily associates the information being processed independently by different cell populations in different cortical regions. This mechanism, as yet unspecified, is called the *binding mechanism* (Kandel et al., 2000, p. 502)

Kandel and Wurtz puzzled over the question of how, for example, adjacent houses and the trees in their gardens are each perceived, *with all their properties bound together*, as separate objects?

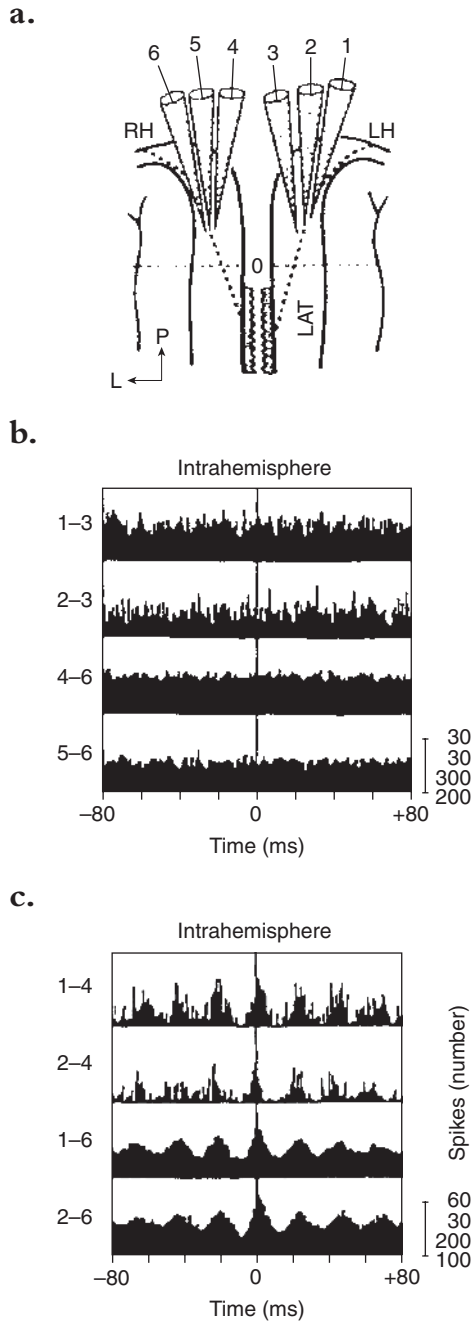
Francis Crick suggested that, at any given moment, any particular object in the visual field is ‘represented by’ (i.e. causally correlated with) the firing of a set of neurones, which are distributed in different ‘visual’ areas (for form, colour, motion, etc.). We perceive the object as a unity. ‘One striking feature of our internal picture of the visual world is how well organized it is. . . we seldom get things jumbled in space when seeing them under ordinary conditions’ (Crick, 1994, p. 232).

In the late 1980s, Singer and his colleagues offered a solution to the binding problem conceived as described above (Singer, 1991). They obtained what they took to be experimental evidence for the proposal that temporal synchrony of neuronal firing patterns may underlie binding (Gray and Singer, 1989), and suggested that ‘synchrony of oscillatory responses in spatially separate regions of the cortex may be used to establish a transient relationship between common but spatially distributed features of a pattern’ (Gray et al. 1990). They discovered that when some components of a visual scene were perceived by an observer as properties of a single object there was synchrony of the temporal impulses in the neurons that subserved each of the different components.

Consider, for example, two vertically oriented light bars moving at the same speed in the same direction past the eyes. Despite the fact that the bars are sufficiently far apart to register on two quite different parts of the retina which project to two distinct neuronal groups in the visual cortex (V1), there is a tendency for these bars to appear as a single object. The photomicrograph in fig. 1.18 shows the position of two such neuronal groups in area V1 of a cat. The bars are sufficiently far apart to be registered by two quite different parts of the retina which project to two neuronal groups in the visual cortex that are 7 mm apart (as indicated on the photomicrograph of the surface of the visual cortex by the white arrows). The black areas in this flat mount of the surface of the cortex indicate columns of neuronal groups, of which only the tops are shown, that are particularly responsive to vertical light contours. Microelectrodes are placed in the vicinity of these two neuronal groups, and the recordings made are shown in fig. 1.18a. The average impulse firing of the neurons in each of the groups (as shown by the field potentials) is oscillatory. Fig. 1.18b shows on an expanded time scale that the oscillation of both of the groups is at 40 Hz, and that they are in phase, despite the fact that they are 7 mm apart. No such coupled firing would be expected for neuronal groups at such a distance, and because of this coupling Singer and his colleagues took it that these two light bars would appear as a single object to the cat. The experience that the two light bars are one object is correlated with the fact that the



**Fig. 1.18.** Synchronized neuronal firing of two different groups of neurons in the visual cortex of a cat (area 17), shown in **a** and on a different time scale in **b**, during the observations of two vertically oriented light bars moving with the same speed and in the same directions, shown above on the surface of the cortex by two white arrows. (Bennett, 1997, fig. 3.5, after Gray and Singer, 1989.)



**Fig. 1.19.** Synchronized neuronal firing of three different groups of neurons in each hemisphere of the visual cortex of a cat. (Engel et al., 1991, fig. 3, p. 1178.)

neuronal groups in the visual cortex which are independently excited by the image on the retina of just one of the bars are now joined in a 'dynamic way' – that is, by a common frequency and phase of neuronal firing. This is an example of transient excitatory coupling of two neuronal groups within the same area of the neocortex, in this case the visual area.

Singer has also shown that there is interhemispheric synchronization of activity in the visual cortex when, he believes, a binding problem is being solved for a visual object. Suppose a single light bar is sufficient to stimulate three different neuronal groups, about 1 mm apart, in the visual cortex of one hemisphere. Electrodes are positioned in each of the groups, numbered 1–6 in area V1 of each hemisphere of the visual cortex of a cat, as shown in fig. 1.19a. The synchronization of the impulse firing and the phase of this firing, as measured by the different electrodes, can be shown by means of what is called a cross-correlogram. Fig. 1.19b depicts intrahemisphere cross-correlograms for the field recordings of pairs of neuronal group activity indicated by the electrode positions 1–3, 2–3, 4–6 and 5–6. If a periodic pattern is discernible in the cross-correlogram, then this indicates that the signals are correlated and gives information as to the common frequency and phase in the correlation. The cross-correlograms for the intrahemisphere recordings show a strong oscillatory modulation in the same frequency range of about 40 Hz, even though the electrodes may be separated by as much as 2 mm (fig. 1.19b). The cross-correlograms for interhemisphere recordings show surprisingly similar correlations, indicating that both hemispheres participate in the solution of the binding problem for the single white bar. This is not the case if the group of axons that join the two hemispheres (the corpus callosum) is cut. The cross-correlogram for recordings from the two hemispheres is now devoid of any periodic pattern and is flat, indicating that the firing of neuronal groups due to the light bar in each of the hemispheres is no longer correlated (fig. 1.19c). Singer takes it that the corpus callosum must mediate the synaptic connections between the two hemispheres that most likely participate in the solution of the binding problem.

Rodriguez and his colleagues (1999) have provided evidence for such long-distance synchronization of impulse activity in cortical modules of humans performing cognitive tasks. They recorded electrical brain activity from subjects who viewed ambiguous visual stimuli (perceived either as faces or as meaningless shapes). In this work they were able to show that face perception is accompanied by a long-distance pattern of synchronization, corresponding to the moment of perception itself and to the ensuing motor response. A period of strong desynchronization marked the transition between the moment of perception and the motor response. They suggest that this desynchronization reflects a process of active uncoupling of the underlying neural ensembles that is necessary to proceed from one cognitive state to another. Plate 1.8a and b shows the ambiguous visual stimuli used in their experiments, which are called 'Mooney' faces: namely, high-contrast pictures of a human face that are easily recognized as human faces when seen upright as in (a), but are difficult to recognize when inverted, as in (b). Subjects were asked to report as quickly as possible whether they had seen a face or not by pressing on one of two different keys. An electroencephalogram was recorded through 30 electrodes placed over the scalp of the subject and a precise time – frequency analysis carried out up to 100 Hz. Plate 1.8c and d show the spectral power following stimulation with the upright and the Mooney faces respectively.

Power peaks at about 230 ms after stimulus onset, and between 33 and 39 Hz. The perception condition elicits a significantly stronger response than the no-perception condition. The second peak lies at about 800 ms and  $40 \pm 5$  Hz; it follows after the reaction time ( $645 \pm 20$  ms for perception;  $766 \pm 22$  ms for no-perception), and no significant differences between conditions are found.

Thus humans as well as other animals show the synchronized oscillations of about 40 Hz over large areas of cortex during perception. This is perhaps more dramatically illustrated by means of Plate 1.9. This shows the average scalp distribution of about 40 Hz activity and phase synchrony. Colour coding indicates the power (averaged in a 34–40 Hz frequency range) over an electrode and during a 180 ms time window, from stimulation onset (0 ms) to motor response (720 ms). The frequency activity is spatially homogeneous and similar between conditions over time. In contrast, phase synchrony is markedly regional and differs between conditions. Synchrony between electrode pairs is indicated by lines. Black and yellow lines correspond to a significant increase or decrease in synchrony, respectively. Compared with the no-perception condition, which shows few synchronous patterns, the perception condition exhibits a sequence of localized spatial patterns that evolve over time. Synchrony first increases in the area between the left parieto-occipital and fronto-temporal regions. Desynchronization is then observed between the parietal and occipito-temporal area bilaterally. Rodriguez and colleagues propose that phase interactions between parietal and occipito-temporal regions are essential in the large-scale integration that is needed for the perception of upright Mooney faces. The second synchrony increase, which is probably linked to the motor response, is predominant between the right temporal and central regions. Phase synchrony, then, is directly involved in human cognition. Indeed, it has been suggested that the long-range character of the phase synchrony indicates that synchrony about 40 Hz and desynchrony may be viewed as a mechanism that subserves large-scale cognitive integration and not just visual-feature binding.

Both this work and that of Singer and his colleagues places emphasis on the solution of the binding problem itself as a necessary condition for us to be aware of something in the visual field – that is, to be conscious of that thing – without reference to an attentional mechanism. We will examine the nature of ‘attention’ and of ‘awareness’ further in chapter 2.

### 1.9.1 Misconceptions concerning the existence of a binding problem

The sense in which separate neural pathways carry information about colour, shape, movement, etc. is not semantic, but, at best, information-theoretic. In neither sense of ‘information’ can information be ‘organized’ into ‘cohesive perceptions’. In the semantic sense, information is a set of true propositions, and true propositions cannot be organized into perceptions (i.e. into a person’s perceiving something). In the engineering sense, ‘information’ is a measure of the freedom of choice in the transmission of a signal, and the amount of information is measured by the logarithm to the base 2 of the number of available choices – and this too is not something that can be ‘organized’ into perceptions. One cannot combine colour, form and dimensions into perceptions, just as one cannot put events into holes – this form of words makes no sense. And, correspondingly, when we see a square purple box,

we do not ‘combine’ purple, squareness and boxhood – for this too is a nonsensical form of words. It is true that in order to see a coloured moving object with a given shape, separate groups of neurons must be active simultaneously. But it does not follow that, in the semantic sense of information, the brain must ‘associate’ various bits of information; nor could it follow, since brains cannot act on the basis of information or associate pieces of information. Whether the brain, in some sense that needs to be clarified, ‘associates’ information in the information-theoretic sense is a further question. But if it does, that is not because the features of the object perceived have to be ‘combined in the brain’, for that is a nonsense.

Above all, to see an object is neither to see nor to construct an image of an object. The reason why the several neuronal groups must fire simultaneously when a person sees a coloured three-dimensional object in motion is not because the brain has to build up a visual image or create an internal picture of objects in the visual field. When we see a tree, the brain does not have to (and could not) bind together the trunk, boughs and leaves, or the colour and the shape, or the shape and the movement of the tree. One may see the tree clearly and distinctly or unclearly and indistinctly, and one may be sensitive to its colour and movement, or one may suffer from one or another form of colour-blindness or visual agnosia for movement. Which neuronal groups must simultaneously be active in order to achieve optimal vision, what form that activity may take, and how it is connected with other parts of the brain that are causally implicated in cognition, recognition and action, as well as in co-ordination of sight and movement, are what needs to be investigated by neuroscientists. Since seeing a tree is not seeing an internal picture of a tree, the brain does not have to construct any such picture. It merely has to be functioning normally so that we are able to see clearly and distinctly. It does not have to take a picture apart, since neither the visual scene nor the light array falling upon the retinae are pictures. It does not have to put a picture back together again, since what it enables us to do is to see a tree (not a picture of a tree) in the garden (not in the brain).

### 1.9.2 On the appropriate interpretation of synchronicity of neuronal firing in visual cortex

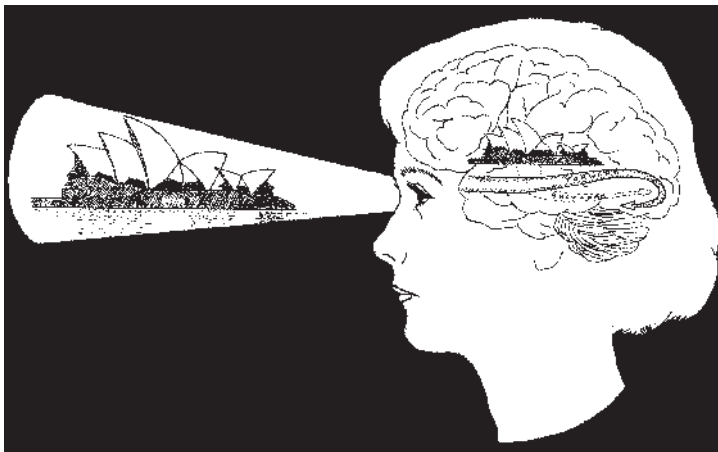
Kandel and Wurtz are not correct in suggesting that the brain ‘constructs a perceived world’; rather, it enables the animal to see a visible scene. Moreover, the brain does not create a three-dimensional perception which is different from the ‘two-dimensional image’ on the retina. It confers depth vision upon the animal, but the ability visually to discriminate depth is neither different from nor the same as an inverted reflection on the retina (which is incidental to vision anyway) – it is categorially distinct. The binding problem arises only if we consider that perceiving involves an internal picture or image of the external scene, so that the picture must be constructed, and the image ‘built up’. And then one might indeed wonder how the brain produces such coherent pictures or images, correctly associating the shape, motion, depth and colour of the perceived object and not ‘jumbling them up’.

To be sure, the cells that respond to motion, those that respond to shape, and those that respond to colour had better be active at (more or less) the same time; otherwise the person or animal will not see a coloured moving object of the relevant shape (or the asynchronicity

will simply be reflected in a corresponding delay in the perception). And presumably the simultaneous activity of these cell groups had better be connected in some way to the centres that control recognition, movement and co-ordination. That much seems obvious. And indeed the first step towards clarifying the processes involved has been taken by the discoveries of Singer and his colleagues on the synchronous 40 Hz oscillations of neuronal firing in different neurons in the different parts of the brain that are involved in seeing.

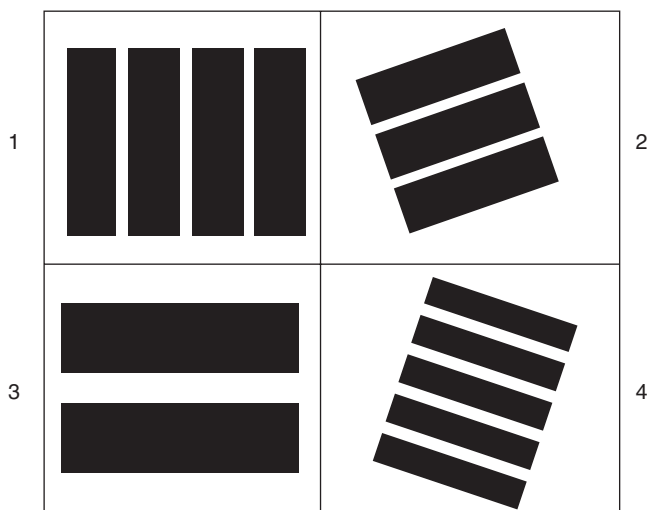
### 1.10 Images and Imagining

The problem of identifying the different parts of the human brain associated with visual experience has been greatly illuminated by the introduction of non-invasive visual imaging techniques such as positron emission tomography (PET) and fMRI, especially in the 1990s. One of the pioneers of these techniques, Kosslyn, has made extensive observations concerning the identification of those cortical areas that are active during the period in which he claims his patients 'visualise things in their imagination'. For example, when one perceives an object – for example, the Sydney Opera House – then the image of the Opera House on the retina is said to be 'reconstructed' in area V1 of the cortex, involving the forward projection from the retina to the lateral geniculate nucleus and from there to V1 in the occipital cortex and beyond as shown by the forward projecting arrow in fig. 1.20. Kosslyn set out to determine if one closed one's eyes and imagined the structure of the Opera House, instead of actually perceiving it, whether area V1 is involved in reconstructing the image of the Opera House again on the basis of information reaching it from higher centres such as those in the temporal lobes. This would require a backward projection to V1 as indicated by the arrow in fig. 1.20. Such a problem would seem to be ideally suited to



**Fig. 1.20.** Forward and backward projections to the primary visual cortex (area 17, V1) when viewing or imagining a scene. (From Bennett, 1997, fig. 5.3.)





**Fig. 1.21.** Illustration of the stimuli used to determine the role of visual area 17 in visual imagery. (Kosslyn et al., 1999, fig. 1.)

study with PET or fMRI. However, although there has been general agreement using non-invasive imaging that both the temporal-occipital and parieto-occipital visual association areas of the cortex are involved in imagining a visual scene such as the Opera House, there has been little agreement as to whether area V1 is involved (see Kosslyn and Ochsner, 1994; Roland and Gulyas, 1994). The probable causes of this disagreement almost certainly reside in what are called the baseline conditions in setting up the PET or fMRI studies. They involve, in some experiments, determination of the areas of the brain that are active during the imagery test compared with the activation present when the subject is not performing a requested imagery task but simply lies motionless with eyes closed. However, this does not guarantee that the subject is not ‘imagining something in visual consciousness’, i.e. day-dreaming accompanied by mental imagery. As the detection of the increased activity in the imaging task requires the subtraction of this background activity, it is easy to see that activation of V1 during a day-dreaming episode could remove the indications of activity in this area of the brain during the imaging task.

However, Kosslyn and his colleagues devised approaches that seemed to avoid this problem and establish the role of area 17 (V1) in imagining. They used two convergent techniques, in one of which subjects closed their eyes during PET while they visualized and compared properties (e.g. relative length) of sets of stripes. Fig. 1.21 shows the stimuli used, which consisted of stripes that vary in length, width, orientation and the amount of space between the bars. The numbers 1, 2, 3 and 4 are used to label the four quadrants, each of which contains a set of stripes. After memorizing the display, the subjects closed their eyes, visualized the entire display, heard the names of two quadrants, and then heard

the name of a comparison term (e.g. 'length'); the subject then decided whether the stripes in the first named quadrant had more of the named property than those in the second, and the response time was noted. The results showed that when people perform this task, area 17 (V1) is activated. The results of the PET scan, showing activation of area V1 (and areas 18/19) during imagery compared with baseline are shown in Plate 1.10, with the strength of the activation given by the colours, with blue, green, yellow and red representing increasingly higher activation.

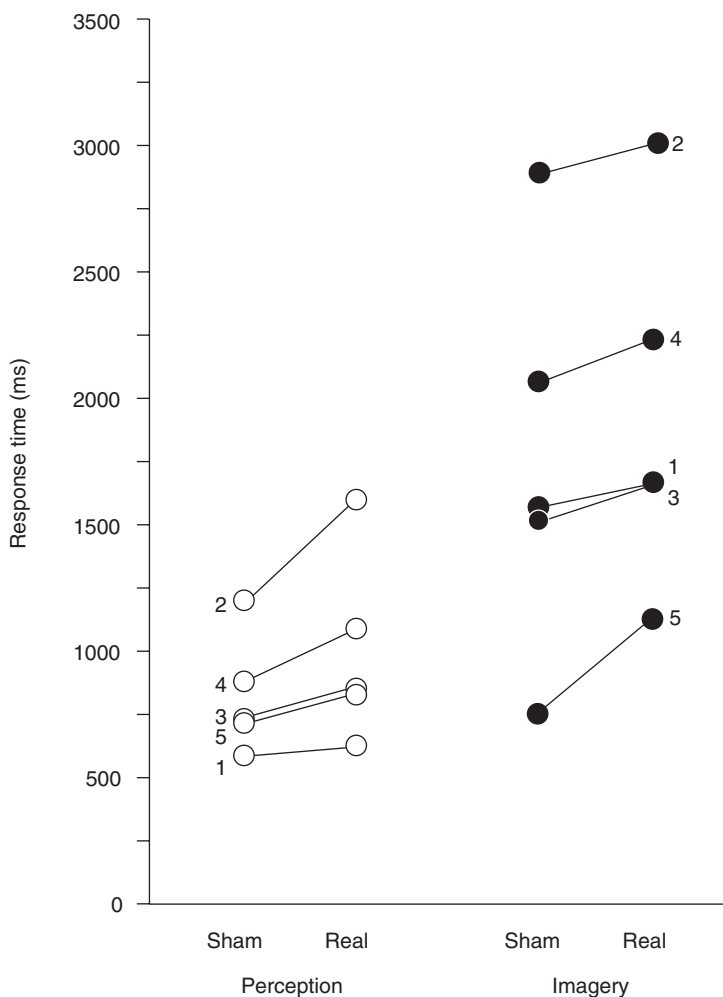
In the other technique used by Kosslyn, repetitive transcranial magnetic stimulation was applied to the medial occipital cortex before presentation of the same task, thus transiently disrupting activity in this area of the cortex. This led to impaired performance after stimulation compared with sham control conditions, as it did when the subjects performed the task by actually looking at the stimuli. These results when stimulation was delivered before the imagery and perception conditions are given in fig. 1.22. In this figure 'Real' refers to when stimulation occurred with the magnetic field directed into area 17 (V1), whereas 'Sham' stimulation occurred when the field was diverted away from this site. The response times during Real stimulation were greater than those during Sham stimulation in both imagery and perception (1945 ms versus 1759 ms, and 1002 ms versus 827 ms, respectively). As shown, this response time increases in all five subjects in both modalities (digits next to each line indicate the subject number).

In other experiments, Kosslyn and his colleagues have shown that many different parts of the brain are involved when one visually imagines each of the letters of the alphabet in turn, compared with just naming the letters of the alphabet to oneself in one's imagination. Kosslyn takes this to provide evidence that the former is a much more complex task than the latter. He claims that the neural network modules involved in the visual imagination are the same as those involved in seeing, and that these modules are also involved in the processes of visual attention (Kosslyn, 1994).

### 1.10.1 Misconceptions concerning images and imagining

The claim of Kosslyn and his colleagues is that visualizing something (i.e. conjuring up visual images of it) involves the excitation of much the same neural systems as would the corresponding visual experience. Whether or not much the same neural systems are involved in the exercise of the faculty for producing eidetic imagery (which we shall call 'fantasia') as are involved in the corresponding perceptual experience is an empirical question which Kosslyn believes he has solved. Before accepting that this is the case, it is important to be clear about differences between seeing and its objects, on the one hand, and between visualizing and its objects, on the other.

Secondly, it is mistaken to suppose that recognition involves comparing a mental image with what one perceives. This is a confusion that we have already encountered in Marr's theory of vision, since he supposed that to generate a 3D model representation in the brain, a 2½ D sketch has to be compared with a stored catalogue of 3D model descriptions. This is indeed required for 'machine vision' – but then machine vision is no more a kind of vision than computers are kinds of mathematicians. It is a fiction that human recognition involves matching a perception with a mental image.



**Fig. 1.22.** Effects of transcranial magnetic stimulation applied to visual cortex (area 17, V1) on the response time for perception and imagery tasks. (Kosslyn et al., 1999, fig. 3.)

Thirdly, to perceive something is not to have, construct, or reconstruct an image of it in the mind. To perceive is not to have or to form images, and what is perceived is not an image save in the cases in which one perceives pictures. To form a mental image of a scene is not to imagine an image of that scene. That is something a painter might do when he is trying to visualize the painting he intends to paint. But to form a mental image of a scene is visually to imagine *that scene* (not an image of it).

Finally, it is incoherent to suppose that one can discover things by reading off visual or spatial information from one's mnemonic image or 'visualized display' of an antecedently

seen object. According to this conception, a mental image is a pictorial representation, akin to a private photograph, from which one can derive information about what it is an image of by observation. It is uncontentious that one might try to recollect the different lengths of stripes in experiments of the kind carried out by Kosslyn. What is problematic, however, is the idea that one might determine the length of the stripes in a particular quadrant in one's mental image by observation. One may visualize a set of stripes in one quadrant, and other sets of stripes in other quadrants. One may imagine the stripes in one quadrant to be longer than those in another. But one cannot *discover* that the stripes in one imagined quadrant are longer or shorter than the stripes in another (after all, one cannot *see* them). One may indeed come to *realize* that one has imagined the stripes in one quadrant to be longer than the stripes in another quadrant, but one cannot *find that out by looking* – since there is no such thing as looking at one's mental images. But one cannot make the comparison between the lengths of the stripes in the different quadrants that one is visualizing in order to find out their comparative length. (Similarly, one cannot measure, but only imagine measuring, the objects one visualizes.) Mental (visual) images are not like private pictures that only the subject can see.