FRAMEWORKS OF ANALYSIS FOR THE NEURAL REPRESENTATION OF ANIMATE OBJECTS AND ACTIONS

By D. I. PERRETT, M. H. HARRIES, R. BEVAN, S. THOMAS, P. J. BENSON, A. J. MISTLIN, A. J. CHITTY, J. K. HIETANEN AND J. E. ORTEGA

Department of Psychology, University of St Andrews, St Andrews, Fife KY16 9JU, Scotland

Summary

A variety of cell types exist in the temporal cortex providing high-level visual descriptions of bodies and their movements. We have investigated the sensitivity of such cells to different viewing conditions to determine the frame(s) of reference utilized in processing. The responses of the majority of cells in the upper bank of the superior temporal sulcus (areas TPO and PGa) found to be sensitive to static and dynamic information about the body were selective for one perspective view (e.g. right profile, reaching right or walking left). These cells can be considered to provide viewer-centred descriptions because they depend on the observer's vantage point. Viewer-centred descriptions could be used in guiding behaviour. They could also be used as an intermediate step for establishing view-independent responses of other cell types which responded to many or all perspective views selectively of the same object (e.g. head) or movement. These cells have the properties of object-centred descriptions, where the object viewed provides the frame of reference for describing the disposition of object parts and movements (e.g. head on top of shoulders, reaching across the body, walking forward 'following the nose'). For some cells in the lower bank of the superior temporal sulcus (area TEa) the responses to body movements were related to the object or goal of the movements (e.g. reaching for or walking towards a specific place). This goal-centred sensitivity to interaction allowed the cells to be selectively activated in situations where human subjects would attribute causal and intentional relationships.

Introduction

This paper describes the response properties of cells in different regions of the temporal association cortex of the macaque monkey. It is the aim of the paper to summarize the sensitivity of cells in this brain area to different types of biologically important visual stimuli. A parallel aim is to consider frameworks for visual processing which are appropriate for making explicit different types of information about animate objects and hence for achieving a more complete understanding of the world.

Key words: temporal cortex, single cell, face, action.

The first section of the paper focuses on cells in one region of the higher association cortex (the upper bank of the superior temporal sulcus, areas TPO and PGa of Seltzer & Pandya, 1978; Pandya & Yeterian, 1985) which appear to be involved in the recognition of individuals (Fig. 1, left column) and how these individuals are moving (Perrett *et al*. 1985*a*,*b*; Baylis *et al*. 1985). This region has received extensive study since it was realised that it contained cells selectively responsive to faces (Bruce *et al*. 1981; Perrett *et al*. 1982, 1984, 1987; Rolls, 1984; Desimone *et al*. 1984; Mikami & Nakamura, 1988). The second section focuses on coding in an adjacent section of cortex in the lower bank of the same sulcus (area TEa of Seltzer & Pandya, 1978). Populations of cells in this region appear to be involved in the recognition of actions, that is, how other individuals are interacting with the environment (Fig. 1, middle column). Studies of action coding have been mainly restricted to actions of the hand but there are indications that the framework for such processing applies to actions of the whole body (Perrett *et al*. 1989*a*,*b*,*e*).

Physiological methods

Standard single-unit recording techniques were employed to study cells in different regions of the temporal cortex of awake, behaving rhesus macaque monkeys (for details see Perrett *et al*. 1985*a*,*b*). A large-aperture shutter was used to present different types of visual stimuli. These included real faces and bodies, two-dimensional slides of monkeys and humans in different postures, videotapes of different actions and a variety of simple and complex three-dimensional stimuli. Responses to these stimuli were measured by analysing the number of action potentials from individual cells in a 0·25 s period beginning 100 ms after the shutter opened. This period of analysis was chosen because it is relatively uncontaminated with eye movements and visual responses in the temporal cortex generally have latencies greater than 100 ms.

Viewer- and object-centred descriptions
General characteristics of face

Within the cortex of the superior temporal sulcus (STS) populations of cells have been studied that respond more to the sight of faces than to a variety of simple stimuli (e.g. bars or gratings) or complex, potentially arousing stimuli (e.g. hands, bananas, pictures of snakes and birds of prey) (Bruce *et al*. 1981; Perrett *et al*. 1982). Most such cells are sensitive to the general characteristics of faces and respond to a variety of faces regardless of their species (human or monkey). These cells also show a remarkable tolerance for changes in viewing conditions and respond to faces despite change in retinal size, orientation and position (Perrett *et al*. 1982, 1984, 1988, 1989*d*; Bruce *et al*. 1981; Rolls & Baylis, 1986). This generalization indicates that the cells' discriminative responses to faces are not dependent directly on simple visual attributes (e.g. position and orientation of local edges, spatial frequency components) which change from display to display.
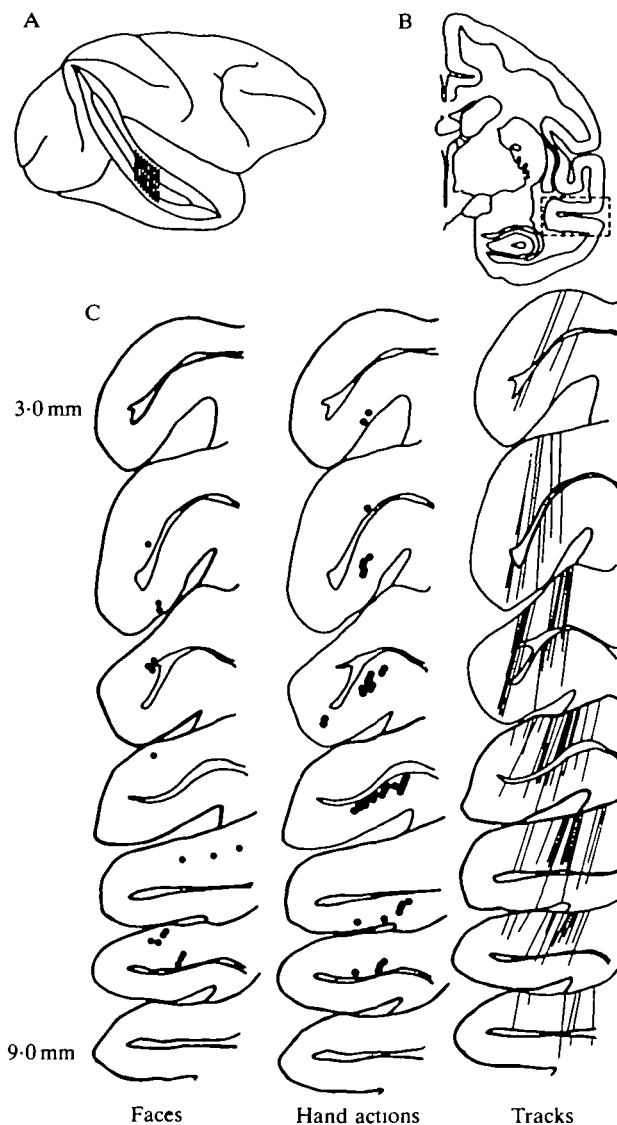
Fig. 1. Anatomical distribution of different cell types within the superior temporal sulcus (STS). (A) Side view of the right hemisphere of the macaque brain with the STS opened out to reveal the recording site. (B) Coronal section of the right hemisphere 9 mm anterior to the interaural plane. The box around the STS indicates the recording area. (C) Reconstruction of the right STS of one monkey displayed in serial sections from 3 to 9 mm anterior to the interaural plane area. The right column gives the trajectories of all recording tracks made in this monkey. The left column marks the position of all cells found to respond selectively to the sight of the face or other views of the head. The centre column marks the position of cells found to be selective for the sight of hand–object interactions. The distribution of cell types between the upper and lower banks was significantly different ($\chi^2 = 68 \cdot 6$; df $= 1$ $P < 0 \cdot 01$). Cells responsive to the head occurred mainly in the upper bank within or adjacent to area TPO and cells responsive to hand–object interactions occurred mainly in the lower bank within area TEa.

Studies presenting parts of the face in isolation, or covering specific regions of the face, reveal that many cells are sensitive to the presence of a single facial feature, e.g. the eye region, and to no other information about the face (Perrett *et al*. 1982). Such studies, however, reveal that other cells can respond independently to two or three regions of the face presented alone (e.g. the hair or mouth or the eye regions) (Perrett *et al*. 1982, 1984, 1989*d*; Bruce *et al*. 1981; Desimone *et al*. 1984). Comparisons of responses to normal and jumbled arrays of facial features provide direct evidence that some cells are also sensitive to configuration (Perrett *et al*. 1982; see also Yamane *et al*. 1989).

Perspective view

Although cells generalize to many different examples of the face, despite changes in illumination, size and orientation (upright/inverted), most have only a limited capacity to generalize across changes in perspective view. Turning the face to profile or rotating it up and down reduces most cells' responses to the face. Studies using different views of the head and body, however, have revealed other cells which were maximally responsive to the profile and to other views of the head and body (Perrett *et al*. 1985*a*; Desimone *et al*. 1984).

Cells tuned to other views of the head were found to have properties equivalent to cells responsive to the face. For example, cells maximally responsive to the profile face generalize their response to many different examples of profile faces, with changing retinal position, orientation and size. Fig. 2 illustrates a receptive field study of one cell selectively responsive to the left profile and demonstrates the invariance of response over different retinal positions. For this study the position of stimuli (presented for 100 ms) is plotted relative to the fovea. For the left profile, presentation anywhere within the cell's large receptive field, which extended some 25° into both visual fields, produced an excitatory response in excess of the spontaneous activity of the cell. Presentation outside this field produced responses no different from spontaneous. The right profile and control stimuli (slides of hands and geometrical shapes) failed to produce responses greater than spontaneous whether presented inside or outside the receptive field.

We have found many cells such as that illustrated in Fig. 2 which are responsive to one profile view of the head but not to the other profile view (Perrett & Mistlin, 1989). This finding argues that arousal is unimportant in explaining the response. One suggestion frequently made is that the apparently visual responses of cells to the sight of faces are due to emotional responses elicited by the sight of the face that are not provoked by other stimuli. One can assume that the left and right profiles are likely to evoke an identical state of arousal/emotional response, hence the observation that cells discriminate between left and right profile argues strongly that the cell's responses are based on a visual analysis of the image rather than occurring as a consequence of some change in emotional state.

We have been interested in how many perspective views of the head need to be coded for the head to be recognized from any view. We have now studied substantial numbers of cells responsive to the sight of the head and selective for
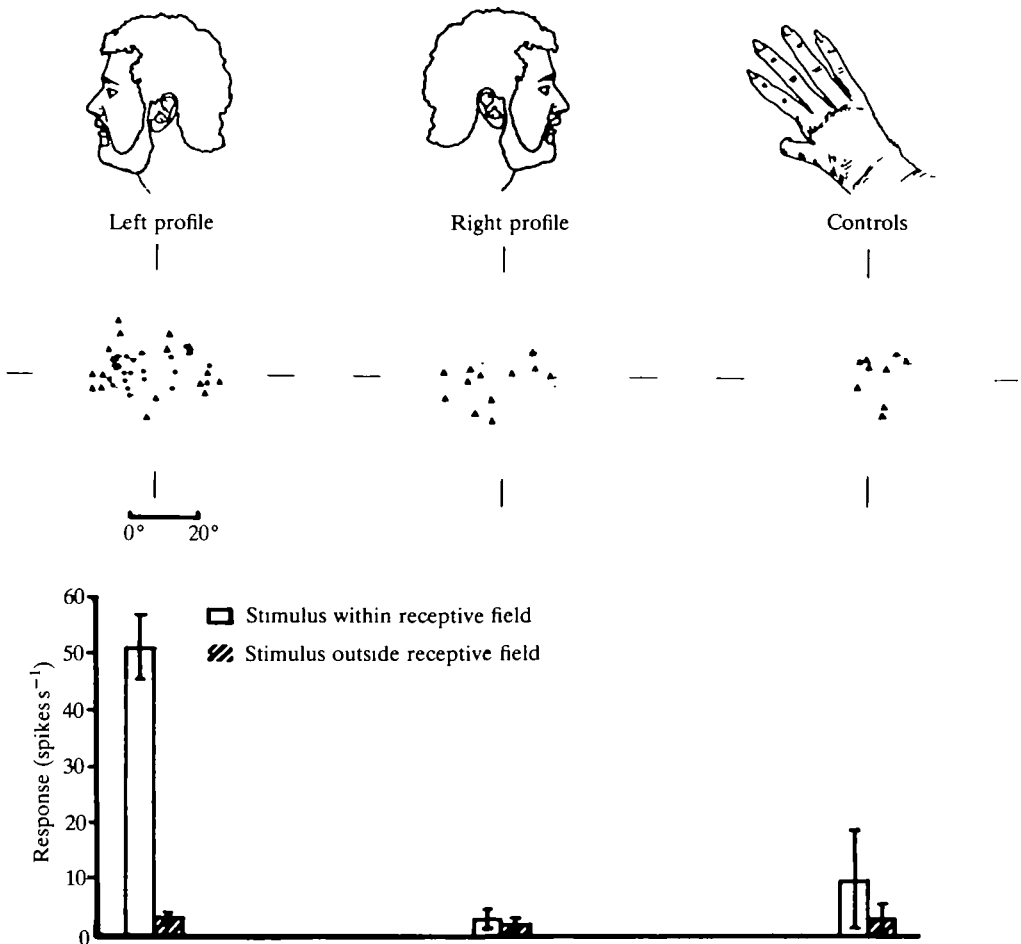
Fig. 2. Receptive fields of one cell selective for the left profile face. Upper, schematic line drawing of stimuli used. Stimuli are not drawn to scale, the size at presentation was 5–10°. Middle, the position of left profile, right profile and control stimuli presented for 100 ms relative to the fovea (intersection of cross wires). The receptive field is defined as that region of the visual field in which responses (filled circles) to the left profile face were greater than the mean spontaneous activity +1·96 standard deviations. Responses to the left profile less than this amount are shown as filled triangles. The histogram illustrates the response (mean ± 1 s.e.) inside (cross-hatched bars) and outside (open bars) the receptive field. (Adapted from Perrett & Mistlin, 1989.)

particular perspective views. The results (R. Bevan M. H. Harries, D. I. Perrett, S. Thomas, P. J. Benson, J. Hietanen & J. Ortega, in preparation) indicate that although the cells are tuned to a whole range of views from the front to the back of the head, the distribution of views preferred by cells shows a clustering around four prototypical views (face, left and right profile and back views). This unequal distribution is in line with the speculation that only a small number of high-level

descriptions of particular views of an object or person are held in memory (Perrett
et al. 1984, 1985*b*).

Cells selective for perspective view show a gradual decline in response as the
head is turned away from the optimal view for the cell. One way of estimating the
breadth of tuning is to compute the angle away from the optimal view that is
required to reduce response to a value mid-way between the best and worst view
responses [(maximum response + minimum response)/2]. In a sample of 34 cells
this value averaged 67° (range 40–110°). Cells coding the face will therefore be
activated to about half-maximal rate by views of the head rotated 60° towards
profile.

A view between prototypical views of the face and profile (e.g. the head turned
45° to profile) may activate a few cells tuned specifically for this view, but it will
also activate many cells tuned either to the face or to the profile views. The level of
activation of cells responsive to the face and profile by this intermediate view will
be reduced but this may be offset by the fact that two populations are activated.
Thus, coding of a small number of prototypical views can cover a wide range of
perspective views, much as the retinal cones cover the complete spectrum of
colours with only three broadly tuned colour pigments.

A simplified account of visual processing

Fig. 3 gives a very simplified scheme of visual processing leading to the visual
recognition of familiar individuals. Early visual processing at the level of the visual
cortex provides information about elementary components of the image such as
constituent edges, their orientation, size and positions. This stage of processing
has been referred to as the primal sketch in David Marr's computational model of
vision (Marr & Nishihara, 1978; Marr, 1982).

The details of the next stages of processing after the visual cortex but before the
structural encoding that has been studied in the temporal cortex are to a large
extent unknown, though they may involve the operation of Gestalt grouping
principles to label particular parts of the image as belonging together. This would
enable the structural encoding at later stages in the temporal cortex to 'identify'
grouped image regions as features or parts of an object. In the case of faces,
separate areas of the image, such as the vertically oriented fine texture, might be
separated by grouping processes and subsequently activate cells selective for the
hair in the temporal cortex. Similarly, a horizontal pair of blobs with concentric
circular structure might be grouped and subsequently activate cells functionally
tuned to the eyes.

Fig. 3 illustrates a parallel analysis for a different perspective view of the head.
In reality there would be at least four views analysed in the horizontal plane and
when changes in elevation (head-up/head-down) are also included there may be a
further 4–8 views coded, since some cells are selective for the face or the profile in
head-down or head-up views (Perrett *et al.* 1985*a*; R. Bevan M. H. Harries, D. I.
Perrett, S. Thomas, P. J. Benson, J. Hietanen & J. Ortega, in preparation).
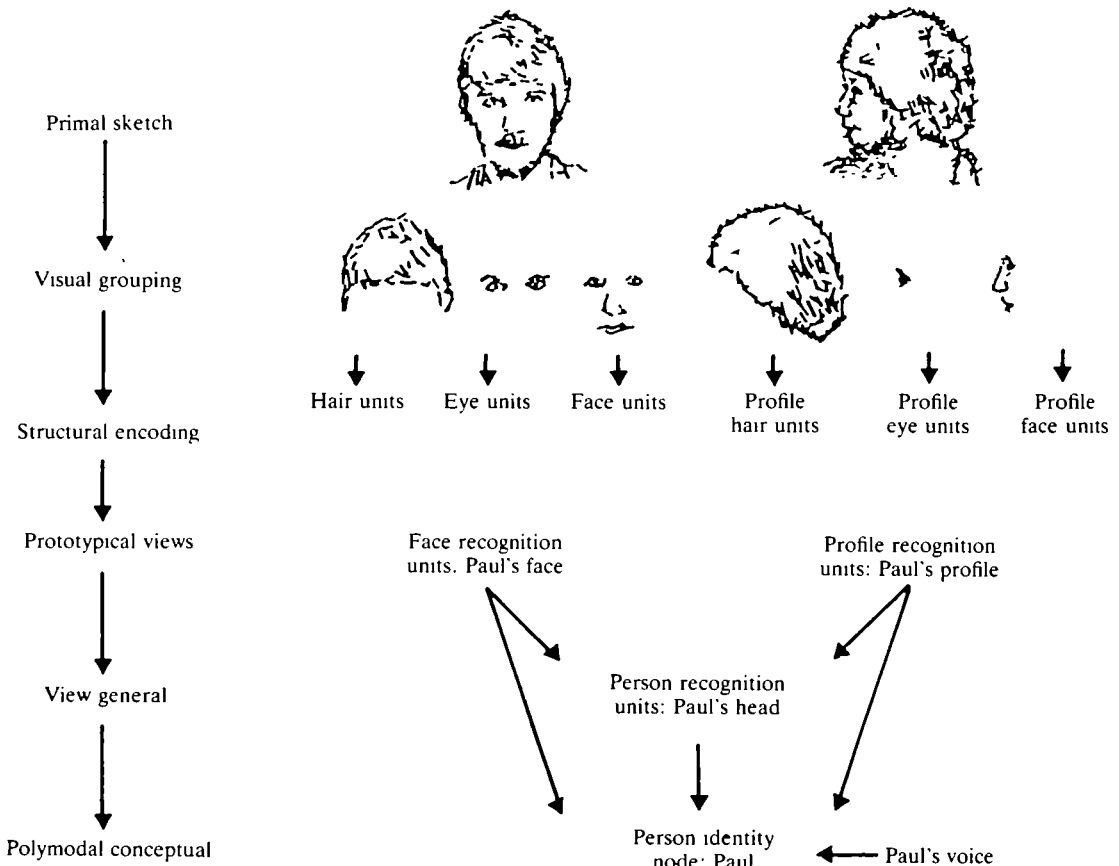Information about different facial features for each view appears to be integrated

Primal sketch

Visual grouping

Structural encoding

Prototypical views

View general

Polymodal conceptual

Hair units   Eye units   Face units   Profile hair units   Profile eye units   Profile face units

Face recognition units. Paul's face

Profile recognition units: Paul's profile

Person recognition units: Paul's head

Person identity node: Paul ◄——— Paul's voice

Fig. 3. Schematic overview of processing leading to recognition of a familiar face. (From Perrett *et al.* 1987.)

since cells are often tuned to the characteristics of several regions of the face, e.g. mouth shape and configuration of the eyes (Perrett *et al.* 1985a, 1989d).

Identity of a given perspective view

Amongst the cells sensitive to different perspective views, some have been found which are selective for identity (Perrett *et al.* 1984, 1987; Baylis *et al.* 1985). Fig. 4 illustrates the responses of one such cell. This cell responded above its spontaneous activity to the sight of the right side of one experimenter (MH). Front, back and left side views of the same individual failed to produce responses larger than spontaneous activity. No views of another experimenter (PB) produced significant responses.

Such cells represent a high-level 'viewer-centred' description of familiar individuals, in that their responses occur only to one perspective view of the familiar individual. (Occasionally cells are found that respond to two views 180°
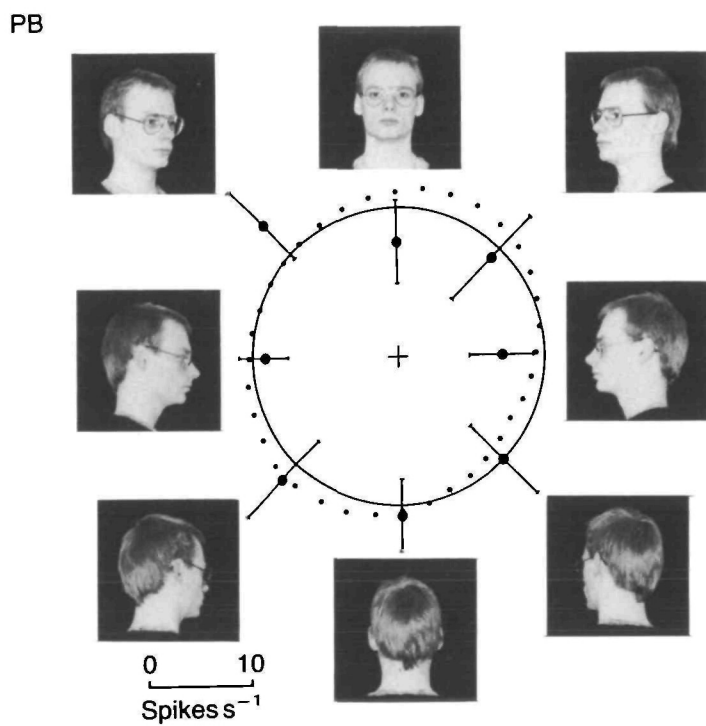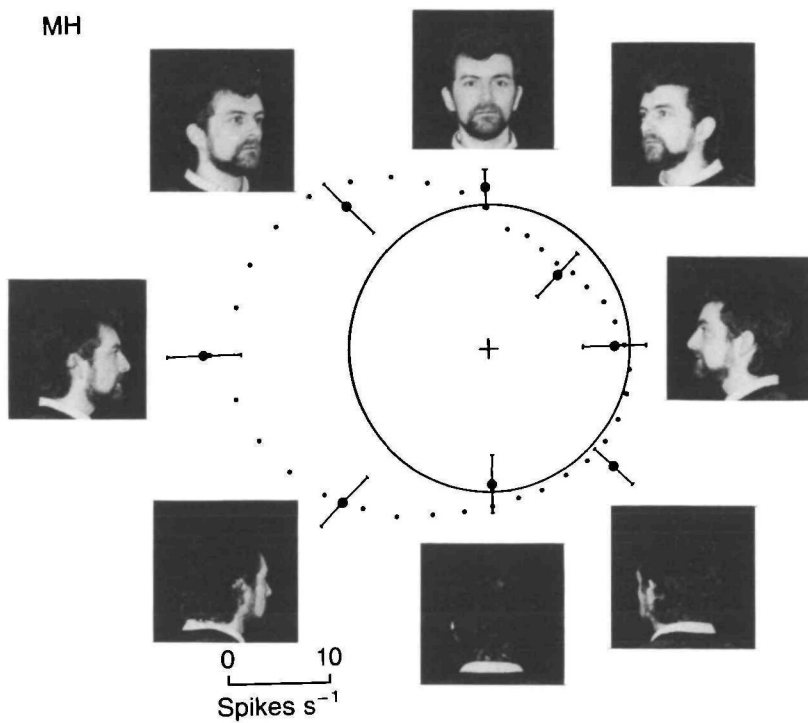
D. I. PERRETT AND OTHERS

Fig. 4

Fig. 4. Tuning for perspective view of cell sensitive to face identity. Upper: the mean and standard error of responses from five trials are illustrated for eight views of one experimenter (MH). Response magnitude to each view is depicted as the distance of the large dot from the central cross. The dotted line is the best fit second-order cardioid function. The cell responded more to the right profile view of Mark Harries than the cell's spontaneous activity (central circle). Lower: the cell did not respond to any of the eight views of a second equally familiar individual (PB).

apart, especially the left and right profiles.) In Fig. 3 this is illustrated in separate cells coding Paul's face and Paul's profile. Responses of such cells must be considered high-level because the cells generalize across different instances of one perspective view (Perrett *et al.* 1984, 1987).

For each familiar individual it would appear likely that there are many cells differentially responsive to that individual and additionally selective for view, some more responsive to the frontal (face) view, some more responsive to the right profile, some to the back view, etc. Cells that we and others have studied vary in their selectivity for individual faces. Some cells are highly selective, being responsive to one of many different faces tested, others are less selective in that they may respond to several but not all individuals tested and most appear unselective for identity (Baylis *et al.* 1985; Yamane *et al.* 1988; Mikami & Nakamura, 1988; Perrett *et al.* 1984).

Identity across different views

There are other cells in the temporal cortex which are responsive to many or all perspective views of the head (Perrett *et al.* 1985a). Analysis of the visual basis of the response of these cells reveals that many are not simply responsive to the presence of one feature, such as hair, which is common to all perspective views of the head. Cells sensitive to identity have been found amongst this population of cells. These cells may respond differentially to different people or monkeys but, unlike the cells described so far, they continue to differentiate between individuals for many different perspective views. This stage of analysis is denoted in Fig. 3 by units responsive to Paul's head.

One example is given in Fig. 5. The cell illustrated here responds equally to all views of one experimenter but no views of a second familiar experimenter. One might suggest that such a differential response could be due to a single feature present in one face and/or absent in the second (such as straight or spiky hair). Large responses from this cell were, however, absent to two-dimensional images and were only detected when both upper torso and head were visible. Thus, it is unlikely that any single visual feature was responsible for the difference. Rather, the results indicate that the cell was receiving multiple sources of visual information about the head and body. The cell can therefore provide information relevant to the discrimination between these two individuals regardless of perspective view.

Deriving object-centred descriptions from viewer-centred codes

Marr (1982) noted the importance of establishing descriptions in which the parts of an object were related not to the viewer but to the main axis of the object itself. These descriptions he termed 'object-centred' and argued that one important property was that they were the same for all vantage points of the viewer. Marr argued that such descriptions were important because only one would need to be stored in memory for recognition from all vantage points.

Cells selective for multiple views of one face can be understood as providing object-centred descriptions of this face. We have suggested that object-centred descriptions can be established by pooling the output of different viewer-centred descriptions. For example, one could recognize an object as being Paul by pooling the outputs from descriptions of Paul's face, Paul's profile and other views of Paul.

The distinction between view-general (object-centred) and view-specific (viewer-centred) is not categorical but is more graded. In a survey of 60 cells the percentage difference in response from best to worst view $[(R_{max} - R_{min})/R_{max} \times 100]$ ranged from 43 % to 100 %. Cells with 100 % change are clearly view-specific, but categorization of cells with less dramatic changes is less clear,
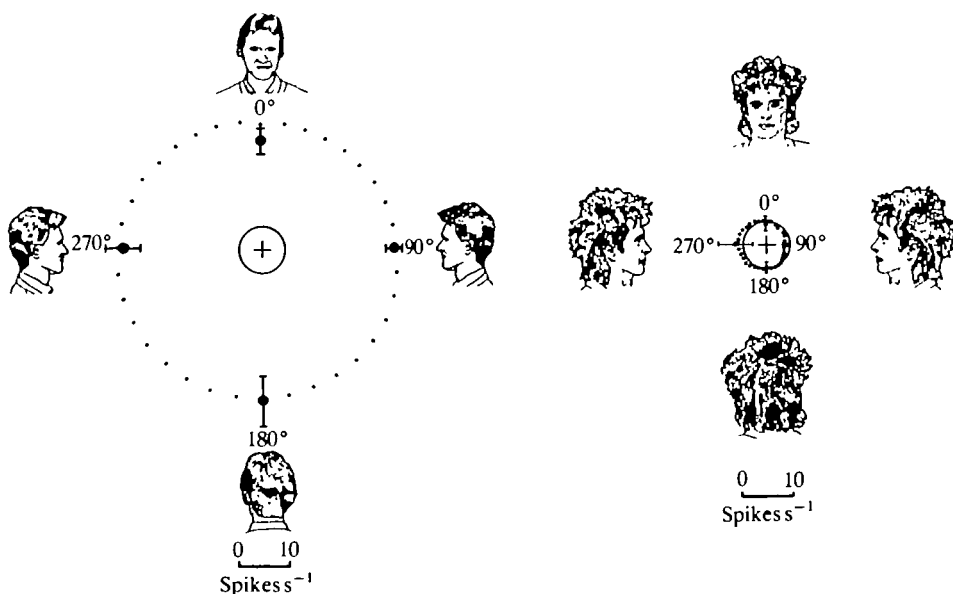

Fig. 5. Generalization of sensitivity to identity across perspective view. Mean and standard error of responses to different views of two experimenters (JH, left and DP, right). Response magnitude to face, left and right profile and back views is depicted as the distance of the large dot from the central cross. The dotted line is the best fit first-order cardioid function. For this cell all views of JH produced responses significantly greater ($P < 0.0001$ Newman–Keuls test, each comparison) than all views of a second experimenter DP and spontaneous activity (central circle). No response to DP was larger than spontaneous activity. (Number of separate trials, $N = 8, 7, 5, 6, 7, 4, 4, 6, 8$; $F = 17.36$; df = 8, 46; $P < 0.0001$.)

particularly when the response to the worst view was still above spontaneous activity. Three cells (in a sample of 60) responded to all views of the head more than to control stimuli (an object-centred property) and yet still showed significant differences between views (a viewer-centred property).

The situation may be similar to the proposed hierarchical formation of hypercomplex (length-sensitive) cells from simple (length-insensitive) cells in the visual cortex (Hubel & Wiesel, 1968). The simple hierarchical model has to be qualified, since the property which confers length-sensitivity (inhibitory end-stopping) is found to vary in a graded and continuous manner (Rose, 1974) with cells at one end of the spectrum behaving more like simple cells, with weak end-stopping, and cells at the other end of the spectrum behaving more like hypercomplex cells with strong end-stopping. With this in mind, it is difficult to maintain a hierarchical model progressing neatly from exclusively viewer-centred to completely object-centred coding. Still, it is consistent with the results to maintain that cells displaying increased tolerance for perspective view may be formed hierarchically through the combination of outputs of cells with limited view tolerance. [For such a statement to be true it would be necessary to exclude from consideration cells responsive to all views because they were sensitive to a single feature common to all views (such as hair).]

Single-cell coding

As has been argued in detail elsewhere, cells with selective responses to individual faces can be interpreted as part of an extensive population code yet at the same time the cells have properties close to the hypothetical and much-derided concept of 'grandmother cells' (Barlow, 1972, 1985; Perrett *et al.* 1987). It is sufficient to note here that the fidelity of coding of individual cells can be very high, much higher than was predicted from early population coding models in which cells could contribute to the coding of many visually different objects and could be very unreliable in their signalling of any one object.

It would be wrong to assume that the activity of only one cell is sufficient to code the presence of one familiar face. It would also be a mistake to assume that grandmother-cell coding models require that only a single cell be selective. Konorski, an original proponent of single-cell coding, suggested that the number of cells tuned for each known individual might be proportional to the familiarity and importance of the individual (Konorski, 1967). If a given single cell was as accurate as the whole observer at discriminating one individual face from others, then it would only be necessary for behaviour to rely on the output of that one cell. Nonetheless, it still could be advantageous for the brain to use a (highly) redundant code to ensure accurate recognition, with many cells tuned to the same individual and possessing the same discriminative capacity. We find several cells with a high degree of selectivity for a given familiar face, which probably means that there are many thousands, since we sampled so few.

It is an assumption of many population coding models that coding can only be understood by reference to an entire ensemble or population of neurones. One

may be led from this assumption to 'throw the baby out with the bathwater'. The fact that many cells may be involved in the coding does not mean that it is impossible to tell from individual cells anything about the coding. Otherwise there would be little insight into higher brain functions from single-cell neurophysiology, much as Uttal (1978) predicted. If a population code is made up from a systematic operation on single-cell activity then it must be the case that the code will be readable from single-cell elements. Of course, one might have to record from several units to decode the message. Indeed, in several situations studied so far it has been possible to predict population codes from single-cell data (Georgopoulos *et al.* 1989; Anderson *et al.* 1985). In almost every brain area studied explicit relationships have been found between firing frequency and external events or internal states. There may be other codes superimposed which modulate firing frequency (e.g. Gray *et al.* 1989), but it is inconceivable that the single-cell frequency code is epiphenomenal.

It is, of course, not easy to determine exactly what a cell is coding; if a cell responds to several faces but not to several others is it part of a population code for determining identity or is it really signalling some combination of features that the faces it 'prefers' have in common (e.g. dark eyes or long noses)? The answer could be both. The cell may well be contributing to population codes for identifying individuals, but doing exactly this on the basis of high-level (grandmother or explicit) coding of a given facial feature.

Levels of representation
Viewer- and object-centred representations

The importance of object-centred descriptions can be realised in the context of learning. If one learns a new piece of information about a person when only one view is visible, then forming a single association between this information and an object-centred description of that person would allow the information to be retrieved subsequently when the person was seen from any other perspective view. If all descriptions of people and objects were viewer-centred, then one would have to relearn associations over and over again for each different perspective view.

Alternatively, they may be some information one wants to associate only with particular views. For instance, one might learn to act differentially while the front view of a dominant individual is present but act otherwise, e.g. defiantly, while their back is turned. Thus, in some circumstances, viewer-centred descriptions may be useful for guiding behaviour and storing associations.

Fig. 6 gives a schematic overview of some possible types of representation in the visual system. The initial processing of the visual image is conducted using the viewer as a frame of reference. This is inevitable because the starting point of analyses is the retinal image which is entirely dependent on where the viewer is looking. Within the temporal cortex high-level viewer-centred descriptions or representations are established. These hold only for particular perspective views (or vantage points) but have the power to generalize across many instances of that

LEVELS OF REPRESENTATION

IMAGE

PRIMAL SKETCH

2+½-D SKETCH

VIEWER-CENTRED

OBJECT-CENTRED

GOAL-CENTRED

Static: profile body
Dynamic: walk right

Static: body
Dynamic: bring arm to chest
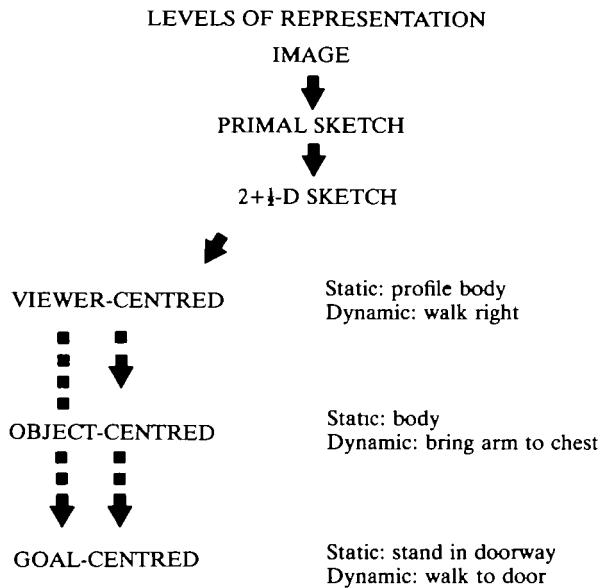
Static: stand in doorway
Dynamic: walk to door

Fig. 6. Levels of representation of information during visual processing. The initial stages of processing follow Marr's scheme whereby the image is first broken down into component local orientations (the primal sketch) then synthesized to make explicit surfaces, their orientation and boundaries (the 2+½-D sketch). High-level descriptions of objects are then built for static and dynamic information from particular vantage points (viewer-centred descriptions). These descriptions hold for changes in retinal position, orientation and size but do not generalize over changes in perspective. Viewer-centred descriptions are combined to form object-centred descriptions which are independent of vantage point. Goal-centred descriptions specify the disposition or movements of an animate object (agent) relative to part of the environment (the goal or object of the action). (From Perrett *et al.* 1989*a*.)

view, where lower-order visual variables such as lighting, position, size, colour and orientation change.

Viewer-centred representations may be used as an intermediary level in the process of establishing object-centred descriptions. The flow of information from viewer- to object-centred representations is seen as hierarchical, but it is also possible, as Marr & Nishihara (1978) suggested, for object-centred representations to be established directly from descriptions of the surface boundaries present in the 2+½-D sketch (Fig. 6). Thus high-level viewer and object-centred representations could be computed to some extent in parallel.

We conceive that the hierarchical sequence is followed in the processing of both static and dynamic information about the body's appearance (e.g. Fig. 7). For dynamic information, object-centred descriptions of locomotion, such as walk forward, where the direction of movement is related to the torso or body itself (following one's nose), could be formed by combining the separate viewer-centred descriptions (e.g. walk towards viewer facing viewer, walk to viewer's right facing
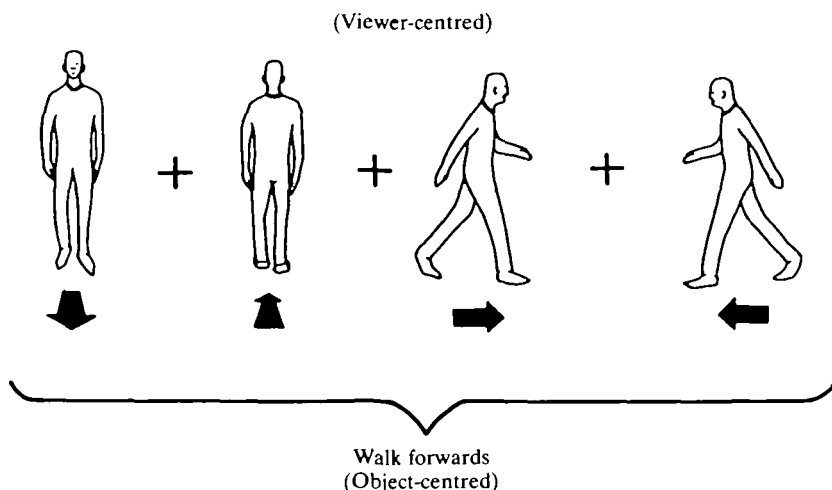
(Viewer-centred)

Walk forwards
(Object-centred)

Fig. 7. Construction of object-centred description of walking forwards. The outputs of four different types of cell with vantage-point-dependent (viewer-centred) responses that are selective for body walking towards viewer facing viewer, walking away facing away, walking right facing right and walking left facing left are pooled to establish the sensitivity of a cell which responds to the body walking forward following its nose). The latter type of cell has the properties of an object-centred description and holds from any vantage point or direction of walking. (From Perrett et al. 1989a.)

viewer's right, walk away from viewer facing away from viewer, and walk left facing left, Perrett et al. 1985a,b, 1989a,b). Thus, descriptions that generalize across vantage point for both static and dynamic information (such as Paul, or body walking forward) can be formed by combining descriptions that are specific to particular vantage points (Paul's face, or body walking right).

We have noted before that high-level viewer-centred descriptions are important in their own right; their sole function need not be seen as an intermediary step in establishing the all-important object-centred descriptions (Perrett et al. 1985a,b). A considerable amount of an organism's behaviour must be guided by information specified relative to that organism. Social interactions are dependent on each participant of an interaction perceiving the communicative signals of the other as directed to itself. It is not sufficient to realize that someone has made a threatening gesture, one needs to know whether it was directed at oneself. In this context, it is of interest to note that cells coding threat expressions are generally more responsive to the frontal face with eye contact than to the profile face or the face with eyes averted (Perrett & Mistlin, 1989; Perrett et al. 1989d). In any predator/prey chase the predator must interpret the prey's movements relative to itself in order to catch the prey. Reciprocally, the prey must interpret the predator's movements relative to itself to avoid being caught. Thus, descriptions of another organism's static posture and dynamic movements relative to the viewer are of the utmost importance in guiding the viewer's behaviour (Perrett & Mistlin, 1989).

The relative importance of analyses using the viewer and observed object as the frame of reference may well be reflected by the frequency of cell types recorded in the temporal cortex. Cells displaying viewer-centred coding are common in the superior temporal sulcus but cells displaying view-general coding, which use the observed object as the frame of reference for the analysis, are rare, particularly for movement.

Goal-centred coding

Viewer- and object-centred descriptions are appropriate for describing what an object is and what its movements are, but these types of descriptions are of little use for understanding actions or in accounting for why an individual is performing a particular movement. For describing actions of an organism one needs a completely different type of description – one which relates the movements of the agent of an action to the goal of the action (Perrett *et al.* 1989*a–d*).

In Fig. 6 descriptions which make this relationship clear have been labelled goal-centred. We define goal-centred descriptions as descriptions in which the disposition or movements of one animate object (the agent) are specified with respect to a second object or part of the environment (the goal).

Actions can be directed at achieving many different types of goals, and there are therefore many varieties of goal-centred descriptions. Actions can also be directed at the viewer which can complicate the classification of descriptions. The distinction between viewer- and goal-centred descriptions with the viewer as the goal can be understood if a description such as 'reach for my hand' is considered here. There is a considerable amount of proprioceptive information needed to specify where my hand is. Viewer-centred descriptions of the same action would not use such extra information, nor would they make explicit the goal of the movement (a viewer-centred description might be 'movement with components left, down and towards me' but this does not predict whether the movement will contact my hand).

Descriptions of hand actions

We have studied a population of cells which are selectively activated by the sight of actions of the hand and whose responses can be understood as providing goal-centred representations. These cells were located within the bottom bank of the STS (predominantly in area TEa, Seltzer & Pandya, 1978). Fig. 1 shows that the population is anatomically distinct from the populations of cells responsive to static views of the face or head which are found mainly in the upper bank of the sulcus.

Selectivity for different actions

So far we have studied 50 cell (12·3 % of the cells sampled in the lower bank), which were found to be responsive to hand–object interactions. None of these cells was found to be responsive to conventional visual stimuli (bars or gratings) or to a variety of more complex three-dimensional stimuli or to a number of

meaningful stimuli (including the sight of faces, body movements, food items, somatosensory and auditory stimuli).

We found that the cells did not respond equivalently to all hand actions. Different actions of the hand activated different subpopulations of cells. So far, we have found cells selective for seven different actions: reach for, retrieve, manipulate, pick, tear, present to the monkey, and hold. (This list will probably increase with further study.) For 34 cells studied with different actions, 12 were found to be highly selective, in that their response to one action was more than four times that to any other action tested. Twelve further cells displayed some selectivity, responding to two or more but not all of the actions tested.

Fig. 8 illustrates three cells selectively responsive to different actions. One is selective for the sight of a manipulatory action but unresponsive to a hand presenting, tearing, picking or rotating an object. A second cell is selective for a picking action, and not responsive to other actions. A third cell is selective for the sight of tearing.

Generalization to different instances of one action

In common with other cells in the temporal cortex, the majority of those found to be selective for actions showed considerable perceptual generalization to preferred stimuli across many different viewing conditions. While all cells responded to actions performed close (0·2 m) to the monkey, 74 % (23/31) of cells were also responsive to the same actions performed at a distance of 4 m. These cells tolerated a 20-fold change in retinal image size and velocity. The responses also generalized over different speeds of the preferred action with the individual hand movements completed briskly within 0·5 s or more slowly with a duration of more than 5·0 s (e.g. Fig. 9). Similarly, cells generalized across vantage point, with 24 out of 26 cells responsive to the front view of hands performing an action also responding to the side view of the hands. This finding indicates that the cells are not providing viewer-centred descriptions of the actions.

Thus, not only are the cells selective for particular actions but they also generalize their responses to different instances of the preferred action. There are considerable changes in the local orientations, velocities and sizes of image components across different instances of one action. The cells must therefore generalize over many low-order visual variables (orientation, etc.) which are important during early visual processing. One can infer from this generalization that the simple visual characteristics are not sufficient to account for the visual selectivity displayed by such cells.

Cell responses were found to be unrelated to auditory cues associated with actions such as tearing, since the sound of the action performed out of sight was ineffective (for 21 out of 21 cells tested), whereas silent video films of hand actions were effective in eliciting responses (for nine out of nine cells tested).

Although all cells studied responded to the sight of hand actions performed by other individuals, we have only just begun to explore whether the cells are operative as the monkey witnesses hand actions performed by itself. Preliminary
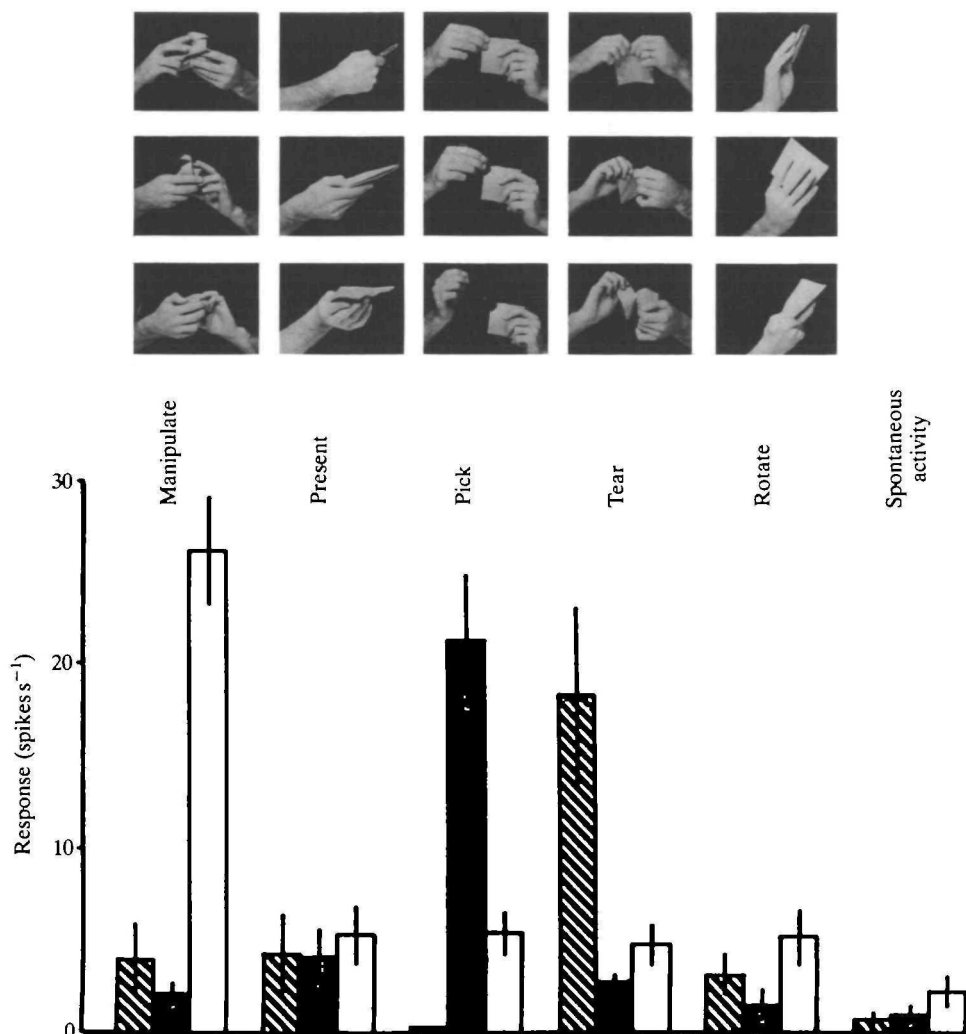
Fig. 8. Selectivity of neuronal responses to the sight of actions of the hand. Upper: three stages of different hand actions. Lower: selectivity of neuronal responses to the sight of actions of the hand. The mean and standard error of response of three cells is illustrated for the five actions. Cell F053 (clear bars) responded more ($P < 0.01$) to the sight of manipulation than other actions and SA ($N = 10, 6, 10, 8, 10, 10$; $F = 23.6$; df = 5, 47; $P < 0.01$). Cell F049 (solid bars) was selective for the sight of a picking action over other actions and spontaneous activity ($P < 0.01$ each comparison, $N = 7, 5, 8, 6, 4, 7$; $G = 13.0$; df = 5, 31; $P < 0.01$). Cell F028 (cross-hatched bars) responded more to a tearing action than to other actions or its spontaneous activity, SA, ($P < 0.05$ each comparison, $N = 7, 8, 4, 8, 4, 6$; $F = 4.43$; df = 5, 31; $P < 0.01$). (From Perrett *et al.* 1989*e*.)

study, however, indicates an affirmative answer to this question; five out of six cells studied were found to be responsive to the sight of the monkey's own hands during appropriate self-produced actions. This sensitivity could allow the cells to
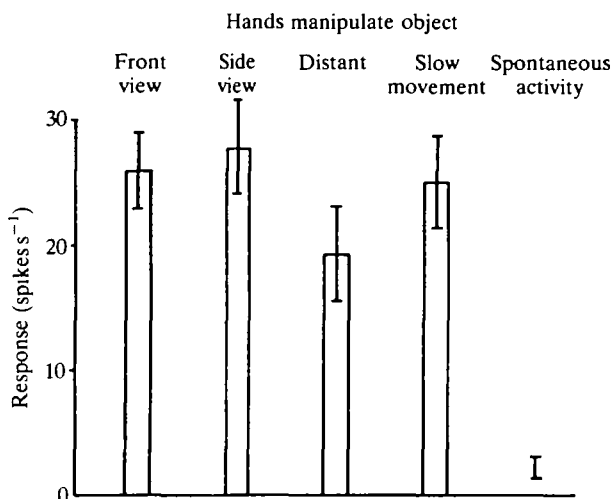
Fig. 9. Generalization of response to different instances of manipulation. With the action performed close to the monkey (0·2 m) and briskly (with individual hand movements completed within 0·3 s) the cell responded with the hands seen from the front (pointing towards the monkey) or from the side (in profile). The cell also responded to the action performed at a greater distance (4·0 m) and with the hand movements executed slowly with individual movements taking more than 5·0 s. Mean response to different instances of the same action were not significantly different ($P > 0.05$) but were greater than the spontaneous activity ($P < 0.05$). (Number of separate trials per condition: $N = 10, 8, 7, 7, 9$; $F = 9.9$; df = 4·36; $P < 0.01$). (From Perrett *et al.* 1989*e*.)

participate in the control of the monkey's hand movements, particularly when dextrous skills are required. Generalization to self-produced visual stimulation is, however, not a universal finding. In the upper bank of the sulcus, particularly in area TPO, movement-sensitive cells apparently lacking any form-selectivity, nonetheless, are found to be unresponsive to the sight of the monkey's own movements (Perrett *et al.* 1989*e*).

Selectivity for agent, object and agent–object interaction

Generalization indicates encoding of visual information invariant to that action rather than encoding of incidental image qualities (e.g. retinal velocity, orientation or size) which change with different viewing conditions. We have therefore studied whether the response selectivity of these cells depended on the visual characteristics of the hand performing the action, the object acted upon or the interaction between agent and object.

Agent of action

Fig. 10 compares the responses for one cell to a hand reaching for an object with responses to a control bar (of similar size to an arm and hand) moving towards the same object. For 14 out of 18 cells for which this type of comparison was made, a
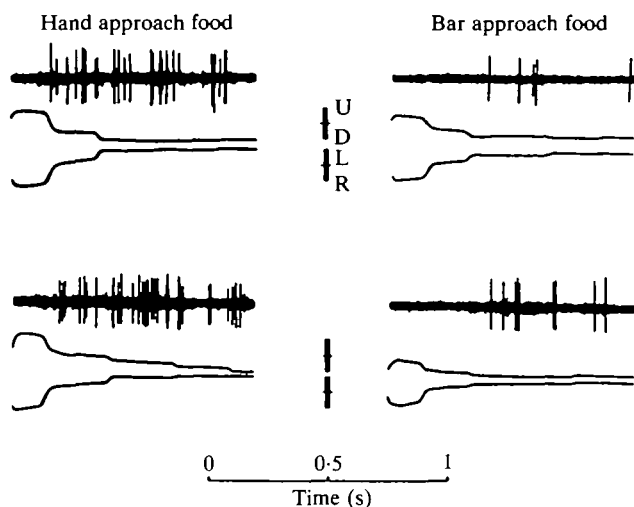
Fig. 10. Two trials with movements of the hand (left column) and control stimuli (right column) towards a target object. For each trial the upper trace represents neuronal activity and lower traces are records of vertical and horizontal eye position monitored simultaneously with an infrared tracking device. Vertical calibration bars for eye movement are 20° (10° up, U, and down, D, and 10° left, L, and right, R, of a position straight ahead). Trial records begin with the onset of stimulus movement; neuronal responses were not related to patterns of eye movements which are triggered by or used to follow given actions, since qualitatively similar patterns of saccades, tracking and fixations occur, despite dissimilar neuronal responses. (Adapted from Perrett *et al.* 1989*e*.)

clear selectivity was found for hand–object interactions compared with object–object interactions, despite similar eye movements accompanying both actions (Fig. 8). Comparisons between an action (such as manipulation) performed bimanually and unimanually for 20 cells also indicated selectivity for some aspect of the agent performing the action. Nine cells responded to an action performed with two hands, but not when it was performed with one hand (for the remaining 11 cells, there was no difference between conditions). We have yet to identify the visual attributes of a hand or pair of hands that are necessary for the responses, but these preliminary results indicate that the cells are to some extent selective for the agent performing the act.

Object of action

Although cells were selective for the agent performing an action, 16 out of 27 tested were found to be unselective for the object acted upon. For these cells it did not matter whether the object was large (30–40 cm) or small (1–2 cm), black, white or coloured; three cells, however, were more responsive to actions involving deformable objects than rigid objects and a further eight cells appeared more responsive to actions involving food than to non-food objects. It is relevant that object properties – such as surface reflectance or colour, size and weight (over a

considerable range) – do not generally constrain the actions which can be performed on them. Rigidity, however, does constrain actions like tearing and manipulation.

Thus, the cells are generally insensitive to the properties of the objects acted upon but are sensitive to the type of interaction between hand and object. The exception to this seems to be for particular food items. Indeed, one of the main functions of the cells may be in recognizing the exploitation of food sources by other monkeys. There are many instances where attention to food preparatory acts of others is of great benefit to animals in a social environment. Benefit may be accrued in the development of a new food-acquisition skill. This can be either by trial and error learning, once attention has been drawn to a food source, by direct imitation of the actions necessary to get the food or even by comprehending the goal of the actions of others and inventing a personal solution (Thrumble, 1987; Thrumble & Perrett, 1987; Whiten, 1989).

Agent–object interaction

Sensitivity to characteristics of the interactions between agent and object was found to be as fundamental as sensitivity to the characteristics of either the agent or the object alone. Indeed, it was a defining characteristic of all cells reported here that their responses were dependent upon the interrelationship of hand and object movements (e.g. Fig. 11). For the 50 cells studied, hand movements alone miming the preferred action elicited reduced neuronal responses compared to hand–object interactions. Similarly, object movements appropriate to the action, but with no hands visible, provoked less response than combined hand–object interactions for all 50 cells. Even the combination of appropriate hand and object movements produced little or no response (for 28 of 28 cells tested) when the movements were performed with a spatial separation of more than 3–4 cm in height or in depth.

The comparison between normal and spatially separate conditions is important in demonstrating a cell's selectivity to interaction, but it is also important in ruling out a great number of variables which might be suggested to account for the cell's responses. There may be some who might wish to account for the cell's responses in terms of visual variables which are important in the early stages of visual processing. Any explanation advanced to account for the visual basis of responses must, however, accommodate situations both ineffective and effective in producing response differences. A variety of simple image parameters (e.g. contrast, spatial and temporal frequencies) change radically across different instances of the same action but change minimally, if at all, between situations where the actions are performed with spatial contact and without spatial contact. Yet the different instances of the same action affect cell responses minimally, whereas spatial separation affects cell responses maximally. It is unlikely, therefore, that the cell responses are related to simple visual attributes. Rather, it is parsimonious to assume (and a more useful working hypothesis) that the responses reflect a high level of interpretation of the interaction between hands and objects.
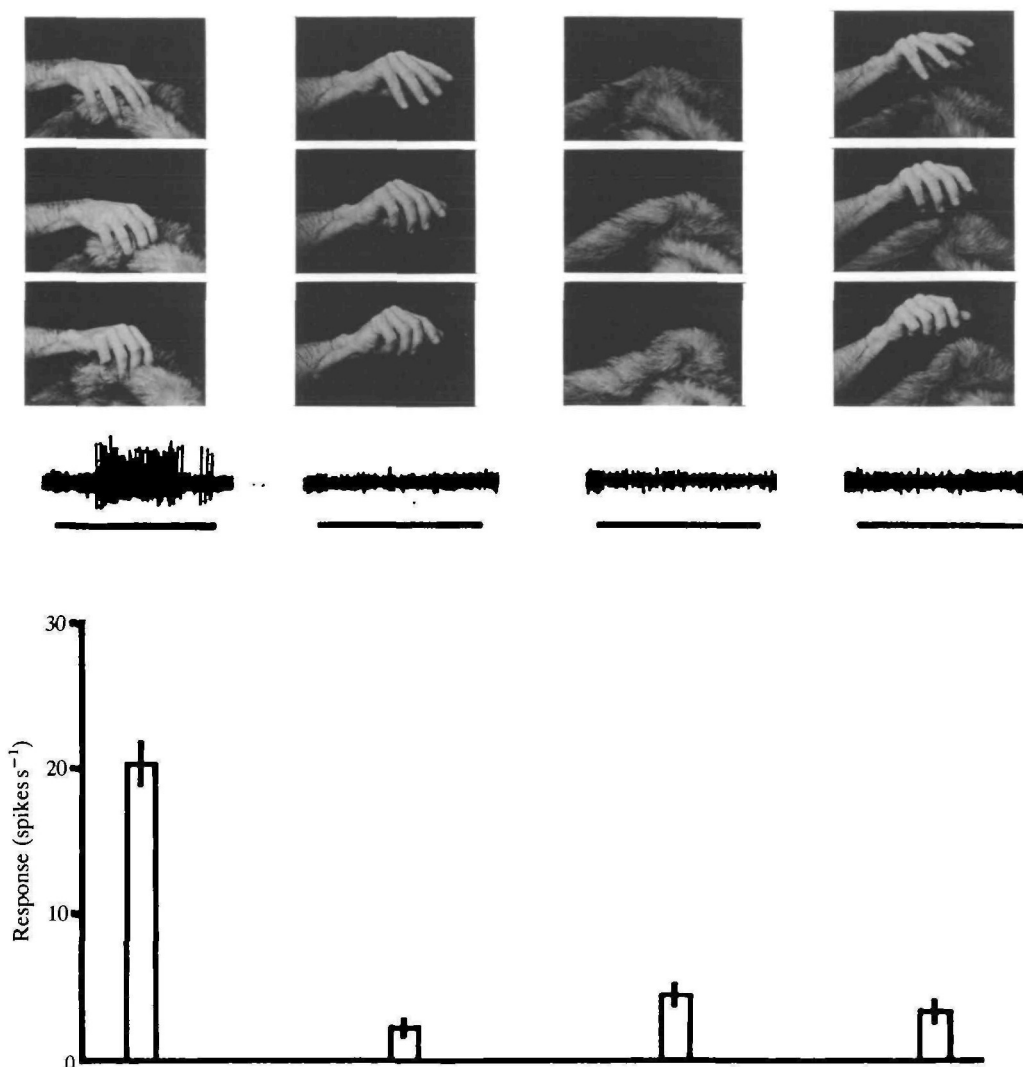
Fig. 11. Visual sensitivity to interaction between hand and object. Upper: three stages of action sequences involving a hand manipulating an object (a large piece of fur), hand movement (no object visible), the object deforming (no hand visible), and a hand and object moving but spatially separate. Middle: traces of neuronal activity during one trial for each action sequence. Stimulus onset and duration (1·0 s) is illustrated by calibration bar beneath activity traces. Lower: mean response (± 1 s.e.) for 10 trials each condition for one cell. The response to the sight of a hand manipulating an object was greater than that to other conditions or spontaneous activity ($P < 0.01$ each comparison; $F = 43.5$; df = 4, 36; $P < 0.01$; SA = 2.1 ± 0.7). (From Perrett *et al.* 1989*e*.)

Goal direction

The sensitivity to interaction between an agent performing an action and the object or goal of an action can be seen clearly for cells selective for reaching
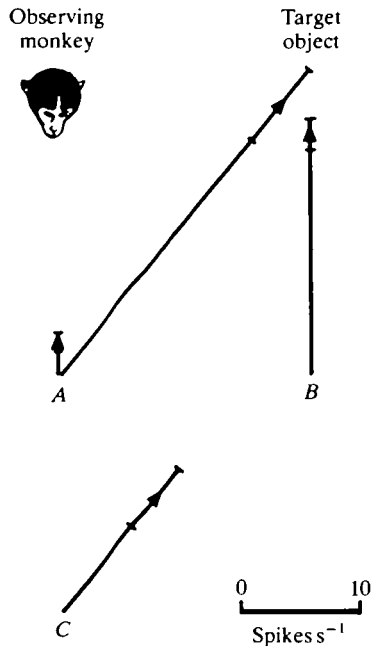
Fig. 12. Neuronal responses to an experimenter reaching in different directions. The lengths of the lines to the arrowhead represent the mean magnitudes of neuronal response (mean ± 1 s.e.) for several trials of the experimenter's reaching. The direction of reaching is given by the arrowheads. The starting position of the reach relative to the observing monkey and to the target is indicated by the origin of the lines (A,B,C). Position A, experimenter 1 m in front of monkey; B, 1 m in front and 1 m to the left; C, 2 m in front. Reaching from A or B in a direction which brought the hand over the target produced responses greater than comparable arm movements from A and C which did not bring the hand towards the target or spontaneous activity ($P < 0.01$ each comparison; $N = 8, 4, 8, 5$; $F = 11.8$; df $= 4, 28$; $P < 0.01$; spontaneous activity $= 2.3 \pm 0.5$; $N = 8$) (From Perrett et al. 1989e.)

movements of the hand. The responses of one such cell are illustrated in Fig. 12. This cell was responsive to the sight of reaching movements which brought the experimenter's arm to a particular spatial location in which a target object was positioned. Reaching in other directions which did not bring the hand to the target were less effective. These statements constitute what we term goal-centred descriptions, because the movements are defined relative to the goal of the movement (which could be an object or a spatial location).

The direction selectivity of this cell cannot be understood in the same way as that of cells in the early visual pathways. This is because effective movements have no consistent direction with respect to the observer's retina; viewer-centred representations are thus inappropriate as a framework for interpreting responses. Similarly, specification of effective arm movements relative to the experimenter's body (object-centred descriptions) change for different positions of the experimenter (e.g. Fig. 12: from position A experimenter reaches 45° to his right, from
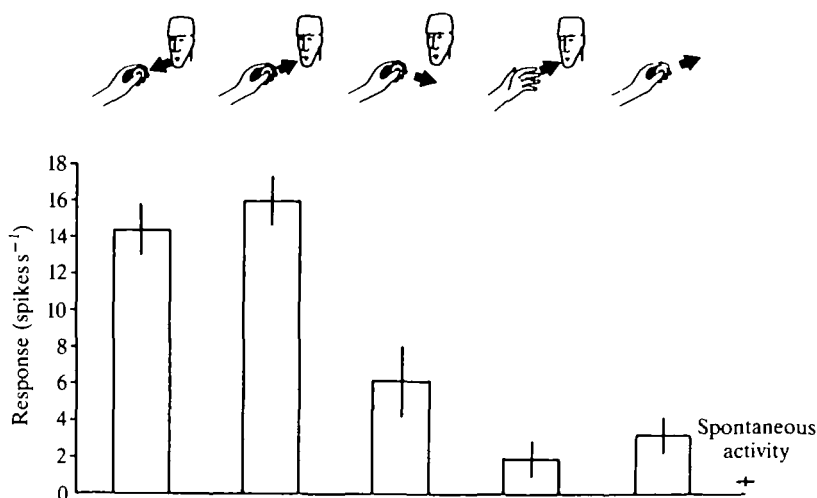
Fig. 13. Visual sensitivity to the act of an experimenter bringing an object to his mouth. The mean (± 1 s.e.) response of one cell is illustrated for different viewing conditions. The cell responds to the sight of the mouth moving towards the hand and object and to the sight of a hand bringing an object towards the mouth. The cell is significantly less responsive in control conditions in which the hand and object move towards the body missing the mouth, the hand alone moves towards the mouth, the hand and object make the appropriate movement but the head is absent or spontaneous activity ($P < 0.001$ each comparison; $N = 10$; $F = 20.9$; df = 5, 46; $P < 0.0001$). (From Perrett *et al.* 1989c.)

position *B* experimenter reaches to his front). To deduce the relationship between reaching and the target, one needs additional information concerning the position of the target relative to the experimenter. Thus many object-centred descriptions are needed for this (or any other) action and none of them makes the interaction between reaching and target explicit.

Goal-centred descriptions, by contrast, code the effective movements economically and make the interaction explicit. The goal-centred framework applies to all the cells sensitive to hand actions and to other cells responsive to whole-body actions (Perrett *et al.* 1989b,e). This is because their responses can only be understood by relating the movements of the hand or body to particular objects or positions (i.e. goals) in the environment.

Achieving goals by different means

Actions can be achieved by a variety of means. In a trivial sense the act of reaching for the target (e.g. Fig. 13) can be achieved from a variety of starting positions using the same type of arm movement, though aimed in different directions. In a more fundamental sense, entirely different body movements can achieve the same goal. Consider the responses of the cell illustrated in Fig. 13. This cell responded to hand movements which carry an object contained in the hand towards the mouth. Hand movements directed to other parts of the body

were less effective – and movement of an empty hand to the mouth was also ineffective. The act of bringing an object and the mouth together can also be achieved by leaving the hand static and moving the whole body and head so that the mouth moves closer to the object. Such movements were also effective in activating the cell depicted. Furthermore, the action could be completed with the entire body remaining stationary and the object moved towards the mouth by a second individual. Again such movements activated this cell (not illustrated). Thus, three qualitatively distinct types of movement achieve the same goal and each caused the cell to increase its rate of signalling. The cell, therefore, appears to be signalling the act of bringing an object and the mouth together, irrespective of the particular means of achieving this end.

The use of goal-centred descriptions

The coding of interrelationships that is inherent in goal-centred descriptions provides a framework through which the visual system can achieve a rich understanding of the world which embodies causation and intentionality.

The sensitivity to spatial contiguity that is manifested by cells' responses confers on them the property of detecting causal relationships. Human observers have a reduced impression of causality in situations where hand and object movements are separated in space or in time (e.g. Fig. 11; Michotte, 1963; Leslie, 1982). With small separations between hand and objects there is a partial sense of causality and an impression of some 'magical' control at a distance, but if the separation is widened then the sense of causality breaks down and the movements are perceived as unrelated.

Relational coding can directly specify intentionality for actions where the goal or object of an action is some distance away from the agent. That an agent reaches towards a target presumes that there is an intention in the reaching movement to attain the target goal. Similarly, for a whole-body action, such as walking towards the door, this description embodies an implicit assumption that the person walking intends to reach the door (Perrett et al. 1989a,b). Of course, there might be varying degrees to which an observer is convinced that a person walking towards or reaching towards the door intends to get there, the impression might depend on the starting distance, but this reduction in impression of intention might correlate with (or be caused by) a reduction in the cellular activity of neurones of the type described here.

Conclusion

In attempting to understand different aspects of higher visual processing it has been important to realize that different frames of reference are suited to different types of recognition (Feldman, 1985). To recognize what an object is, viewer- and object-centred descriptions are appropriate, but to recognize what an organism is doing one needs to employ goal-centred descriptions. To contrast the three types of description referred to here, take a scene containing two monkeys. A viewer

centred description of the scene might be: monkey A turns head to my left and moves arm to my left. Here the viewer (myself) is the frame of reference and the monkey's limb movements are specified relative to me. An object-centred description might be: monkey A turns its head over its shoulder and moves arm to its right. Here the monkey's movements are related to its own body. A goal-centred description might be: monkey A turns to face monkey B and reaches for monkey B. Goal-centred descriptions are thus important because they provide a much richer account of what is going on in the environment than other types of description considered to date.

The studies of cells sensitive to hand and body actions have revealed visual encoding of interactions in the environment. This type of coding has not been reported at the physiological level and is generally lacking in discussion of computational frameworks for vision. Yet this visual encoding of interaction is important for the provision of a meaningful and causal account of the world.

The finding of cells in the monkey brain that are selective for the sight of actions and which are unaffected by auditory cues associated with actions indicates the extent to which meaningful relationships can be derived purely within the visual modality, without the reliance on the capacity for language. Moreover, the identification of cell types specific for actions provides an opportunity for direct study of the mechanisms by which the brain computes interactions and determines causal and intentional relations within actions.

References

ANDERSON, R. A., ESSICK, G. K. & SIEGLE, R. M. (1985). The encoding of spatial location by posterior parietal neurons. *Science* **230**, 456–458.

BARLOW, H. B. (1972). Single units and sensation: A neuron doctrine for perceptual psychology. *Perception* **1**, 371–394.

BARLOW, H. B. (1985). The twelfth Bartlett Memorial Lecture: The role of single neurons in the psychology of perception. *Q. Jl Psychol.* **37A**, 121–145.

BAYLIS, G. C., ROLLS, E. T. & LEONARD, C. M. (1985). Selectivity between faces in the responses of a population of neurons in the cortex in the superior temporal sulcus of the monkey. *Brain Res.* **342**, 91–102.

BRUCE, C. J., DESIMONE, R. & GROSS, C. G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *J. Neurophysiol.* **46**, 369–384.

DESIMONE, R., ALBRIGHT, T. D., GROSS, C. G. & BRUCE, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *J. Neurosci.* **8**, 2051–2062.

FELDMAN, J. A. (1985). Four frames suffice: A provisional model of vision and space. *Behav. Brain Sci.* **8**, 265–289.

GEORGOPOULOS, A. P., LURITO, J. T., PETRIDES, M., SCHWARTZ, A. B. & MASSEY, J. T. (1989). Mental rotation of the neuronal population vector. *Science* **243**, 234–236.

GRAY, C. M., KONIG, P., ENGEL, A. K. & SINGER, W. (1989). Oscillatory responses in cat visual cortex exhibit inter columnar synchronization which reflects global stimulus properties. *Nature, Lond.* (in press).

HUBEL, D. H. & WIESEL, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *J. Physiol., Lond.* **195**, 215–243.

KONORSKI, J. (1967). *Integrative Activity of the Brain.* Chicago, Illinois: University of Chicago Press.

LESLIE, A. M. (1982). The perception of causality in infants. *Perception.* **11**, 173–186.

MARR, D. (1982). *Vision.* San Francisco: W. H. Freeman & Co.

MARR, D. & NISHIHARA, N. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proc. R. Soc.* B **200**, 269–294.

MICHOTTE, A. M. (1963). *The Perception of Causality.* London: Methuen.

MIKAMI, A. & NAKAMURA, K. (1988). Selective neuronal activities in the superior temporal sulcus of macaque monkey. *Soc. Neurosci. Abstr.* **8**, 4.

PANDYA, D. & YETERIAN, E. H. (1985). Architecture and connections of cortical association areas. In *Cerebral Cortex*, vol. 4 (ed. Peters & Jones), pp. 3–64. New York: Plenum Press.

PERRETT, D. I., HARRIES, M. H., BENSON, P. J., CHITTY, A. J. & MISTLIN, A. J. (1989a). Retrieval of structure from rigid and biological motion; an analysis of the visual response of neurons in the macaque temporal cortex. In *AI and The Eye* (ed. T. Troscianko & A. Blake). Chichester: J. Wiley. (in press).

PERRETT, D. I., HARRIES, M. H., CHITTY, A. J. & MISTLIN, A. J. (1989b). Three stages in the classification of body movements by visual neurones. In *Images and Understanding* (ed. H. B. Barlow, C. Blakemore & M. Weston-Smith), Cambridge: Cambridge University Press.

PERRETT, D. I., HARRIES, M. H., MISTLIN, A. J. & CHITTY, A. J. (1989c). Recognition of objects and actions: frameworks for neural computation and perceptual experience. In *Higher Order Sensory Processing and Artificial Sensing* (ed. D. M. Guthrie). Manchester University Press. (in press).

PERRETT, D. I. & MISTLIN, A. J. (1989). Perception of facial characteristics by monkeys. In *Comparative Perception* (ed. M. Berkeley & W. Stebbins). New York: J. Wiley.

PERRETT, D. I., MISTLIN, A. J. & CHITTY, A. J. (1987). Visual cells responsive to faces. *Trends Neurosci.* **10**, 358–364.

PERRETT, D. I., MISTLIN, A. J., CHITTY, A. J., SMITH, P. A. J., POTTER, D. D., BROENNIMANN, R. & HARRIES, M. H. (1988). Specialised face processing and hemispheric asymmetry in man and monkey: Evidence from single unit and reaction time studies. *Behav. Brain Res.* **29**, 245–258.

PERRETT, D. I., MISTLIN, A. J. & HARRIES, M. H. (1989d). Seeing faces: The representation of facial information in temporal cortex. In *Seeing Colour and Contrast* (ed. J. Kulikowski). Oxford: Pergamon Press. (in press).

PERRETT, D. I., MISTLIN, A. J., HARRIES, M. H. & CHITTY, A. J. (1989e). Understanding the visual appearance and consequences of hand actions. In *Vision and Action: the Control of Grasping* (ed. M. A. Goodale), Norwood, NJ: Ablex Publishing Corporation (in press).

PERRETT, D. I., ROLLS, E. T. & CAAN, W. (1982). Visual neurons responsive to faces in the monkey temporal cortex. *Expl Brain Res.* **47**, 329–342.

PERRETT, D. I., SMITH, P. A. J., MISTLIN, A. J., CHITTY, A. J., HEAD, A. S., POTTER, D. D., BROENNIMANN, R., MILNER, A. D. & JEEVES, M. A. (1985a). Visual analysis of body movements by neurons in the temporal cortex of the macaque monkey: A preliminary report. *Behav. Brain Res.* **16**, 153–170.

PERRETT, D. I., SMITH, P. A. J., POTTER, D. D., MISTLIN, A. J., HEAD, A. S., MILNER, A. D. & JEEVES, M. A. (1984). Neurones responsive to faces in the temporal cortex: Studies of functional organization, sensitivity to identity and relation to perception. *Human Neurobiol.* **3**, 197–208.

PERRETT, D. I., SMITH, P. A. J., POTTER, D. D., MISTLIN, A. J., HEAD, A. S., MILNER, A. D. &

JEEVES, M. A. (1985*b*). Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proc. R. Soc.* B **223**, 293–317.

ROLLS, E. T. (1984). Neurones in the cortex of the temporal lobe and in the amygdala of the monkey with responses selective for faces. *Human Neurobiol.* **3**, 209–222.

ROLLS, E. T. & BAYLIS, G. C. (1986). Size and contrast have only small effects on the responses to faces of neurons in the cortex of the superior temporal sulcus of the monkey. *Expl Brain Res.* **65**, 38–48.

ROSE, D. (1974). The hypercomplex cell classification in cat striate cortex. *J. Physiol., Lond.* **242**, 123P–125P.

SELTZER, B. & PANDYA, D. N. (1978). Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey. *Brain Res.* **149**, 1–24.

THRUMBLE, K. (1987). An examination into the acquisition and temporal modification of tool use strategies by captive stimptail macaques. Unpublished BSc thesis, St Andrews University.

THRUMBLE, K. & PERRETT, D. I. (1987). The acquisition of tool use behaviour in a social colony of stumptail macaques. Paper presented to the Association for the Study of Animal Behaviour.

UTTAL, W. R. (1978). *The Psychobiology of Mind.* Hillsdale, New Jersey: Lawrence Earlbaum.

WHITEN, A. (1989). Transmission mechanisms in primate cultural evolution. *Trends Evol. Ecology* **4**, 61–62.

YAMANE, S., KAWANO, K., KAJI, S. & KOMATSU, H. (1988). Face neurons and neurons responsive to human faces in inferotemporal cortex of the monkey. *Soc. Neurosci. Abstr.* **85**, 21.