How parallel is visual processing in the ventral pathway?

Guillaume A. Rousselet[1], Simon J. Thorpe[2] and Michèle Fabre-Thorpe[2]

[1]McMaster University, Department of Psychology, Hamilton, Ontario L8S 4K1, Canada
[2]Centre de Recherche Cerveau & Cognition (UMR 5549, CNRS-UPS), Faculté de Médecine de Rangueil, 133 route de Narbonne, 31062 Toulouse, France

Visual object perception is usually studied by presenting one object at a time at the fovea. However, the world around us is composed of multiple objects. The way our visual system deals with this complexity has remained controversial in the literature. Some models claim that the ventral pathway, a set of visual cortical areas responsible for object recognition, can process only one or very few objects at a time without ambiguity. Other models argue in favor of a massively parallel processing of objects in a scene. Recent experiments in monkeys have provided important data about this issue. The ventral pathway seems to be able to perform complex analyses on several objects simultaneously, but only during a short time period. Subsequently only one or very few objects are explicitly selected and consciously perceived. Here, we survey the implications of these new findings for our understanding of object processing.

Our capacity to identify visual objects relies on a large neural network, from the retina to the prefrontal cortex, which includes the so-called 'ventral pathway', a set of posterior cortical areas extending from the primary visual cortex (V1) to the infero-temporal (IT or TE) cortex. The ventral pathway can be characterized by a hierarchical architecture in which neurons in higher areas code for progressively more complex representations by pooling information from lower areas [1,2]. Hence, neurons in V1 code for relatively simple features such as local contours and colors, whereas neurons in TE fire in response to whole complex objects (Figure 1).

Another important characteristic of this hierarchical organization is the size of the receptive fields (RFs; the area of the visual field to which a neuron responds). In V1, neurons respond to stimuli appearing in a small patch of the visual field; that is, they have small RFs and thus provide precise information about the position of the stimulus in the visual field. However, because TE neurons receive massive convergent inputs from lower areas, they integrate information from larger areas of the visual field and consequently have large RFs. This might underlie our capacity to recognize an object independently of its size and position (Figure 1). But large RFs could be a problem because it is commonly assumed that they do not preserve

spatial information, leading to the so-called 'binding problem'. Assuming that TE neurons do not preserve spatial information, a neuron with a large RF responding selectively to the conjunction of two elements 'AB' could potentially be activated by the simultaneous presentation of 'AC' and 'DB' [3]. This raises a fundamental question for models of objects processing: is the ventral pathway able to represent multiple objects simultaneously without ambiguity? Recent experiments reviewed here provide new key ideas on this topic. They suggest that a great deal of spatial information might in fact be preserved in TE, strengthening the case for parallel processing of visual objects. However, other constraints seem to limit such parallel processing to a short period of time after stimulus onset.

Isolated objects and the binding problem

In laboratory conditions, when the visual system is stimulated with a single object, it seems that there is no binding problem, because all low-level elements in the visual field belong to the same object. However, one still needs a mechanism to organize correctly those elements in space. Indeed, a large number of TE neurons appear to be sensitive to specific element configurations such that a neuron might respond to element A above B (e.g. a circle and a square), but not to B above A [4,5]. It is striking that such units can code for conjunctions of spatially organized elements with very short latencies (on the order of 80–100 ms) suggesting that a feedforward pass through the ventral pathway might very often be enough to trigger selective responses [6,7] (Figures 1 and 2).

Computational models have suggested how this feedforward selectivity could be established through learning by progressively increasing feature complexity and pooling information from larger parts of the visual field [8–10] (Figure 2). Fast coding of element conjunctions is possible with neurons sensitive to the spatial organization of their inputs [11] and a biologically plausible way to achieve this goal has been suggested recently [12]. This hypothesis is reinforced by two key properties of neurons in V1. First, they encode simultaneously low-level properties such as orientation, color and spatial frequency [13–15]. Second, they integrate information from outside their classical RF, making inferences about contours for example [16]. Hence, the binding of the different elements composing an object starts already in V1. Then, in V2, V4, and

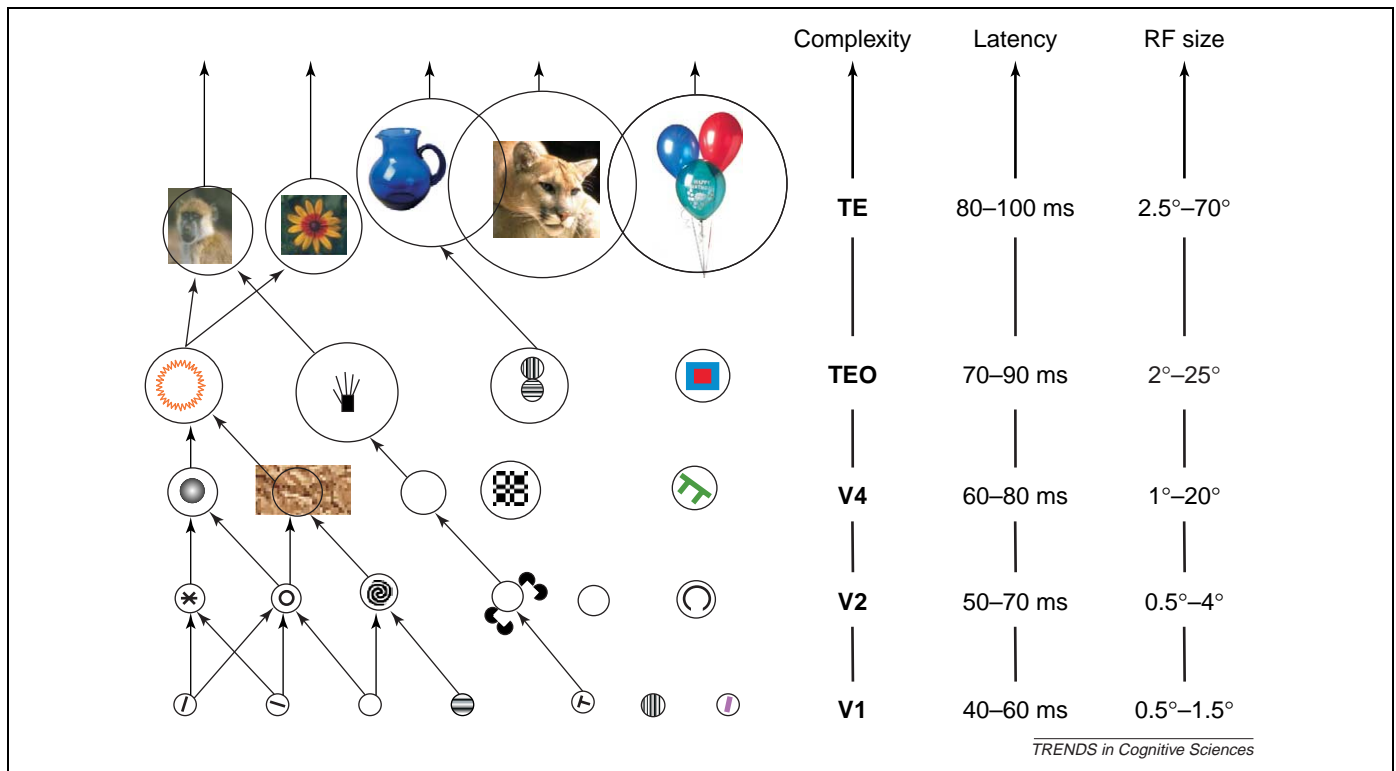*Corresponding author:* Guillaume A. Rousselet (rousseg@mcmaster.ca).

Figure 1. Schematic representation of the ventral visual pathway. Through a hierarchy of cortical areas, from V1 through temporal-occipital cortex (TEO) to temporal cortex (TE), complex and invariant object representations are progressively built by integrating convergent inputs from lower levels. Examples of elements for which neurons respond selectively are represented inside receptive fields (RFs; represented by circles) of different sizes. Feedback and horizontal connections are not shown but are often essential to build up object representations. The first column of bottom-up arrows on the right indicates the progressive increase in the 'complexity' of the neuronal representations. In the second column, figures are an estimate of the minimum and the average response latencies reported in the literature [7]. Although they suggest a serial organization, there is considerable overlap in response latencies at different levels. In the third column are estimates of the smallest and largest RF sizes reported in the literature. As RF size increases (presumably to allow translation and size invariance), neurons at higher levels typically receive inputs from more than one object at a time during natural scene perception. This is generally considered as a problem for representing several objects without ambiguity. However, many TE neurons in fact have small RFs at different positions, and neurons along the ventral pathway do not code for one characteristic of the input at a time but typically encode conjunctions of elements, thus limiting 'binding' problems. Finally, it is important to note that area TE can be subdivided in smaller areas (not shown), based on anatomical and functional properties, suggesting the existence of sub-pathways within the ventral stream.

temporal lobe areas TEO and TE, neurons code for more and more complex conjunctions of elements processed in the previous stages.

Spatial information in TE

The ventral pathway is able to build very rapidly the representation of an isolated object, finally encoded in TE neurons. But because TE neurons have large RFs, ranging from about $10°$ to $30°$ and often more (e.g. [17–19]), one might question whether the ventral pathway is able to deal with two or more objects simultaneously. Indeed, it is tempting to assume that spatial information is lost in the large RFs of TE neurons, making it impossible to combine the elements from different objects without errors. However, large RFs are not inconsistent with the coding of spatial information (Figure 3). Indeed, TE responses typically decrease in amplitude with the stimulus distance from the RF center, suggesting that these neurons at least encode the distance between the object and the RF center. Some TE neurons even provide a strong response only in a sub-part of their RF, or 'hot-spot', which might define a computationally effective RF smaller than the full RF. Having a variety of RF positions provides another source of spatial information. Indeed, although the position of a majority of RFs covers the center of vision, including the fovea, they can be shifted up to about $8°$ in the

parafovea [19]. Thus, even with large RFs, position information can be recovered by combining the responses of several neurons with overlapping RFs, at least in the central part of the visual field. RFs are also strongly biased towards the contralateral hemifield, suggesting a division of labor between the two hemispheres (Box 1). Finally, during the fixation of a target, many TE units modulate their responses according to the position of the eyes [20]. These different sources of spatial information can potentially be combined to provide reliable position information.

Although early studies have described RFs in area TE as being large, it is only recently that a fully systematic study has been performed [19]. This study found a high degree of variability in RF size, with a mean of about $10° \pm 5°$ (SD), a minimum of $2.8°$ and a maximum of $26°$. Other teams have also reported a large variability in RF size (mean $\pm$ SD): $24.5° \pm 15.7°$ [21]; $16.5° \pm 6.1°$ [18]; and $13.6° \pm 7.3°$ [22].

Therefore, some RFs are actually rather small. For instance, after training with a set of new stimuli, neurons were recorded with RF sizes ranging from around $4°$ to $15°$ [23]. More dramatically, DiCarlo and Maunsell trained monkeys to discriminate small stimuli ($<1°$) and found that almost all TE neurons recorded had strong sensitivity to small ($1.5°$) position changes [24]. They estimated RF
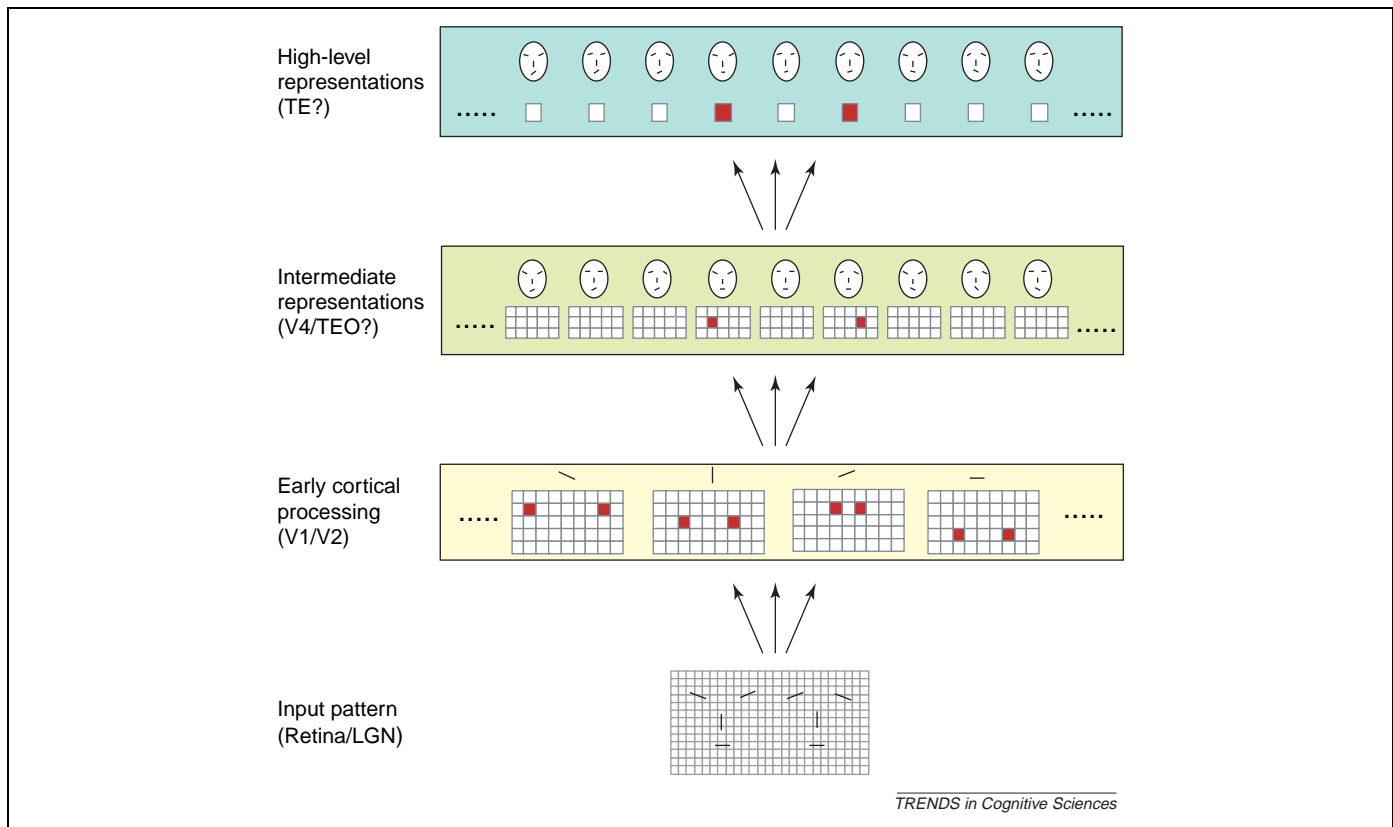
Figure 2. Multiple representations in the visual system. A simple hierarchical architecture illustrating how the visual system might represent multiple objects and avoid the binding problem. The input pattern at the retina is composed of 8 oriented segments. These segments are then represented at the next processing level (roughly equivalent to V1 and V2) by the activity of units in 4 maps of units, selective to left oblique, vertical, right oblique and horizontal orientations. At this point, no hypothesis has been made concerning the nature or number of objects present. However, if the next stage (roughly corresponding to V4/TEO) contains neurons that have learnt to respond to particular diagnostic configurations of activity, the presence of two face-like objects can be rapidly determined. These neurons are still at least partially spatially selective. If necessary, we can add a final level (perhaps corresponding to TE) where the neurons pool together responses from neurons sharing similar selectivities, to produce object specific responses that are fully invariant to changes in position. Such a system is able to handle multiple objects simultaneously, but could be configured to select one particular output if required, by introducing competitive interactions between neurons at each level.

Box 1. Two hemispheres, two ventral pathways?

The ventral pathway is often described and modeled as a single entity. It is actually split between two hemispheres and very little is known about how they cooperate to give rise to visual scene perception. In macaques, after section of the corpus callosum and the anterior commissure, TE neurons no longer respond to stimuli presented in the ipsilateral hemifield, indicating that both structures are indispensable for inter-hemispheric integration [72,73].

When presented with two stimuli, one in each hemifield, TE neuronal responses in the normal macaque brain are strongly driven by the contralateral stimulus with very little influence from the ipsilateral one [30,32]. This could be explained by faster activation of TE neurons by contralateral stimuli, ipsilateral afferents being delayed by the trans-callosal conduction time. Similarly, studies in humans have suggested that during bilateral stimulation each hemisphere analyzes mainly the contralateral hemifield [52,53,71].

In split-brain patients with transection of the corpus callosum, bilateral search arrays can be inspected twice as fast than by control subjects [74]. This indicates that each hemisphere has its own limited pool of computational resources. In normal subjects, the interference between the two hemispheres might occur in the visual system itself or in other structures. In keeping with data from recordings in monkeys, it seems that each ventral pathway can analyze half of the visual scene on its own, in large part independently of what is going on in the other one. Information from each ventral pathway might then be combined in such structures as the prefrontal cortex [52,53]. Finally, with one object at the fovea, neuronal activations are also split between the two hemispheres but the mechanisms leading to the perception of a unified object are not yet fully understood [15].

size to be about 2.5°. The time course of the neuronal responses indicated that position information was available in the early part of the response, suggesting it could be essentially determined using feedforward mechanisms.

Small RFs might be explained by the existence of a correlation between RF size and stimulus size [24]. This suggestion does not seem to fit with the classic description of TE neurons as being insensitive to changes in stimulus size. However, although true for many neurons, some actually modulate their response as a function of object size and others respond selectively to a stimulus at a particular size [22,25], with optimal sizes ranging from 1.7° to 30° for instance [21]. More importantly, one study reported larger RFs with larger stimuli [19], a result consistent with the finding that RF structure is dynamically shaped by inhibitory and excitatory interactions within TE [26]. Therefore, it might be that RF size is adjusted to meet the demands of a particular task, as has been demonstrated in parietal cortex [27].

There is also evidence that learning is position specific unless the observer has already some expertise with the stimuli, suggesting that position tolerance is learned (see [24]). Similarly, view-invariant neurons are found intermixed with view-selective neurons, indicating that such invariance might be constructed within TE [2]. Thus, TE might contain a heterogeneous population of cells with

TRENDS in Cognitive Sciences

Figure 3. Coding of spatial information in area TE. In contrast to the small RFs in areas V1 and V2 (white circles) that provide reliable information about stimulus position, the large RFs in area TE (yellow circles) are typically stimulated by more than one object during natural scene perception. This could potentially lead to ambiguous representations. However, non-linearities in the visual system might allow TE neurons to avoid such ambiguities even when they are stimulated by several objects (see also Box 2). Furthermore, there are three ways by which spatial information might be computed in TE. By combining the responses from several neurons with large overlapping RFs (e.g. the 3 large circles on the right side of the figure) it is theoretically possible to reconstruct the position of an object. Some neurons also have a sub-area (or 'hot spot'; dashed circle on the left) within their RF that provides a stronger response than other parts in the RF, thereby reducing the functional size of the RF. Finally, recent experiments have shown that many TE RFs are actually very small (lower middle circle), providing efficient coding of one object without interference from neighboring objects in the scene.

a large panel of RF sizes to cope with various kinds of perceptual problems. Quantitatively, this point also suggests that contrary to area V1 where all cells have small RFs, the spatial information in TE might be strongly dependent on learning and on the task at hand.

Receptive fields in natural scenes

The literature reviewed above concerned isolated stimuli presented on uniform backgrounds. Very few experiments have looked at properties of TE neurons in natural scenes. One such study recently showed that the RF size of TE neurons is greatly reduced for stimuli presented over complex rather than uniform backgrounds [28]. Furthermore, RFs appeared to shrink to fit approximately the size of an object, a result accounted for by a model with competitive interactions between object and background elements in each layer of the ventral stream [29].

Representation of two or more stimuli

Overall, the responses of TE neurons might incorporate enough spatial information to be able to encode the identity of several objects in parallel. Several studies have investigated the behavior of TE neurons in response to two stimuli (e.g. [28,30–32]). Two stimuli typically enter into competitive interactions following two important rules. First, at the level of a single neuron, competition is most likely to occur when both stimuli appear in the RF of a neuron and not when one stimulus is inside the RF and the other outside. Second, the response to two stimuli is generally non-linear, being equal or inferior to the strongest response with one stimulus. These two rules

apply also to areas V2 and V4, and might thus be a general property of neurons in the ventral pathway [33–36].

With small RFs, there would be no competition, strengthening the case for parallel processing. With large RFs, in some cases the response to the two stimuli falling in a neuron's RF is the weighted average of the response to each of them presented alone [30,34]. In other cases, neurons follow a MAX response function; that is, the response to two stimuli is equal to the response elicited by the most effective stimulus of the pair [32,36]. By filtering out the influence of the less effective stimulus, such non-linear response functions are particularly efficient at preserving response selectivity in clutter (Box 2). Although these two sets of results seem to be at odds, even when neurons perform an average on their inputs they do so after a delay, the initial part of their response being close to a MAX function [37]. The fact that the initial part of the neuronal response is identical regardless of whether one or two objects are presented in the neuron's RF has been taken as evidence for late selection models in which high-level object representations are first built before entering into competition [38]. Indeed, a few tens of ms of neuronal activity in a feedforward network might be sufficient to transfer significant amounts of information about an object [2,39,40].

Lateral inhibition

The evidence reviewed so far suggests that scene processing might be achieved massively in parallel in a feedforward network. However, all the information about a scene cannot be extracted with a single feedforward pass through the ventral pathway. Feedforward networks

Box 2. Neuronal coding of information in the ventral pathway

The way information is encoded in neurons sets to a large extent the computational limits of any sensory system. With RFs of increasing size, the ventral pathway would not be able to represent several objects concurrently if neurons performed a linear combination of the inputs. However, neurons present stimulus-response functions with various kinds of non-linearities. For instance, there are neurons in areas V4 [36] and TE [32] that respond according to a MAX function, that is, they respond to the most active of their afferents, ignoring other inputs. The MAX function is thus very interesting computationally because it provides an efficient way to select a complex stimulus in clutter, even if neurons have large RFs [37,75]. Such non-linearities might considerably increase the capacity of the ventral pathway to encode objects in parallel.

One important question is how the MAX (and other MAX-like) functions might be implemented [76]. It has been suggested that once activated by an input, a neuron might be desensitized, for instance by shunting inhibition, so that its response to the next inputs are reduced or even completely blocked [77]. In a network including fast-spiking inter-neurons [78], shunting inhibition can be used to make neurons sensitive to the order of occurrence of the incoming spikes, hence implementing a MAX function in the time domain [40,77] (Figure I). Here, the crucial aspect is the timing of arrival of the inputs. Indeed,

neurons selective to a given input will have a stronger and earlier output for this input compared with other inputs. Thus, latency provides a way to bind object representations – neurons that fire together belong together [79]. But in a natural scene, a neuron will not necessarily be reached first by the input it is sensitive to. With objects of different contrasts, a non-effective object with a high contrast might have a temporal advantage in the ventral pathway and, by mean of lateral connections, inhibit the responses of neighboring neurons to their preferred input arriving few milliseconds later. This is exactly what has been found recently in V4: stimuli with higher contrast drive the competition in neuronal RFs, strongly exciting neurons sensitive to them, and strongly inhibiting those that are not [35]. Thus, by setting response latencies, object contrast is a major low-level constraint in natural scene processing that should be systematically manipulated in computational and empirical studies.

Finally, the visual system seems to use a sparse coding of the input, that is, only a relatively limited number of neurons are active at the same time. Contrary to a massively distributed representation, sparse coding potentially allows the representation of several objects in parallel [80]. Sparse coding might be a way both to exploit the non-accidental statistical properties of natural scenes and to reduce energy consumption [51].
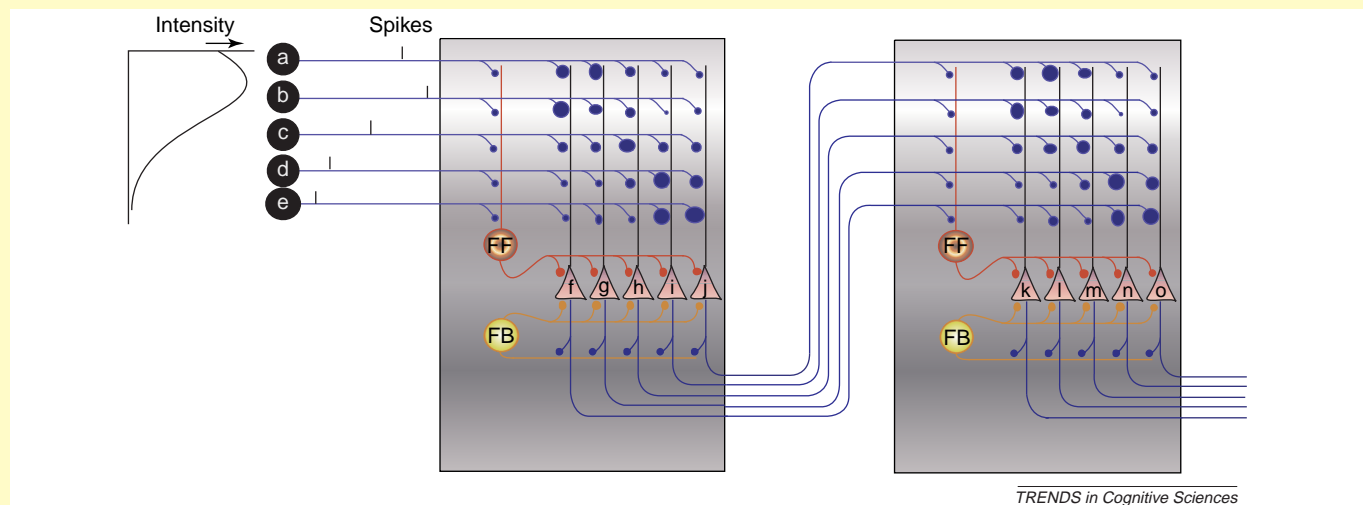

Figure I. Using temporal coding to generate selectivity. A cascade of two cortical processing stages that each contain feedforward excitatory connections onto relay neurons as well as feedforward (FF) and feedback (FB) inhibitory circuits. Input neurons a–e generate a wave of spikes in response to an input pattern. The first inputs to fire produce a maximal effect at the next stage, but spikes arriving later are progressively shunted by the increase in feedforward inhibition, with the result that neurons are sensitive to the order of firing. In this case, relay cell f will be most strongly activated because the strengths of its excitatory inputs matches the order of input firing (b > a > c > d > e). The degree of activation of the other cells depends on how close their weights match the input order. Next, the feedback inhibition from the output level of each stage can be used to perform a form of Winner-Take-All operation because as soon as one cell has fired, inhibition can be used to prevent other cells from firing. If more than one cell in the first stage is allowed to fire, the next stage can decode the order in which they are activated. The plausibility of such a scheme is demonstrated by the SpikeNet image processing software system, which uses these principles to produce highly selective responses to particular visual patterns. (A demonstration of the system can be downloaded from http://www.spikenet-technology.com.)

have serious problems with ambiguous patterns. Such processing could require the use of feedback and lateral connections [16,41,42]. Two pieces of evidence suggest that the complete processing of an object requires time. First, object representations in TE are refined over time, the first part of the response providing coarser information (e.g. it is a face) about the stimulus compared with the later part (e.g. the face is smiling) [43,44]. Second, complex objects seem to be represented by the combined activity of several distant cortical columns [45], through lateral GABAergic inhibition [46,47]. These lateral connections are relatively slow compared with feedforward and feedback connections [48]. With their large and highly spiny dendritic arbors, pyramidal cells in TE might

potentially receive more lateral inhibitory inputs than neurons in more posterior areas [49]. The same lateral inhibition might explain the competitive interactions between spatially nearby object representations [30]. It is thus possible that the competition between objects and the construction of complex object representations in TE are mediated by the same neuronal mechanisms. These data have far reaching consequences. They strongly suggest that TE might be able to encode several coarse object representations simultaneously during a short time after stimulus onset, but the number of objects represented would then rapidly decay as representations are progressively refined through lateral inhibition.

On the other hand, the interactive inhibition between

cortical columns might be a way to code spatial and shape relationships between objects [31]. For instance, a neuron responding to object A could modulate its response depending on how close it is to object B, but not to object C, thus providing a way to encode several objects and their relationships at once. This might be possible with neurons encoding full objects rather than object parts. In keeping with this idea, the response of many TE neurons appears to be tuned to diagnostic elements of stimuli [50] and some neurons can be highly selective to one or very few objects among a large set (e.g. [44]). Thus, well-learned stimuli could be detected in cluttered scenes using diagnostic conjunctions of elements, without relying on highly detailed representations which might require further processing time. Indeed, what is important for object recognition is to associate objects with distinct patterns of activation [8], an idea very close to Barlow's original idea of 'cardinal' cells [51]. The existence of such cells could explain why some complex objects can be detected so efficiently [52–55]. Reciprocally, efficient object detection does not imply that rich representations are formed for all objects in a scene.

Low-level and high-level constraints

The evidence reviewed so far suggest that the ventral stream might be able to encode high-level diagnostic information about objects in parallel, but competitive interactions would then strongly limit the number of detailed representations available. In this section, we consider two other important constraints that might limit parallel object processing.

Low-level factors

Low-level factors include: (1) the information provided by the stimulus itself and its context (such as contrast, see Box 2); (2) the optics of the eye; (3) RF size and distribution of the neurons in the retina, the lateral geniculate nucleus and V1; and (4) noise in neuronal responses.

The effects of low-level factors are far from trivial and need to be taken into consideration in realistic models of visual processing. For example, even a discrimination system that performs optimally, using all the information available in the stimulus without limitations will be sensitive to such factors as the number of distractors or the similarity between target and distractors [56,57].

A fundamental example of how low-level factors constrain object processing is the discovery that during visual search, targets at the fovea are processed more efficiently than targets in the periphery [58]. This eccentricity effect can be largely accounted for by cortical magnification and cortical image density in V1 [59,60]. The cortical magnification in V1 might also passively bias competition between object representations in TE [29], so that the responses of TE neurons are biased towards the stimulus at the fovea over those in the periphery [28,61]. This could be explained by faster responses to foveal than extra-foveal stimuli [25]. The foveal bias might considerably limit the capacity of area TE to represent several objects accurately. It is therefore crucial that future experiments determine the amount of information available for objects in the periphery with and without an object at the fovea. Finally, because the spatial resolution in the retina declines dramatically with eccentricity, information about objects in the periphery might be so poor that no decisions can be made about them.

High-level constraints

The ventral pathway is integrated in a network of cortical areas allowing goal-oriented behaviors. Both prefrontal cortex and the parietal cortex are thought to provide high-level, top-down constraints on computations performed in the ventral pathway [38]. They both represent stimulus information in the context of the current task as opposed to the ventral pathway that is more tuned to the immediate physical aspects of the stimulus [7,62–64].

Contrary to a widespread belief, the parietal cortex does not appear to be necessary to bind complex object representations [65]. Instead, it is needed to filter out the influence of distractors during object discrimination [66]. Areas V4 and TEO in the ventral pathway might receive top-down bias from parietal cortex, allowing a target to be explicitly reported [66]. Furthermore, spatial selection in areas V4 and TE has a very high resolution, suggesting that competitive interactions between objects in those areas is in large part achieved in earlier areas V1 and V2 through feedback connections [29,30,34]. The complex interactions between the different levels of the ventral pathway and the parietal cortex might therefore constrain the number of objects that can be perceived simultaneously and consciously reported (see also Box 3).

Similar conclusions might apply to the prefrontal cortex. In addition to visual information, prefrontal neurons represent categorical information about stimuli; for example, whether they are targets or distractors for the task [63,64]. For instance, the frontal eye field (FEF) is a visuo-motor structure involved in target selection and saccade generation. Neurons in the FEF discriminate a target less efficiently when it is more similar to the distractors, in keeping with the behavior of monkey and human observers [64]. Thus, in addition to the ventral pathway itself, the FEF might be another site in which competition takes place between objects. The outcome of

---

Box 3. Questions for future research

• What is the information processing capacity of the population of TE neurons as a whole? This could be tentatively evaluated by the simulation of very large neuronal networks based on the most recent data about connectivity and functional properties. How many solutions can such a system hold simultaneously?

• Is the primary function of area TE to recognize objects or to represent them in detail? This would have implications on the number of objects that can be represented by TE.

• How many objects can be explicitly processed in parallel? Is the number of objects that can be explicitly processed inversely proportional to the complexity of the representations necessary to perform a given task? How many objects can be implicitly processed?

• How does foveal bias constrain parallelism in the ventral pathway?

• What is the effect of stimulus contrast on the competitive interactions between objects in the ventral pathway? Combining the approaches used by Reynolds [35] and Rolls [28] might provide the answer to this question.

• How do lesions and stimulation of V4, TEO, FEF and parietal areas affect the responses of TE neurons?

this competition might be sent back to the ventral pathway, as suggested by the finding that before a saccade is generated towards a target, the activity of FEF neurons encoding the position of the target enhances the response of V4 neurons with the same retinotopic preferences [67].

A recent model has been suggested in which 'selection for action' in the FEF affects object representations in the ventral pathway [68]. This proposition has been strengthened by evidence that in macaques and humans there is considerable overlap in neuronal response latencies in the ventral pathway with those in parietal and prefrontal cortices [41,69].

Overall, mechanisms of selection for action in the prefrontal and parietal cortices could constrain computations in the ventral pathway, in keeping with models in which object representations compete to a large extent at the output selection stage [38,53]. Hence, behavioral responses might systematically underestimate the capacity of the ventral pathway.

Conclusions

Recent advances in monkey neurophysiology and computational neuroscience highlight the sophistication of the mechanisms implemented in the ventral pathway. Contrary to the claim that the large RFs in TE prevent object-recognition mechanisms from dealing with more than one object at a time, there is now clear evidence that the ventral pathway is well equipped for parallel processing in cluttered scenes. However, local competition, and low-level and high-level constraints considerably limit the amount of information that can be processed and explicitly accessed at once, so that serial selection of objects is often necessary.

Recent experiments in humans have provided striking evidence in favor of serial selection of objects in some circumstances [70,71]. This review points to a kind of model in which parallelism is limited to a very short period of time after stimulus onset. The dynamic network comprising prefrontal and parietal cortices seems to select an object very rapidly in order to move the eyes towards it, without processing all objects in a scene in detail. Thus, in normal viewing conditions, a scene is probably processed in cycles (e.g. there might be a reset every saccade), starting with a short high-level parallel analysis and then zooming in on one or a few objects. Future experiments will need to determine not only the behavioral conditions in which object processing in scenes is limited, but also how the dynamics of the underlying cortical network limits behavioral performance.

References
1 Logothetis, N.K. and Sheinberg, D.L. (1996) Visual object recognition. *Annu. Rev. Neurosci.* 19, 577–621
2 Rolls, E.T. (2000) Functions of the primate temporal lobe cortical visual areas in invariant visual object and face recognition. *Neuron* 27, 205–218
3 von der Malsburg, C. (1999) The what and why of binding: the modeler's perspective. *Neuron* 24, 95–125
4 Baker, C.I. *et al.* (2002) Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. *Nat. Neurosci.* 5, 1210–1216
5 Tanaka, K. (2003) Columns for complex visual object features in the inferotemporal cortex: clustering of cells with similar but slightly different stimulus selectivities. *Cereb. Cortex* 13, 90–99
6 Oram, M.W. and Perrett, D.I. (1992) Time course of neural responses discriminating different views of the face and head. *J. Neurophysiol.* 68, 70–84
7 Thorpe, S.J. and Fabre-Thorpe, M. (2001) Seeking categories in the brain. *Science* 291, 260–263
8 Riesenhuber, M. and Poggio, T. (2002) Neural mechanisms of object recognition. *Curr. Opin. Neurobiol.* 12, 162–168
9 Wersing, H. and Korner, E. (2003) Learning optimized features for hierarchical models of invariant object recognition. *Neural Comput.* 15, 1559–1588
10 Wallis, G. and Rolls, E.T. (1997) Invariant face and object recognition in the visual system. *Prog. Neurobiol.* 51, 167–194
11 Edelman, S. and Intrator, N. (2000) (Coarse coding of shape fragments) + (retinotopy) approximately = representation of structure. *Spat. Vis.* 13, 255–264
12 Elliffe, M.C. *et al.* (2002) Invariant recognition of feature combinations in the visual system. *Biol. Cybern.* 86, 59–71
13 Bullier, J. and Nowak, L.G. (1995) Parallel versus serial processing: new vistas on the distributed organization of the visual system. *Curr. Opin. Neurobiol.* 5, 497–503
14 Leventhal, A.G. *et al.* (1995) Concomitant sensitivity to orientation, direction, and color of cells in layers 2, 3, and 4 of monkey striate cortex. *J. Neurosci.* 15, 1808–1818
15 Schiller, P.H. (1997) Past and present ideas about how the visual scene is analyzed by the brain. In *Extrastriate Visual Cortex in Primates* (Vol. 12) (Rockland, K.S. *et al.*, eds), pp. 59–90, Plenum Press
16 Lee, T.S. (2003) Computations in the early visual cortex. *J. Physiol. (Paris)* 97, 121–139
17 Boussaoud, D. *et al.* (1991) Visual topography of area TEO in the macaque. *J. Comp. Neurol.* 306, 554–575
18 Kobatake, E. and Tanaka, K. (1994) Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J. Neurophysiol.* 71, 856–867
19 Op De Beeck, H. and Vogels, R. (2000) Spatial sensitivity of macaque inferior temporal neurons. *J. Comp. Neurol.* 426, 505–518
20 Nowicka, A. and Ringo, J.L. (2000) Eye position-sensitive units in hippocampal formation and in inferotemporal cortex of the macaque monkey. *Eur. J. Neurosci.* 12, 751–759
21 Ito, M. *et al.* (1995) Size and position invariance of neuronal responses in monkey inferotemporal cortex. *J. Neurophysiol.* 73, 218–226
22 Tanaka, K. *et al.* (1991) Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *J. Neurophysiol.* 66, 170–189
23 Logothetis, N.K. *et al.* (1995) Shape representation in the inferior temporal cortex of monkeys. *Curr. Biol.* 5, 552–563
24 DiCarlo, J.J. and Maunsell, J.H.R. (2003) Anterior inferotemporal neurons of monkeys engaged in object recognition can be highly sensitive to object retinal position. *J. Neurophysiol.* 89, 3264–3278
25 Lueschow, A. *et al.* (1994) Inferior temporal mechanisms for invariant object recognition. *Cereb. Cortex* 4, 523–531
26 Wang, Y. *et al.* (2002) Contribution of GABAergic inhibition to receptive field structures of monkey inferior temporal neurons. *Cereb. Cortex* 12, 62–74
27 Maravita, A. and Iriki, A. (2004) Tools for the body (schema). *Trends Cogn. Sci.* 8, 79–86
28 Rolls, E.T. *et al.* (2003) The receptive fields of inferior temporal cortex neurons in natural scenes. *J. Neurosci.* 23, 339–348
29 Deco, G. and Rolls, E.T. (2004) A neurodynamical cortical model of visual attention and invariant object recognition. *Vision Res.* 44, 621–642
30 Chelazzi, L. *et al.* (1998) Responses of neurons in inferior temporal cortex during memory-guided visual search. *J. Neurophysiol.* 80, 2918–2940

31 Missal, M. *et al*. (1999) Shape interactions in macaque inferior temporal neurons. *J. Neurophysiol.* 82, 131–142

32 Sato, T. (1989) Interactions of visual stimuli in the receptive fields of inferior temporal neurons in awake macaques. *Exp. Brain Res.* 77, 23–30

33 Chelazzi, L. *et al*. (2001) Responses of neurons in macaque area V4 during memory-guided visual search. *Cereb. Cortex* 11, 761–772

34 Reynolds, J.H. *et al*. (1999) Competitive mechanisms subserve attention in macaque areas V2 and V4. *J. Neurosci.* 19, 1736–1753

35 Reynolds, J.H. and Desimone, R. (2003) Interacting roles of attention and visual salience in V4. *Neuron* 37, 853–863

36 Gawne, T.J. and Martin, J.M. (2002) Responses of primate visual cortical V4 neurons to simultaneously presented stimuli. *J. Neurophysiol.* 88, 1128–1135

37 Rousselet, G.A. *et al*. (2003) Taking the MAX from neuronal responses. *Trends Cogn. Sci.* 7, 99–102

38 Desimone, R. and Duncan, J. (1995) Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* 18, 193–222

39 Keysers, C. *et al*. (2001) The speed of sight. *J. Cogn. Neurosci.* 13, 90–101

40 VanRullen, R. and Thorpe, S.J. (2002) Surfing a spike wave down the ventral stream. *Vision Res.* 42, 2593–2615

41 Bullier, J. (2003) Communications between cortical areas of the visual system. In *The Visual Neurosciences* (Vol. 1) (Chalupa, L.M. and Werner, J.S., eds), pp. 522–540, MIT Press

42 Roelfsema, P.R. *et al*. (2000) The implementation of visual routines. *Vision Res.* 40, 1385–1411

43 Sugase, Y. *et al*. (1999) Global and fine information coded by single neurons in the temporal visual cortex. *Nature* 400, 869–873

44 Tamura, H. and Tanaka, K. (2001) Visual response properties of cells in the ventral and dorsal parts of the macaque inferotemporal cortex. *Cereb. Cortex* 11, 384–399

45 Tsunoda, K. *et al*. (2001) Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. *Nat. Neurosci.* 4, 832–838

46 Tamura, H. *et al*. (2004) Presumed inhibitory neurons in the macaque inferior temporal cortex: visual response properties and functional interactions with adjacent neurons. *J. Neurophysiol.* 91, 2782–2796

47 Wang, Y. *et al*. (2000) Neuronal mechanisms of selectivity for object features revealed by blocking inhibition in inferotemporal cortex. *Nat. Neurosci.* 3, 807–813

48 Angelucci, A. and Bullier, J. (2003) Reaching beyond the classical receptive field of V1 neurons: horizontal or feedback axons? *J. Physiol. (Paris)* 97, 141–154

49 Elston, G.N. (2002) Cortical heterogeneity: implications for visual processing and polysensory integration. *J. Neurocytol.* 31, 317–335

50 Sigala, N. and Logothetis, N.K. (2002) Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature* 415, 318–320

51 Barlow, H. (2001) Redundancy reduction revisited. *Network* 12, 241–253

52 Rousselet, G.A. *et al*. (2002) Parallel processing in high-level categorization of natural images. *Nat. Neurosci.* 5, 629–630

53 Rousselet, G.A. *et al*. (2004) Processing of one, two or four natural scenes in humans: the limits of parallelism. *Vision Res.* 44, 877–894

54 Li, F.F. *et al*. (2002) Rapid natural scene categorization in the near absence of attention. *Proc. Natl. Acad. Sci. U. S. A.* 99, 9596–9601

55 VanRullen, R. *et al*. (2004) Visual search and dual tasks reveal two distinct attentional resources. *J. Cogn. Neurosci.* 16, 4–14

56 Geisler, W.S. and Chou, K.L. (1995) Separation of low-level and high-level factors in complex tasks: visual search. *Psychol. Rev.* 102, 356–378

57 Palmer, J. *et al*. (2000) The psychophysics of visual search. *Vision Res.* 40, 1227–1268

58 Carrasco, M. *et al*. (1995) The eccentricity effect: target eccentricity affects performance on conjunction searches. *Percept. Psychophys.* 57, 1241–1261

59 Carrasco, M. and Yeshurun, Y. (1998) The contribution of covert attention to the set-size and eccentricity effects in visual search. *J. Exp. Psychol. Hum. Percept. Perform.* 24, 673–692

60 Motter, B.C. and Holsapple, J.W. (2000) Cortical image density determines the probability of target discovery during active search. *Vision Res.* 40, 1311–1322

61 Sheinberg, D.L. and Logothetis, N.K. (2001) Noticing familiar objects in real world scenes: the role of temporal cortical neurons in natural vision. *J. Neurosci.* 21, 1340–1350

62 Gottlieb, J. (2002) Parietal mechanisms of target representation. *Curr. Opin. Neurobiol.* 12, 134–140

63 Freedman, D.J. *et al*. (2003) A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *J. Neurosci.* 23, 5235–5246

64 Schall, J.D. (2002) The neural selection and control of saccades by the frontal eye field. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 357, 1073–1082

65 Driver, J. and Vuilleumier, P. (2001) Perceptual awareness and its loss in unilateral neglect and extinction. *Cognition* 79, 39–88

66 Friedman-Hill, S.R. *et al*. (2003) Posterior parietal cortex and the filtering of distractors. *Proc. Natl. Acad. Sci. U. S. A.* 100, 4263–4268

67 Moore, T. and Armstrong, K.M. (2003) Selective gating of visual signals by microstimulation of frontal cortex. *Nature* 421, 370–373

68 Hamker, F.H. (2003) The reentry hypothesis: linking eye movements to visual perception. *J. Vis.* 3, 808–816

69 Foxe, J.J. and Simpson, G.V. (2002) Flow of activation from V1 to frontal cortex in humans. A framework for defining 'early' visual processing. *Exp. Brain Res.* 142, 139–150

70 Bricolo, E. *et al*. (2002) Serial attention mechanisms in visual search: a direct behavioral demonstration. *J. Cogn. Neurosci.* 14, 980–993

71 Woodman, G.F. and Luck, S.J. (2003) Serial deployment of attention during visual search. *J. Exp. Psychol. Hum. Percept. Perform.* 29, 121–138

72 Rocha-Miranda, C.E. *et al*. (1975) Visual activation of neurons in inferotemporal cortex depends on striate cortex and forebrain commissures. *J. Neurophysiol.* 38, 475–491

73 Tomita, H. *et al*. (1999) Top-down signal from prefrontal cortex in executive control of memory retrieval. *Nature* 401, 699–703

74 Luck, S.J. *et al*. (1994) Independent attentional scanning in the separated hemispheres of split-brain patients. *J. Cogn. Neurosci.* 6, 84–91

75 Riesenhuber, M. and Poggio, T. (1999) Are cortical models really bound by the 'binding problem'? *Neuron* 24, 87–125

76 Yu, A.J. *et al*. (2002) Biophysiologically plausible implementations of the maximum operation. *Neural Comput.* 14, 2857–2881

77 Thorpe, S. *et al*. (2001) Spike-based strategies for rapid processing. *Neural Netw.* 14, 715–725

78 Swadlow, H.A. (2003) Fast-spike interneurons and feedforward inhibition in awake sensory neocortex. *Cereb. Cortex* 13, 25–32

79 Gawne, T.J. *et al*. (1996) Latency: another potential code for feature binding in striate cortex. *J. Neurophysiol.* 76, 1356–1360

80 Földiák, P. (2002) Sparse coding in the primate cortex. In *Handbook of Brain Theory and Neural Networks* (Arbib, M.A., ed), pp. 1064–1068, MIT Press