Evidence accumulation in cell populations responsive to faces: an account of generalisation of recognition without mental transformations

D.I. Perrett*, M.W. Oram, E. Ashbridge

Psychological Laboratory, St Andrews University, St Andrews, KY16 9JU, UK

Abstract

In this paper we analyse the time course of neuronal activity in temporal cortex to the sight of the head and body. Previous studies have already demonstrated the impact of view, orientation and part occlusion on individual cells. We consider the cells as a population providing evidence in the form of neuronal activity for perceptual decisions related to recognition. The time course of neural responses to stimuli provides an explanation of the variation in speed of recognition across different viewing circumstances that is seen in behavioural experiments. A simple unifying explanation of the behavioural effects is that the speed of recognition of an object depends on the rate of accumulation of activity from neurones selective for the object, evoked by a particular viewing circumstance. This in turn depends on the extent that the object has been seen previously under the particular circumstance. For any familiar object, more cells will be tuned to the configuration of the object's features present in the view or views most frequently experienced. Therefore, activity amongst the population of cells selective for the object's appearance will accumulate more slowly when the object is seen in an unusual view, orientation or size. This accounts for the increased time to recognise rotated views without the need to postulate 'mental rotation' or 'transformations' of novel views to align with neural representations of familiar views. © 1998 Elsevier Science B.V. All rights reserved

Keywords: Mental rotation; Face recognition; Cell populations

1. Introduction

One phenomenon that has featured heavily in accounts of the way object recogni-

---

* Coresponding author. Fax: +44 1334 463042; e-mail: dp@st-andrew.ac.uk

tion processes cope with changes in view and orientation is that of 'mental rotation'. The basic finding is that the time to match two views of an unfamiliar object differing by a rotation in depth or orientation in the picture plane is linearly related to the 3-D angular difference between the views (Shepard and Metzler, 1971). Introspectively, subjects feel that they are mentally rotating one version of the object and then checking whether the rotated image matches the other. A consciously-driven process has been envisaged whereby an internal representation of one object is subjected to a continuous, analogue transformation until it matches the image of the second object. The increasing reaction times (RT) with larger angular differences between views is taken as reflecting the greater degree of internal rotation subjects need to perform to achieve a match.

An effect which appears analogous is the extra time taken to recognise familiar or recently learnt objects when they are rotated in the picture plane from their usually experienced orientation (Cooper and Shepard, 1973; Jolicœur, 1985, 1990; Tarr and Pinker, 1989, 1991). Here again the amount of extra time required for recognition is related to the angular rotation from upright. The increase in reaction time with orientation is not necessarily linear as with the original experiments on matching of Shepard and Metzler (1971). Indeed the function relating RT to orientation is usually monotonic but can have anomalous inflections (e.g. for inverted images; see Tarr, 1995). However the similarity in the dependence of RT on orientation has led to the assumption that one mechanism (mental rotation) underlies both phenomena.

Similar accounts have been suggested for the effect of perspective view on the recognition of familiar objects. Palmer et al. (1981) found that subjects could most easily bring to mind one 'prototypical' or 'canonical' view of everyday objects (e.g. a model car, shoe or house). Photographs taken from this same view were matched to a category name (e.g. 'car') more quickly than photographs taken from different perspective views. Reaction time to match names with pictures depended on how much the objects were rotated in perspective from the canonical orientation. These findings are often interpreted as suggesting that, for each object, a single canonical view is represented in memory and that all other views and orientations of objects are mentally rotated to bring them into alignment with the single stored view in its normal orientation.

Analogous effects are present when subjects learn the appearance of a novel object from one perspective view and then are tested with views of the same object rotated in perspective. The time required for recognition increases as the angular difference between trained and test views increases up to a maximum of 180° (Tarr and Pinker, 1989; Humphreys and Kahn, 1992; Edelman and Bülthoff, 1992) or to a 'virtual' view (Vetter et al., 1994; Logothetis et al., 1994, 1995). Several authors account for the monotonic increase in processing time by postulating a process which is engaged to 'align', 'normalise', 'transform' or 'mentally rotate' the image of the novel test view of the object so that it matches the representation of the view previously experienced.

The effect of image size on simultaneous and successive matching of stimuli follows a similar pattern, with subjects taking progressively longer the greater the

difference in size between stimuli (Bundesen and Larsen, 1975; Besner, 1983; Jolicœur, 1987; Jolicœur and Besner, 1987; Ellis et al., 1989). To account for these findings, a process referred to as 'mental adjustment of size' or 'size normalisation' (Shepard and Metzler, 1971; Bundesen and Larsen, 1975; Besner, 1983; Larsen, 1985) has been suggested in which the internal representation of a stimulus is size specific and any image of the same stimulus in a non-standard size must first be size transformed to match the internal representation. The greater the size difference between image and internal representation, the greater the degree of mental size change before a match can be successful and recognition accomplished.

More recent theoretical accounts of recognition suggest that multiple views are represented in memory for each familiar object (Koenderink and van Doorn, 1979; Perrett et al., 1984, 1991; Poggio and Edelman, 1990; Edelman and Weinshall, 1991; Logothetis et al., 1994, 1995). Most models postulate that novel viewpoints of familiar objects are recognised by transformation, or alignment to the nearest stored view or by matching the input view to a view constructed as an interpolation between stored views (Tarr and Pinker, 1989, 1990, 1991; Ullman, 1989; Bülthoff and Edelman, 1992; Tarr, 1995; Edelman, 1997). Such models extend to generalisation across unusual orientations and sizes.

1.1. Problems with the mental rotation account

The main evidence for an involvement of 'mental rotation' or 'mental size change' in recognition is circumstantial. It is assumed that a monotonic increase in the time required to process images that are rotated in the picture plane or in 3-D perspective, or changed in size from a training stimulus, is a reflection of the operation of continuous transformation processes which are executed until completion. The increase in reaction times is not in itself evidence for mental rotation since it is also consistent with any process that takes more time the greater with greater differences between the test stimulus and the training stimulus.

There are several findings that make mental rotation implausible in accounting for object recognition. The rate of mental rotation reported in matching experiments of Shepard and Metzler (1971) was between 60 and 460°/s. This appears too slow to account for the recognition of misoriented objects. With this speed of mental rotation, an inverted object would take an extra 0.4 to 3 s to recognise compared to an upright object. Yet in some experiments, the extra time taken to process misoriented objects is only a few milliseconds and thus rotation speeds of 12 000°/s would be necessary to account for the identification (Humphreys and Quinlan, 1987). In such experiments subjects may employ recognition strategies that do not involve mental rotation but still give rise to a dependency of RT on orientation. This dependency is, however, the hallmark of mental rotation. Explaining different RT-orientation slopes with different mechanisms lacks consistency.

When researchers use the same objects for both recognition tasks and for left-right comparisons (equivalent to the Shepard and Metzler original mental rotation tasks) then RT orientation functions are comparable across tasks (Tarr and Pinker, 1989; Tarr, 1995). However, such consistency does not explain the radically different

speeds of mental rotation that are supposed to underlie recognition of rotated objects in different circumstances.

More problematic is the finding that the extra time required for recognising rotated objects depends on the task and decreases with practice (Jolicœur, 1985; Tarr and Pinker, 1989; Perrett et al., 1988b). One might assume that practice increased the efficiency of the transformational process and hence the speed of mental rotation but the practice effect is selective for the training pattern and the training orientation. Training at a particular novel orientation or view speeds up recognition selectively at that view and orientation (Perrett et al., 1988b; Tarr and Pinker, 1989; Tarr, 1995; Logothetis et al., 1994). If more efficient mental transformation to the canonical view accounted for the improvement with practice, then one would expect an improvement for all orientations and views between trained and canonical conditions.

Models postulating multiple stored representations account for practice effects by assuming that new representations are formed as a result of experience (Bülthoff and Edelman, 1992; Tarr and Pinker, 1989). Subjects, frequently exposed to an object in an inverted orientation, will form a representation of the object in the inverted orientation. Such subjects will no longer need to transform the inverted orientation to upright but horizontal orientations will still need transforming to match either the canonical upright representation or the new inverted representation.

Problems with the mental rotation account of recognition are not restricted to recognition by normal subjects. Mental rotation does not provide a parsimonious account of the impact of brain lesions on object recognition. Farah and Hammond (1988) describe a patient who was able to recognise objects presented in various orientations but yet was impaired in the mental rotation tasks of Shepard and Metzler (1971). Perrett et al. (1988a) described a patient who was able to discriminate face configuration at any orientation yet (unlike normal subjects) showed no increase in reaction time with rotation of stimuli from upright. Thus orientation invariant object recognition can occur in individuals who do not show signs of mental rotation processes.

Similarly, a patient (L.G.) described by Turnbull et al. (1995) could recognise line drawn objects without being able to indicate the canonical orientation of the drawings. Turnbull et al. (1995) suggest that there are two pathways in normal object recognition, one displaying object-centred (view-independent) properties and a second displaying view-dependent properties. The second pathway would be selectively damaged in patient L.G. Such an explanation follows the account of Goodale and Milner (1992) of two streams of cortical visual processing. They suggest that the dorsal (parietal) pathway employs view-dependent visual descriptions to guide motor reactions and that the ventral (temporal) pathway employs view-independent descriptions to support recognition of what an object is. Such accounts overlook the extensive view-dependent processes within the temporal cortex (for review see Perrett and Oram, 1993). The patient described by Turnbull et al. (1995) could also be interpreted as having intact orientation-dependent descriptions in the ventral system. These descriptions may access limbic structures containing view-independent descriptions and semantic associations but for the

patient the descriptions may have no access to the parietal systems necessary for defining orientation with respect to egocentric or gravitational axes.

1.2. A physiological explanation

The thesis advanced here is that mental rotation is not necessary to account for the effect of view or orientation on recognition of familiar objects. An alternative explanation relies on two assumptions. First, one stage in the processing leading to recognition involves neurones tuned to the appearance of objects in an orientation, view and size-specific manner and, second, the number of neurones tuned to an object at a given orientation, view and size depends on the amount of experience for the object in that circumstance. It follows that evidence for the presence of an object (in terms of neuronal activity) will accumulate more slowly when the object is presented at an unusual orientation, view or size.

If an object's appearance is learned in a view-specific manner, then more cells will become tuned to the view or views of the object that are commonly experienced. Fewer cells will respond to the object the more it is rotated from the experienced view(s). If we assume that any 'decision making' neural process guiding behavioural output acts on the basis of evidence in the form of input neuronal activity, then behavioural output will take longer and longer for progressively more unusual views. This will be the case simply because visual neuronal populations weakly activated by unusual views will take longer to provide sufficient input for the neurones in the 'decision apparatus' to act upon. Physiological data support this account of the generalisation across unusual viewing conditions with respect to perspective view, orientation, size and part occlusion. Manipulations of perspective view will be considered in the next section.

Explanations in which signals of different strengths require different durations for detection are familiar to Experimental Psychology and follow from information theory. Weak signals can give rise to longer reaction times compared with strong signals because, in a noisy background, sampling of weak signals takes longer to achieve a given level of evidence. Previous studies have presented such an account for sensory processing and for the way reaction time relates to simple stimulus parameters such as visual intensity (e.g. Vaughan et al., 1966; Lennie, 1981). We argue that the same applies to higher stages of visual processing and can account for what has been assumed to be a cognitive operation within visual recognition. We show that the responses of populations of neurones processing faces collectively exhibit the properties equivalent to those assumed in previous models of signal processing.

2. View

In the mental rotation tasks of Shepard and Metzler, the effects of change in view and change in orientation were comparable. The psychological impact of view and orientation change on recognition was also noted to be similar in the introduction.

Stimulus rotation in depth, however, changes the visibility of features while rotation in the picture plane changes orientation with respect to gravity but does not change feature visibility. In principal, the mechanisms by which the nervous system cope with the two types of rotation may be fundamentally distinct. Therefore, the physiological impact of the effect of view and orientation are considered separately in Sections 1 and 2. Section 4 also addresses changes in view that affect feature visibility.

From the studies so far, the vast majority of cells in temporal cortex that respond selectively to complex objects are sensitive to perspective view (Perrett et al., 1985, 1991; Logothetis et al., 1995). Different cells are maximally responsive to different views. View sensitivity is apparent for cells processing the appearance of socially significant stimuli such as the face and body and for cells processing arbitrary objects (amoeboid or angled wire frame shapes) that have acquired significance in the context of behavioural tasks. For cells responsive to the head and body, the distribution of view coding is uneven; more cells are tuned to particular characteristic views (full face, profiles and back) and more cells are tuned to the frontal views than to the back views (Perrett et al., 1991, 1994). There are probable behavioural consequences of these anisotropies (Harries et al., 1991; Perrett et al., 1994). The prevalence of cells tuned to frontal views presumably reflects greater experience and attention to these views and might mean that front views are detected more efficiently than back views. Stronger predictions for behavioural data can be made on the basis of cell tuning for view.

Most cells show a gradual decline of activity for views progressively more rotated from one (the cells preferred) view. The sharpness of the tuning function varies from cell to cell. For cells selectively responsive to the head, a rotation of 60° from optimal views reduces cell responses by approximately 1/2 and a rotation of 180° from optimal generally elicits responses barely different from background activity or the responses to other classes of control object. This type of tuning function characterises 83% of all cells responsive to the head (Perrett et al., 1991).

Tuning curves are usually calculated on the basis of the neural activity over extended periods of time, 250–500 ms (Hasselmo et al., 1989; Perrett et al., 1991; Logothetis et al., 1995). Responses of individual cells are assessed over these time periods because fixation of the stimulus is constant for the initial 250 ms and therefore one can consider cell responses as reflecting processing of the same stimulus. To make predictions for behavioural reaction times during recognition it is appropriate to look in detail at how the neuronal responses develop over time. After cells in temporal cortex have commenced firing they provide continuous output to other brain systems. Control of behaviour can be generated in such systems as soon as sufficient sensory evidence is available. This may occur well before temporal cortex cells have been firing for 250–500 ms. The view tuning of cells responsive to the sight of the face and other views of the head have been well characterised and the methods of study detailed extensively elsewhere (Perrett et al., 1985, 1991; Oram and Perrett, 1992, 1994; Wachsmuth et al., 1994). Here we present an analysis of the time course of cells' responses recorded in area STPa of the temporal cortex and characterised in these other studies.
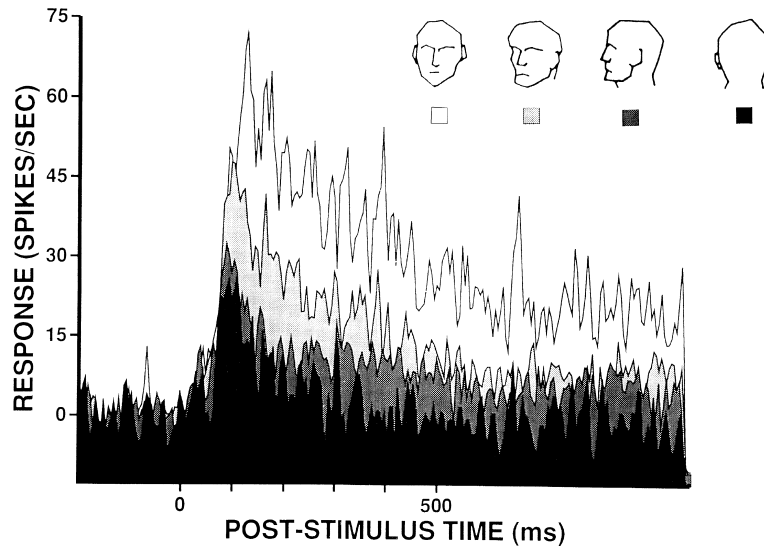
Fig. 1. Post-stimulus time histogram of responses showing effect of rotation from face view. Average response of a population of 20 cells selective for the face view of the head plotted as a function of time relative to stimulus presentation (0 ms) in successive 5 ms bins. View is expressed as the angle of rotation (45, 90 and 180°) away from the face (optimal) view. Data from views (5–10 trials per cell per view) rotated clockwise and anticlockwise by an equal angle from the face view have been pooled. At the population level, responses to views rotated an equal amount from face view were statistically indistinguishable (protected least significant difference post-hoc testing: $P>0.5$ each comparison; overall effect of condition: $F_{(9,171)} = 35$, $P < 0.0005$). The angle of rotation from face view is symbolised by schematic head views. S.A., spontaneous activity.

Fig. 1 displays the average responses of a population of 20 cells to different views of the head (Perrett et al., 1991; Oram and Perrett, 1992, 1994). The population is comprised of cells that were selective for the face view with a declining response function for views rotated from the face.

Fig. 1 gives the total number of spikes recorded in successive 5 ms time bins before and after presentation of different views of the head. Non-face control stimuli (not shown in Fig. 1) produced responses equivalent to the view rotated by 180° from best view. This peri-stimulus time histogram (PSTH) format is a conventional way for neurophysiologists to display the time course of neuronal activity. Most visually responsive neurones are spontaneously active even in the absence of visual stimuli. For cells in the temporal cortex this spontaneous activity occurs at a rate of about 10 spikes/s. During the presentation of an optimal stimulus the firing rate rises to approximately 50–250 spikes per second (the exact magnitude of response and spontaneous firing rate varies from cell to cell).

To compare activity of cells across a population all firing rates can be expressed as a % of the difference between the cell's maximum rate (set to 100%) and the cell's spontaneous activity (set to 0%). This normalisation allows each cell to make an equivalent contribution to the estimate of the population activity displayed in Fig. 1 (Oram and Perrett, 1992, 1994). Since results of analysis with and without normal-

isation were equivalent, data from different cells have been combined without normalisation in subsequent analysis (Figs. 2–4).

For the cells tested in Fig. 1, all visual stimuli were presented at time 0 ms after a 0.5 s signal tone and continued to be visible for the following one second. Cells in the temporal cortex commence stimulus related activity some 70–150 ms after the presentation of a visual stimulus (time 0 ms on the abscissa). It is apparent in the figure that the population activity rises after 70 ms for all stimuli; firing rate increases to peak levels over the following 100–300 ms and then declines to lower rates. In Fig. 1 it is evident that the overall activity declines as the head is rotated from the face view. Cell responses to clockwise and anticlockwise rotations of the head from the face view have been combined because the population responses to rotations through an equivalent angle (clockwise or anticlockwise) were statistically indistinguishable ($P > 0.5$ each comparison). The precise nature of the relationship between firing rate and view is, however, difficult to visualise because of the moment by moment fluctuations in spike frequency due to the fine time resolution of the data collection, the stochastic way neurones fire and because of the relatively limited sample of cells and test trials.

Neural activity in conventional PSTHs is illustrated as the average firing rate at each instant prior to and after stimulus onset. An alternative mode of display of the same data is one in which the spike activity that has occurred recently has a cumulative impact on the response assessed at any instant. Fig. 2 (upper) displays this cumulative difference between firing rate for each stimulus type and spontaneous activity. This cumulative analysis is like a statistical procedure in which a statistician does not want to base judgements on input occurring in a limited time window but prefers to assess activity over the entire time period from sample onset up until the present moment; this includes the current activity level and any evidence of activity that has occurred previously. The cumulative neural response, in this analogy, becomes equivalent to the level of certainty or weight of evidence for the presence of the cells' preferred stimulus in the visual scene.

Such evidence accumulation is biologically plausible in that it is equivalent to temporal summation by neurones receiving multiple synaptic inputs at different times. The postsynaptic membrane acts like a capacitor; the depolarisation caused by input excitatory postsynaptic potentials therefore accumulates over time. Fig. 2 (upper) would be equivalent to the membrane potential of a cell receiving the input activity displayed in Fig. 1. For simplicity, it is assumed here that the cells have a long memory (>1.0 s) and that membranes act like perfect capacitors. In reality membranes act like leaky capacitors and lose charge slowly over time. The time constant for this loss varies though for pyramidal cells the effect of synaptic activity at the site of spike generation can persist for 200 ms (Stratford et al., 1989). For clarity, cumulative analysis is presented in the following sections without modelling information loss. Analysis incorporating a rate of 'forgetting' caused by membranes discharging with an appropriate time constant, however, produces qualitatively similar results.

In Fig. 2 (upper), prior to stimulus presentation, the firing rate fluctuates around the level of spontaneous activity so there is on average, zero cumulative difference

from spontaneous activity. When responses to stimuli commence, the cumulative response curves show a positive deflection. The evidence for each view being present in the image rises over time and this rise is fastest for the face view. The rate of accumulation of response decreases in proportion to the angular rotation of the head from the face. This difference occurs despite the onset of responses to the different views occurring at the same latency (see also Oram and Perrett, 1992).

Neural populations upstream (i.e. those receiving output from populations equivalent to those whose activity is displayed) will begin their own responses when the
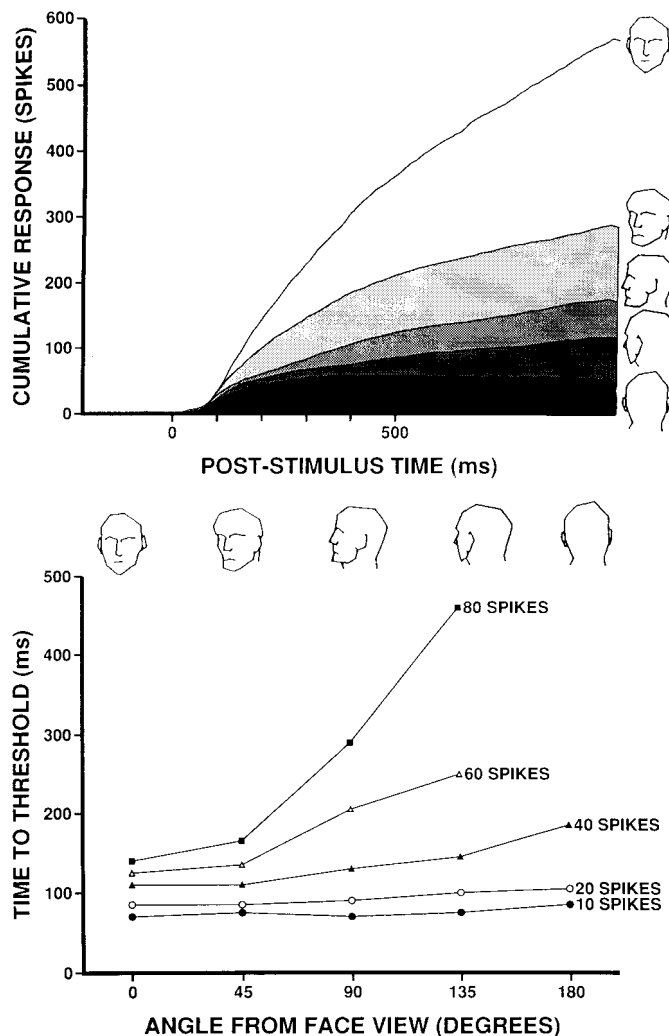
Fig. 2. Effect of rotation from face view on the time course of cell responses. Upper: the data from Fig. 1 plotted as cumulative neuronal activity (spikes) above spontaneous activity. Lower: the time for cumulative responses to exceed different threshold levels.

cumulated input activity exceeds a certain level above background. This is equivalent to input excitatory post-synaptic potentials exceeding the threshold voltage for action potential generation in the post-synaptic elements. Responses upstream will be triggered fastest for the face view, and progressively slower for views rotated in depth 45, 90, 135, 180° away form the face.

Another way of looking at the situation is to think of the differences between activity in cell populations. If two types of object are to be discriminated (e.g. face 1 vs. face 2 or view 1 vs. view 2 of the same object) then a decision can only be reached when the cumulative difference between the activity in two populations exceeds a threshold. At the cellular level, this may correspond to excitatory post-synaptic inputs from one population and inhibitory post-synaptic inputs from a second population. Note for such a comparison speed of output or decision depends not on the absolute rate of firing but on the difference between activity rates. Again if the difference in activity accumulates rapidly then reaction can be quick. In general, detecting a weak signal in noise takes longer detecting a strong signal imbedded in the same noise.

With the assumption that learning about an object from one view leads to cells optimally tuned to that view, then these results can be used to account for the impact of view on recognition. If a previously unfamiliar object is experienced extensively at a 'training' view, then it is assumed here that the visual appearance of the object will be 'stored in memory', in the sense that the experience will establish neural populations that are selective for the feature configuration of the object at the training view, orientation and image size. In subsequent recognition testing, the order in which individual test views produce a supra-threshold level of accumulated activity will depend on the angle of rotation between test and training views. This level will be exceeded fastest when the test view exactly matches the trained view and progressively larger angles of rotations will exceed the threshold criterion at longer and longer intervals after stimulus presentation.

While data are presented here for cells responsive to the sight of the head, view tuning has been found for other classes of cell in the temporal cortex responsive to other types of object. Logothetis et al. (1994) trained monkeys to discriminate 3-D wire frame and amoeboid objects similar to the stimuli used in psychophysical studies of humans (Bülthoff and Edelman, 1992; Edelman and Bülthoff, 1992; Rock and DiVita, 1987). Like humans, the monkeys displayed view sensitivity at the behavioural level with efficiency of recognition declining as the test stimuli were rotated in perspective away from the trained views (Logothetis et al., 1994). Neurophysiological studies of the trained monkey subjects revealed cells in the temporal cortex that were tuned to particular familiar wire frame stimuli (Logothetis et al., 1995). These cells displayed view sensitivity comparable to those selective for the sight of the head and body. Thus the results presented in Figs. 1 and 2 will apply equally to wire frame or amoeboid objects and presumably most other classes of 3-D object.

The notion that the cumulative response rate must exceed a threshold level is equivalent to a test of statistical significance or a decision criterion being applied at a cognitive level such that behavioural output is made only when sufficient sensory

evidence has accumulated. A statistical or decision criterion can be set at various levels of evidence. The account provides no indication of what that level should be but it should be apparent that the results are qualitatively similar over a range of thresholds. This is confirmed in Fig. 2 lower which shows that the time for cumulative responses to different views to exceed different threshold levels. Over a range of arbitrarily defined threshold values (between 10 and 80 spike above spontaneous activity), the time to threshold increases monotonically and approximately linearly as the view is rotated from the best view. Highly rotated views (e.g. 180°) do not evoke sufficient response to exceed thresholds greater than 40 spikes above spontaneous activity. The ordering effects may well be qualitatively similar for very low thresholds (<10 spikes) but limitation in the sample size and time resolution of sampling (5 ms time bins) prevent the rank ordering from being apparent for small angles of rotation. Indeed, when responses are calculated for cells with the same response latency, then the rank order of response to views is the same during the first 5 ms of response as it is several hundred milliseconds after response onset (Oram and Perrett, 1992). This analysis implies that the rank ordering of effectiveness is not dependent on integration over large time scales but is present at response onset.

At the stage of processing studied here, responses to unusual views exhibit slower rise times but do not have longer latencies compared to typical views. At subsequent stages of processing, differences between typical and unusual views may amplify and slower rise times for unusual views will translate into increased response latencies. Cells receiving outputs from the visual representations in temporal cortex will act as described, pooling information (input excitatory post-synaptic potentials) over time until their threshold level for firing is reached. This will be slower for weak inputs from the temporal cortex. By the time stages controlling behavioural output are reached, neuronal responses (related to muscle contraction and limb motion) will commence later for unusual views than for typical views since differences in the latency of muscle operation will underlie RT differences.

The nature of generalisation (and the function relating RTs to view) will depend on the recognition task and the view tuning functions of cell responses. These functions could be steeper or more shallow depending on the degree of familiarity with the object class and the discriminability of pattern features that differentiate exemplars within the class. Cells tuned to wire frame objects show tighter view tuning functions (Logothetis et al., 1995); behavioural performance in generalising recognition from one trained view of such objects also shows a steeper decline with perspective rotation (Logothetis et al., 1994).

Note that there is no need to postulate that an image of a novel view needs transforming (Edelman, 1997), normalising or mentally rotating to establish alignment with the training view. The increase in recognition time arises simply from the view tuning of neurones whose selectivity has been established from a training experience with an object. This neuronal population will be the most appropriate for identifying the trained object. By listening to the activity of this cell population, evidence for the presence of the trained object in the image will accrue with a rate proportional to the similarity between test and training views.

Recognition depends only on how well the input image falls within the tolerance of neural representations of familiar objects. The speed of classification of an unfamiliar exemplar of a familiar class of objects (e.g. recognising a new car model as a car) should depend only on the novel item's similarity or resemblance to familiar exemplars (Moses et al., 1996; Gauthier and Tarr, 1997a). While this prediction is straightforward, the point is that recognition does not need to rely on a continuous imagined transformation of the visual appearance of the novel exemplar to match that of familiar exemplars. To recognise a new car model we do not need to imagine the new car gradually undergoing a change in shape until it conforms to the shape of a familiar car model or a generalised prototype of all cars.

2.1. Object symmetry

Generalisation functions are likely to be affected by the view chosen for training. Fig. 1 displays the response functions for cells whose optimal view was symmetric (i.e. face) and for rotations about the axis of symmetry. The situation is different for cells maximally responsive to a profile or 1/2 profile view of the head. Such cells are often maximally responsive to the profile pointing to one side of the viewer (e.g. left profile) but not all show a single peak of view sensitivity, a substantial number (25% of cells tuned to profile images) show a bimodal tuning function with two peaks of responsiveness, one for each of two mirror symmetric views (Perrett et al., 1991). In general terms this cellular sensitivity means that the behavioural performance after training on an asymmetric view (of a bilaterally symmetric object) will not be a monotonic function of the angle of perspective rotation between training and test. As the object is rotated away from the training view generalisation will first decline but then performance may show a subsidiary peak as the object is rotated to a test view that is the mirror image of the training view. Such effects have been seen at the behavioural level (Logothetis et al., 1994; Hill et al., 1997). Indeed humans and monkeys find it relatively hard to learn to discriminate between mirror symmetric patterns and damage to temporal cortex (removing neurones generalising across mirror reflection) can improve discrimination performance between mirror symmetric patterns (Holmes and Gross, 1984a).

3. Orientation

Studies of temporal cortex reveal that the majority of cells selectively responsive to faces are orientation-sensitive (Tanaka et al., 1991; Wachsmuth and Perrett, 1997; Wachsmuth et al., in preparation) as are the vast majority of cells sensitive to simpler features (Tanaka et al., 1991). Cell sensitivity to head orientation appears to show very similar tuning functions to the tuning functions for head view (Perrett et al., 1991; Wachsmuth et al., submitted). Changes in view or in orientation of approximately 60° from optimal conditions reduce cell responses by half. Particular cells in the superior temporal sulcus do generalise across all orientations and remain selectively responsive for face patterns throughout 360° of rotation in the picture plane

(Perrett et al., 1982, 1984, 1985, 1988b). Such orientation invariant cells, however, constitute a minority of those responsive to the head and body (16%, 4/25 of cells tested, Wachsmuth and Perrett, 1997; Wachsmuth et al., submitted).

The similar tuning function for orientation and view means that the above description of the neural responses to unusual views will apply to unusual orientations as well. For orientation, however, the physiological data provide an additional direct link between recognition and neural processing of unusual viewing conditions. More cells (15/21) are found tuned to the upright orientation than tuned to non-upright orientations (Wachsmuth and Perrett, 1997; Wachsmuth et al., submitted; Tanaka et al., 1991). This bias presumably arises because upright orientations are experienced more often than non-upright orientations.

The consequence of greater numbers of cells responsive to upright faces is evident in Fig. 3. This figure plots the activity of cells responsive to the sight of the head and body that were recorded in the superior temporal sulcus (STS) (Wachsmuth et al., 1994; Wachsmuth et al., submitted). Cells were included in the sample independent of their orientation sensitivity. Fig. 3 represents a limited estimate of the relative activity in the STS evoked by the sight of the head and body in different orientations. Fig. 3 displays how the cumulative response of this population varies with orientation. It is not surprising that the response to the upright stimuli is largest as this simply reflects the greater number of cells responsive to the normal orientation. The rate of accumulation of response declines as the stimuli are rotated from upright. This occurs despite the fact that the responses to different orientations commence at approximately the same latency. The lower part of Fig. 3 shows the time for the cumulative responses to exceed different threshold levels. Over a range of levels (>15 spikes), the time to exceed threshold increases as a function of angle of rotation from upright.

The curves plotted in Fig. 3 are an estimate of the population activity from a limited sample of cells. An analysis of a larger sample of cells tested with fewer orientations revealed equivalent results (the cumulative response to upright rising faster than horizontal and inverted orientations). Reduced numbers of cells responsive to inverted orientations have been revealed in recordings from 25 cells (Wachsmuth and Perrett, 1997) and in a previous study of 40 cells (Perrett et al., 1988a). Modelling based on the empirically defined orientation tuning function of the population (including its greater response to upright than inverted) and the distribution of cell response latencies, shows that the rank ordering of responses with rotation from upright to inverted is independent of the number of cells sampled and occurs within 5 ms of population response onset (Oram et al., submitted).

3.1. Behavioural effects of orientation

Most experience of faces or any mono-oriented object (Rock, 1973; Yin, 1969, 1970; Carey, 1992) occurs in the gravitationally upright orientation. As a result of this preferential experience it is suggested that more cells in the temporal cortex become selectively tuned for the upright orientation of these objects. This numerical bias in cell responsiveness to the upright orientation is sufficient to account for the
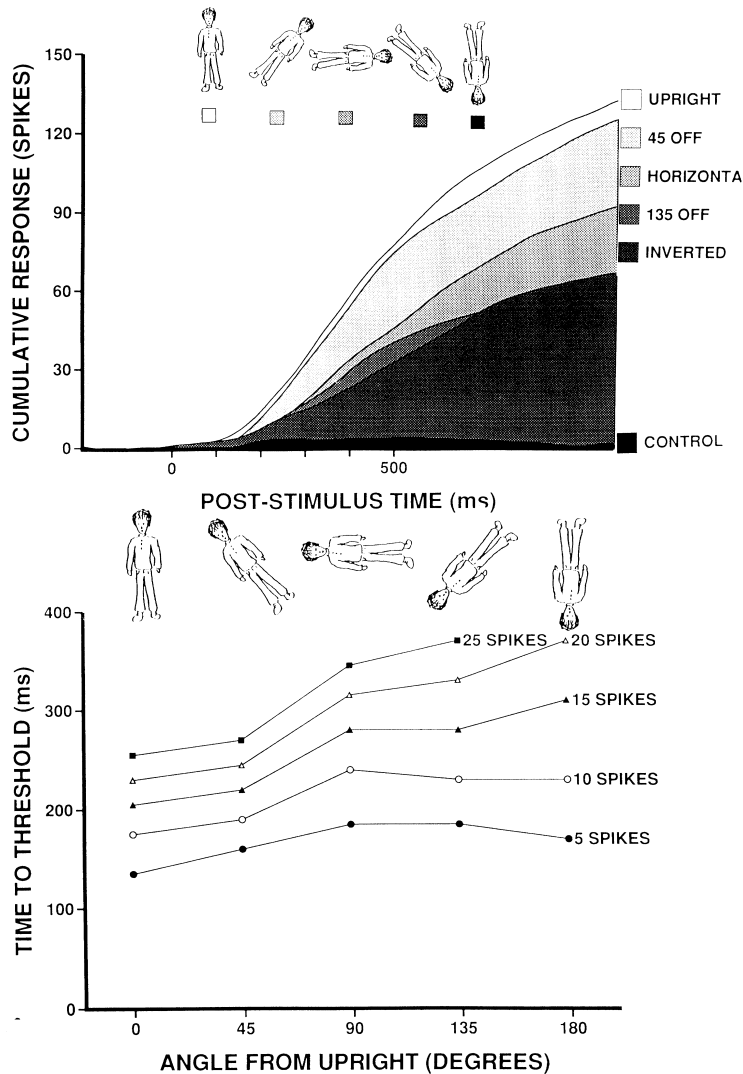
Fig. 3. Effect of orientation on the time course of cell responses. Upper: the cumulative difference in response from spontaneous activity of 11 cells selectively activated by the sight of the head and body (Wachsmuth et al., 1994; submitted). The 11 cells were selected on the basis that they had each been tested with eight face orientations and responded to one or more orientation at rates above control stimuli and spontaneous activity. The cells were chosen independent of their orientation tuning (most responded best to upright orientations but some responded best to non-upright orientations). Data are plotted for five stimulus orientations expressed as the angle from upright and control stimuli. Lower: the time for cumulative responses to exceed different threshold levels. For the horizontal, 45 and 135° orientations, responses to stimuli rotated clockwise and anticlockwise from upright have been pooled. Conventions as for Fig. 2.

greater ease of recognition of upright objects and the slower, less accurate recognition of more unusual orientations.

There are numerous studies of the effect of orientation change on recognition. The vast majority of studies report a reduced efficiency (increased reaction times and decreased accuracy) for recognising rotated images (Yin, 1969; Rock, 1973; Holmes and Gross, 1984b; Weiskrantz and Saunders, 1984; Jolicœur, 1985, 1990; Larsen, 1985; Corballis, 1988; Perrett et al., 1988b; Flin and Dziurawiec, 1989; Tarr and Pinker, 1989, 1990, 1991; Carey, 1992; Bülthoff and Edelman, 1992; Edelman and Bülthoff, 1992; McMullen and Jolicœur, 1992; Phelps and Roberts, 1994; for review see Wachsmuth and Perrett, 1997; Tarr, 1995).

Under special conditions it is possible that rotated images are processed with equal efficiency to upright images. One would not expect rotation to affect performance in tasks requiring only the detection of simple features that can differentiate target from non-targets at any orientation (e.g. colour or textural cues). Even without such obvious cues there are face recognition tasks which do not require coding of a face configuration per se but require only the detection of the presence of a feature (e.g. the presence of hair or the presence of a circular iris) which will be equally salient at any image orientation (Perrett et al., 1988b). The orientation dependence of recognition performance in object classification tasks also depends on similarity between stimuli (Edelman, 1995). Entry level or superordinate classification where subjects allocate the class name (car, face, chair, etc.) are situations where orientation (and view) effects are less likely to occur (Tarr, 1995) because they do not require discrimination of feature configuration. Superordinate tasks requiring face/non-face classification which include judgements of feature configuration do show sensitivity to stimulus orientation (e.g. Perrett et al., 1988b). The cells described here could participate in such tasks since they are sensitive to configuration, exhibiting reduced responses to displays of jumbled face or body parts (Perrett et al., 1982, 1988b).

Tasks which require differentiation of exemplars within the same class (subordinate classification) are ones in which the effects of orientation are more prevalent (for discussion see Tarr, 1995). The results described here are also appropriate for subordinate-class discriminations because the same view and orientation tuning is apparent for cells which show high degrees of selectivity amongst exemplars within a class, discriminating between different faces (Perrett et al., 1984, 1988b, 1991; Hasselmo et al., 1989; Young and Yamane, 1992) or different wire frame objects (Logothetis et al., 1995).

Even within these subordinate-level tasks the slope relating RT to stimulus orientation from behavioural studies can show various functions (linear, non-linear) depending on the paradigm and depending on the subjects' experience. For the cellular account described here, the nature of the RT-orientation function is determined only by the relative responsiveness of the cell population to different orientations. While this should parallel experience exceptions may arise due to the tuning of cells (e.g. cells tuned to horizontal eyes in an upright face may respond to eyes in an inverted face). Non-linear slopes are more of a problem for mental transformation models. A non-linear slope implies that the speed of transformation (mental rota-

tion) depends on the object orientation. This does not appear to explain orientation-dependent processing but rather restates the phenomenon.

4. Size

Studies of cells responsive to the sight of the face and the body indicate that the majority are sensitive to image size (Wachsmuth and Perrett, 1997; Wachsmuth et al., submitted). While the cells generalise over a range of image sizes (Perrett et al., 1982, 1984), when unusually small images are presented the population response declines. Using a constant projection distance of 4 meters (the greatest distance our subjects see human faces), we have found that cells respond maximally to life-sized projected images of humans and have monotonically declining responses to 3/4, 1/2 and 1/4 sized images. Analysis, with the methods described above, reveals that the more unusual the image size, the smaller the population response and the longer the cumulative population response takes to exceed threshold levels. By similar arguments to those presented for view and orientation, it is, therefore, not necessary to postulate 'mental transformations of size' to account for the extra time taken to recognise objects seen at unusual image sizes (Shepard and Metzler, 1971; Bundesen and Larsen, 1975; Besner, 1983; Larsen, 1985).

5. Recognition from the whole or part of an object

Neuropsychological tests of object recognition from unusual views often include views of objects in which particular parts are occluded from sight or placed in less salient positions (Warrington and James, 1986; Humphreys and Riddoch, 1984; Warrington and Taylor, 1973). The analysis of the generalisation of temporal cortex neural responses across perspective view and orientation described so far extends easily to situations where object parts are occluded from sight.

Analysis of cells responsive to faces in the temporal cortex has revealed a spectrum of sensitivity to facial parts. Some cells respond only to one facial region (e.g. eyes). Most cells, however, exhibit independent tuning to several facial parts (e.g. eyes or mouth) and some cells respond only when multiple features are presented simultaneously (Perrett et al., 1982; Oram and Perrett, 1994). The face is just one component of a larger and more complex object (the body) and recent studies have analysed the contribution of body parts for cells in the STS that respond to the sight of the whole body (Wachsmuth et al., 1994; Oram et al., submitted).

Fig. 4 plots the responses of cells in the STS studied previously for sensitivity to two parts of the body (the head and the rest of the body; Wachsmuth et al., 1994). This study again revealed a spectrum of cell sensitivity to the two parts. Thirty-eight percent of cells responded to only one part. These included 28% which responded to the head presented in isolation but not to the rest of the body when the head was occluded from sight and 10% responded conversely to the body but not to the head. These cells continued to respond to images of the whole body that contained the

effective and ineffective parts. A more common pattern of activity amongst the cell population studied (43%) was for cells to respond to the head or to the rest of the body. A final group of cells (19%) responded only when both the head and body were simultaneously visible. Fig. 4 displays the responses of all the cells in this study that were responsive to the whole body. The population therefore includes those
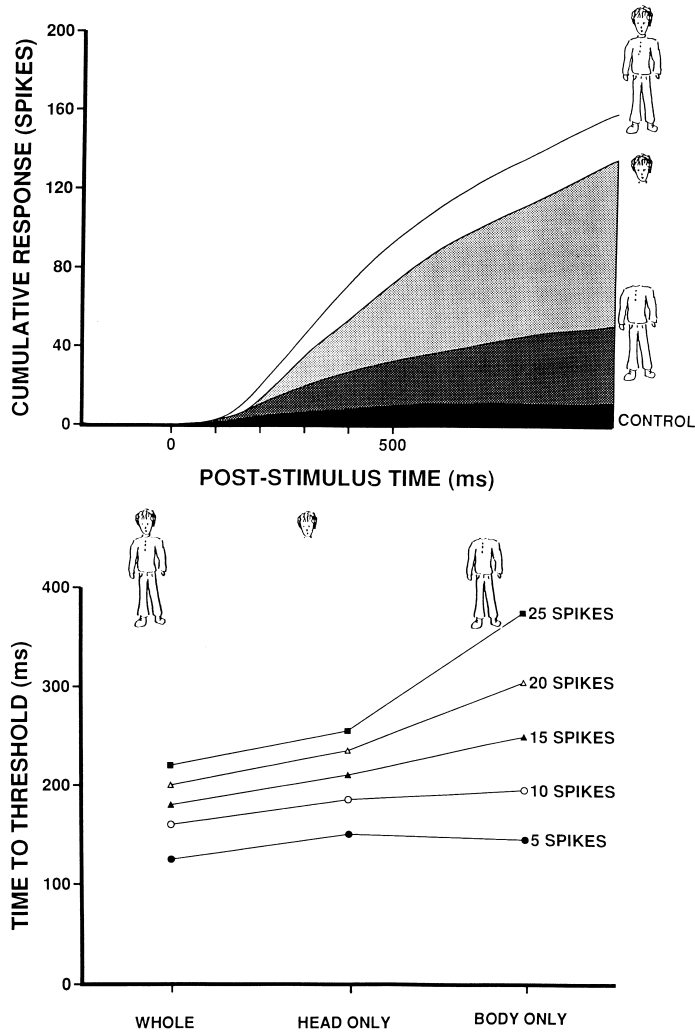

Fig. 4. Responses to an entire object and to object components. Data from a population of 35 cells responsive to the whole body (Wachsmuth et al., 1994). Cumulative response of the cell population activity is displayed for four types of stimulus: whole body (WHOLE), the head alone with the rest of the body occluded from sight (HEAD ONLY), the body alone with head occluded from sight (BODY ONLY) and objects other than faces and bodies (CONTROL). Lower: the time for cumulative responses to exceed different threshold levels. Conventions as for Fig. 2.

requiring the whole body before response, those responsive to the one component and those responsive independently to the head or the rest of the body.

From Fig. 4 it is evident that the population response to the head presented in isolation was greater than that to the rest of the body with the head covered. This reflects the greater number of cells responsive to the head alone (28%) than to the body alone (10%). It is also evident in the figure that the response to the whole body accumulates faster than the response to either part presented in isolation. The greater response occurs for two reasons. First, a sub-population of 19% of cells responds only when the whole body is presented and, second, the sight of the whole body will activate both sub-populations tuned to one part (i.e. both the 28% of cells responsive to the head alone and the 10% responsive to the body alone). (More generally, if one considers an object made up of two parts (A and B) and that 10% of the cells respond only to part A and 10% respond only to part B, then presenting the whole object will activate twice as many cells as presenting just one part). For these reasons it is apparent that the total brain activity will always be greater in response to the visual image of an entire object than to the image of one component of the object presented in isolation.

One can speculate from the data illustrated in Fig. 4 that the rate of accumulation of neural evidence for the parts and whole will have an impact at the behavioural level. Cumulative responses to the whole object will exceed a threshold level faster than responses to parts (Fig. 4, lower). Reaction times to detect an entire object (e.g. a human body) will be quicker for images containing the whole object than images displaying just one part of the object. The finding of Biederman (1987), that the more components visible in an object the quicker the time to name the object, follows this pattern. Similarly it is expected that recognition of famous individuals would be superior with whole face photographs than with photographs of isolated facial regions (eyes).

This relationship of speed of detection of parts and the whole may not be surprising but it suggests caution in inferring the order of stages of processing from differences in behavioural reaction times. Many theories of object recognition propose separate stages of processing for the definition of object parts and for access to the overall structure of the whole object (e.g. Biederman, 1987). Some theories of object recognition propose that the overall configuration of an object is processed first before the details of the object's parts (e.g. Baker Cave and Kosslyn, 1993). Quicker reaction times for naming whole objects as compared to object parts could be taken as evidence consistent with models in which the order of processing is 'whole' before 'parts'. Such inferences from the reaction time ordering as to the ordering of processing could be misleading. The data presented in Fig. 4 shows that the rate of accumulation of neuronal activity can be faster for the entire object even when cell responses to the entire object and to its separate parts begin at approximately the same latency.

It is interesting that a progressive ordering of reaction times for different types of judgements is often taken as evidence for a progressive ordering of processing stages in cognitive psychological models. For example, the finding that retrieving the name of a face takes longer than retrieving other forms of personal information (e.g.

profession) has been taken as supporting a serial processing model in which processes underlying naming are only initiated after accessing other person identity information (Bruce and Young, 1986). Similarly subordinate judgements between exemplars of the same object class (e.g. face identity) may take longer than super-ordinate judgements of object class (that a face is human). Smaller numbers of cells tuned to specific identities (Perrett et al., 1984, 1989; Young and Yamane, 1992) than tuned to the general structure of faces could produce this ordering of reaction times, even if cells responsive to general and specific facial characteristics were activated at the same time.

In this section, the neural responses to part and whole objects provide straight-forward predictions for the speed of behavioural responses to part and whole objects. What is more interesting is the realisation that neural responses can accumulate at different rates when they have the same onset in time. Thus the relationship between behavioural reaction times and underlying neural stages of processing is not straightforward. One cannot assume that because two types of judgement take different lengths of time the judgements necessarily reflect two different stages of processing activated in serial order. Moreover, there is again no need to postulate an extra transformational process, such as 'mental completion' (extrapolating the presence of occluded parts), to explain slower recognition of parts than whole objects.

The terms word, object or face advantage (or superiority effect) have been used to describe the greater efficiency in recognition of normal configurations of words, objects or faces as compared to jumbled configurations (Purcell and Stewart, 1988). The difference between neural responses to whole objects and object parts can provide a basis for understanding the advantages conferred by normal configurations.

Cells responsive to one facial feature presented in isolation (and receiving no inputs from cells sensitive to other face parts) will respond equivalently to that feature in a jumbled face or a face with a correct configuration of features. Cells sensitive to multiple features in a face will respond preferentially to the normal configuration since jumbled configurations disrupt some local feature information (Perrett et al., 1982, 1988b; Perrett and Oram, submitted). For example moving the nose up in the face (a configuration change) will increase the gap between nose and mouth (a change in the local upper lip feature). Sensitivity to combined visual features imbues cells with selectivity for the configuration of face parts (Perrett et al., 1984; Yamane et al., 1988; Perrett and Oram, submitted). A normal face will excite more cells than a jumbled face; the normal configuration excites both cells responsive to whole faces and cells responsive to individual face parts but the jumbled face activates only the latter cell type. Therefore, the total response in the temporal cortex to a normal configuration of a face will be greater than the response to a jumbled face or an isolated part (Rolls et al., 1994). This is evident in local evoked potentials to faces in the temporal cortex of humans (Allison et al., 1994). From this one can predict that, provided subjects employ a search strategy utilising configuration, the latency to detect a normal face will be shorter than that required to detect a jumble (e.g. Perrett et al., 1988b). Likewise, if the visual patterns

are degraded (e.g. by short exposure and backward masking), one can predict that the normal face pattern will become detectable before the jumbled configuration (Purcell and Stewart, 1988). The differences in recognition or perception times can be attributed to different numbers of cells activated at the same time by different stimuli.

6. Discussion

6.1. Recognition of unusual views following brain damage

One question which is left unanswered by the account here is why brain damage in the right parietal area should disrupt object recognition at all (Warrington and James, 1986; Warrington and Taylor, 1973). If viewer-centred representations in the ventral temporal cortex account for object recognition from typical and unusual views, then damage to ventral but not dorsal systems should produce impairment in tests of unusual views.

The impact of damage to the dorsal system on recognition may arise indirectly because any complex mental operation must have contributions from many brain systems. Outputs from temporal cortex may well provide evidence that a particular object is present in a particular orientation but if the amount of evidence is small (because the image is degraded, the view unfamiliar, or because early visual processing is defective), then the observer may need to check and form a considered opinion before reacting. If so, an entire strategy for gathering more visual information needs to be executed. For example, on the basis of the first fixation there may be evidence for a horizontal rabbit ear shape or a bird beak shape. Given the orientation and size of the object part and the current position of the eyes, a prediction can be made that moving the eyes to a new location should bring rabbit paws or a bird tail into central vision and window of attention. This prediction needs translating into action and the visual image resulting from the second fixation will need analysing. This cycle of operations might just be the first in a series of evaluations extending over several predictions and fixations. A single inspection saccade may fail to result in confirming evidence and a new test location would need to be chosen, alternatively the image interpretation and object hypothesis (Gregory, 1980) may need modifying after the first inspection saccade. As objects are presented at increasingly unusual orientations one expects the information processing requirements to increase. This could be reflected in number of inspections or their duration. Indeed, in the Shepard and Metzler matching paradigm, Carpenter and Just (1978) found that the time spent looking at each shape increased as the angle of rotation between the stimuli increased.

Visual search strategies of the type just described require planning (perhaps involving frontal cortex), awareness of position of eye fixation, decisions as to how to direct eye movements to selected locations (involving a whole variety of systems including parietal cortex, frontal eye fields and superior colliculus) in addition to information about the nature of visual features and object parts from ventral

occipito-temporal cortex. One could add to the list of requirements the need for a short term memory to remember the object hypothesis from preceding fixations and the need for semantic associations (or object-centred representations) to predict the relationships between object parts and features and the capacity to translate these spatial relations into image co-ordinates. Damage in parietal or frontal regions could produce deficits in recognising unusual views for a variety of reasons, including difficulties in implementing an organised visual search.

Normal subjects may make decisions in speeded reaction tasks after a single pass through the ventral cortical system that engages the relevant detailed viewer-centred description of objects. Recognition under degraded viewing conditions or on the first trial with an object experienced in a completely novel view may involve many brain systems even for normal subjects with checks being made to assess the viability of the current interpretation. Damage, particularly in parietal systems, would impact on recognition not so much because this system contains detailed viewer-centred descriptions of objects but more because damage here disrupts search strategies, translation of spatial relations to action plans, or control of fixation and attention. All these processes are essential for gathering sufficient evidence for recognition in difficult circumstances.

This explanation might account for some aspects of the original mental rotation tasks of Shepard and Metzler (1971). The larger the difference between stimuli to be matched the more difficult and time consuming the checking processes. It should be noted, however, that the accumulation of cell responses described here does not depend on sequences of eye movements made after an image is presented. The effects on cell responses are seen before the onset of saccades and in situations in which the eyes remain on fixation points (e.g. Oram and Perrett, 1992; Logothetis et al., 1995).

6.2. The limits of generalisation

One question which arises concerns how generalisation is possible for totally novel objects or for a familiar object experienced in a totally novel view/orientation. If selective viewer-centred neural codes in temporal cortex are formed only after experience, then they cannot be used for generalisation on the initial trial with a totally novel object.

After becoming familiar with an object in just one specific circumstance (one view, orientation and retinal size), one can ask how does recognition occur after a dramatic rotation in depth or change in viewing circumstances? There are several mechanisms which might contribute to generalisation of recognition in this circumstance.

First, view sensitive cells, that we presume are formed through experience with an object from one view, may well aid recognition even after a dramatic change in viewing circumstance. The success of recognition depends on where the novel test view falls relative to the view, orientation and size tolerance functions of the relevant cells. If the test view falls within the tuning functions, then the speed of recognition will be a function of the difference between novel and familiar views.

When the novel view is dramatically different but still within the tuning function, the reaction time will be dramatically increased, and the recognition accuracy dramatically reduced (but still above chance). If the test view falls completely outside the tolerance, then the cells we describe will be irrelevant.

It should be realised that the pattern sensitive representations described here are tolerant to quite an extensive range of viewing conditions. Thus although the cells depicted in Fig. 1 are selective for view, their response remains above baseline even after 135° of rotation (baseline response to controls was approximately equal to the 180° rotated view). Generalisation on the initial trial after 135° of rotation should be possible for some familiar objects, although the rate of recognition would be very slow because of the slow rate of accumulation of evidence in the relevant cell populations. As noted the tuning functions of cells selective for a complex wire frame (trained in a specific view) are narrow. Recognition may fail when the test view is rotated 40° from a training view.

Second, some authors have argued that generalisations on initial and subsequent trials with a new view employ different mechanisms (Jolicœur, 1985, 1990); the account here may apply more to trials subsequent to the initial trial. It is relevant here to ask, 'what experience with an object from one view is necessary to derive selective neural populations in temporal cortex?' The answer will depend on how much experience the observer has already had with a particular class of stimuli. Evidence shows that cells in temporal cortex can show very quick adaptations to new patterns of input that are versions of familiar objects and faces (Dolan et al., 1997). Cells already responsive to faces can show modifications in selectivity to new faces within a relatively short time course of 1–2 trials (Rolls et al., 1989). Of course it may take far longer for temporal cortex cell populations to become selectively tuned to subtleties of new object classes (Logothetis et al., 1994, 1995; Gauthier and Tarr, 1997b). Even when the relevant cells are wired up and 'ready to go' in temporal cortex, one can assume that synaptic connections allowing such cells to have influence over neural control systems for arbitrary motor responses (such as those chosen by experimental psychologists as an index of recognition, e.g. 'press right button if the stimulus is familiar') take time and practice to optimise.

Third, the significance of an object (and semantic codes more generally) may be accessed through multiple routes, they may be reached not only from the high level view-centred pattern-selective cells described here but also from 'elaborate feature' sensitive cells in inferior temporal cortex or prestriate areas (Tanaka et al., 1991). These cells code simple characteristics that may be sufficient to differentiate objects on a coarse level. Some elaborate feature sensitive cells code features (e.g. brown and spiky, or spotted texture and round shaped) that are relatively unaffected by changes in orientation and view and size.

Recognition should be seen as a graded ability where performance depends on the similarity of the target to alternative distracters (non-targets). If the distracters are very dissimilar, then recognition is simple and can be made on the basis of simple view independent features (colour, texture, or idiosyncratic features visible from virtually any view).

6.3. Measures of recognition: detection, discrimination, and categorisation

To account for recognition of novel test views of an object, one needs to consider: (1) the angle between training and test views, (2) the tuning functions of cell populations that become adapted to the training view or views and (3) the task. After training on one view or orientation alone, then as the angle between training and test view increases so will the reaction time in detection tasks. This increase would continue up to the point where the training view falls outside the tuning function of cells selective for the training view. At this point reaction time would increase to infinity and recognition would fail.

For faces, view tuning at the cellular level is relatively broad: 60° of rotation is required to reduce cell responses to 1/2 maximal response (Perrett et al., 1991). After extensive training with paper clip and amoeboid figures the view tuning curves of cells in temporal cortex are narrower (Logothetis et al., 1995). Thus, one would expect generalisation across view to be more limited for paper clips than faces which is what is observed psychophysically (Edelman, 1995; Logothetis et al., 1994).

In the situation where two views of the same object have become equally familiar through training, we presume that two populations of cells would become tuned to the object, each population specialised for one view and each showing a tuning function. For recognition of the object class, activity in either population will provide evidence. One again predicts an increase in time for detection in proportion to the difference in angle between the test view and nearest training view. Unlike the mental rotation account, there is no need to define which training view is closest to the test view; recognition is successful when activity in either population exceeds a particular threshold.

If the two views have been associated with the same response or reward (i.e. the subject 'knows' that two views are known to belong to the same object), then cells higher up the processing stream will receive a convergent input from the two-view-specific populations. In this case, recognition of the object class will be successful when the summed activity of both view-specific populations exceeds a certain threshold. Such association would have the effect of smoothing out the reaction time dependence on view between the two trained views.

If a subject experiences two views of an object, the front view (0°) and the side or profile view rotated by 90°, experience would result in cells that become tuned to the features of the object. Amongst these cells there would be two populations of cells, tuned to the 0 and 90° views. If we assume that the view tuning is broad (equivalent to that observed for heads), seeing a novel 45° view (lying directly between the front view and side views) will activate both cell populations tuned to familiar front and side views. The novel view is not more detectable than training views because it will be sub-optimal for both front and side view-tuned cell populations. The combined activity in both populations to the new view will not exceed that evoked by one of the original training views.

It may help this description of generalisation to provide an analogy for colour vision. We have three cone mechanisms tuned to short (blue), middle (green) and long (red) wavelengths. The cones have wide spectral adsorption characteristics or

tuning functions. If we take the red and green cones to stand for two populations of view tuned cells, then the novel intermediate 45° test view is equivalent to light of mid-long (yellow) wavelengths which will excite both red and green cones. Yellow light is sub-optimal for red and green cones since it does not match the peak spectral sensitivity of either cone. The total cone output to intermediate yellow light will not exceed that produced separately by red or green light. Hence detection will not be superior at stimulus values that fall between the peak sensitivities of two populations of neuronal detectors.

In some circumstances, object and face discrimination performance can be superior for the half profile view (with the front and side vies at 45° to the line of subject's sight, Palmer et al., 1981; Harries et al., 1991; Perrett et al., 1994). In these tasks the most efficiently recognised views may well be intermediate to the views represented most numerously by visual cells in the subjects' nervous system. This must, however, remain conjecture because the subjects' perceptual history of views of familiar objects and faces is usually unknown (though see Perrett et al., 1994).

In tasks where subjects are required to perform same-difference discriminations, performance can depend on how activity in neural populations is modified by small changes in the stimuli. Activity change in a neural population is greatest when a stimulus falls on the slope of a tuning function, rather than close to the peak sensitivity.

Supra-threshold wavelength discrimination is most efficient in the yellow region of the spectrum, intermediate to wavelengths matched to the peak spectral sensitivities of green and red cones. This is because a small change in wavelength in the yellow region of the spectrum will modulate the responses of red and green cones more than an equivalent change at wavelengths close to the peak sensitivities of the cones in the red or green regions of the spectrum. Thus, even for elementary sensory processing, 'intermediate stimuli' can produce superior or inferior psychophysical performance compared to stimuli matched to peak detector sensitivity depending on the nature of the task (detection or discrimination) and how the neural codes are being read (detecting any modulation above background activity or discriminating changes in the level of population activity when activity is already above background).

In higher level visual processing, detecting stimulus change is more efficient (just noticeable differences are smaller) at the boundary between two categories than detecting change for stimuli at the centre of categories. Indeed superior discrimination between categories is the hallmark of categorical judgements (Young et al., 1997). Superior discrimination at category boundaries has also been found for discriminating facial expression (Etcoff and Magee, 1992; Calder et al., 1996), facial identity (Beale and Keil, 1995) and discriminating object class (F. Newall, personal communication). Categorical judgements are assumed to reflect the existence of specialised neural populations with peak tuning matched to the centre of categories (i.e. the underlying neural 'detectors' are tuned to the prototypical or canonical examples of categories). Stimuli falling at the boundary between two categories will therefore fall on the slopes of the tuning functions of neurones tuned to the category prototypes. Superior same-difference judgements (in successive or simul-

taneous discriminations) at category boundaries can be interpreted as resulting from the fact that stimulus change at such boundaries will induce the greatest change in neuronal activity amongst the cells tuned to the category prototypes.

Classification of stimuli as examples of one or other of two categories will suffer further if the stimuli are moved away from the category prototypes. This will be because the stimuli will fall further from the peak sensitivity of cells tuned to the category prototypes, and evidence for the presence of either category prototype will accumulate progressively more slowly. Object categorisation like object detection will therefore show costs rather than benefits as one moves towards stimuli intermediate to familiar categories (Young et al., 1997). Both the costs and benefits for recognising intermediate stimuli in different types of recognition tasks can be seen to arise from consideration of the response functions of detectors tuned to category prototypes (Young et al., 1997).

For cells tuned to the front view of the head (the face or 0°), a 10° rotation of the head starting with the half profile view (45–55°) will produce a greater change in population activity than the same rotation starting with the face view (0–10°). Predicting performance for same-different discrimination for head views is complicated by unknown perceptual history with faces in general and by the fact that same-different judgements could be performed with other strategies (e.g. judging degree of symmetry). Predictions can be made from above considerations for the recognition of an asymmetric novel object (e.g. the paper clip stimuli of Poggio and Edelman, 1990; Logothetis et al., 1994; or symmetric 'greeble' stimuli of Gauthier and Tarr, 1997b) trained extensively at two views separated by a rotation equal to twice the band width of view sensitivity function for one view (1/2 width at 1/2 height). With such training, a task requiring same-difference discrimination between view pairs should peak in efficiency for views intermediate between the two training views. By contrast, a task requiring detection of the object from any view, or a task requiring categorical labels 'Training view 1' vs. 'Training view 2', should show the highest cost function (lowest accuracy and longest reaction time) for intermediate views.

6.4. Population vector hypothesis

Georgopoulos et al. (1982) and Georgopoulos (1995) have used 'population vector analysis' to read the activity of cell populations in motor cortex. Such analysis has also been used to analyse cell responses in the visual system (Young and Yamane, 1992; Földiák, 1993; Oram et al., 1998). Under this analysis each cell 'votes' for its preferred direction with strength proportional to the activity evoked in a test situation. Georgopoulos et al. (1989) observed that the population vector for cells in motor cortex rotated over time, during the interval between the presentation of a visual cue and an arm movement response made 90° to the direction of the cue. At the beginning of the trial the direction of the population vector pointed towards the sensory cue and later pointed in the direction of the monkey's reaching movement. The gradual rotation of the population vector direction through intermediate directions was likened to mental rotation (see also Lurito et al., 1991). Whatever

neural mechanisms underlie this sensory-motor phenomenon in motor cortex, they need not necessarily relate to the recognition of familiar objects from unusual views and the processing in temporal cortex.

To study recognition of objects from unusual views, we have been recording activity in visual areas rather than in motor cortex since the subjects are not required to make movements. In our test situation objects were presented in a variety of views and orientation. When unusual views of familiar objects were presented, subjects in our test situation were not required to make any imagined rotations, though mental rotation accounts of recognition would suggest that such rotations take place whether they are consciously imagined or implemented without subjects awareness.

It is relevant to review the time course of activity in cells when an object is presented at some intermediate angle to the cells' preferred view. In this situation mental transformation models of recognition would predict that activity in such cells would commence at an intermediate time. Our findings show that cells tuned to views intermediate to the test stimulus are activated at intermediate rates, but not at intermediate times.

When an unusual orientation or view is presented our recordings show that cells tuned to a range of different views or orientations are activated. Each cell is activated in proportion to the angle between the test view and the cell's preferred view/ orientation. For an inverted stimulus orientation, cells tuned to inverted orientations are activated most, those tuned to upright are activated least and those tuned to intermediate orientations at activated at intermediate amounts. The same is true for a back view with cells tuned to the back activated most, those tuned to the front are activated least and those tuned to intermediate side views activated to intermediate amounts. Our analysis shows that cells tuned to different orientations and views begin firing at the same latency but with different rates. We find that the cells tuned at intermediate views or orientations are activated at intermediate rates but not at intermediate latencies; they begin activity at the same time as those tuned to the test view or orientation. The relative degree of activation amongst cells (tuned to the test view, intermediate view or opposite view) does not change with time since stimulus appearance.

We hypothesis that the speed of recognition of the object is determined by the time for activity in the entire population of differently tuned cells to exceed a given criteria. This time to threshold reflects the frequency of cells with different orientation and view tuning. For unusual test views the small numbers of cells tuned to these unusual views will fire at high rates while larger numbers of cells tuned to more typical views will fire at medium to low rates. By contrast, for frequently experienced test views, large numbers of cells tuned to the typical views will fire at high rates, and the rare cells tuned to unusual views will fire at low rates.

Consideration of population vector coding with the present data from temporal cortex also suggests that there is no mental rotation in the recognition of faces. The rank order of stimulus effectiveness is apparent at response onset and is maintained throughout test periods of 1 s duration (Oram and Perrett, 1992). Moreover, the tuning of cells is independent of their preferred view/orientation (Oram and Perrett, 1992). Therefore the population vector analysis applied to the entire population of

cells responsive to the face (from any view or orientation) points in the same direction from response onset to end of trial. If an inverted face image were presented, the population vector would indicate an inverted orientation throughout the trial period. While the direction of the population vector can be used to estimate stimulus identity (including head view) with limited accuracy (Sanger, 1996; Földiák, 1993; Oram et al., 1998), it does not predict changes in behavioural changes in recognition time from temporal cortical cell responses. The magnitude of the vector is related to the total population response and might be more appropriate for explanations of reaction times.

6.5. Evidence for mental transformations

The physiological explanation advanced here, accounts for the costs to object recognition (in time and accuracy) that are associated with change in view. It should be noted that the account does not address mechanisms underlying imagery. Imagery and imagined sensory-motor transformations may well have a role in a variety of tasks (e.g. subjects imagining moving their own hand to match an unusual view of a hand, Bonda et al., 1995; Parsons et al., 1995).

The cellular mechanisms within the ventral cortical stream that are described here are appropriate for the recognition of familiar objects encountered in unusual views. Other brain mechanisms may be involved in matching unfamiliar objects, in sensory-motor transformations or in the use of proprioceptive information to classify images of body parts (Bonda et al., 1995; Parsons et al., 1995). Indeed Cohen et al. (1996) note that classical mental rotation tasks do not engage the ventral visual stream but rather areas of the dorsal cortical pathway and frontal lobes.

The phenomena most often cited as evidence for a transformational process underling recognition of rotated stimuli include experiments in which prior information about stimulus orientation can null the extra time required to recognise rotated stimuli (e.g. Cooper and Shepard, 1973). These authors propose that 'the subject carries out a purely mental rotation of.... a mental image of the anticipated stimulus'. To account for these effects without invoking transformations, we propose that prior information about the upcoming stimulus (priming) acts on the same representation (i.e. cell populations) as the visual stimulus itself. We assume that the efficiency of priming is proportional to the degree to which the stimulus is therefore represented in the nervous system. The effectiveness of priming will parallel the effectiveness of visual activation during normal unprimed recognition. That is, priming will be less effective for more unusual views. Hence an increasing amount of time will be needed for prior information to prime increasingly rotated stimuli.

This explanation is post-hoc, as we do not have data on how cells are influenced by prior information during tasks equivalent to the Shepard and Metzler tasks. Some insight is available from 'delayed matching to sample' paradigms. With attention directed to stimuli within a cell's receptive field, then when the match stimulus is the same as the sample stimulus, cell responses are augmented during the match phase compared to when the sample and match differ. The augmentation is proportional to the cell's response to the stimulus when unprimed (McAdams and Maunsell, 1996).

In other words, the enhancement is greatest for cells optimally tuned to the stimulus in question. In this situation, priming due to prior information about the stimulus (seeing the sample) acts on the same cell population that represents the visual appearance of the stimulus itself. Furthermore, the magnitude of the priming is proportional to the strength of the stimulus representation in the nervous system.

A further effect interpreted as supporting internal transformation processes includes the impression that alternating presentations of differently oriented stimuli give a compelling demonstration of continuous rotation. Indeed, subjects can place a marker to indicate where the image passes at successive moments during the inter-polated rotation. This phenomenon need not require the cellular explanation we have provided relating RTs to neural population responses. The perceptual impression of a smooth trajectory may be more related to the illusion of apparent motion. Early work by Wertheimer (1912) showed that apparent motion of objects can extend over large angles of rotation. Furthermore Kolers (1972) showed that the transition between start and end shapes during apparent motion is perceived to be a smooth transformation even when shape changes occur.

The perceptual history of apparent motion can influence the perceived trajectory. For example seeing a bar rotating back and forward around one end with rotations occurring over a gradually increasing angle, starting with an acute angle but slowly increasing to an obtuse angle, can induce apparent rotation through 180° rather than motion through the shorter path with than an angle less than 180°. In mental rotation tasks too when subjects are given enough biasing they appear to rotate stimuli 'the long way around'.

Mirror image stimuli are recognised according to reaction time (RT) functions that are predicted by the assumption that subjects mentally rotate stimuli along the most efficient axis. For some 2-D stimuli this would appear to include a flip in depth, since some 2-D stimuli rotated through 180° can produce faster RTs than stimuli rotated 90° (Tarr and Pinker, 1989; Tarr, 1995). Such phenomena appear to lie outside the physiological explanation advanced here. The unusually fast reaction times when matching mirror equivalent stimuli may arise in part from equivalence in orientation of some features across 180° rotation or mirror reflection. In general, the resolution of mechanism will depend on obtaining comparable physiological and psychological measures in the tasks requiring processing of the special stimulus classes and orientations that produce anomalous results.

The mental rotation explanation of the paths taken, including flips in depth is, however, not without problems. To rotate mentally the image (or its representation) along the optimal path, it is necessary to know the start and end points of the path. A difficulty with mental rotation accounts of object recognition concerns recognising which way to rotate the object when encountered in an unusual view. Information about an object might be pre-processed to establish which way to rotate it to bring it upright, prior to a matching process for identification. Admittedly, there are some cues from distinctive object features which might point to the relation of an object's axis to gravity and hence guide rotation (Ullman, 1989) but if one can recognise which way an image needs to be rotated, then to some extent the process of rotation is redundant since recognition has already been accomplished (Corballis, 1988).

In the account provided here, there is no need to pre-process an object to realise its likely orientation so that a mental rotation can take place subsequently. The system just accepts evidence in proportion to the activity of the relevant visually responsive cells in temporal cortex and elsewhere. Cells responsive to the face in temporal cortex will have outputs to brain systems capable of accessing semantic associations for faces (including verbal labels in the case of humans) and activating appropriate behavioural outputs. These associations will be more likely to be retrieved when face stimuli are presented in a frontal view, upright orientation and at a normal size, because under these conditions there will be more activity from cells in the temporal cortex already tuned to faces. Mental rotation has been invoked as an additional process to cope with unusual views. By contrast it is argued here that both typical and unusual views can be recognised using viewer-centred cells which operate in the same manner; unusual views take longer simply because there is less machinery dedicated to their processing.

Exposure to one particular view of an object increases the number of cells tuned to that view (Logothetis et al., 1995). This will decrease the time taken for the cumulative population response to exceed a threshold for that training view. This accounts for the selective improvement in recognition seen at the trained view or orientation (Logothetis et al., 1994; Perrett et al., 1988a; Tarr, 1995). Experience at all views during a task will increase the proportion of cells tuned to the previously unfamiliar views/orientations and may reduce the threshold of the population response required for behavioural output. This neural population account covers the wide variation in the RT-orientation slopes in different tasks without invoking different recognition strategies.

6.6. Explaining behavioural generalisation from physiological mechanisms

The proposed account suggests that people take longer to recognise an object in unusual orientations because the neural representation of that object is less strongly activated. This does not provide a full explanation of recognition of unusual views since it does not explain how and why the neurones fire as they do. This section briefly considers why the neural representation of familiar objects is sensitive to viewing conditions.

One can ask how individual cells exhibit response selectivity for complex stimuli and simultaneously show sensitivity to orientation, size and view. We suggest that the orientation and size sensitivity of cells at progressively higher levels of pattern processing is inherited from preceding stages, as cells in each level pool the activity of particular orientation and size specific feature detectors of lower levels (Perrett and Oram, 1993; Oram and Perrett, 1994; Fukushima, 1980). This process could establish the Elaborate feature detectors described by Tanaka et al. (1991) in the inferior temporal cortex from simpler features in prestriate cortex. Increased complexity of pattern selectivity and sensitivity to configuration can be achieved by combining several the outputs of several elaborate feature detectors which are themselves conjunctions of elements. Rearranged configurations (such as jumbled face parts) yield images lacking some of the 2-D visual features present in the

normal configuration (Perrett and Oram, submitted). Sensitivity to the configuration of one view of a face can be generated in this way even when elaborate feature detection exhibits translation invariance over the parafovea.

While this scheme may explain the orientation sensitivity of individual cells to faces, it does not account for the overall tuning of the cell population. The population of cells responsive to faces exhibits tuning with respect to orientation and is biased to upright orientations. The bias can be explained by postulating that, through learning, the tuning of cells comes to reflect the statistics (e.g. frequency) of the images that we see. We see more upright than inverted faces. There is growing evidence that temporal cortex cell selectivity is indeed biased towards images experienced as an adult (Logothetis et al., 1995; Miyashita and Chang, 1988; Perrett et al., 1984). Competitive learning mechanisms (e.g. Földiák, 1991) coupled with visual experience may account both for cell selectivity for objects that are frequently seen and for the cell sensitivity to the viewing conditions in which the objects are seen.

What is important for the current paper is the inheritance of orientation and size specificity from early cortical processing. This specificity coupled with experience-dependent plasticity will produce bias in the population activity towards frequently seen views of faces. The dependence of RTs on orientation at the behavioural level is therefore explicable just from considering cellular mechanisms. There is no need to postulate additional mental operations to account for generalisation functions.

Examination of Fig. 1 shows that there is no systematic variation in onset latency of responses and viewing condition. Indeed, examination of cell responses to stimuli of different effectiveness shows that response latencies are coincident (within the 5 ms accuracy of measurement, Oram and Perrett, 1992, 1994). Previous reports of increased response latencies to rotated images of faces may reflect an artefact of measurement. Most, though not all, cell responses to non-upright faces are smaller and have slower rise times than responses to upright faces (Perrett et al., 1984, 1985, 1988b; Wachsmuth and Perrett, 1997). The smaller response magnitude means that with limited samples of trials it takes longer for a cell's response to a rotated face to exceed any statistical criterion (Perrett et al., 1984, 1985, 1988b).

If cell responses reflected mental transformations, then one would predict a systematic change in latency of response onset with change in view, orientation, size and part visibility. For example, a cell tuned to an upright face should not respond to an inverted face until the image has been mentally rotated by 180°. Thus inverted images should take longer to activate cells. We report here that cells do not show this predicted change in onset latency. Separation in the cumulative population response to different viewing circumstances occurs after response onset and amplifies over time. So the cells are not themselves implementing transformations nor are they reflecting the outputs of transformation processes occurring elsewhere. Since the behavioural RT-orientation functions can be predicted from the neural operations we describe, Occam's razor can be applied to conclude that there is no need to postulate additional transformation mechanisms for recognising familiar objects.

Acknowledgements

References

Allison, T., McCarthy, G., Belger, A., Puce, A., Luby, M. Spencer, D.D., Bentin, S., 1994. What is a face: electrophysiological responsiveness of human extrastriate visual cortex to human faces. Society for Neuroscience Abstracts 20, 316.

Baker Cave, C., Kosslyn, M., 1993. The role of parts and spatial relations in object identification. Perception 22, 229–248.

Beale, J.M., Keil, F.C., 1995. Categorical effects in the perception of faces. Cognition 57, 217–239.

Besner, D., 1983. Visual pattern recognition: size pre-processing re-examined. Quarterly Journal of Experimental Psychology 35, 209–216.

Biederman, I., 1987. Recognition by components: a theory of human image understanding. Psychological Review 94, 115–145.

Bonda, E., Petrides, M., Frey, S., Evans, A., 1995. Neural correlates of mental transformations of the body-in-space. Proceedings of the National Academy of Science USA 92, 11180–11184.

Bricolo, E., Poggio, T., Logothetis, N.K., 1998. 3D object recognition: a model of view tuned neurons. In: Advances in Neural Information Processing Systems, Vol. 9. Morgan Kaufman, Los Altos, CA, in press.

Bruce, V., Young, A., 1986. Understanding face recognition. British Journal of Psychology 77, 103–327.

Bundesen, C., Larsen, A., 1975. Visual transformation of size. Journal of Experimental Psychology: Human Perception and Performance 1, 214–220.

Bülthoff, H.H., Edelman, S., 1992. Psychophysical support for a two-dimensional view interpolation theory of object recognition. Proceedings of the National Academy of Science USA 89, 60–64.

Calder, A.J., Young, A.W., Perrett, D.I., Etcoff, N.L., Rowland, D., 1996. Categorical perception of morphed facial expressions. Visual Cognition 3, 81–117.

Carey, S., 1992. Becoming a face expert. Philosophical Transactions of the Royal Society of London B 335, 95–103.

Carpenter, P.A., Just, M.A., 1978. Eye fixations during mental rotation. In: J.W. Senders, D.F. Fisher, R.A. Monty (Eds.), Eye Movements and Higher Psychological Functions. Erlbaum, Hillsdale, NJ.

Cohen, M.S., Kosslyn, S.M., Breiter, H.C., Digirolamo, G.J., Thompson, W.L., Anderson, A.K., Bookheimer, S.Y., Rosen, B.R., Belliveau, J.W., 1996. Changes in cortical activity during mental rotation – a mapping study using functional MRI. Brain 119, 89–100.

Cooper, L.A., Shepard, R.N., 1973. In: W.G. Chase (Ed.), Chronometric Studies of the Rotation of Mental Images. Academic Press, New York, pp. 75–176.

Corballis, M., 1988. Recognition of disoriented shapes. Psychological Review 95, 115–123.

Dolan, R.J., Fink, G.R., Rolls, E.T., Booth, M., Holmes, A., Frackowiak, R.S.J., Friston, K.L., 1997. How the brain learns to see objects and faces in an impoverished context. Nature 389, 596–599.

Edelman, S., Bülthoff, H.H., 1992. Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. Vision Research 32, 2385–2400.

Edelman, S., Weinshall, D., 1991. A self-organising multiple-view representation of 3D objects. Biological Cybernetics 63, 209–219.

Edelman, S., 1995. Class similarity and viewpoint invariance in the recognition of 3D objects. Biological Cybernetics 72, 207–220.

Edelman, S., 1997. Representation is representation of similarities. Behavioural and Brain Sciences, in press.

Ellis, R., Allport, D.A., Humphreys, G.W., Collis, J., 1989. Varieties of object constancy. Quarterly Journal of Experimental Psychology 41, 775–796.

Etcoff, N.L., Magee, J.J., 1992. Categorical perception of emotion. Cognition 44, 227–240.

Farah, M.J., Hammond, K.M., 1988. Mental rotation and orientation-invariant object recognition: dissociable processes. Cognition 29, 29–46.

Flin, R., Dziurawiec, S., 1989. Development factors in face processing. In: A.W. Young, H.D. Ellis. Handbook of Research on Face Processing. North Holland, Amsterdam.

Földiák, P., 1991. Learning invariance from transformation sequences. Neural Computation 3, 194–200.

Földiák, P., 1993. The 'ideal homunculus': statistical inference from neural population responses. In F.H. Eeckman, J.M. Bower (Eds.), Computation and Neural System. Kluwer Academic Publishers, Norwell, MA, pp. 55–60.

Fukushima, K., 1980. Neocognition: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biological Cybernetics 36, 193–202.

Gauthier, I., Tarr, M.J., 1997. Orientation priming of novel shapes in the context of viewpoint dependent processing. Perception 26, 51–73.

Gauthier, I., Tarr, M.J., 1997b. Becoming a 'Greeble' expert: exploring the mechanisms for face recognition. Vision Research, in press.

Georgopoulos, A.P., Kalaska, J.F., Caminiti, R., Massey, J.T., 1982. On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. Journal of Neuroscience 2, 1527–1537.

Georgopoulos, A.P., Luroto, J.T., Petrides, M., Schwartz, A.B., Masey, J.T., 1989. Mental rotation of the neuronal population vector. Science 243, 234–243.

Georgopoulos, A.P., 1995. Current issues in directional motor control. Trends in Neuroscience 18, 506–510.

Goodale, M.A., Milner, A.D., 1992. Separate visual pathways for perception and action. Trends in Neurosciences 15, 20–25.

Gregory, R.L., 1980. Perceptions as hypotheses. Philosophical Transactions of the Royal Society of London B 290, 181–197.

Hasselmo, M.E., Rolls, E.T., Baylis, G.C., Nalwa, V., 1989. Object centred encoding by face-selective neurons in the cortex of the superior temporal sulcus of the monkey. Experimental Brain Research 75, 417–429.

Harries, M.H., Perrett, D.I., Lavender, A., 1991. Preferential inspection of views of 3-D model heads. Perception 20, 669–680.

Hill, H., Schyns, P.G., Akamatsu, S., 1997. Information and viewpoint dependence in face recognition. Cognition 62, 201–222.

Holmes, E.J., Gross, C.G., 1984. Stimulus equivalence after inferior temporal lesions in monkeys. Behavioural Neuroscience 98, 898–901.

Holmes, E.J., Gross, C.G., 1984. Effects of inferior temporal lesions on discrimination of stimuli differing in orientation. Journal of Neuroscience 4, 3063–3068.

Humphreys, G.W., Kahn, S.C., 1992. Recognizing novel views of three-dimensional objects. Canadian Journal of Psychology 46, 170–190.

Humphreys, G.W., Riddoch, M.J., 1984. Routes to object constancy: implications from neurological impairments of object constancy. Quarterly Journal of Experimental Psychology 36, 385–415.

Humphreys, G.W., Quinlan, P.T., 1987. Normal and pathological processes in visual constancy. In: G.W. Humphreys, M.J. Riddoch (Eds.), Visual Object Processing: a Cognitive Neuropsychological Approach. Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 281–301.

Jolicœur, P., 1985. The time to name disorientated natural objects. Memory and Cognition 13, 289–303.

Jolicœur, P., 1987. A size-congruency effect in memory for visual shape. Memory and Cognition 15, 531–543.

Jolicœur, P., 1990. Identification of disoriented objects: a dual-system theory. Mind and Language 5, 387–410.

Jolicœur, P., Besner, D., 1987. Additivity and interaction between size ratio and response category in the comparison of size-discrepant shapes. Journal of Experimental Psychology 38, 80–93.

Koenderink, J.J., van Doorn, A.J., 1979. The internal representation of solid shape with respect to vision. Biological Cybernetics 32, 211–216.

Kolers, P., 1972. Aspects of Motion Perception. Pergamon Press, Elmsford, NY.

Larsen, A., 1985. Pattern matching: effects of size ratio, angular difference in orientation, and familiarity. Perception and Psychophysics 38, 63–68.

Lennie, P., 1981. The physiological basis of variations in visual latency. Vision Research 21, 815–824.

Logothetis, N.K., Pauls, J., Bülthoff, H.H., Poggio, T., 1994. View-dependent object recognition by monkeys. Current Biology 4, 401–414.

Logothetis, N.K., Pauls, J., Poggio, T., 1995. Shape representation in the inferior temporal cortex of monkeys. Current Biology 5, 552–563.

Lurito, J.T., Georgopoulos, T., Georgopoulos, A.P., 1991. Cognitive spatial-motor processes. 7. The making of movements at an angle from stimulus direction: studies of motor cortical activity at the single cell and population level. Experimental Brain Research 87, 562–580.

McAdams, C.J., Maunsell, J.H.R., 1996. Attention enhances neuronal responses without altering orientation selectivity in macaque area V4. Society for Neuroscience Abstracts 22, 1197.

McMullen, P.A., Jolicœur, P., 1992. The reference frame and effects of orientation on finding the top of rotated objects. Journal of Experimental Psychology: Human Perception and Performance 18, 807–820.

Miyashita, Y., Chang, H.-S., 1988. Neuronal correlate of pictorial short-term memory in the primate temporal cortex. Nature 331, 68–70.

Moses, Y., Ullman, S., Edelman, S., 1996. Generalization to novel images in upright and inverted faces. Perception 25, 443–462.

Oram, M.W., Perrett, D.I., 1992. Time course of neural responses discriminating different views of the face and head. Journal of Neurophysiology 68, 70–84.

Oram, M.W., Perrett, D.I., 1994. Modelling visual recognition from neurobiological constraints. Neural Networks 7, 945–972.

Oram, M.W., Wachsmuth, E., Perrett, D.I., Emery, N.J., submitted. Visual coding of body-part and whole body motion in the temporal cortex of the macaque monkey. Cerebral Cortex.

Oram, M.W., Foldiak, P., Perrett, D.I., Sengpiel, F., 1998. The ideal Homunculus: decoding neural population signals. Trends in Neuroscience, in press.

Palmer, S.E., Rosch, E., Chase, P., Baddely, D., 1981. Canonical perspective and the perception of objects. In: J. Long, A. Baddeley (Eds.), Attention and Performance IX. Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 135–151.

Parsons, L.M., Fox, P.T., Downs, J.H., Glass, T., Hirsch, T.B., Martin, C.C., Jerabek, P.A., Lancaster, J.L., 1995. Use of implicit motor imagery for visual shape discrimination as revealed by PET. Nature 375, 54–58.

Perrett, D.I., 1996. View-dependent coding in the ventral stream and its consequences for recognition. In: R. Caminiti, K.-P. Hoffmann, F. Lacquaniti, J. Altman (Eds.), Vision and Movement Mechanisms in the Cerebral Cortex. HFSP, Strasbourg, pp. 142–151.

Perrett, D.I., Oram, M.W. Visual recognition based on temporal cortex cells: viewer-centered processing of pattern configuration. Zeilschrift für Naturforschung, in press.

Perrett, D.I., Oram, M.W., Hietanen, J.K., Benson, P.J., 1994. Issues of representation in object vision. In: M. Farah, G. Ratcliff (Eds.), The Neuropsychology of Higher Vision: Collated Tutorial Essays. Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 33–61.

Perrett, D.I., Oram, M.W., 1993. The neurophysiology of shape processing. Image and Visual Computing 11, 317–333.

Perrett, D.I., Oram, M.W., Harries, M.H., Bevan, R., Hietanen, J.K., Benson, P.J., Thomas, S., 1991. Viewer-centred and object-centred coding of heads in the macaque temporal cortex. Experimental Brain Research 86, 159–173.

Perrett, D.I., Mistlin, A.J., Chitty, A.J., Smith, P.A.J., Potter, D.D., Brönnimann, R., Harries, M.H., 1988. Specialised face processing and hemispheric asymmetry in man and monkey: evidence from single unit and reaction time studies. Behavioural Brain Research 29, 245–258.

Perrett, D.I., Mistlin, A.J., Chitty, A.J., Harries, M., Newcombe, F., De Haan, E., 1988a. Neuronal mechanisms of face perception and their pathology. In: C. Kennard, F. Clifford Rose (Eds.), Physiological Aspects of Clinical Neuro-opthalmology. Chapman and Hall, London, pp. 137–154.

Perrett, D.I., Harries, M.H., Bevan, R., Thomas, S., Benson, P.J., Mistlin, A.J., Chitty, A.J., Hietanen, J.K., Ortega, J.E., 1989. Frameworks of analysis for the neural representation of animate objects and actions. Journal of Experimental Biology 146, 87–114.

Perrett, D.I., Smith, P.A.J., Potter, D.D., Mistlin, A.J., Head, A.S., Milner, A.D., Jeeves, M.A., 1985. Visual cells in the temporal cortex sensitive to face view and gaze direction. Proceedings Royal Society London B 223, 293–317.

Perrett, D.I., Smith, P.A.J., Potter, D.D., Mistlin, A.J., Head, A.S., Milner, A.D., Jeeves, M.A., 1984. Neurons responsive to faces in the temporal cortex: Studies of functional organization, sensitivity to identity and relation to perception. Human Neurobiology 3, 197–208.

Perrett, D.I., Rolls, E.T., Caan, W., 1982. Visual neurons responsive to faces in the monkey temporal cortex. Experimental Brain Research 47, 329–342.

Phelps, M.T., Roberts, W.A., 1994. Memory for pictures of upright and inverted primate faces in humans (*homo sapiens*), squirrel monkeys (*saimiri sciureus*), and pigeons (*columba livia*). Journal of Comparative Psychology 108, 114–125.

Poggio, T., Edelman, S., 1990. A network that learns to recognise three-dimensional objects. Nature 343, 263–266.

Purcell, D.G., Stewart, A.L., 1988. The face-detection effect: configuration enhances detection. Perception and Psychophysics 43, 355–366.

Rock, I., 1973. Orientation and Forms. Academic Press, New York.

Rock, I., DiVita, J., 1987. A case of viewer-centred object perception. Cognitive Psychology 19, 280–293.

Rolls, E.T., Baylis, G.C., Hasselmo, M.E., Nalwa, V., 1989. The effect of learning on the face selective properties of neurons in the cortex in the superior temporal sulcus of the monkey. Experimental Brain Research 76, 153–164.

Rolls, E.T., Tovee, M.J., Purcell, D.G., Stewart, A.L., Azzopardi, P., 1994. The response of neurons in the temporal cortex of primates and face identification and detection. Experimental Brain Research 101, 473–484.

Sanger, T.D., 1996. Probability density estimation for the interpretation of neural population codes. Journal of Neurophysiology 76, 2790–2793.

Shepard, R.N., Metzler, J., 1971. Mental rotation of three-dimensional objects. Science 171, 701–703.

Stratford, K., Mason, A., Larkman, A., Major, G., Jack, J., 1989. The modeling of pyramidal neurones in the visual cortex. In: R. Durbin, C. Miall, G. Mitchison (Eds.), The Computing Neurone. Addison–Wesley, New York, pp. 296–321.

Tanaka, K., Saito, H., Fukada, Y., Moriya, M., 1991. Coding visual images of objects in the inferotemporal cortex of the macaque monkey. Journal of Neurophysiology 66, 170–189.

Tarr, M.J., Pinker, S., 1989. Mental rotation and orientation dependence in shape recognition. Cognitive Psychology 5, 233–282.

Tarr, M.J., Pinker, S., 1990. When does human object vision use a viewer-centred reference. Psychological Science 2, 207–209.

Tarr, M.J., Pinker, S., 1991. Orientation-dependent mechanisms in shape-recognition – further issues. Psychological Science 2, 207–209.

Tarr, M.J., 1995. Rotating objects to recognise them: a case study on the role of viewpoint dependency in the recognition of three-dimensional objects. Psychonomic Bulletin and Review 2, 55–82.

Turnbull, O.H., Laws, K.R., McCarthy, R.A., 1995. Object recognition without knowledge of object orientation. Cortex 31, 387–395.

Ullman, S., 1989. Aligning pictorial descriptions: an approach to object recognition. Cognition 32, 193–254.

Vaughan, H.G. Jr., Costa, L.D., Gilden, L., 1966. Functional relation of visual evoked response and reaction time to stimulus intensity. Vision Research 6, 645–656.

Vetter, T., Poggio, T., Bülthoff, H.H., 1994. The importance of symmetry and virtual views in three-dimensional object recognition. Current Biology 4, 18–23.

Wachsmuth, E., Oram, M.W., Perrett, D.I., 1994. Recognition of objects and their component parts: responses of single units in the temporal cortex of the macaque. Cerebral Cortex 4, 509–522.

Wachsmuth, E., Perrett, D.I., Oram, M.W., submitted. Effect of image orientation and size on cells responsive to faces in the cortex of the superior temporal sulcus of macaque. Cognitive Neuropsychology.

Wachsmuth, E., Perrett, D.I., 1997. Generalising across object orientation and size. In: V. Walsh, S. Butler (Eds.), Perceptual Constancies. Oxford University Press, Oxford, in press.

Weiskrantz, L., Saunders, R.C., 1984. Impairments of visual object transforms in monkeys. Brain 107, 1033–1072.

Warrington, E.K., James, M., 1986. Visual object recognition in patients with right hemisphere lesions: axes or features? Perception 15, 355–366.

Warrington, E.K., Taylor, A.M., 1973. The contribution of the right parietal lobe to object recognition. Cortex 9, 152–164.

Wertheimer, M., 1912. Experimentelle Stuidien uber das Sehen von Beuegung. Zeitschrift fuer Psychologie 61, 161–265.

Yamane, S., Kaji, S., Kawano, K., 1988. What facial features activate face neurons in the inferotemporal cortex of the monkey. Experimental Brain Research 73, 209–214.

Yin, R.K., 1969. Looking at upside-down faces. Journal of Experimental Psychology 81, 141–145.

Yin, R.K., 1970. Face recognition by brain-injured patients, a dissociable ability? Neuropsychologia 8, 395–402.

Young, A.W., Rowland, D., Calder, A.J., Etcoff, N. L, Seth, A., Perrett, D.I., 1997. Facial expression megamix: tests of dimensional and category accounts of emotional recognition. Cognition 63, 271–313.

Young, M.P., Yamane, S., 1992. Sparse population coding of faces in the inferotemporal cortex. Science 256, 1327–1331.