The Quarterly Journal of Experimental Psychology
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t716100704

Seeing the future: Natural image sequences produce "anticipatory" neuronal activity and bias perceptual report

David I. Perrett [a]; Dengke Xiao [a]; Nick E. Barraclough [a]; Christian Keysers [a]; Mike W. Oram [a]
[a] University of St Andrews, St Andrews, UK

First published on: 23 June 2009

To cite this Article Perrett, David I., Xiao, Dengke, Barraclough, Nick E., Keysers, Christian and Oram, Mike W.(2009) 'Seeing the future: Natural image sequences produce "anticipatory" neuronal activity and bias perceptual report', The Quarterly Journal of Experimental Psychology, 62: 11, 2081 — 2104, First published on: 23 June 2009 (iFirst)

To link to this Article: DOI: 10.1080/17470210902959279

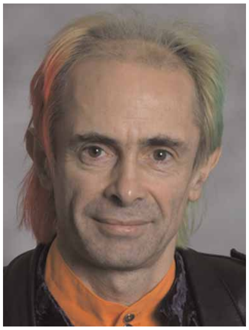
URL: http://dx.doi.org/10.1080/17470210902959279

THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY
2009, 62 (11), 2081–2104

Psychology Press
Taylor & Francis Group

EPS Mid-Career Award 2008

Seeing the future: Natural image sequences produce "anticipatory" neuronal activity and bias perceptual report



David I. Perrett
Dengke Xiao, Nick E. Barraclough, Christian Keysers, and Mike W. Oram
University of St Andrews, St Andrews, UK

This paper relates human perception to the functioning of cells in the temporal cortex that are engaged in high-level pattern processing. We review historical developments concerning (a) the functional organization of cells processing faces and (b) the selectivity for faces in cell responses. We then focus on (c) the comparison of perception and cell responses to images of faces presented in sequences of unrelated images. Specifically the paper concerns the cell function and perception in circumstances where meaningful patterns occur momentarily in the context of a naturally or unnaturally changing visual environment. Experience of visual sequences allows anticipation, yet one sensory stimulus also "masks" perception and neural processing of subsequent stimuli. To understand this paradox we compared cell responses in monkey temporal cortex to body images presented individually, in pairs and in action sequences. Responses to one image suppressed responses to similar images for ~500 ms. This suppression led to responses peaking 100 ms earlier to image sequences than to isolated images (e.g., during head rotation, face-selective activity peaks before the face confronts

Correspondence should be addressed to David I. Perrett, School of Psychology, University of St Andrews, St Andrews KY16 9JP, UK. E-mail: dp@st-andrews.ac.uk

Paper prepared for Experimental Psychology Society (EPS) Mid-Career Award, 2008, to D. Perrett. The Wellcome Trust and EU Framework grants supported this work. We thank Mary Latimer for assistance in histology and R. Passingham, S. Soto, M. Reisenhuber, M. Barr, I. Thornton, M. Giese, T. Poggio, P. Földiák, A. Whiten, P. Hibbard, J. Williams, and Z. Kourtzi for discussion of results and text. Nick E. Barraclough is now at the University of Hull, UK. Christian Keysers is now at the BCN Neuro-Imaging-Center, University Medical Center, University of Groningen, Groningen, The Netherlands.

http://www.psypress.com/qjep
DOI:10.1080/17470210902959279

the observer). Thus forward masking has unrecognized benefits for perception because it can transform neuronal activity to make it predictive during natural change.

Keywords: Prediction; Sequence; Masking; Face; Single cell.

In everyday life our visual world is full of changing scenes. For the most part the changes are continuous and gradual; the scenes do not come in the chaotic jumble of photos taken on a digital camera. The scenes we encounter normally are more like the frames of a movie in which events unfold gradually with each new frame changing slightly from the last. Despite our preoccupation with watching video and cinema, let alone all our experience of real life, there is little visual science informing us of how visual processing at any moment is affected by the progressively changing visual context.

The purpose of this paper is to address the nature of visual form processing in naturally changing scenes. To that end the paper describes the way cells "handle" visual images in temporal cortex. It first presents a brief retrospective of our studies on the organization and selectivity of cell responses for particular meaningful image configurations (such as faces and bodies) and how these cells respond to brief images. Then the paper details studies we have made over the past 9 years of cell responses to face and body postures occurring momentarily during natural actions.

Functional organization of cells processing faces

Since the review deals with cells responding to faces and bodies, a sense of location and organization is useful by way of orientation. Studies in the late 1960s and early 1970s reported one cell responding to a hand in the inferior temporal cortex of the monkey. There was also mention of similar cell selectivity for faces but no detailed description (for historical perspective, see Gross, 2008). Subsequent work on the rhesus monkey, setting out to describe cells responsive to faces systematically, showed that their distribution within one temporal lobe region (the cortex of the superior temporal sulcus, STS; see Figure 1a) was clumped (Perrett, Rolls, & Caan, 1979, 1982; Perrett et al., 1984). Just as early visual cortex was functionally organized so was temporal cortex. Moving across the cortex in a particular

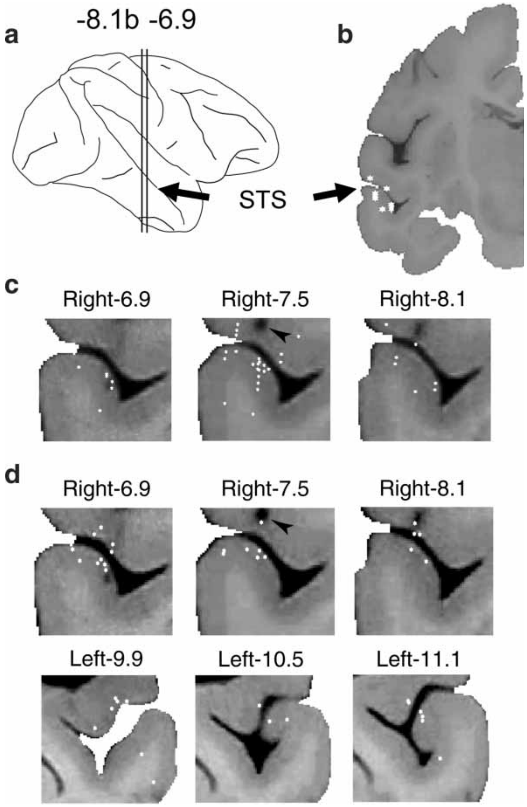


Figure 1. *Location of cells within the superior temporal sulcus (STS). (a) Schematic of the lateral surface of the macaque brain with STS and positions of sections marked. (b) Photograph of a coronal section through the right hemisphere 8 mm posterior to the anterior commissure. (c) Serial sections (every 0.6 mm) expanding the region of right STS with cell locations (white stars) tested with image pairs for masking and position of a microlesion used for reconstruction (black arrow section −7.5). (d) Serial sections in left and right hemispheres recording position of cells tested for responses during image sequences.*

direction revealed a high probability of encountering cells with similar selectivity for distances of up to 3 mm (Perrett et al., 1984). Sampling further away showed a change in the pattern type for which cells were selective, and occasionally sampling even further away established a recurrence of similar selectivity (a new patch). This clumping made studies of selectivity sometimes easy, because cells in the same clump could be accessed repeatedly, but most often clumping meant frustration because such small patches responsive to faces could not be found or repeatedly accessed. Indeed the clumped nature meant that experiments had to be tailored to the selectivity present, which changed the nature of studies focusing in turn on head view, hand actions, body posture, walking bodies, head movement, and torso flexion (Jellema & Perrett, 2006; Perrett et al., 1989; Wachsmuth, Oram, & Perrett, 1994).

Since the early single-cell mapping studies, the functional organization of temporal cortex in terms of columns and patches has been documented with cell recording and optical imaging (Fujita, Tanaka, Ito, & Cheng, 1992; Wang, Tanaka, & Tanifuji, 1996), and functional magnetic resonance imaging (fMRI; Tsao, Freiwald, Tootell, & Livingstone, 2006).

Our studies showed that the anatomical outputs of the region were also segregated with 3–5-mm patches connecting to posterior parietal cortex (Harries & Perrett, 1991; for recent confirmation see Rozzi et al., 2006). One patch site, midway along the STS in the upper bank lying next to the lateral geniculate nucleus, we suggested might be conveying information (from gaze, face view and body posture) about the direction of attention of others to attention control mechanisms in the parietal lobe (Harries & Perrett, 1991; Perrett, Hietanen, Oram, & Benson, 1992; Leekam, Baron-Cohen, Perrett, Milders, & Brown, 1997). Examination of the visual selectivity using combined fMRI and single-cell recording has confirmed that >95% of the cells within a patch at this location are indeed selective for faces (Tsao et al., 2006).

The temporal cortex contains cells sensitive to a variety of parameters (Tanaka, Saito, Fukada, & Moriya, 1991). Experience with particular objects increases the proportion responsive to the learned objects (Logothetis, Pauls, & Poggio, 1995). Nonetheless, the temporal cortex region remains inhomogeneous with cells coding faces and cells coding other objects of experience largely separated from one another. Cells coding hands are segregated from those coding faces (Perrett et al., 1989) though some of the latter cells may process information about both the face and hand action (Jellema, Baker, Wicker, & Perrett, 2000). Likewise those showing selectivity for trained patterns tend to occur in more anterior and ventral temporal lobe regions (e.g., Sakai & Miyashita, 1991).

The “speed of sight” for sequences of unrelated images

The majority of visual neurophysiology has assessed cell responses to visual images presented individually and for an appreciable time (0.5–5 s). We began to explore the nature of responses at shorter durations. With this approach we could measure cell performance in detecting images presented briefly and compare this to human performance detecting the same images. To this end we recorded from neurons within the STS and chose eight stimuli spanning a range of effectiveness for each cell. We then tested the cells with the eight stimuli in pseudorandom order at different rates of image presentation (Keysers, Xiao, Földiák, & Perrett, 2001). Fast presentation showed that cells were capable of responding to very briefly presented stimuli (see Figure 2). We measured a given cell’s responses after each image and used signal detection to assess how often an ideal observer listening to the response of this cell would be able to determine whether or not the target stimulus (the most effective of the eight test stimuli) was present in a section of the random sequence (for details of assessment methods, see Keysers et al., 2001). We found that cells maintained their selectivity amongst images and could signal the target image presence even at the fastest rate of testing where the target was present for 14 ms.

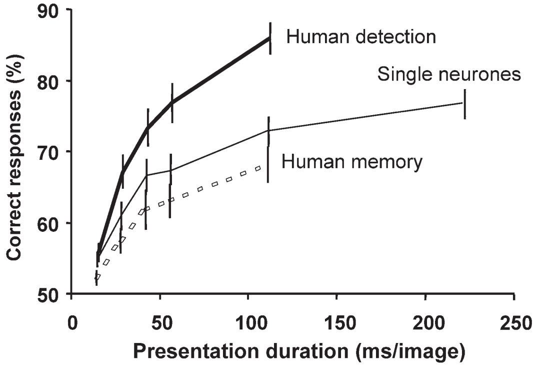


Figure 2. *Comparison between human perception, memory, and single-cell responses in "detecting" stimuli presented at different rates. Vertical axis gives the accuracy of recognition performance for human observers and single-cell responses for detecting the presence of an image in the central position of a sequence of seven visual stimuli. Performance of humans and single cells declines as presentation rate increases but remains above chance even at the highest presentation rates of 11 ms per image. From Figure 6a, p. 98, "The speed of sight", by C. Keysers, D.-K. Xiao, P. Földiák, and D. I. Perrett, 2001,* Journal of Cognitive Neuroscience, 13, *pp. 90–101. Copyright © 2001 by MIT Press Journals. Adapted with permission.*

This line of research showed that briefly exposed images reach high levels of processing in temporal cortex even at presentation rates as fast as 70 different images presented in one second. Not surprisingly the cell responses to such brief stimuli are transient in nature. Human perceptual performance is equivalently ephemeral. It was possible but difficult for observers to register and report on the presence of each image. We tested perception by presenting target images to be detected, followed by a sequence of 7 images, containing the target or not as the 4th image. Perceptual experience was demonstrable, and observers scored better than chance in detecting the targets when they were present (Figure 2, thick line).

We also tested the impact on memory by presenting sequences of 7 images first and then a target for the observers to report whether or not they remembered the target's presence in the sequence. When a few hundred milliseconds had lapsed, the observer's memory for particular images was poor, though again still above chance (Figure 2, dashed line). Had we forced observers to wait longer before presenting them with a target and asking them to recall the presence or absence in sequences presented several seconds previously, their memory would probably have declined to chance level.

We can conclude from these studies that perception and neural responses are equivalently graded. When cell activity is high and prolonged, perception is emphatic; the smaller and shorter the duration of the cell response, the less secure the perceptual experience (Perrett, Benson, Hietanen, Oram, & Dittrich, 1995; Perrett et al., 1984). In the grey area close to perceptual threshold, pictures are processed but images are not seen clearly. If a difficult perceptual or memory test is forced on observers, then experimenters will conclude that there is an absence of conscious awareness. Any effects of the targets on behaviour are then considered subliminal (Kouider & Dehaene, 2007). It may be wrong to think that if perception and memory in some tasks are at chance then any stimulus-related effects on behaviour are due to subliminal processing. For example, in Figure 2 we show that memory performance for images of 11-ms duration is close to chance level, so it could be argued that any influence of those brief stimuli on subsequent behaviour was subliminal. A change in task (to detection) shows clearly that the same 11-ms stimuli should also be considered supraliminal since they were detected at levels appreciably above chance. The brief activity in temporal cortex to 11-ms images is likely to be essential for any later behavioural effects: Whether such behavioural effects are considered subliminal may depend critically on the task given to observers.

Selectivity for faces in cell responses

The selectivity of cells in temporal cortex has always been a debated subject since 1972 when Gross, Rocha-Miranda, and Bender (1972) indicated that cells might respond preferentially to particular complex stimuli. Given that cells responded reliably to images presented briefly

(Figure 2), we decided to test individual cells with very large collections of images to define their selectivity more fully than had been attempted in any prior experiment (Földiák, Xiao, Keysers, Edwards, & Perrett, 2003). This way we could define selectivity systematically rather than serendipitously. For 23 cells in the STS cortex, we explored a large set of up to 1,200 images containing faces, patterns, scenes, and objects. Cell selectivity within the image collection was high so that few images provoked substantial responses. Typically for a given cell, 95% of stimuli would produce responses, weaker than one third of the magnitude of the most effective stimuli. So what stimuli were effective in eliciting large responses for given cells?

Of those tested, we isolated seven cells that were almost exclusively responsive to face images; that is, virtually all stimuli evoking a response significantly above baseline contained a facial image that human observers judged to be clearly visible. This was true despite the fact that depictions of faces varied considerably. For example, Figure 3 depicts the responses of a cell responding to a wide variety of face images: Each response occurs despite the facial images being presented only briefly for 50 ms and embedded in a continuously changing stream of unrelated images. This cell and the six others like it did not respond to all faces images tested, but virtually all of the images they responded to did contain a face. Of course one can set different criteria; if one selects a low response rate just above background then the cells appear "slightly" less selective. If one were to set an efficient criterion of, say, 50% of maximum firing rate observed, then particular cells were exclusively selective for faces. For example, in ranking the effectiveness of stimuli, the 73 most effective images for one cell all contained clearly visible faces (Figure 4).

This selectivity might be thought of as a chance finding; test enough cells, and one might find one by chance apparently selective for faces, but this claim does not bear scrutiny. If our image collection contained 50% faces, the probability of finding one of the 23 cells that "by chance" was tuned so that its top 73 images were all faces would be less than 23 in a thousand, million, million, million ($23/2^{73}$). Our image test set contained fewer clear faces, making it even less likely. Moreover, we did not find just one cell tuned in this way; for a second cell the top 41 images were all clear faces (a chance of $23/2^{41}$). The chance of finding selectivity in both cells is $23/2^{114}$. A claim that is frequently made is that cells responsive to faces might respond to some other images if only enough were tested. Again the evidence and consideration of probability invalidate this claim.

Certainly there were cells responsive both to faces and to other objects but their frequency was much lower than chance predicts. If we consider cell tuning for or against faces (i.e., a category excluding faces), then the distribution of cells was bimodal and not random. Of 32 cells tested, 6 were highly selective for image categories that excluded faces, and, as already stated, 7 were highly selective for faces.

These studies confirmed both the selectivity of temporal lobe cells for particular categories of stimuli and the reliability of responses. Kiani et al. (Kiani, Esteky, Mirpour, & Tanaka, 2007), using a similar approach in a more extensive study, confirm temporal cortex cell selectivity for face and body categories. In humans too, neurons in the amygdala and hippocampus show highly selective responses for various visual categories including familiar faces, buildings, animals, and food (Quiroga, Reddy, Kreiman, Koch, & Fried, 2005). While the response latency of activity in such cells recorded in humans indicates a later stage of semantic processing, the selectivity between categories and tolerance of different examples of the same preferred category parallels that seen in the monkey temporal cortex responses to faces and to particular individuals (Perrett et al., 1989; Perrett et al., 1992; Perrett et al., 1984).

Interactions between stimuli backwards in time

In sequences of unrelated stimuli, our physiological data indicated cellular competition. Responses to successive brief stimuli competed with one

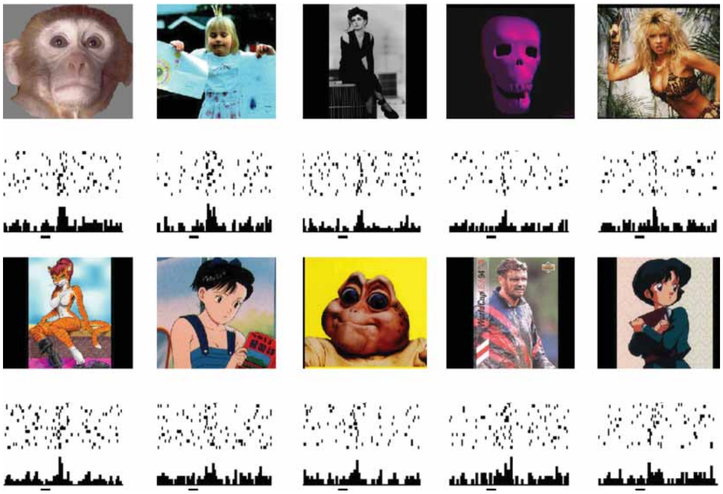


Figure 3. *Sensitivity of cell response to diverse face patterns during rapid image presentation. Results of testing with images presented in random order at 20 images/s. A total of 10 images from the set tested are displayed together with responses evoked from one cell. Horizontal rows of dots record the action potentials from the cell in relation to 15 occurrences of the test image (image duration 50 ms indicated by horizontal line); poststimulus histograms record the average response to the image across 15 trials of presentation. The cell responded reliably to images containing a wide variety of face patterns from the set of >1,000 images tested but did not respond well to images without faces.*

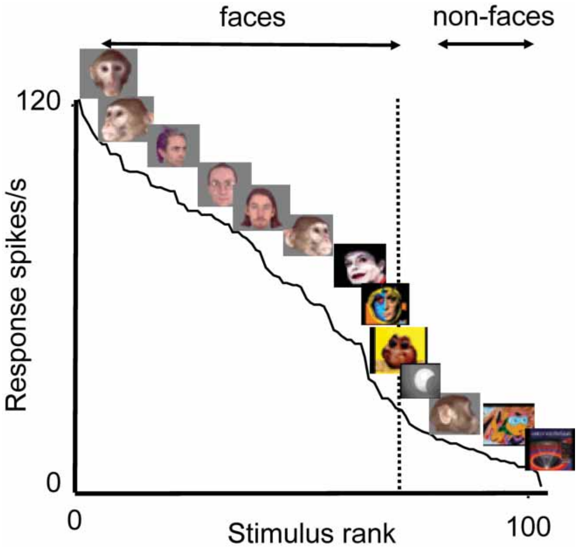


Figure 4. *Selectivity in cellular responses to faces. The rank order of effectiveness of stimuli for a cell after an automated narrowing search through a large collection of 1,200 images. Iterations of the search presented each image once and maintained the most effective 25% and least effective 25% images for the next search iteration. The resulting rank order of stimulus effectiveness is shown. For this cell the most effective 73 stimuli (up to dotted vertical line) all contained clear images of faces. The left-hand edge of example images is aligned with their stimulus rank effectiveness. From Figure 5, "Rapid serial visual presentation for the determination of neural selectivity in area STSa", by P. Földiák, D.-K. Xiao, C. Keysers, R. Edwards, and D. I. Perrett, 2003,* Progress in Brain Research, 144, *pp. 107–116. Copyright © 2003 by Elsevier. Adapted with permission.*

another; each new stimulus by virtue of its heightened transient response onset competed favourably and effectively abolished the ongoing responses to other stimuli (Keysers & Perrett, 2002; Keysers, Xiao, Földiák, & Perrett, 2005). If no new stimulus was presented for 84 ms, and a blank screen occurred, the response to the last image continued for 84 ms (Figure 5). Processing in this gap between stimuli continued unabated, just as if the stimulus was still physically present in the retinal image. The neural discharge during the gap or blank screen represents an iconic memory for the last image.

We were able to show that human perception for images in sequences with or without gaps paralleled cell responses. The ability of human participants to detect the presence of target images was examined in equivalent tests to that described above (Figure 1). The presence of blank screens between successive images did not detract from the ability of human observers to make a perceptual report on particular images, but the presence of other images did (see Figure 5). Thus recognition was compromised only if new stimuli appeared, and the cell response rate dropped. The results were quite surprising given that the gaps between stimuli make the presentation flicker horribly as the contrast between image and blank screen varies wildly. Thus despite the degradation of the graphic appearance of images from the no-gap to 84-ms-gap conditions, detection of information about pictorial content remained unaffected.

New visual stimuli have an impact on the processing of a prior stimulus, curtailing the information available about it from neural firing rates in temporal cortex. This disruptive phenomenon is referred to as backward masking (Keysers & Perrett, 2002). In the remaining sections of the paper we consider how an image presented at one moment in time affects cell responses to future images. Such forward interactions are usually described in negative terms (forward masking or suppression) from the detrimental effects seen in detection tasks like those used to

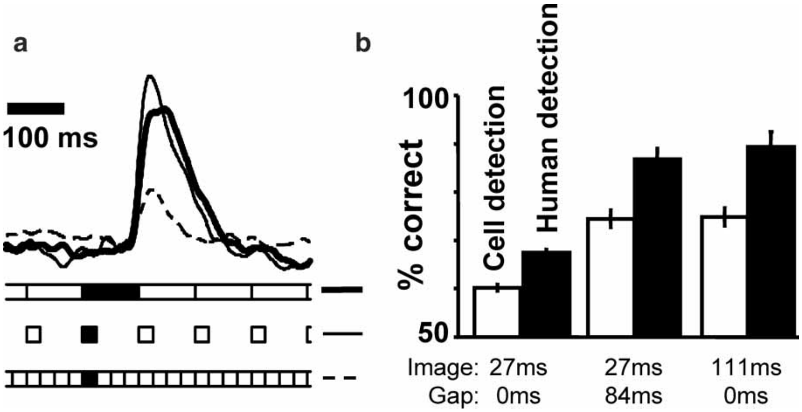


Figure 5. *Backward masking: competitive interaction between cell responses. (a) Average normalized population response of the single neurones to the best stimulus in sequences of different timing. The* x*-axes represent time, the* y*-axes the mean averaged latency-aligned population response. The horizontal "ladders" under the response represent the stimulus timing, with the filled squares representing the timing of the best stimulus (to which responses were aligned), while the open squares represent other stimuli in the random sequence. The space between two squares represents the gaps in gap sequences (thin line). Responses during testing at 111 ms/image (thick line) were equal to responses during testing with 27-ms images and 84-ms gaps (thin line); both of these responses were greater than those during testing with 27 ms/image without gaps between images (dashed line). (b) Stimulus timing and the corresponding accuracy with which a single stimulus can be detected in image sequences by human observers (black bars) and by an ideal observer of recorded cell activity (open bars). From Figure 1, "Out of sight but not out of mind: The neurophysiology of iconic memory in the superior temporal sulcus", by C. Keysers, D.-K. Xiao, P. Földiák, and D. I. Perrett, 2005,* Cognitive Neuropsychology, *22, pp. 316–332. Copyright © 2005 by Psychology Press. Adapted with permission.*

explore backward masking but, as we describe below, the same forward interactions may also have beneficial effects in establishing "anticipatory" neural activity and expectations in perception.

FORWARD INTERACTIONS BETWEEN SUCCESSIVE IMAGES

The study of perception and its neural basis typically focuses on the analysis of "snap-shots" with one or two stimuli presented in isolation. Yet we experience the world as a stream of events where previous visual scenes help us anticipate future states (Freyd & Finke, 1984; Guo et al., 2004; Verfaillie & Daems, 2002). So far the studies reviewed document the processing of unrelated stimuli and interactions where responses to a new image can interfere or mask the processing of prior images. We now turn to forward interactions between images that may help explain the effects of context in facilitating recognition.

In visual (Dragoi, Sharma, Miller, & Sur, 2002; Felsen et al., 2002; Grill-Spector, Henson, & Martin, 2006; Kourtzi & Kanwisher, 2001; Macknik & Livingstone, 1998; Sawamura, Orban, & Vogels, 2006; Turvey, 1973), auditory (Wehr & Zador, 2005), and somatosensory systems (Khatri, Hartings, & Simons, 2004), one stimulus is also found to mask perception of and neural response to a following stimulus but little is known about the functional role of this suppression or how it affects cell tuning in higher brain areas.

To bridge the gap between the complexity of real-world events and the simplified situations studied previously, we sought to determine how the processing of isolated visual images relates to the processing of coherent image sequences in which images transform sequentially and predictably. We began investigations with minimal sequences of two images and progressed to longer sequences presented rapidly in random and in natural order. This approach allows us to determine whether or not we can understand continuous sensory processing of real life (or video and film) in terms of brain responses to discrete sensory inputs (or pictures).

Physiological methods

Standard techniques (Földiák et al., 2003; Keysers et al., 2001) were used (in accordance with institutional guidelines and under UK Home Office licence) to record cells from 3 rhesus macaques (age 6–9 years), trained to sit in a primate chair with head restraint. Eye position (±1° monitored with IView, SMI), spike arrival, and stimulus on/offset times were recorded with CED1401 interface. Coronal and parasagittal X-ray photographs recorded the trajectory of each microelectrode. Cell positions were mapped (using microlesion references) to coronal brain sections after transcardial perfusion and histology. Reconstruction confirmed that tested cells occurred in the cortex surrounding the rostral STS (see Figure 1 for reconstruction of recording in one monkey).

Colour images (24 bits) stored on an Indigo2 Silicon Graphics workstation were presented (size 19 × 19°, distance 57 cm) on a Sony GDM-20D11 monitor (resolution 25.7 pixels/degree, refresh rate 72 Hz) in random order with a 500-ms interstimulus interval commencing when the subjects fixated (± 3°) a central dot for 500 ms. Fixation was rewarded with the delivery of fruit juice. Fixation breaks >100 ms stopped recording and stimuli. A search set of 50 images was used to activate STS neurons (Földiák et al., 2003).

For 39 cells, two images were selected from the search set: the most effective (target) and an effective "similar" mask (on average >50% target response). For 26 cells an ineffective "dissimilar" mask image (<50% target response) was additionally selected. Cells were tested with target-similar and -dissimilar masks presented alone for 125 ms and together as mask–target pairs (stimulus onset asynchrony, SOA = 125–458 ms; 500-ms blank screen between stimuli). For 13 further cells, mask duration equalled SOA for 28–83-ms SOAs. For an additional 57 cells, effective mask duration and SOA equalled 55 ms.

Actions of the body (walking), hand (e.g., grasping, tearing, picking) and head (e.g., rotation, vocalization) were filmed (Panasonic NV-DX100) and were used to test 61 cells separately from image pair studies. Digitized images (range 5–20, average 7) were presented in isolation or in sequence with frame duration = 42 or 55 ms (except 5 cells where duration = 111 ms). A total of 24 cells were selected for analysis when responses to separate images declined monotonically. A total of 37 different cells were selected for analysis when (a) responses to the sequence and one or more separate frames were reliable (>5 spikes/s above background), (b) the maximally responsive single frame occurred at least two frames in the sequence after the first detectable response to a single frame, and (c) each intervening frame produced a response greater than that for the preceding frame. This ensured that tuning increased steadily from minimum to peak level before declining.

Directional selectivity (Oram & Perrett, 1996) could affect sequence sensitivity, so motion-sensitive cells were specifically excluded from analysis. Five cells had responses to images in sequence before the first isolated image producing a response was reached (response to sequence > isolated images, $p < .001$). These cells were excluded to avoid potential sensitivity to movement.

A total of 53 further cells were tested with sequences of 8–30 images presented for 56 ms in continuous random order (rapid serial visual presentation, RSVP; Földiák et al., 2003; Keysers et al., 2001). We chose stimuli producing large responses as targets (>75% of the best stimulus response). Mask stimuli produced >50% response to the best stimulus. Measurement of the separate mask and target responses was performed in sequences with at least 1 prior and 1 following image producing weak responses (<25% best response). Measurement of mask–target pair responses was performed with 1 prior image producing a weak response. To examine whether masks affected target responses despite intervening stimuli, these intervening stimuli were also selected to produce weak responses.

Single-cell activity was isolated offline using template matching and principal components analysis (*Spike2*). For each stimulus a spike density function (SDF) was calculated by averaging across trials (1-ms time bins, Gaussian smoothing, $SD = 10$ ms). Background activity was measured for 100 ms before stimulus onset. Image pair and RSVP response latency = the first 1-ms time bin where the SDF exceeded background (+ 2.58 *SD*s) for 25 ms following stimulus onset. For paired stimuli, cell responses were normalized in magnitude (to the difference between the cell's peak response to the target alone and background) and time shifted (Keysers et al., 2001) so the response latency to the target alone = target onset + 100 ms. Sequence analysis was similar except cell latency = time to half peak response for optimal isolated image. All statistical tests reported use two-tailed probabilities. For sequences, cell responses were normalized in magnitude to the cell's average response to the most effective separate image (measured for the duration of image presentation).

Physiological results

Processing of stimulus pairs

With sequences of two images we varied stimulus onset asynchrony (SOA) and recorded responses from STS cells tuned to complex static patterns. Figure 6a shows the response of one cell to the second "target" stimulus of a pair presented after the first "mask" stimulus. At an interval of 125 ms, the response to the second stimulus is suppressed to less than half of the response to the same stimulus presented at an interval of 250 ms. For this cell, suppression is evident for intervals up to 400 ms. Spike density functions (Figure 6b) show that the first stimulus has a dual effect on subsequent target responses, both reducing the response magnitude and increasing the time to the peak response at short SOAs of 125 ms. Peak suppression in cell population data is very prominent at short intervals of 55 ms (Figure 6c). Thus suppression, which is apparent in sensory processing generally, persists to very high levels of image processing in the temporal cortex (Sawamura et al., 2006) and at image or frame rates typical of television or cinema.

Time course of forward suppression

To assess how this "forward suppression" of cell responses might impact on the processing of image sequences, it is necessary to measure the

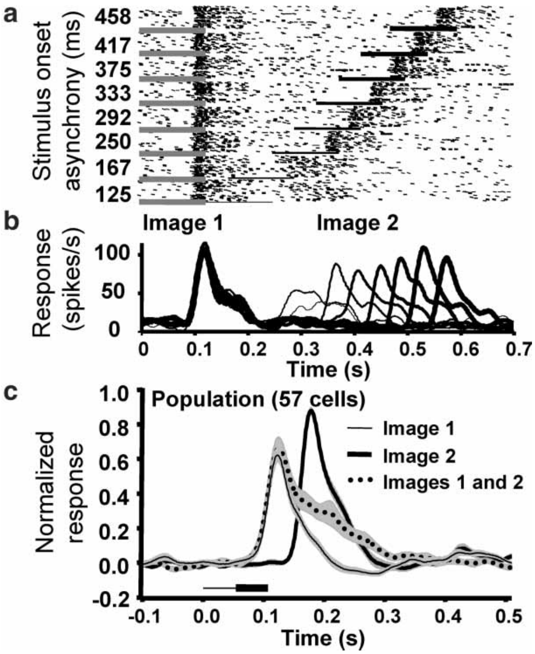


Figure 6. *Cell responses to image pairs. (a) Activity of one cell on 24 trials to a pair of different stimuli (with stimulus onset asynchronies, SOAs, of 125–458 ms); dots record spike times, horizontal thick grey and thin-to-thick black bars show duration of Images 1 and 2 in the pair (mask and target, respectively). (b) Spike density functions for the same cell data. (c) Population responses (mean and shaded* SEM *of spike density functions for 57 cells) to image pairs at 55-ms SOA (dotted line), Image 1 alone (thin line), and Image 2 alone (thick line).*

time course of suppression. We compared responses to the target stimulus in paired and isolated presentations. Since the response to the mask can extend into the period of target response over short intervals, we subtracted the response to the mask tested alone from the paired response, although the effects reported here remain without this subtraction.

A total of 26 cells were tested with 125–458-ms SOAs, and a further 13 cells were tested with shorter SOAs (28–250 ms). For the cell population as a whole, the magnitude and latency of the peak target response following the mask recovered from suppression systematically with mask–target interval (Figure 7, solid symbols). For the cell population, response magnitude to the target recovered with SOA (regression of natural logarithm of the SOA, lnSOA, 125–458 ms against response magnitude, $R^2 = .957$, $p < .001$). Delay of peak response also decayed with SOA (lnSOA regressed against peak delay, $R^2 = .944$, $p < .001$).

The reduction of suppression over time was also evident in the responses of individual cells. A total of 22 of the 26 cells showed a significant ($p < .05$) regression of peak response with natural logarithm of the SOA. For these cells, the duration of the influence of the mask stimulus, estimated from the time taken for the regression line to return to the isolated target response, was 647 ± 104 ms (mean ± *SEM*). From these results it is evident that the forward suppression between one image and following images lasts over an appreciable timescale, longer than half a second. One would therefore expect suppression to affect cellular and perceptual processing of movies over a similar timescale. The duration of suppression may match intervals over which apparent movement can be seen between successive images of people (Stevens, Fonlupt, Shiffrar, & Decety, 2000).

Forward suppression and image similarity

The relative magnitude of responses of temporal cortex cells to different images generally depends on their similarity (Desimone, 1991; Földiák et al., 2003; Logothetis et al., 1995; Tanaka et al., 1991). Interactions between successive images are also likely to depend on similarity (Dragoi et al., 2002; Felsen et al., 2002; Sawamura et al., 2006). For 26 cells, we therefore compared the target response after mask stimuli that produced similar (large) responses to target stimuli to the target response occurring after other mask stimuli that produced dissimilar (small) responses. Note that our starting definition of similarity uses a metric derived from cell response rate. Two images that both produce a large response from a given cell are deemed "similar" for that cell whereas a pair of images producing widely discrepant responses in the cell (one image producing a high response rate and second a low response rate) are deemed dissimilar for that cell. To check that this definition was concordant with perceptual similarity we asked human participants to judge similarity

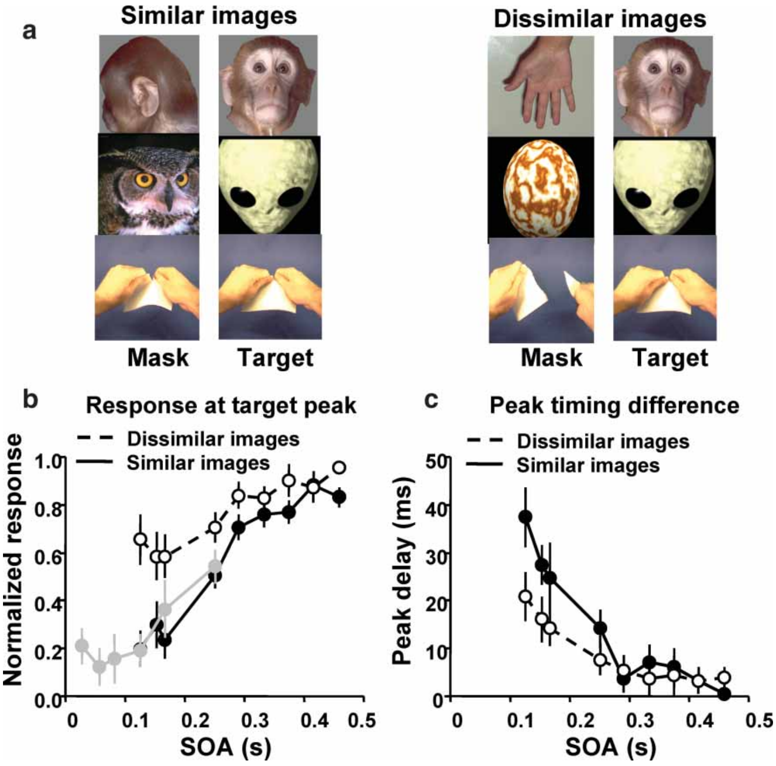


Figure 7. *The time course of suppression of cell responses to similar and dissimilar images. (a) Three rows illustrate examples of image pairs used for three different cells. The left image of each pair formed the first or "mask" and the right image the second or "target" stimulus. For pairs of "similar images" the mask stimulus produced >50% of the cell's response to the target stimulus when tested in isolation. For pairs of "dissimilar images" the mask stimulus produced <50% of the target response. (b) Population response magnitude for the image pair at the time of the peak response to the target alone for long (125–458 ms,* n = *26, black solid circles) and short (28–250 ms,* n = *13, grey solid circles) mask–target stimulus–onset asynchronies (SOAs). Open circles record the target response following dissimilar mask images. (c) Delay in the peak responses to target following mask (SOA 125–458 ms). For SOAs <125 ms the peak of the target response could not be measured reliably.*

between pairs from the trios of images used experimentally to measure cell responses (two images producing large responses and the third image producing a low responses for each cell tested). We confirmed that image pairs eliciting similarly large cell responses were judged to be visually more similar by 11 human observers than image pairs producing differential cellular responses, $t(25) = 8.41$, $p < .00001$. Thus for each cell tested for the impact of similarity on suppression, our measure of similarity in terms of cell response conformed to the perceptual similarity judged by human observers.

We found dissimilar mask stimuli reduced and delayed peak responses to the target but less so than similar mask stimuli (Figure 7, open symbols). Dissimilar images produced less effect on response magnitude—main effect of image similarity, analysis of covariance, ANCOVA, $F(1, 14) = 28.11$, $p < .001$, Similarity × LnSOA interaction, $F(1, 14) = 22.48$, $p < .001$—and on time to peak—similarity main effect, $F(1, 14) = 25.54$, $p < .001$, Similarity × LnSOA interaction, $F(1, 14) = 22.61$, $p < .001$. Thus the forward suppression of neural responses at short intervals is related to image similarity.

Similar images may produce similar responses when tested individually but when presented as a pair, if the first image produces a large cell response, the response to the second image is suppressed. In movies and real life, at each moment the visual image is likely to be highly similar to that occurring in the preceding moments. Since similar images cause maximum forward suppression, we can expect that forward suppression will have a profound effect on the cellular processing (and perception) of movie sequences and unfolding scenes in real life.

Suppression and cell fatigue

The influence of the mask stimulus on the target response could reflect fatigue of the recorded neurons after the mask response; this would explain the greater suppression between similar images. A large response to the first image might induce a refractory period in which the cell is unable to respond well to a following image. The effects on the peak and latency of target responses were, however, found to be similar for trials with "good" (>mean response) and "poor" (<mean response) to the same mask stimuli ($p > .7$). Therefore suppression does not arise from fatigue.

Studies of suppression in other brain areas have also concluded that suppression does not arise from fatigue in the recorded cells (e.g., Movshon & Lennie, 1979). Instead suppression is thought to reflect depression of inputs from neighbouring cells responsive to similar stimuli (Felsen et al., 2002).

The suppression we describe is a prevalent characteristic of temporal lobe cells responsive to complex visual images but it may not be the only form of suppression operating at a high level of visual processing. The relationship of the short-term suppression we describe to long-lasting "repetition suppression" (Li, Miller, & Desimone, 1993; Xiang & Brown, 1998) and priming over extended delays (Grill-Spector et al., 2006) is unclear; however, like all of the suppressive effects it does depend on image resemblance.

Suppressive effects during random image sequences

In random sequences (55 ms/image of 8–30 images) we compared situations where suppression should be more prevalent with situations where suppression should be less prevalent. Specifically, we contrasted responses to effective target stimuli (Figure 8, thin line) following stimuli that also produced substantial responses (masks) with responses to the target stimuli (Figure 8, dotted line) following other stimuli that elicited only weak responses. In the former situation we expect suppressive effects of the response to the mask on the response to target, whereas in the latter situation we expect less suppression.

Despite the continuous nature of rapid serial visual presentation (RSVP) testing with hundreds of successive images (Földiák et al., 2003; Keysers et al., 2001), interactions between mask and target stimuli were still apparent. Responses to consecutive mask and target stimulus pairs initially show temporal summation but then decline significantly ($p < .001$) below that to the target preceded by

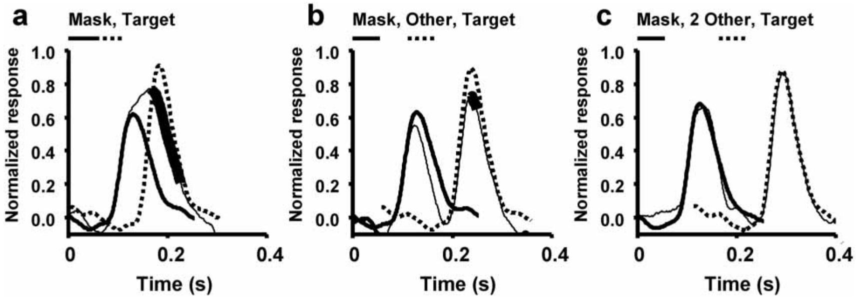


Figure 8. *Population responses to stimulus pairs in random, rapid serial presentation sequences. (a) Average spike density functions (47 cells) to mask (solid line), target (dotted line), and successively paired mask–target (thin line) images. Horizontal solid and dashed bars illustrate the duration of mask and target stimuli (55 ms). The mean response to the target is significantly reduced when it follows the mask (very thick line region),* $p < .001$. *(b–c) Conventions as (a) except the mask–target pair of images (thin line) is separated by one (b, 48 cells) or two other images (c, 53 cells) that produce a weak response (<25% target).*

ineffective stimuli (Figure 8a, thickened region of thin line).

Masking from one image could extend over periods of time in which other images are presented. Alternatively masking from one stimulus could be reset to zero by the presentation of a new image (whether or not this image causes a response). Analysis of the RSVP sequence data with random sequences of images provided evidence that masking extends over successive images. Our measurements revealed significant suppression of the mask stimuli on the target responses despite one intervening (ineffective) image between the mask and target stimuli (Figure 8b). Response suppression was not evident with two intervening images between target and mask (Figure 8c). It is likely that the persistence of masking is underestimated because of the repeated testing of a small set of images in these RSVP studies. Nonetheless they demonstrate that, as found in primary visual cortex (Felsen et al., 2002), the suppressive action of a masking stimulus on target responses in temporal cortex does persist over time filled with another image.

Suppression and image sequence processing

The fact that forward suppression is strongest when successive stimuli are similar would seem to detract from the processing of natural sequences where successive stimuli are necessarily similar. To investigate this paradox further we recorded neural responses to specific scenes embedded in biologically plausible sequences such as movies of walking, head rotations, and hand actions. Cell responses when an image was presented in isolation were compared to the responses when the same image was presented in a sequence.

For a given cell, sequences could start with effective or ineffective images. When sequences commenced with an effective image and continued with progressively less effective images, then responses decayed faster for sequences than did responses to separate images (Figure 9a). For 24 cells tested this way responses to the second and subsequent images in a sequence were smaller ($p < .05$) than responses to the same images presented individually. Such results follow the suppression observed for paired stimuli. An initially large response to a mask image suppresses the response to a target image whether this occurs as the second image of a pair or as a subsequent image in a sequence.

For sequences beginning with images that by themselves produced weak responses, going on to more effective and finally less effective images, then responses to isolated and sequential images were initially equivalent. For individual cell responses (e.g., Figure 9b), however, the peak response for images in sequence occurred earlier than during individual frame testing.

The peak advance was apparent at the population level for 37 cells tested. Alignment of each cell's response to the image that elicited the largest response when presented in isolation (Figures 10a and 10b, fourth stimulus or image number 0) shows that the peak response in the sequence (Figures 10a and 10b, dashed line) occurs before the most effective isolated image is presented. Across different cells, the peak sequence latency occurred 97.1 ± 24.9 ms (mean ± *SEM*) before the peak latency of the most effective separate frame, $t(36) = 3.90$, $p < .001$. Thus the peak response occurred on average 100 ms earlier in sequences although the latency shift varied across different cells.

Eight cells were tested with the same images in forward and reverse sequence directions. For these cells the earlier peak was independent of sequence direction. That is, the maximum activity in the sequence occurred before the most effective single image was reached independent of whether the image sequence was played forward or in reverse. Note that image $+2$ in the forward becomes image -2 when the sequence is reversed.

Although sequence responses appear diminished in Figures 10a and 10b, different cells had different breadths of tuning for individual images, taking 2–5 images to reach the peak. Response alignment in Figures 10a and 10b therefore blurs the peak of the sequence response. Alignment to the most responsive image during the sequence (Figures 10c and 10d) shows that the maximum response attained during sequences was comparable (mean = 89%) to the response to the most

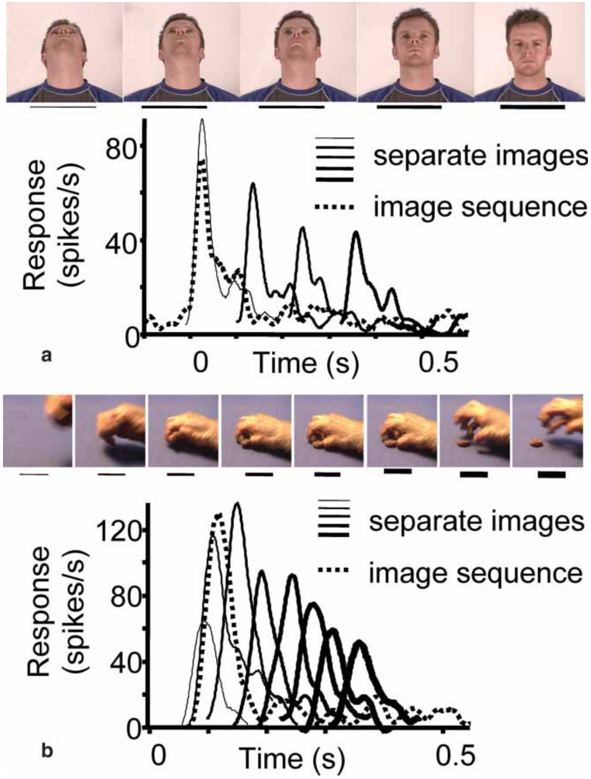


Figure 9. *Individual cell tuning to isolated images and image sequences. (a) Example of single-cell responses to images of declining effectiveness. Upper: test images of a rotation from head up to face the camera. Lower: The mean cell response to the sequence (dotted line) is aligned to onset of first image in the sequence. For comparison the mean cell responses to individual images (thin–thick solid lines) are also aligned to time of occurrence in the sequence. (b) Example of single-cell responses to images that increase and then decrease in effectiveness. Upper: test images of a hand–object interaction. Lower: responses to images presented separately (thin–thick solid lines) or in sequence (dotted line).*

effective single frame, $t(36) = -0.084$, $p = .93$. Thus sequential presentation changes peak latency but not magnitude. During movies, activity is therefore maximal amongst cells tuned to images that are likely to occur in the very near future.

The advance in peak sequence response can be explained by the suppression measured in responses to image pairs. In sequences, neural responses build up but once activity reaches a high level response to subsequent images is curtailed. Sequences of stimuli approaching a cell's optimal input set up two processes, mounting excitation and mounting suppression. Critically, suppressive effects are operating before the maximally effective (isolated) stimulus is reached.

MODELLING NEURONAL RESPONSES IN SEQUENCE

Our explanation of the role of forward suppression in shaping responses to images sequences can be tested quantitatively. Our measurements of the time course and magnitude of suppression between pairs of images provide a basis for predicting the cell responses that we observed in the sequence tests. We therefore assessed the suppression explanation of sequence responses quantitatively. The results show that the main characteristics of earlier response peak during sequences can be explained by forward suppression between pairs of images.

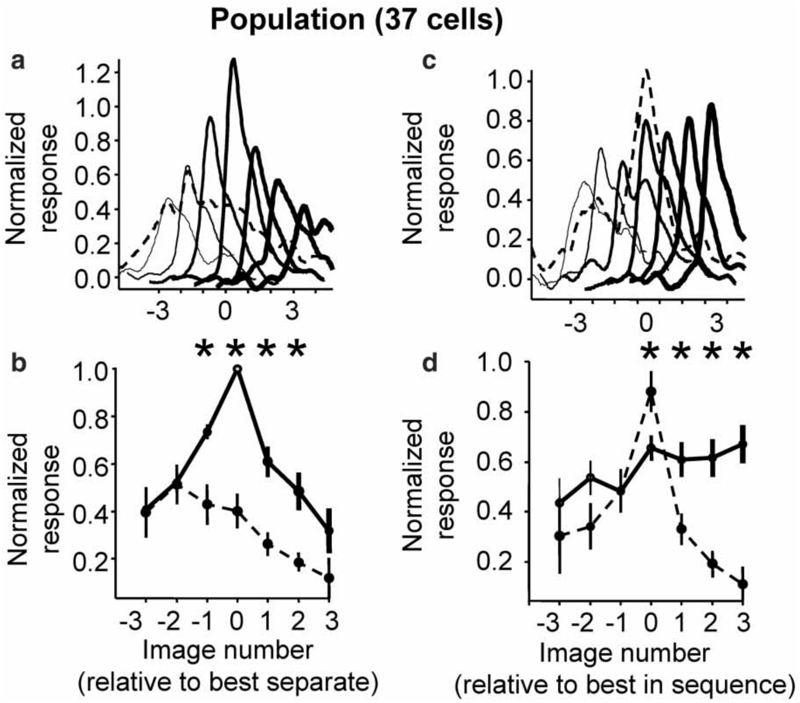


Figure 10. *Cell population tuning to isolated images and image sequences. Combined responses of 37 cells to images of increasing and decreasing efficacy. (a) Spike density functions to test images presented separately (thin–thick lines) or in sequence (dashed line). Responses are aligned to the image producing largest response when tested separately and are normalized in magnitude to the average response for the duration of stimuli. (b) Response magnitude (mean* ± SEM*) calculated for the duration of isolated (solid line) and sequenced images (dashed line). Ordinate (a, b) response magnitude is normalized to each cell's mean response for the duration of presentation of the most effective isolated image. The time course of responses is expressed in terms of image frames (duration* = *56 or 42 ms). (c, d) Conventions as (a, b). Responses are aligned to the image producing the largest response when tested in sequence. Mean responses to sequence and separate images are significantly different* *p < *.005.*

Modelling methods

For each cell we calculated a sequence response on a frame-by-frame basis. The calculated response at any stage in the sequence was obtained from the sum of the ongoing response to the prior images and the response to the current image (scaled by the masking from prior images, see below).

Two main characteristics of suppression were established from physiological experiments (peak suppression and delay, Figure 7). These characteristics defined a masking function (or scaling factor), which was used to compute the suppression of response to the current image based on the magnitude of prior response(s) and interval between prior response(s) and current image. Note that the terms "target" and "mask" derive from consideration of the first and second images in a pair but when considering a sequence of images the term "target" refers to the current image while the term "mask" refers to any prior image.

The modelled response to the first image of the sequence = the response to the same image presented in isolation. The response to the second image = the continuing response to the first image + the response to the second image presented in isolation × the scaling function determined by the magnitude of the response to the first image. The response to the third image in sequence was calculated similarly as the sum of the continuing responses to prior images (Images 1 and 2) + the response to the current image

(Image 3 tested separately) multiplied by suppression from the first image at the interval between first and third images and suppression from the "suppressed" response to the second image. This iterative process is repeated for each image in sequence.

From Figure 7b, we define two levels of mask effectiveness: a similar (effective) mask and a dissimilar (ineffective) mask. We estimate that at zero SOA the average effective mask reduces the target response magnitude from 1.0 to ~0.2; accordingly we set the scaling factor to 0.2. By contrast the average ineffective mask reduces or scales target response to ~0.6. For masks producing responses between the level of effective and ineffective masks the scaling factor was estimated by linear interpolation between 0.2 and 0.6. For very weak masks producing responses less than the average ineffective mask, the scaling factor was proportionally interpolated between 0.6 and 1.0 (scaling of 1.0 representing no masking). The scaling factor for masks producing responses larger than the effective mask was limited to 0.2 (see Figure 11a). Response suppression shows an approximately linear decline with increasing SOA, reaching zero at ~500 ms SOA. The scaling factor was therefore modulated by the interval between mask and target. The modulation resulted in the scaling factor having full impact (i.e., full fractional value) at 0-ms SOA and a zero impact (=1.0) at 500-ms SOA (see Figure 11b).

Figure 7c documents the impact of masks on the delay in the timing of the target response, the peak being delayed most at short SOAs (100 ms), and the delay being most prevalent following large mask responses. Accordingly, we modelled the impact of an effective mask as delaying the peak response to target stimuli by 40 ms and an ineffective mask as introducing a delay of 20 ms. Again, we used a bilinear interpolation to estimate the delays introduced by masks of intermediate effectiveness (see Figure 11c). The delay for masks producing greater responses than the average effective mask was limited to 40 ms. The delay in time for cells to reach peak response to the target declined with increasing interval between mask and target. In simulation, we modulated the peak delay with a linear function maximal at 100-ms SOA and declining to 0 at 300 ms. For mask−target intervals >300 ms no delay in the peak was introduced. For intervals <100 ms the delay was set to that for 100 ms (see Figure 11d).

To calculate the full time course of suppression on the target response, the estimate of the magnitude of neuronal activity at the time of the target response peak for the target frame presented in isolation (a) needs to be combined with the delay calculated for the peak of the suppressed target response (b). This was done linearly. The firing rate of the suppressed response was modelled to rise from the neuron's onset latency, pass through the response amplitude calculated at the time of the peak of the isolated unmasked response (a), and continue until the peak time was reached (b). From this value the suppressed target response was set to trace decay of the target response measured in isolation. This procedure effectively removes the transient component of the target response at short SOAs. Such abolition of the response transient was evident in cell responses (e.g., Figure 6).

Modelling results

The observed response to sequences (Figure 12, dotted line) is the average of observed cell responses to image sequences where effectiveness of separate images increments and then declines. The modelled response to sequences (Figure 12, thin line) is the average of the cells' predicted responses to the image sequences. The predicted sequence response, based on modelling the suppressive action between image pairs, has the main features of the data obtained empirically from image sequences. The modelling[1] accounts

[1] The modelling appears to ignore backward masking (Figure 5a). We measure response to Image 2 (targets) after subtracting the isolated response to Image 1 (masks). This procedure means that the estimates of forward masking on Image 2 include any backward masking effects on the Image 1 response. Therefore the calculation of sequence responses includes both forward and backward masking.

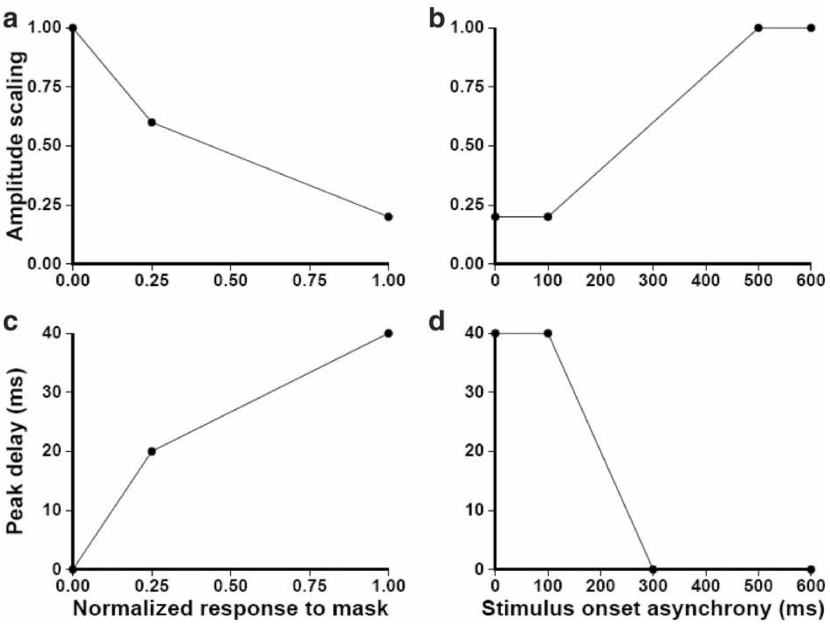


Figure 11. *Parameters of cell response suppression chosen for modelling. Response to a target stimulus was scaled in amplitude as a function of (a) the magnitude of the response to a mask stimulus and (b) the time interval between onsets of mask and target stimuli. The peak response to a target stimulus was delayed in time as a function of (c) the magnitude of the response to mask stimuli and (d) the time interval between onsets of mask and target stimuli.*

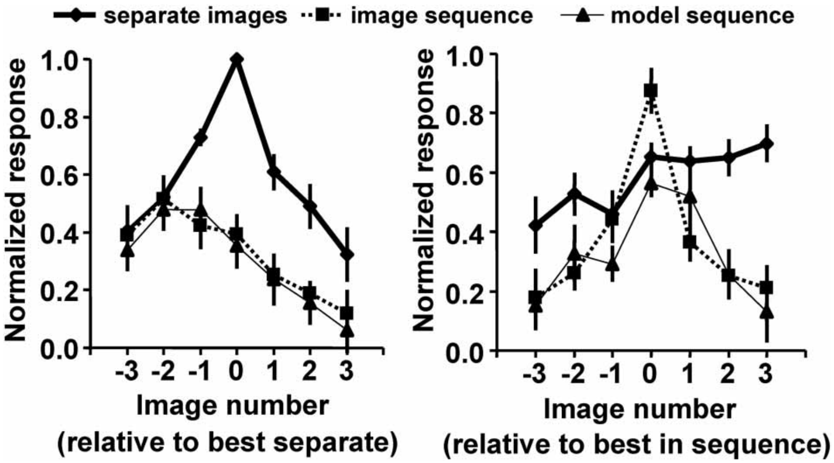


Figure 12. *Modelled cellular responses to sequences. Average responses of 37 cells to images where effectiveness of separate images increments and then declines. Response magnitude (mean ± SEM) was measured for the duration of separate (thick line) and sequenced images (dotted line). Responses are aligned to the image producing the largest response when tested separately (data and conventions as Figure 10b). Component cell responses were normalized to the average response during presentation of most effective separate images. Abscissa: Time course is expressed in terms of image frames (duration = 56 or 42 ms). The modelled response to sequences (thin line) is computed from the responses to separate images and a masking function established empirically from the interaction between pairs of images.*

for the earlier peak response during image sequences and a faster decline of response than for separate images. Thus what we have learned from interactions between pairs of images can explain cell responses to long sequences of images that change naturally as actions unfold.

Discrepancies between predicted and measured responses during sequences may arise from several sources. We have not measured suppression at a range of intermediate levels of mask effectiveness.

HUMAN FACE DETECTION DURING IMAGE SEQUENCES

The cellular studies indicate that activity in cells representing faces will be heightened before the target face view is reached during an ordered rotation of head views. This should lead to faster report of faces when present but also a bias to report face presence over absence in natural rotation sequences compared to isolated view presentations.

One can also make predictions of faster and biased reports for detecting a target face in a natural head view rotation sequence compared to random sequences of head view images, or indeed situations where an ordered sequence of numerical digits terminates in a target face. The faster and biased reporting of face targets in natural sequences should occur because in the run up to the target view, the images (from random head views or digits) are on average less face-like than those occurring in an ordered head view sequence. We tested and confirmed these predictions for sequence perception.

Psychophysical methods

Stimuli including eight head views (in 45° rotation steps from full face) of 9 individuals and digits 1,2,3,4 at 256-pixel height were presented on a mid-grey background. Test sequences (Figure 13) contained six images presented at 50 ms/image (a presentation rate corresponding to the majority of physiological testing; different image presentation rates are to be considered

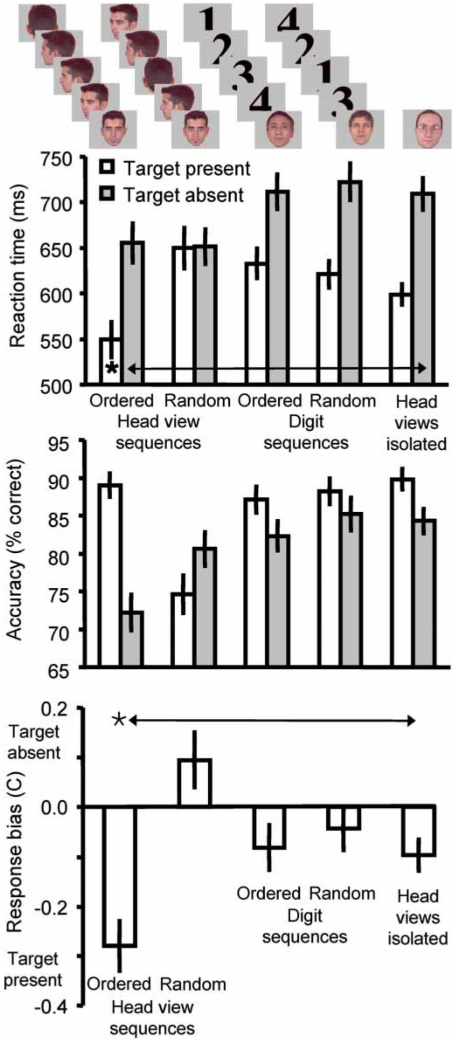


Figure 13. *Sequence effects on target face detection. Upper: stimuli consisting of a "target" face or a "nontarget" alternative head view (not shown) preceded by sequences of four head views rotating in 45° steps from the back view towards the face (ordered); the same views shuffled (random); digits 1–4 in ascending order; digits in random order; or no prior images (isolated). A scrambled face (not shown) followed targets and nontargets. (a) Reaction time to give correct reports of face views (target present) and alternative head views (target absent). Reaction times were shortest for ordered head view sequences compared to all other conditions (✱* p < *.05). (b) Accuracy of performance; conventions as (a). (c) Sensitivity independent criterion (*C*) as a function of stimulus type (negative values reflect a bias to report face presence; positive values a bias to report face absence). Bias to report the faces was greatest in ordered head view sequences compared to all other conditions (*⋆ p < *.0005).*

elsewhere). The first four images were a clockwise or anticlockwise rotation from the back view towards the face, the same images randomized, and digits 1–4 in ascending or random order. The fifth image was a "target" face view or an equally frequent "nontarget" head view (randomly chosen from the first three images of the preceding sequence). The terminal sixth image was a scrambled face split into 16 blocks and rearranged. Isolated head views were presented followed by a scrambled face.

A total of 88 participants were tested with head view sequences, digit sequences, and isolated head views in equal frequency. Participants practised with on-screen error feedback until 80% correct before beginning the experiment without feedback. Participants were encouraged to respond as quickly as possible for target presence (left hand pressing the \ key) or absence (right hand, / key) and completed three or more trials in each condition. Responses were discarded from speed analysis if the reaction time was >3 standard deviations from the participant's mean. A total of 10 participants did not maintain 80% accuracy, and their data were discarded. Speed, accuracy, sensitivity (d'), and bias (C) in performance were compared across conditions using repeated measures analysis of variance (ANOVA) with within-subject factors of sequence type (order, random, ordered digit, random digit sequences, and single isolated image).

Psychophysical results

Overall, reaction time varied with sequence type: single, ordered digits, random digits, $F(4, 308) = 10.5$, $p < .005$, and target-present versus target-absent trials, $F(1, 77) = 47.1$, $p < .005$, the effect of sequence type depending on whether the target was present or absent: interaction $F(4, 308) = 6.9$, $p < .005$ (see Figure 13a). Reaction times following ordered head view sequences in both target-present and target-absent trials were faster than those seen to single image trials or following digit sequences (simple effect analysis, all ps < .05). Reaction times to target-present trials following ordered head view sequences were also faster than those following random sequences of head views (549 ± 22 ms vs. 650 ± 25 ms, simple effect analysis, $p < .05$), but this reaction time advantage was not seen in target-absent trials (656 ± 24 ms vs. 651 ± 21 ms). The speed advantage for behavioural detection of target following ordered head view sequences compared to random sequences (100 ± 17 ms) is of the same order of magnitude as the peak latency shift observed in cell responses (97 ± 25 ms), supporting the idea that the propensity to anticipate a stimulus in predictable sequences results at least in part from the cell activity changes.

While reaction times to single images and following digit sequences (both ordered and random) showed faster reaction times to target-present trials than target-absent trials, reaction times did not depend on these three sequence types—separate ANOVA: target present/absent, $F(1, 77) = 71.2$, $p < .0005$; effect of sequence type, $F(2, 154) = 1.8$, *ns*; interaction, $F(2, 154) = 1.3$, *ns*. Hence, the speed gain in reacting to targets does not depend on timing cues that could be gained from any ordered sequence; the advantage appears to depend on the visual similarity of the prior images to face targets.

Accuracy, like reaction time, varied with sequence type, $F(4, 308) = 9.1$, $p < .005$, target-present versus target-absent trials, $F(1, 77) = 6.3$, $p < .02$, and their interaction, $F(4, 308) = 9.4$, $p < .005$ (see Figure 13b). Increased detection speed following ordered sequences of head views compared to single image presentation did not reflect lower accuracy on face target-present trials (ordered: 89 ± 2%; single image 90 ± 2%, *ns*) but was accompanied by decreased accuracy (increased false positives) on target-absent trials (ordered: 72 ± 3%; single image 84 ± 2%, simple effects analysis $p < .05$). Accuracy to target-present trials following random sequences (75 ± 3%) was reduced compared to all other sequence types (simple effects analysis, all ps < .05), which were in turn equivalent (all ps > .05). A random sequence of head views did not influence target-absent accuracy compared to single frames (81 ± 3% vs. 84 ± 2%, respectively) nor following the digit sequences (all ps > .05).

Accuracy to single images and following ordered or random digit sequences was higher on target-present trials than on target-absent trials, and there was no effect of sequence type—separate ANOVA: target present/absent, $F(1, 77) = 4.3$, $p < .05$; effect of sequence type: single, ordered digits, random digits, $F(2, 154) = 1.1$, *ns*; interaction, $F(2, 154) = 0.4$, *ns*. Thus, as with reaction times, preceding sequences of images influenced accuracy of perception only if the preceding sequence was of related images.

The differential impact of sequence type on accuracy with target present/absent is captured by the bias statistic *C*. One-way ANOVA—effect of sequence type: single, ordered, and random head views, ordered and random digits, $F(4, 308) = 9.6$, $p < .005$ (Figure 13c)—revealed that, compared to single frames, ordered head views produced a greater bias to report target present whereas random head views produced a greater bias to report target absent (planned comparison, all *p*s < .005). Sequences of digits, whether ordered or random, did not induce a change in bias compared to single frame presentations (all *p*s > .2). This analysis confirms predictions that ordered sequences bias observer to confirm the presence of targets that naturally follow from the prior sequence.

While this bias is accompanied by a faster detection of targets when present, it occurs at a cost to accuracy in performance when targets are absent. The combined accuracy for target-present and target-bsent trials is specified by the sensitivity index d'. One-way ANOVA showed sensitivity depended on sequence type—comparing d' for single, ordered, and random head views, ordered and random digits trial types; $F(4, 308) = 9.2$, $p < .005$—being reduced following sequences of head views (both ordered, 1.9 ± 0.1, and random, 1.7 ± 0.1) compared to all other conditions (single frames, 2.4 ± 0.1; ordered digits, 2.2 ± 0.1; random digits, 2.3 ± 0.1). For random sequences, the reduction in sensitivity is from a combination of reduced accuracy to both target-present and target-absent trials. Such lowered accuracy can be taken to reflect "forward masking" seen at the cellular level. We note that for ordered sequences, the reduction in sensitivity is due to increased false positives on target-absent trials, not a reduction in accuracy to reporting targets when present.

In the real world, the face view would almost invariably be seen after witnessing the head turning towards the face. An object rotating may decelerate, stop, and progress to an adjacent view, but there is no possibility of a large and random view change. In other words, the decreased accuracy in target-absent trials is unlikely to have major detrimental consequences in situations outside experimental psychology where large, random view changes are possible.

To summarize our perceptual studies, we confirm that human observers show anticipation within the same image sequences as those used in our cell studies. Anticipation was apparent as a faster report of faces when present but also a bias to report face presence over absence in natural rotation sequences compared to isolated view presentations and unnatural sequences. Face detection following ordered and random digit sequences and single frames did not differ in speed, sensitivity, or bias, which suggests that the anticipation in sequences reflects visual and conceptual continuity that is absent when a series of numbers ends in a face.

GENERAL DISCUSSION

The first part of the paper reviewed the organization and selectivity of cells responsive to faces. Studies of such cells have presented simple accounts of diverse psychological phenomena. For example, recognition speed for unusual views of familiar objects can be accounted for by the activity of these "little grey" cells without the need to postulate elaborate processes such as "mental rotation" or "size zooming" (Perrett, Oram, & Wachsmuth, 1998). Study of the cell activity in temporal cortex has also advanced our understanding of attention (e.g., Duncan, 2006).

We focused on the ability of cells to respond to images presented very briefly. Here, knowledge of cell activity helps the understanding of perception.

The responses to brief stimuli allowed clarification of the extent of cell selectivity for faces. It also allowed human perceptual performance to be compared to cellular performance under conditions of rapid serial visual presentations of randomly related images. These comparisons show marked parallels in (a) the degradation of cell responses and human perceptual report as image presentation rate increases and (b) the preservation of cell responses and perception despite gaps between successive images. These studies give insight into the graded nature of visual awareness down to levels close to psychophysical threshold. They inform us about the biological basis of our conscious experience and our iconic memory across brief gaps in experience, for example while we blink.

In the heart of the current paper we collated results from studies of how cell processing of one image affects processing of subsequent images. Again our studies attempt to relate cell activity to human perceptual performance. In so doing we hope to begin an account of the experience that we refer to as “anticipation”.

Our study has focused on the nature of neural representations of complex meaningful images occurring in succession. We find that a neural response to one image suppresses responses to similar images for a short period of time. We show that a consequence of short-term suppression (Felsen et al., 2002) in a natural sequence is a relative increase in the activity of cells tuned to inputs about to occur. During sequences the distorted pattern of neural activity will represent probable future inputs rather than the current input. In other words, temporal interactions amongst neurons not only change sensitivity (Dragoi et al., 2002; Felsen et al., 2002) but also produce a bias in representations that is consistent with anticipation (Freyd & Finke, 1984; Verfaillie & Daems, 2002).

Short-term suppression or adaptation of neural responses is present in different forms throughout the nervous system (Dragoi et al., 2002; Felsen et al., 2002; Grill-Spector et al., 2006; Hosoya, Baccus, & Meister, 2005; Khatri et al., 2004; Kourtzi & Kanwisher, 2001; Macknik & Livingstone, 1998; Sawamura et al., 2006; Turvey, 1973; Wehr & Zador, 2005), and thus anticipatory coding will occur at each level of sensory analysis, from the retina (Berry, Brivanlou, Jordan, & Meister, 1999) through to the highest levels of cortical elaboration.

Observers may predict the outcomes of actions of others by “simulating” the motor programmes involved (Rizzolatti, Fogassi, & Gallese, 2001; Wolpert, Ghahramani, & Jordan, 1995) but prediction in perceptual systems need not require simulation (Fogassi et al., 2005). The anticipatory effects we describe in temporal cortex arise from interactions between successive views of the world. Taking response latency into account, we find that at any instant during the sight of natural actions the maximally responding cells are those selective for postures that have yet to occur. While watching movies, brain activity will therefore be maximal for images shortly to appear. Modelling shows that this anticipatory effect is attributable to the forward suppression of neural responses without the need for motor simulation. This sheds new light on the functional consequence of forward masking and reveals a simple mechanism for how the brain implements predictive computations.

The world has natural sequences and does not have a succession of random visual events. We presume that cell properties evolved to reflect natural contingencies in the world. Under this view the paradox of masking disappears. A cellular mechanism that causes suppression of response to successive similar images will inevitably affect temporal perception of sequences and produce anticipatory behaviour. The focus on the disruptive effect of masking between a pair of images has prevented appreciation of the advantage conferred by suppression in longer, naturally ordered sequences of images. Far from degrading sensory performance, the cellular mechanisms underlying masking actually benefit perception through anticipation.

Our physiological studies have general implications for sensory coding. Bell-shaped tuning functions for isolated stimuli are common to neurons in both biological and artificial sensory systems (Desimone, 1991; Giese & Poggio, 2003; Logothetis et al., 1995; Riesenhuber & Poggio, 1999; Tanaka et al., 1991). Our results, however, suggest that notions of neural receptive fields and tuning for discrete stimuli may need reformulating to reflect a predictive function during the processing of more natural and continuously changing stimuli. We conclude that in a changing perceptual world, the values represented by cells' activities reflect a predicted future state rather than the present reality.

Original manuscript received 31 March 2008
Accepted revision 8 April 2009
First published online 23 June 2009

REFERENCES

Berry, M. J., Brivanlou, I. H., Jordan, T. A., & Meister, M. (1999). Anticipation of moving stimuli by the retina. *Nature*, *398*, 334–338.

Desimone, R. (1991). Face-selective cells in the temporal cortex of monkeys. *Journal of Cognitive Neuroscience*, *3*, 1–8.

Dragoi, V., Sharma, J., Miller, E. K., & Sur, M. (2002). Dynamics of neuronal sensitivity in visual cortex and local feature discrimination *Nature Neuroscience*, *5*, 883–891.

Duncan, J. (2006). EPS Mid-Career Award 2004: Brain mechanisms of attention. *Quarterly Journal of Experimental Psychology*, *59*, 2–27.

Felsen, G., Shen, Y., Yao, H., Spor, G., Li, C., & Dan, Y. (2002). Dynamic modification of cortical orientation tuning mediated by recurrent connections. *Neuron*, *36*, 945–954.

Fogassi, F., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., & Rizzolatti, G. (2005). Parietal lobe: From action organization to intention understanding. *Science*, *308*, 662–667.

Földiák, P., Xiao, D.-K., Keysers, C., Edwards, R., & Perrett, D. I. (2003). Rapid serial visual presentation for the determination of neural selectivity in area STSa. *Progress in Brain Research*, *144*, 107–116.

Freyd, J. J., & Finke, R. A. (1984). Representational momentum. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 126–132.

Fujita, I., Tanaka, K., Ito, M., & Cheng, K. (1992). Columns for visual features of objects in monkey inferotemporal cortex. *Nature*, *360*, 343–346.

Giese, M. A., & Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nature Reviews. Neuroscience*, *4*, 179–192.

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: Neural models of stimulus-specific effects. *Trends in Cognitive Science*, *10*, 14–23.

Gross, C. G. (2008). Single neuron studies of inferior temporal cortex. *Neuropsychologia*, *46*, 841–852.

Gross, C. G., Rocha-Miranda, C. E., & Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the macaque. *Journal of Neurophysiology*, *35*, 96–111.

Guo, K., Nevado, A., Robertson, R. G., Pulgarin, M., Thiele, A., & Young, M. P. (2004). Effects on orientation perception of manipulating the spatio-temporal prior probability of stimuli. *Vision Research*, *44*, 2349–2358.

Harries, M. H., & Perrett, D. I. (1991). Modular organization of face processing in temporal cortex: Physiological evidence and possible anatomical correlates. *Journal of Cognitive Neuroscience*, *3*, 9–24.

Hosoya, T., Baccus, S. A., & Meister, M. (2005). Dynamic predictive coding by the retina. *Nature*, *436*, 71–77.

Jellema, T., Baker, C. I., Wicker, B., & Perrett, D. I. (2000). Neural representation for the perception of the intentionality of hand actions. *Brain and Cognition*, *44*, 280–302.

Jellema, T., & Perrett, D. I. (2006). Neural representations of perceived bodily actions using a categorical frame of reference. *Neuropsychologia*, *44*, 1535–1546.

Keysers, C., & Perrett, D. I. (2002). Visual masking and RSVP reveal neural competition. *Trends in Cognitive Science*, *6*, 120–125.

Keysers, C., Xiao, D.-K., Földiák, P., & Perrett, D. I. (2001). The speed of sight. *Journal of Cognitive Neuroscience*, *13*, 90–101.

Keysers, C., Xiao, D.-K., Földiák, P., & Perrett, D. I. (2005). Out of sight but not out of mind: The neurophysiology of iconic memory in the superior temporal sulcus. *Cognitive Neuropsychology*, *22*, 316–332.

Khatri, V., Hartings, J. A., & Simons, D. J. (2004). Adaptation in thalamic barreloid and cortical barrel neurons to periodic whisker deflections varying in frequency and velocity. *Journal of Neurophysiology*, *92*, 3244–3254.

Kiani, R., Esteky, H., Mirpour, K., & Tanaka, K. (2007). Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *Journal of Neurophysiology*, *97*, 4296–4309.

Kouider, S., & Dehaene, S. (2007). Levels of processing during non-conscious perception: A critical review of visual masking. *Philosophical Transactions of the Royal Society of London*, *B*, *362*, 857–875.

Kourtzi, Z., & Kanwisher, N. (2001). Representation of perceived object shape by the human lateral occipital complex. *Science*, *293*, 1506–1509.

Leekam, S., Baron-Cohen, S., Perrett, D. I., Milders, M., & Brown, S. (1997). Eye-direction detection: A dissociation between geometric and joint attention skills in autism. *British Journal of Developmental Psychology*, *15*, 77–95.

Li, L., Miller, E. K., & Desimone, R. (1993). The representation of stimulus familiarity in anterior inferior temporal cortex. *Journal of Neurophysiology*, *69*, 1918–1929.

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

Macknik, S. L., & Livingstone, M. S. (1998). Neural correlates of visibility and invisibility in the primate visual system. *Nature Neuroscience*, *1*, 144–149.

Movshon, J. A., & Lennie, P. (1979). Pattern-selective adaptation in visual cortical neurones. *Nature*, *278*, 850–852.

Oram, M. W., & Perrett, D. I. (1996). Integration of form and motion in the anterior superior temporal polysensory area (STPa) of the macaque monkey. *Journal of Neurophysiology*, *76*, 109–129.

Perrett, D. I., Benson, P. J., Hietanen, J. K., Oram, M. W., & Dittrich, W. H. (1995). When is a face not a face? In R. Gregory, J. Harris, P. Heard, & D. Rose (Eds.), *The artful eye* (pp. 95–124). Oxford, UK: Oxford University Press.

Perrett, D. I., Harries, M. H., Bevan, R., Thomas, S., Benson, P. J., Mistlin, A. J., et al. (1989). Frameworks of analysis for the neural representation of animate objects and actions. *Journal Experimental Biology*, *146*, 87–114.

Perrett, D. I., Hietanen, J. K., Oram, M. W., & Benson, P. J. (1992). Organization and functions of cells responsive to faces in the temporal cortex. *Philosophical Transactions of the Royal Society of London*, *335*, 23–30.

Perrett, D. I., Oram, M. W., & Wachsmuth, E. (1998). Evidence accumulation in cell populations responsive to faces: An account of generalisation of recognition without mental transformations. *Cognition*, *67*, 111–145.

Perrett, D. I., Rolls, E. T., & Caan, W. (1979). Temporal lobe cells of the monkey with visual responses selective for faces. *Neuroscience Letters*, *3*, S358.

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

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

Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*, *435*, 1102–1107.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition. *Nature Neuroscience*, *2*, 1019–1025.

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews. Neuroscience*, *2*, 661–670.

Rozzi, S., Calzavara, R., Belmalih, A., Borra, E., Gregoriou, G. G., Matelli, M., et al. (2006). Cortical connections of the inferior parietal cortical convexity of the macaque monkey. *Cerebral Cortex*, *16*, 1389–1417.

Sakai, K., & Miyashita, Y. (1991). Neural organization for the long-term-memory of paired associates. *Nature*, *354*, 152–155.

Sawamura, H., Orban, G. A., & Vogels, R. (2006). Selectivity of neuronal adaptation does not match response selectivity: A single-cell study of the fMRI adaptation paradigm. *Neuron*, *49*, 307–318.

Stevens, J. A., Fonlupt, P., Shiffrar, M., & Decety, J. (2000). New aspects of motion perception: Selective neural encoding of apparent human movements. *Neuroreport*, *11*, 109–115.

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.

Tsao, D. Y., Freiwald, W. A., Tootell, R. B. H., & Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. *Science*, *311*, 670–674.

Turvey, M. T. (1973). On the peripheral and central processes in vision: Inferences from an information processing analysis of masking with patterned stimuli. *Psychological Reviews*, *80*, 1–52.

Verfaillie, K., & Daems, A. (2002). Representing and anticipating human actions in vision. *Visual Cognition*, *9*, 217–232.

Wachsmuth, E., Oram, M. W., & Perrett, D. I. (1994). Recognition of objects by their component parts: Responses of single units in the temporal cortex of the macaque. *Cerebral Cortex*, *4*, 502–522.

Wang, G., Tanaka, K., & Tanifuji, M. (1996). Optical imaging of functional organization in the monkey inferotemporal cortex. *Science*, *272*, 1665–1668.

Wehr, M., & Zador, A. M. (2005). Synaptic mechanisms of forward suppression in rat auditory cortex. *Neuron*, *47*, 437–445.

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science*, *269*, 1880–1882.

Xiang, J. Z., & Brown, M. W. (1998). Differential neuronal encoding of novelty, familiarity and recency in regions of the anterior temporal lobe. *Neuropharmacology*, *37*, 657–676.