



Brain Mechanisms Associated with Top-Down Processes in Perception

Chris Frith; Raymond J. Dolan

Philosophical Transactions: Biological Sciences, Vol. 352, No. 1358, Knowledge-based Vision in Man and Machine. (Aug. 29, 1997), pp. 1221-1230.

Stable URL:

<http://links.jstor.org/sici?sici=0962-8436%2819970829%29352%3A1358%3C1221%3ABMAWTP%3E2.0.CO%3B2-D>

Philosophical Transactions: Biological Sciences is currently published by The Royal Society.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://uk.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://uk.jstor.org/journals/rsl.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

Brain mechanisms associated with top-down processes in perception

CHRIS FRITH AND RAYMOND J. DOLAN

Wellcome Department of Cognitive Neurology, Institute of Neurology, Queen Square, London WC1N 3BG, UK

SUMMARY

Perception arises through an interaction between sensory input and prior knowledge. We propose that at least two brain areas are required for such an interaction: the 'site' where analysis of afferent signals occurs and the 'source' which applies the relevant prior knowledge. In the human brain, functional imaging studies have demonstrated that selective attention modifies activity in early visual processing areas specific to the attended feature. Early processing areas are also modified when prior knowledge permits a percept to emerge from an otherwise meaningless stimulus. Sources of this modification have been identified in parietal cortex and in prefrontal cortex. Modification of early processing areas also occurs on the basis of prior knowledge about the predicted sensory effects of the subject's own actions.

Activity associated with mental imagery resembles that associated with response preparation (for motor imagery) and selective attention (for sensory imagery) suggesting that mental imagery reflects the effects of prior knowledge on sensory processing areas in the absence of sensory input. Damage to sensory processing areas can lead to a form of sensory hallucination which seems to arise from the interaction of prior knowledge with random sensory activity. In contrast, hallucinations associated with schizophrenia may arise from a failure of prior knowledge about motor intentions to modify activity in relevant sensory areas. When functioning normally, this mechanism permits us to distinguish our own actions from those of independent agents in the outside world. Failure to make this distinction correctly may account for the strong association between hallucinations and paranoid delusions in schizophrenia; the patient not only hears voices, but attributes (usually hostile) intentions to these voices.

1. THE PHYSIOLOGICAL BASIS OF PERCEPTION

We take perception to refer to our conscious experience of sensory input. For perception to occur incoming sensations must be imbued with meaning on the basis of our past experience and prior knowledge. In this case the neural activity which most directly relates to our perceptual experience will not be the afferent signals which reflect accurately the physical properties of the visual scene, but activity reflecting integration of these signals from the outside world with prior knowledge stored in our brains. Our purpose in this paper is to review experiments which have identified brain regions where activity corresponds to perception and to suggest a framework for investigating how afferent signals interact with prior knowledge and expectations.

How can we study perceptual processes in the human brain? In the last few years new functional brain imaging techniques have made considerable advances. The two major techniques are positron emission tomography (PET) and functional magnetic resonance imaging (fMRI). Both these techniques depend on the observation that neural activity consumes energy, and consequently, is followed by local increases in blood flow in order that energy stores can be replenished. We

can use measures of local changes in blood flow to provide a sensitive index of local changes in neural activity. The spatial resolution of these techniques is between 5 and 15 mm, and the temporal resolution between 10 and 30 s. This permits the study of brain activity at the systems level (i.e. where the fundamental unit of activity is not the single cell, but the pattern of activity in a large population of neurons). This raises the exciting possibility of identifying the physiological underpinnings of cognitive theories of brain function (for an account of PET and fMRI see Frith & Friston (1997)). Although we are still at the beginning of this research programme, much has been learned about the physiological basis of perception.

2. WHERE SENSORY INPUT IS NOT ENOUGH

(a) *Brain regions where neural activity reflects perception*

One situation in which prior knowledge has a marked effect is in the perception of impoverished pictures. If a full grey-scale photograph is sufficiently fragmented or impoverished, then the object shown in the picture becomes extremely difficult to recognize (figure 1). However, if the viewer is shown the original picture

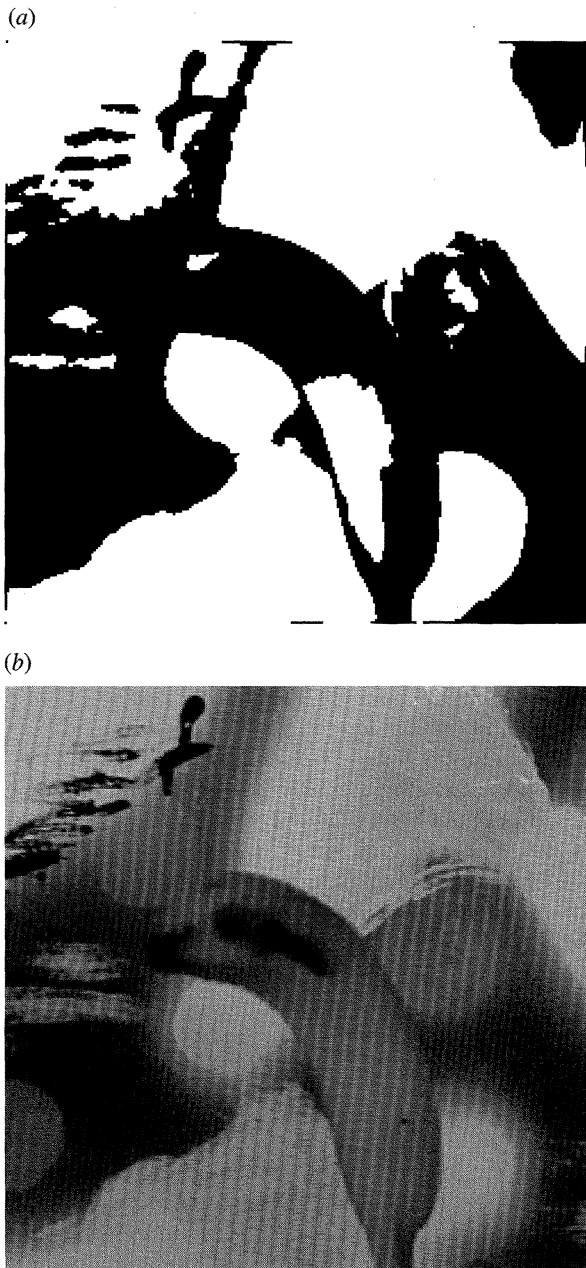


Figure 1. (a) An impoverished view of an object. (b) A recognizable view of the object shown in (a).

then, when the impoverished picture is viewed again, the object can now be perceived in the sense that the viewer recognizes what is depicted (Ramachandran 1994).

This phenomenon naturally lends itself to a brain imaging study. Brain activity can be measured when the person is viewing the meaningless arrangement of blobs and this can be compared with the activity elicited when the viewer, on the basis of prior knowledge, can perceive the object (Dolan *et al.* 1997). In both these conditions the visual input is identical. Activity in brain areas which reflect purely visual input will be the same. However, regions which reflect perception, rather than visual input, will show different activity in the two conditions. Using this approach we included two classes of pictures; one of objects and one of faces. First the volunteers viewed the impoverished picture without prior knowledge and perceived

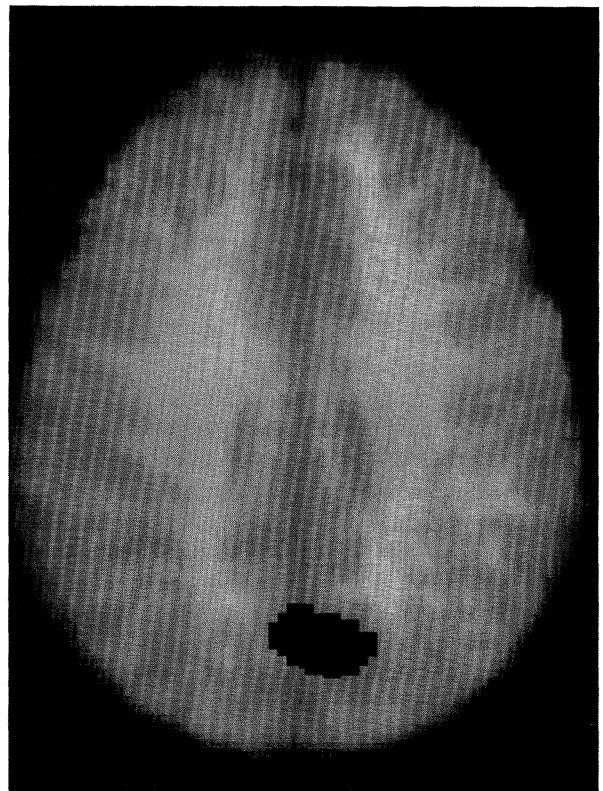


Figure 2. A horizontal slice through the brain (the average magnetic resonance image of the volunteers) with the front at the top and the left on the left. A significant area of activation in the medial parietal lobe is displayed in black. This activity was associated with the recognition of an object or a face presented in an impoverished view.

nothing meaningful. The knowledge about the pictures was then supplied by showing the volunteers the full grey scale version prior to the scan in which they saw the impoverished picture for the second time. There was no change of activity in primary visual areas, confirming that these areas were reflecting visual input rather than perception. When the object or face in the picture was now perceived significant increases in activity were detected in two brain regions: in medial parietal cortex (the precuneus; figure 2) and in the fusiform gyrus (inferotemporal cortex; figure 3).

The fusiform gyrus is part of extra-striate cortex (the so-called visual association areas). This area was more active on the left for objects and on the right for faces. There is much previous evidence that the fusiform gyrus (inferotemporal cortex) is specialized for the perception of objects of various kinds. For example, Tanaka *et al.* (1991) showed that cells in this region responded best to various complex figures including hands and faces. Similar evidence in humans comes from brain imaging studies (Haxby *et al.* 1991; Clark *et al.* 1996). In particular, the right fusiform gyrus seems to be specialized for the perception of faces (De Renzi 1986; Sergent *et al.* 1992). It is plausible, therefore, that the perception of objects and faces is associated with activity in the fusiform gyrus.

The modulation of this activity by prior knowledge of the stimulus could also occur in this area. The presentation of the full grey scale picture could leave

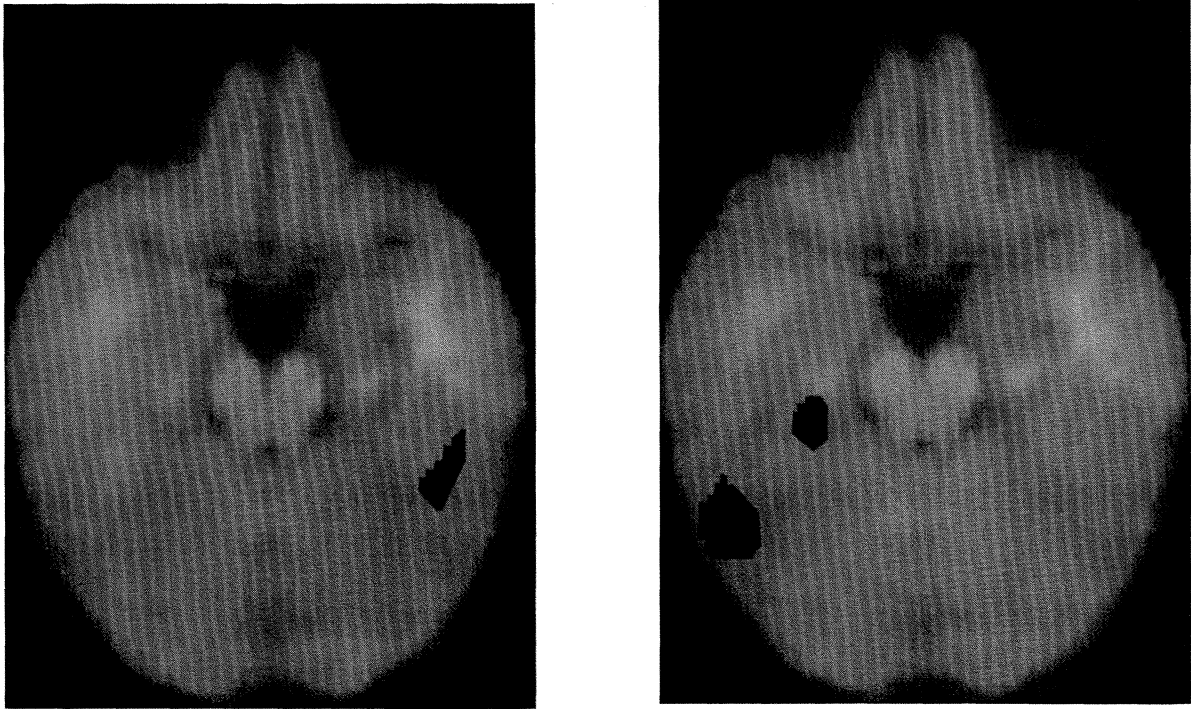


Figure 3. Horizontal slices as in figure 2 at a much lower level through the inferior temporal cortex. The left-hand part shows activity in the right inferior temporal cortex associated with face perception. The right-hand part shows activity in the left inferior temporal cortex associated with object perception.

traces (priming) in terms of modified synaptic connections so that the presentation of an impoverished version of the picture would be sufficient to re-elicited the activity previously associated with the complete version of the picture. However, this account leaves no role for the activation we observed in medial parietal cortex. In contrast to what happened in the fusiform gyrus, the activity in this area did not distinguish between the type of object perceived. Little is known about the function of the medial parietal area identified in this study. Previous brain imaging studies suggest that this region is involved in a crude form of imagery associated with long-term memory (Fletcher *et al.* 1995). It is activated during retrieval from memory, particularly when the items being retrieved have a high degree of imageability. Detailed analysis of the pattern of activity across all twelve scans in our study of impoverished pictures yielded evidence of an interaction between the precuneus and the right fusiform area when perceiving faces and between the precuneus and the left fusiform area when perceiving objects. On the basis of these interactions we propose that reactivating a memory representation of the picture requires the precuneus which can 'tune' the response of the fusiform area so that the gaps can be filled in when the impoverished version of the picture is presented for a second time. This leads to perception of the object or face in spite of the impoverished input. In this account the fusiform gyrus is a *site* where afferent inputs are modified by prior knowledge so as to give rise to the perception and recognition of the stimulus. The precuneus is a *source* from where these modulatory signals arise. In this example our suggestion of the need for a separate site and source remains somewhat speculative, since knowledge of the full picture might remain as

traces in the region where the visual input was analysed. However, in the next section we shall describe paradigms in which priming of activity in a single region by prior visual input is not sufficient to explain changes in perception. These paradigms are all concerned with attention.

(b) *Modification of perception by attention*

Figure 4 shows a large letter (S) made up of small letters (L). If we are asked to report what letter is presented we can respond either at the *global* level (S) or at the *local* level (L). In this paradigm the phenomenon of priming by a previous visual input is insufficient to determine whether we attend to global or local features. Our attention can be directed by verbal instructions or determined by an act of will. In these cases there is no prior visual input available to prime the region concerned with perception. The priming signal must have arrived from some other brain region. We (Fink *et al.* 1996) showed volunteers hierarchically organized visual figures in which a large letter was constructed from small letters (as in figure 4).

In one condition volunteers saw a sequence of such figures and attended to the local level throughout the scan. In another condition they attended to the global level. It should be noted that the set of stimuli presented were the same in both conditions. Attention at the global level was associated with increased activity in the right lingual gyrus, while attention to the local level was associated with activity in left inferior occipital cortex (see figure 5).

The precise function of these areas is unknown, but they are part of extra-striate cortex ('visual association' areas). As far as we can tell, the focus in the left inferior

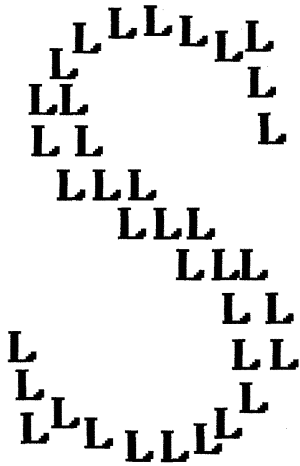


Figure 4. A figure which, at the global level is an S, while at the local level it is a set of Ls.

occipital cortex (local processing) falls between V1/V2 and V2/V3 boundaries while the focus in the right lingual gyrus (global processing) lies anterior to the V2/V3 boundaries. Almost certainly both these areas are concerned with different aspects of form processing. The laterality of these effects is consistent with findings from previous studies in patients with circumscribed lesions (Robertson & Lamb 1991). Patients with left-sided lesions have problems with local processing, while patients with right-sided lesions have problems with global processing. The sites of the modulatory effects we observed are, however, some distance from the lesion sites which lead to dysfunctions in local or global processing.

The brain regions concerned with local and global features have yet to be studied in detail. However, the direction of attention to simpler features has received detailed investigation. A key study in this respect is that of Corbetta and his colleagues (1991). Volunteers were shown complex visual displays in which variations occurred in colour, form and movement. When volunteers attended to changes in colour, neural activity (indexed by changes in blood flow) increased in part of extra-striate cortex centred on the lingual gyrus. When the volunteers observed the same display, but now attended to changes in movement, activity increased in a more lateral part of the extra-striate cortex at the occipito-temporal junction. Similar areas had previously been activated in a very different type of study by Zeki *et al.* (1991). In this study volunteers passively observed different stimuli. Their state of attention remained constant, but the stimulus varied. When coloured stimuli were contrasted with black and white stimuli, increased activity was seen in the lingual and fusiform gyrus. When moving stimuli were compared with stationary ones the more lateral region at the occipito-temporal junction became active. These areas have been identified as homologues of the colour and motion areas (V4 & V5), identified in monkey extra-striate cortex on the basis of single cell recording studies. Although colour and motion are basic visual features, these colour and motion areas are in extra-striate (the so-called visual association areas), rather



Figure 5. Horizontal slices through the brain showing activity associated with global or local processing. On the left, the activity associated with global processing is seen in the right extra-striate cortex. On the right, the activity associated with local processing is seen in the left extra-striate cortex.

than primary visual cortex. We know from other studies that signals in these areas do not simply reflect visual input. For example, activity in V4 (the colour area) represents perceived colour rather than wavelength and therefore reflects visual inputs modified by the processes that give rise to colour constancy (Zeki 1983). The results of these experiments by Zeki and by Corbetta show that we can activate the colour area either by presenting coloured stimuli to a passive subject (bottom-up processing) or by asking volunteers actively to attend to colour (top-down processing). In

the second case, prior knowledge about what to attend to has modified the response of a region concerned with a specific aspect of early visual processing.

On the basis of this and other studies of selective attention we can conclude that prior specification of the focus of attention causes a modification of the neural response in brain areas specialized for the analysis of the appropriate features. (There may also be inhibitory effects on areas concerned with irrelevant or distracting stimuli; Haxby *et al.* (1994)). Activity in these areas correlates with perception rather than sensory input. These studies have identified the *site* of modulation associated with attention and perception, but where is the source of this modulation?

(c) Identifying the source

Information about brain areas concerned with the processing of global and local features has previously been obtained from the study of patients with circumscribed lesions. These studies suggest that impairments in attending to global or local features of stimuli are associated with damage to the right or the left hemisphere, respectively. However, the most common location for this damage is in lateral temporo-parietal areas (Robertson & Lamb 1991). As already mentioned, these areas are a considerable distance from the extra-striate areas identified in our brain imaging study. This discrepancy is not surprising. Problems with attending specifically to global or local features, while retaining the ability to discriminate such features, occur because the patient is no longer able to *modify* activity in the target areas (the sites) even though these areas can still function adequately enough to ensure that the appropriate shape analysis occurs. This implies that the damage has affected the source of the modulation (or the connections between site and source). Damage to the target areas would lead to fundamental problems with form perception rather than problems with attention to different aspects of the scene.

How can we use functional imaging to identify the source of the signals by which prior knowledge affects perception? As we have already shown, the site of modulation is determined primarily by the object of attention. These sites will be in different locations determined by the nature of the attended object or specific features of an object. However, we propose that there is only one source for this modulation, directing its signal to one or other of the sites. If there were two sources, then the problem would remain as to how one source, rather than the other, would 'know' that it was supposed to be active in a given situation.

In our study of global and local processing there were two conditions; attending to global or attending to local features. The single source of the relevant modulatory signal would be active in both conditions and would therefore be rendered 'invisible' by categorical comparison of the two conditions. In order to identify the source we designed a second study in which volunteers had to switch attention between global and local processing throughout the scan (Fink *et al.* 1996, experiment 2). In this paradigm the task of the volun-

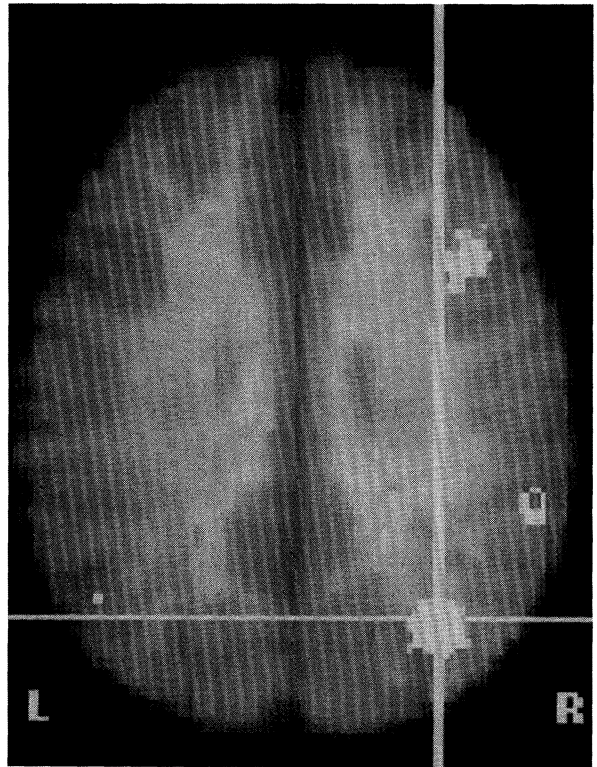


Figure 6. A horizontal slice through the brain. The cross hairs pick out an area in the right temporo-parietal cortex where activity was associated with the requirement to sustain attention to either the global or the local level.

teer was to report whether a target letter (e.g. A) was present at the global or the local level. In one condition the number of switches required between the local and the global level was low; the target letter was presented at the local level for the first six scans, and at the global level for the second six scans. In another condition the rate of switching between the local and the global level was high. Across the 12 scans the number of switches was varied systematically between one and 35 per minute, while the time spent at either the global or the local level was kept constant at 50% during the 30 s scanning window. We found that the longer attention was sustained at one or other level (i.e. the fewer the switches) the greater the activity in a region of the right temporal-parietal-occipital junction (see figure 6). The identification of this area as the source of the modulation is entirely consistent with the results from studies of patients with lesions.

In this level switching paradigm the direction of attention is not determined by prior instructions, but by a combination of the visual input and a memory of the specified target letter. While attention is sustained at one level, we presume that knowledge about the target letter is used to 'tune' areas concerned with detection of features at either the local or the global level. Does the temporo-parietal-occipital junction contain some abstract representation of the target letter that can be used for this purpose, or is activity in this area required to 'hold' the early visual areas in a particular state of 'tuning'?

Switching between levels in this local/global task was elicited by the stimulus. The prespecified target could

appear at either the local or the global level and the volunteer had to switch his attention accordingly. Presumably a higher source of modulation is engaged when attention is switched endogenously (i.e. the volunteer decides for himself where to direct attention). It is far more difficult to design functional imaging experiments to visualize such high level sources. Essentially we have to compare two conditions; in one, volunteers have to attend selectively to some object or feature; in the other, they must not attend to these features or, ideally, to anything else. Clearly the mental state of a volunteer in such a condition of 'inattention' is never going to be well controlled. Nevertheless, 'passive' viewing has been a commonly used condition in functional imaging studies and can give useful results.

For example, in a recent study (using fMRI) volunteers either passively viewed a field of white dots on a black background which appeared to move towards them (motion in depth), or attended to this field of dots in order to detect subtle changes in the rate of movement (Büchel *et al.* 1997). There were two other conditions; one in which the volunteers passively viewed a stationary field of dots and a baseline condition in which they viewed the black background without the dots. When brain activity was compared across these conditions, three distinct patterns of activity could be observed, as follows.

(i). In primary visual cortex (V1) there were highly significant activations when viewing dots was contrasted with viewing a black background. There was, however, little further increase in activity in this area when the dots moved or when the volunteers attended to this movement. This area reflects aspects of visual input, but is not much affected by attention.

(ii). When moving dots were contrasted with stationary dots, significant increases in activity were observed in a number of extra-striate visual areas, including the occipito-temporal junction (V5). When volunteers attended to motion, activity in these areas increased still further, replicating the results of Corbetta and his colleagues. We identify these areas as the *site* of modulation by attention.

(iii). A number of areas were more active when attention was compared with passive viewing. These areas included right prefrontal cortex, premotor cortex (including the frontal eye fields), anterior cingulate cortex, the insula and the thalamus. These areas did not show any difference in activity between moving and stationary dots viewed passively.

This result is in agreement with a wide range of studies suggesting that there is a 'network' of regions that are activated during tasks that require attention (Mesulam 1990). These include prefrontal cortex, premotor cortex and anterior cingulate cortex. It is thought that these different areas have different specific functions in relation to attention. For example, the frontal eye fields might be concerned with preparation for an appropriate response. We assume the source of the signals that modulate activity in the extra-striate areas is to be found somewhere in this attentional network. However, it is not possible from the simple comparisons between conditions described here to

decide which of the many areas might be considered to be the source. On anatomical grounds, posterior parietal and dorsolateral prefrontal cortex are the most likely candidates, since these areas have dense connections to the relevant sites of modulation (prefrontal cortex-posterior parietal cortex, Petrides & Pandya (1984); posterior parietal cortex-V5, Cavada & Goldman-Rakic (1989)). We assume that dorsolateral prefrontal cortex modulates posterior parietal cortex which, in turn, modulates V5. Following Posner, we would suggest that the posterior parietal cortex is concerned with exogenous shifts of attention caused by external stimuli, while the prefrontal cortex is concerned with voluntary, endogenous shifts of attention. Currently, new methods of analysis are being developed which are based on correlations between activity in different regions rather than simple comparisons (effective connectivity, Friston (1994); structural equation modelling, McIntosh & Gonzalez-Lima (1994)). These methods will permit direct assessment of interactions between brain areas and should provide better means for identifying sources of modulation.

(d) *Prior knowledge about action*

So far we have discussed situations in which the prior knowledge that is applied comes from the past. On the basis of images from the past (e.g. the full grey scale picture) the relevant perceptual areas are modified. However, such modulations can also be based on prior knowledge which enables us to predict what sensory changes will happen in the future. We are implicitly making such predictions whenever we perform an action. For example, if we move a limb, there will be perceptual changes consequent upon this movement, such as our feeling of the new position of the limb, and possibly the sight of our limb, moving across the visual field. The most famous example is supplied by von Helmholtz (1866) who wondered how it could be that the world did not appear to move about every time we moved our eyes. Clearly the image of the world must move across our retina, but yet we perceive the world as stationary. Helmholtz proposed that information from the command signal to move the eye could be used to cancel the movement signals from the retina. In other words we can predict the sensory changes caused by our movements on the basis of efferent signals arising in the motor system which determine the movement. These predicted sensory changes, which are based on motor commands, are referred to as the 'forward model' (Wolpert *et al.* 1995). This prior knowledge based on intended actions can be used to modify our perception.

There are several studies in which action has been shown to affect perception. For example, elevated thresholds for the perception of a touch to the finger are observed when the finger is being moved (Rushton *et al.* 1981). Creutzfeldt *et al.* (1989) recorded from electrodes implanted in the temporal cortex (auditory association areas) of patients undergoing neurosurgery, and found areas where activity decreased when the patient was vocalizing. With regard to the visual system, Paus *et al.* (1996) showed that eye movements

affected activity in the visual cortex. Volunteers were scanned while they moved their eyes, at different rates, in the dark. As rate of eye movement increased, there was increased activity in the frontal eye fields, but reduced activity in primary visual cortex (V1). A similar effect was observed by Wenzel *et al.* (1996) who observed reduction of activity in V1 during involuntary eye movements (nystagmus), induced by vestibular stimulation. These observations may well relate to the phenomenon of saccadic suppression. However, we can hardly claim that these effects of eye movements on the visual system are due to the application of a forward model. Application of such a model would involve more complex modulation of perception than the simple reduction of activity in primary visual cortex (V1). Furthermore, in the study by Wenzel *et al.* (1996) the eye movements were involuntary, so that any forward model could not have been based on an intention to move the eyes. The mechanisms by which intentions influence subsequent perception have yet to be examined with functional imaging.

(e) A framework for distinguishing site and source

A major theme of this paper has been the distinction between site and source. We have used the term 'site' to refer to the brain regions where activity underpins perception. In these regions signals reflecting sensory input are modified by signals reflecting prior knowledge and expectations. In contrast we have used the term 'source' to refer to the brain regions in which these modifying signals arise. 'Site' regions can be identified because activity here is altered by changing sensory input (e.g. stationary versus moving dots) *and* by changes in perception or attention (e.g. attended versus unattended). 'Source' regions are not affected by changing sensory input, but reflect level and, possibly direction, of attention. A feature of regions that are the site of perceptual modifications not examined in the studies reviewed so far, concerns the time course of the changes in activity. In those regions directly concerned with sensory input, we would expect to see a transient increase in activity each time the stimulus was presented. We would not expect to see such transient responses in source regions. Functional imaging techniques such as PET integrate activity over several seconds, and thus cannot distinguish transient from sustained activity directly. However, we can get an indirect estimate of transient activity (i.e. the increase in activity associated with presentation of a single stimulus) by varying the rate of presentation of stimuli. Using PET, a number of studies have shown that there is a linear relationship between the number of stimuli presented during the scanning window and the amount of activity in relevant brain areas (e.g. Price *et al.* 1992). For example, the total activity in the auditory cortex is directly proportional to the number of tones presented during the scan. This relationship indicates that there is a transient increase in activity elicited by each tone. The slope of the line is a measure of the amount of activity elicited by each tone. Thus, by varying the rate of stimulus presentation we can distinguish between areas which show transient or sustained activity. We have used this principle to identify

sites and sources in experiments on selective attention (Frith & Friston 1996) and object categorization (Rees *et al.* 1997).

3. PERCEPTION WITHOUT SENSORY INPUT

(a) *Mental imagery*

A normal percept arises from an interaction between afferent signals and prior knowledge. This formulation suggests the possibility of two kinds of extreme occurrence. In one situation the percept is entirely determined by prior knowledge. In the other, the percept fails to be appropriately modified by prior knowledge. We suggest that both cases can lead to hallucinations, but that these two kinds of hallucinatory experience will be qualitatively different. However, we must also accept that normal mental imagery may also be an example of a percept that is entirely dependent upon prior knowledge. Such images are not hallucinations since we are aware that their origin is 'in our heads' and not in the outside world.

To date more is known about motor imagery than visual imagery. There is good evidence that motor imagery engages similar brain regions as those involved in motor preparation (Jeannerod 1994). Thus, asking someone to imagine making a movement activates very much the same areas as asking someone to prepare to make that movement (Stephan *et al.* 1995). Consequently, we can infer that the same processes are involved. We suggest that there is a close analogy between motor preparation and selective attention. Thus, preparing to perceive a particular stimulus or feature (selective attention) is analogous to preparing to make a particular movement. A number of brain imaging studies confirm that mental imagery in various modalities is accompanied by increased activity in the appropriate modality-specific brain regions. For example, imagining the sound of someone's voice increases activity in superior temporal cortex (auditory association cortex) in addition to frontal regions, including Broca's area, concerned with speech production (McGuire *et al.* 1996). Performing tasks requiring visual imagery activates the occipital cortex, possibly including the primary visual cortex (Kosslyn 1996). We would expect imaging colour or movement to activate the same regions as those observed when volunteers attend to colour or movement (as in the study of Corbetta and colleagues). We are not aware of any brain imaging studies in which volunteers were simply asked to imagine colour or motion. In one study, however, volunteers were asked to name the colours of objects presented in the form of achromatic line drawings or words (Martin *et al.* 1995). No activity was seen in the posterior part of the fusiform gyrus believed to be the homologue of the colour area, V4. However, activity was seen in inferior temporal cortex 2 or 3 cm anterior to V4. This is an area in which cells are responsive to colour in combination with other object related features (Tanaka *et al.* 1991). Activity was also seen at the parietal-occipital junction. This area does not seem to contain colour responsive cells, but lesions of this area in neurological patients have been

associated with 'amnesia' for the colours of objects (Varney & Digre 1983).

(b) *The impoverished nature of imagery*

Most of the time mental imagery is strikingly different from our perception of the real world. For one thing imagery is far less vivid. However, this difference seems not just to be quantitative; mental images lack more than just vividness. There is a qualitative difference that Gregory (1996) proposes results from a lack of 'qualia' in mental images. In the mental image we know that something is blue. In the percept of the real world we experience it as blue. Gregory suggests that percepts of the real world have qualia precisely to enable us to distinguish them from mental images. There is another feature of mental images (which we shall discuss in more detail below) which concerns their 'deadness'. They are like still life (*nature morte*) rather than real life. This is because they are generated solely by our brain without unexpected or uncontrolled inputs from the outside world. This property of images has been demonstrated experimentally by Riesberg (1987) using the Necker cube. It is well known that if we look at a drawing of a cube in which all the 12 lines joining the vertices are visible then our percept will spontaneously reverse from time to time (see figure 7). Riesberg showed naive subjects this cube very briefly so that they did not experience a reversal. They were then asked to imagine the cube. No subject ever reported a reversal while imagining the cube. The image of the cube had only one form on the basis of prior knowledge. The lack of incoming stimuli about the cube prevented this prior knowledge from being over-ruled and the percept was frozen.

This observation has interesting implications for creativity (Fish & Scrivener 1990). Creation that takes place entirely 'in the head' is likely to be impoverished since there is no escape from the constraints imposed by prior knowledge and brain structure. To avoid these constraints most people externalize their creative processes by the use of sketches, exercises and other preliminaries.

(c) *Hallucinoses: percepts imposed by prior knowledge*

In certain neurological cases it seems that prior knowledge can interact with 'random' sensory input to create a percept. A striking example of this phenomenon is called a functional experiential hallucination. In one such case (Lam *et al.* 1997) a patient described how she heard excerpts from television programmes, including voices and background music, whenever she started her washing machine. The experience stopped as soon as the washing machine was switched off. In a visual version of this phenomenon a patient, watching a film in which the screen was filled with the image of a rough sea, saw, superimposed upon this, an image of Ginger Rogers dancing (Critchley 1951). These hallucinations are triggered (or carried) by sensory inputs and may consist of memories that are sufficiently vivid to be experienced as real (and believed to be real at least for the first few times that they occur). However, the patient does not develop the sorts of delusions about these experiences that are typically asso-

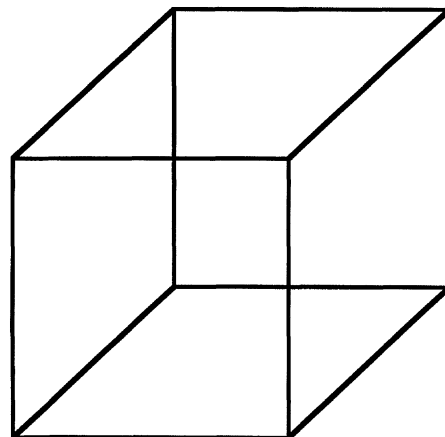


Figure 7. The Necker cube. This figure has two possible interpretations in three dimensions. Over time our perception of the figure spontaneously reverses from one interpretation to the other.

ciated with hallucinations in schizophrenia (see below). Visual hallucinations associated with brain damage are rare, and we are unaware of any experimental studies of this phenomenon. In some cases visual hallucinations are observed in the blind field of patients with damage to early visual areas of the brain. In these cases the experience is obviously not driven by sensory stimuli, but seems to be associated with random neural activity in the damaged area. Such hallucinations often cease when the patient is treated with antiepileptic drugs (Lance 1976). Other cases are associated with damage in the thalamus and the parietal cortex (Noda *et al.* 1993; Critchley 1951). Both these areas are part of the circuit revealed by brain imaging to be involved in the modulation of sensory processes by attention. There seem to be qualitative differences in the hallucinations associated with damage to different regions, but too few cases have been reported to reach any definite conclusions.

(d) *Hallucinations: misperception of action*

In the case of hallucinoses, we suggest that the sites of perception receive aberrant sensory input which is then modified by prior knowledge, giving rise to false perceptions. The hallucinations associated with psychosis seem to reflect a different mechanism. Many patients with schizophrenia hear voices in the absence of any auditory input (discussed in Frith (1996)). There are many differences in these experiences from the 'hallucinoses' we described in the previous section. The patients hear formed voices (rather than unformed sounds) that usually talk to, or about the patient. In many cases the voices may command the patient to do things against the patient's own wishes. In other words the voices seem to come from an external 'agent' who is communicating with the patient and is trying to influence his/her behaviour. In most cases the patient also has 'delusions' about the voices involving elaborate beliefs about where the voices come from and what they are trying to achieve (e.g. 'MI5 is using magnetic rays to broadcast messages. They prevent me from approaching my cousin, the

Queen'). The pattern of brain activity associated with the experience of hallucinations seems to involve left hemisphere language areas, including inferior frontal cortex (Broca's area) and superior temporal cortex (Wernicke's area) (McGuire *et al.* 1993; Silbersweig *et al.* 1995). These areas are also activated when normal people imagine someone speaking. That verbal hallucinations 'share resources' with various aspects of language has also been shown in behavioural studies (David 1994). Listening to reasonably complex sounds reduces the severity and frequency of hallucinations as does verbal articulation. Clearly these hallucinations are in some sense 'self-generated' since there is no auditory input involved. In some cases there is direct evidence of such self-generation. For example, some patients describe an experience (thought broadcasting, thought echo) in which they hear their own thoughts spoken aloud. Many patients mutter and this muttering can sometimes be amplified sufficiently to make out the content of the speech (Gould 1949). This content often corresponds precisely to what the patient reports the 'voices' to be saying.

Why do these patients experience this self-generated verbal material as if it came from the outside world? We have suggested that this misperception arises because the patient's intentions no longer produce the appropriate effects on perceptual areas of the brain (Frith 1996). One line of evidence comes from studies in which patients have to monitor their own intentions to act. Difficulty with such monitoring is suggested by failure to correct simple limb movement errors in the absence of visual feedback (Frith & Done 1989). In addition, patients have difficulty remembering their own actions when these were carried out without visual feedback (Mlakar *et al.* 1994). In both cases these results suggest that patients' knowledge about their own actions depends more on seeing what they have done rather than knowing in advance what they will do. Direct evidence for a misperception of their own actions comes from an experiment in which the sound of the patient's own voice (fed back without delay through earphones) was distorted. When patients in an acute phase of their illness experienced such feedback, they made comments such as 'someone else is speaking when I speak. There is an evil spirit speaking when I speak' (Cahill *et al.* 1996). The speech sounds they heard, although generated by themselves, were readily experienced as coming from an independent agent in the outside world.

At the physiological level we have demonstrated 'functional' disconnections in schizophrenic patients between frontal and temporal areas in the language system in the left hemisphere of the brain (Friston & Frith 1995). Such disconnections could certainly account for a failure of frontal areas concerned with speech production to modulate temporal areas concerned with speech perception. However, much work needs to be done before we have a full explanation of experiences such as thought broadcasting.

(e) Perception of the self

In terms of our framework, disconnections between sites and source would lead to a site of perception that

was no longer appropriately modified by sources of prior knowledge and expectations, derived in this case from intended actions. This lack of modification leads to false perceptions. We still understand little of how action influences perception. Nevertheless, it has long been recognized that this interaction is of fundamental importance for our understanding of the world (Russell 1996). For example, Kant (in the second analogy section of the *Critique of Pure Reason*) made the distinction between self-determined and world-determined sequences of perception. We can move our gaze from the ship anchored in the harbour to the distant horizon and we can then reverse our gaze back to the ship. However, when we watch the ship sailing over the distant horizon we cannot reverse the sequence of perception. The ability to make this distinction between changes in perception caused by our own actions and those that occur because of independent events in the outside world is fundamental to the development of a sense of ourselves as agents (Russell 1996). By making this distinction we can learn to separate those aspects of the world we can control from those we can not. When something goes wrong very early in life with the mechanism that permits this distinction our sense of self can not properly develop (as may be the case in autism). If the mechanisms goes wrong later in life we may start misperceiving independent agents in the world where there are none (as may the case in schizophrenia). We propose that a fundamental component of this mechanism involves the modification of perception by prior knowledge of action.

Given that perception is, in general, the result of an interaction between sensory signals and prior knowledge, it is useful at the physiological level to distinguish between sources of signals reflecting prior knowledge and sites where these signals modified sensory inputs. Functional brain imaging provides a means for localizing sources and sites and has the potential to reveal some aspects of the mechanisms underlying perception. However, the temporal resolution of the technique is far from perfect at between 10 and 30 seconds. The time for signals to reach early visual processing areas is of the order of 30 ms and an 80 ms presentation time is sufficient for an object to be recognized. Brain imaging studies will need to be supplemented by techniques such as EEG and MEG. Only these techniques can achieve the millisecond time scale at which perception occurs.

REFERENCES

- Büchel, C., Josephs, O., Turner, R., Frith, C. D. & Friston, K. J. 1997 The functional anatomy of attention to visual motion. (Submitted.)
- Cahill, C., Silbersweig, D. & Frith, C. D. 1996 Psychotic experiences induced in deluded patients using distorted auditory feedback. *Cogn. Neuropsych.* **1**, 201–211.
- Cavada, C. & Goldman-Rakic, P. S. 1989 Posterior parietal cortex in rhesus monkey. I. Parcellation of areas based on distinctive limbic and sensory cortico-cortical connections. *J. Comp. Neurol.* **287**, 393–421.
- Clark, V. P., Keil, K., Maisog, J. Ma., Courtney, L. G., Ungerleider, L. G. & Haxby, J. V. 1996 Functional magnetic

- resonance imaging of human visual cortex during face matching. *Neuroimage* **4**, 1–15.
- Corbetta, M., Miezin, F. M., Dobmeyer, S., Shulman, G. L. & Petersen, S. 1991 Attentional modulation of neural processing of shape, color, and velocity in humans. *Science* **248**, 1556–1559.
- Creutzfeldt, O., Ojeman, G. & Lettich, E. 1989 Neuronal activity in the human lateral temporal lobe II. Responses to the subject's own voice. *Exp. Brain Res.* **77**, 476–489.
- Critchley, M. 1951 Types of visual preservation: 'paliopsia' and 'illusory visual spread'. *Brain* **74**, 267–299.
- David, A. S. 1994 The neuropsychological origin of auditory hallucinations. In *Neuropsychology of schizophrenia* (ed. A. S. David & J. Cutting), pp. 26–313. Hove: Lawrence Erlbaum Associates.
- De Renzi, E. 1986 Current issues in prosopagnosia. In *Aspects of face processing* (ed. H. D. Ellis, M. A. Jeevs, K. F. Newcombe & A. Young), pp. 243–252. Dordrecht: Neijhoff.
- Dolan, R. J., Fink, G. R., Rolls, E., Booth, M., Frackowiak, R. S. J. & Friston, K. J. 1997 How the brain learns to see objects and faces in an impoverished context. (Submitted.)
- Fink, G. R., Halligan, P. W., Marshall, J. C., Frith, C. D., Frackowiak, R. S. J. & Dolan, R. J. 1996 Where in the brain does visual attention select the forest and the trees? *Nature* **382**, 626–628.
- Fish, J. & Scrivener, S. 1990 Amplifying the mind's eye: sketching and visual cognition. *Leonardo* **23**, 117–126.
- Fletcher, P. C., Frith, C. D., Baker, S. C., Shallice, T., Frackowiak, R. S. J. & Dolan, R. J. 1995 The mind's eye—precuneus activation in memory-related imagery. *Neuroimage* **2**, 195–200.
- Friston, K. J. 1994 Functional and effective connectivity in neuroimaging: a synthesis. *Hum. Brain Mapping* **2**, 56–78.
- Friston, K. J. & Frith, C. D. 1995 Schizophrenia: a disconnection syndrome? *Clin. Neurosci.* **3**, 89–97.
- Frith, C. D. 1996 The role of the prefrontal cortex in self-consciousness: the case of auditory hallucinations. *Phil. Trans. R. Soc. Lond. B* **351**, 1505–1512.
- Frith, C. D. & Done, D. J. 1989 Experiences of alien control in schizophrenia reflect a disorder in the central monitoring of action. *Psychol. Med.* **19**, 359–363.
- Frith, C. D. & Friston, K. J. 1996 The role of the thalamus in 'top down' modulation of attention to sound. *Neuroimage* **4**, 210–215.
- Frith, C. D. & Friston, K. J. 1997 Studying brain function with neuroimaging. In *Cognitive Neuroscience* (ed. M. D. Rugg), pp. 169–196. London: UCL Press.
- Gould, L. N. 1949 Auditory hallucinations and subvocal speech. *J. Nerv. Ment. Dis.* **109**, 418–427.
- Gregory, R. L. 1996 What do qualia do? *Perception* **25**, 377–379.
- Haxby, J. V., Grady, C. L., Horwitz, B. *et al.* 1991 Dissociation of object and spatial visual processing pathways in human extrastriate cortex. *Proc. Natn. Acad. Sci. USA* **88**, 1621–1625.
- Haxby, J. V., Horwitz, B., Ungerleider, L. G., Maisog, J. Ma., Pietrini, P. & Grady, C. L. 1994 The functional organisation of human extrastriate cortex: a PET-rCBF study of selective attention to faces and locations. *J. Neurosci.* **14**, 6336–6353.
- Helmholtz, H. von 1866 *Handbuch der physiologischen optik*. Leipzig: Voss.
- Jeanerod, M. 1994 The representing brain: neural correlates of motor intention and imagery. *Behav. Brain Sci.* **17**, 187–245.
- Kosslyn, S. M. 1996 *Image and brain: the resolution of the imagery debate*. Cambridge, MA: MIT Press.
- Lam, L. C. W., Leung, S. F. & Chow, L. Y. 1997 Functional experiential hallucinosis after radiotherapy for nasopharyngeal carcinoma. (Submitted.)
- Lance, J. N. 1976 Simple formed hallucinations confined in the area of a specific visual field defect. *Brain* **99**, 719–734.
- Martin, A., Haxby, J. V., Lalonde, F. M., Wiggs, C. L. & Ungerleider, L. G. 1995 Discrete cortical regions associated with knowledge of color and knowledge of action. *Science* **270**, 102–105.
- McGuire, P. K., Shah, P. & Murray, R. M. 1993 Increased blood flow in Broca's area during auditory hallucinations in schizophrenia. *Lancet* **342**, 703–706.
- McGuire, P. K., Silbersweig, D. A., Murray, R. M., David, A. S., Frackowiak, R. S. J. & Frith, C. D. 1996 Functional anatomy of inner speech and auditory verbal imagery. *Psychol. Med.* **26**, 29–38.
- McIntosh, A. R. & Gonzalez-Lima, F. 1994 Structural equation modelling and its application to network analysis in functional brain imaging. *Hum. Brain Mapping* **2**, 2–22.
- Mesulam, M.-M. 1990 Large-scale neurocognitive networks and distributed processing for attention, language and memory. *Arch. Neurol.* **10**, 309–325.
- Mlaker, J., Jensterle, J. & Frith, C. D. 1994 Central monitoring deficiency and schizophrenic symptoms. *Psychol. Med.* **24**, 557–564.
- Noda, S., Mizoguchi, M. & Yamamoto, A. 1993 Thalamic experiential hallucinosis. *J. Neurol. Neurosurg. Psychiat.* **56**, 1224–1226.
- Paus, T., Marrett, S., Worsley, K. & Evans, A. C. 1995 Imaging motor to sensory discharges in the human brain. *Neuroimage* **4**, 78–86.
- Petrides, M. & Pandya, D. N. 1984 Projections to the frontal cortex from the posterior parietal region in the rhesus monkey. *J. Comp. Neurol.* **228**, 105–116.
- Price, C., Wise, R. J. S., Ramsay, S. *et al.* 1992 Regional response differences within the human auditory cortex when listening to words. *Neurosci. Lett.* **146**, 179–182.
- Ramachandran, V. S. 1994 In *The artful eye* (ed. R. L. Gregory & J. Harris). Oxford University Press.
- Rees, G., Frackowiak, R. S. J. & Frith, C. D. 1997 Two modulatory effects of attention mediating object categorisation in human cortex. *Science* **275**, 835–838.
- Riesberg, D. 1987 External representations and the advantages of externalising one's own thoughts. *Proc. of the IXth Ann. Conf. of the Cognitive Science Society*. Hillsdale, NJ: Lawrence Erlbaum.
- Robertson, L. C. & Lamb, M. R. 1991 Neuropsychological contributions to theories of part/whole organization. *Cogn. Psychol.* **23**, 299–330.
- Rushton, D. N., Rothwell, J. C. & Craggs, M. D. 1981 Gating of somatosensory evoked potentials during different kinds of movement in man. *Brain* **104**, 465–491.
- Russell, J. 1996 *Agency*. Sussex: Erlbaum (UK) Taylor & Francis.
- Sergent, J., Ohta, S. & Macdonald, B. 1992 Functional neuroanatomy of face and object processing. *Brain* **115**, 15–36.
- Silbersweig, D. A., Stern, E., Frith, C. D. *et al.* 1995 A functional neuroanatomy of hallucinations in schizophrenia. *Nature* **378**, 176–179.
- Stephan, K. M., Fink, G. R., Passingham, R. E. *et al.* 1995 Functional anatomy of the mental representation of upper extremity movements in healthy subjects. *J. Neurophysiol.* **73**, 373–386.
- Tanaka, K., Saito, H., Fukada, Y. & Moriya, M. 1991 Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *J. Neurosci.* **6**, 134–144.
- Varney, N. R. & Digre, K. 1983 Colour 'amnesia' without aphasia. *Cortex* **19**, 545–550.
- Wenzel, R., Bartenstein, P., Dieterich, M. *et al.* 1996 Deactivations of human visual cortex during involuntary ocular oscillations. *Brain* **119**, 101–110.
- Wolpert, D. M., Ghahramani, Z. & Jordan, M. I. 1995 An internal model for sensorimotor integration. *Science* **269**, 1880–1882.
- Zeki, S. 1983 Colour coding in the cerebral cortex: the reaction of cells in monkey visual cortex to wavelengths and colours. *Neuroscience* **9**, 741–756.
- Zeki, S., Watson, J. D. G., Lueck, C. J., Friston, K. J., Kennard, C. & Frackowiak, R. S. J. 1991 A direct demonstration of functional specialisation in human visual cortex. *J. Neurosci.* **11**, 641–649.