Neural mechanisms of object recognition

Maximilian Riesenhuber* and Tomaso Poggio†

Single-unit recordings from behaving monkeys and human functional magnetic resonance imaging studies have continued to provide a host of experimental data on the properties and mechanisms of object recognition in cortex. Recent advances in object recognition, spanning issues regarding invariance, selectivity, representation and levels of recognition have allowed us to propose a putative model of object recognition in cortex.

Addresses
McGovern Institute for Brain Research, Department of Brain & Cognitive Sciences, Center for Biological and Computational Learning and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA
*e-mail: max@ai.mit.edu
†e-mail: tp@ai.mit.edu

Current Opinion in Neurobiology 2002, **12**:162–168

0959-4388/02/$ – see front matter

Published online 4th March 2002

Abbreviations

FFA	fusiform face area
fMRI	functional magnetic resonance imaging
IT	inferotemporal cortex
Max	maximum
PFC	prefrontal cortex
RBF	radial basis function
V1	primary visual cortex
V2	secondary visual cortex

Introduction

Object recognition is fundamental to the behavior of higher primates. It is also the most remarkable achievement of the visual cortex and one that probably greatly influences its functional architecture. The visual system rapidly and effortlessly recognizes a large number of diverse objects in cluttered, natural scenes — a very difficult computational task. Here, we review progress in this field over the past two years. We do so in the context of a recent quantitative model, which helps us summarize and organize existing data as well as interpret contradictory, and occasionally ill-defined, claims. We organize the discussion of the new data around the four key issues of object recognition: invariance, selectivity, object representation and levels of recognition.

Invariance

Simple cells in primary visual cortex (V1) have small receptive fields and respond preferentially to oriented bars. Progressing along the ventral stream — thought to play a central role in object recognition in cortex [1,2] — neurons show an increase in receptive field size and in the complexity of their preferred stimuli [3]. At the top of the ventral stream, in the inferotemporal cortex (IT), cells are tuned to complex stimuli such as faces [4–7]. A hallmark of these IT cells is, in addition to selectivity, the robustness of their firing to stimulus transformations, such as scale and position changes [1,2,8,9]. In contrast, later studies [8,10–12] have shown that most neurons show specificity for a certain object view or lighting condition. In particular, Logothetis *et al.* [8] trained monkeys to perform an object recognition task with isolated views of novel objects (paperclips). When recording from the animals' IT, they found that the great majority of neurons selectively tuned to the training objects showed tight tuning to a specific view of one of the training objects (a few units showed greater tolerance, in agreement with earlier predictions [13]). The view-tuned neurons also showed an average scale invariance of two octaves. That is, the neurons still responded at a higher level to the scaled image of their preferred paperclip than to other paperclips, even when stimulus size was varied over two octaves. Furthermore, the view-tuned neurons had an average translation invariance of 4° (for typical stimulus sizes of 2°) [14], which is much smaller than previous reports, but large for any computational mechanism. A very recent study (JJ DiCarlo, JHR Maunsell, personal communication), using different stimuli and training paradigms, reports translation invariance from one view of less than 3°, pointing to a possible influence of training history and object shape on invariance ranges. Human functional magnetic resonance imaging (fMRI) data have shown a similar pattern of invariance properties for the lateral occipital cortex, a brain region in human visual cortex central to object recognition and believed to be the homolog of monkey area IT [15–17].

From a computational point of view one might ask the question: which object transformations can be estimated from one versus several object views? It is well known that only a very small number of views are required to generalize object recognition across different uniform transformations [18• and references therein]. Scaling and translation in the image plane, for instance, solely require a single object view, as they preserve the original information of an image. In this case, it is possible to dispense with the need for additional examples of different sizes or positions in the field of view. In sharp contrast, multiple views are generally required to recognize objects subjected to three-dimensional shape transformations, whether actual — such as the rotation of objects in depth — or induced — such as those resulting from illumination changes. The frontal view of a novel face, for instance, does not contain sufficient information to predict the profile of that face.

Computational considerations such as these lead to a hierarchical architecture of a system for object recognition that instantiates the basic facts about the ventral pathways of the brain [18•]. The model shown schematically in Figure 1 reflects the general organization of visual cortex in a series of layers from V1 → IT → prefrontal cortex (PFC). Invariance properties emerge from the functional

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
2002		**00-00-2002 to 00-00-2002**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Neural mechanisms of object recognition	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Massachusetts Institute of Technology,Center for Biological and Computational Learning,77 Massachusetts Avenue,Cambridge,MA,02139**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

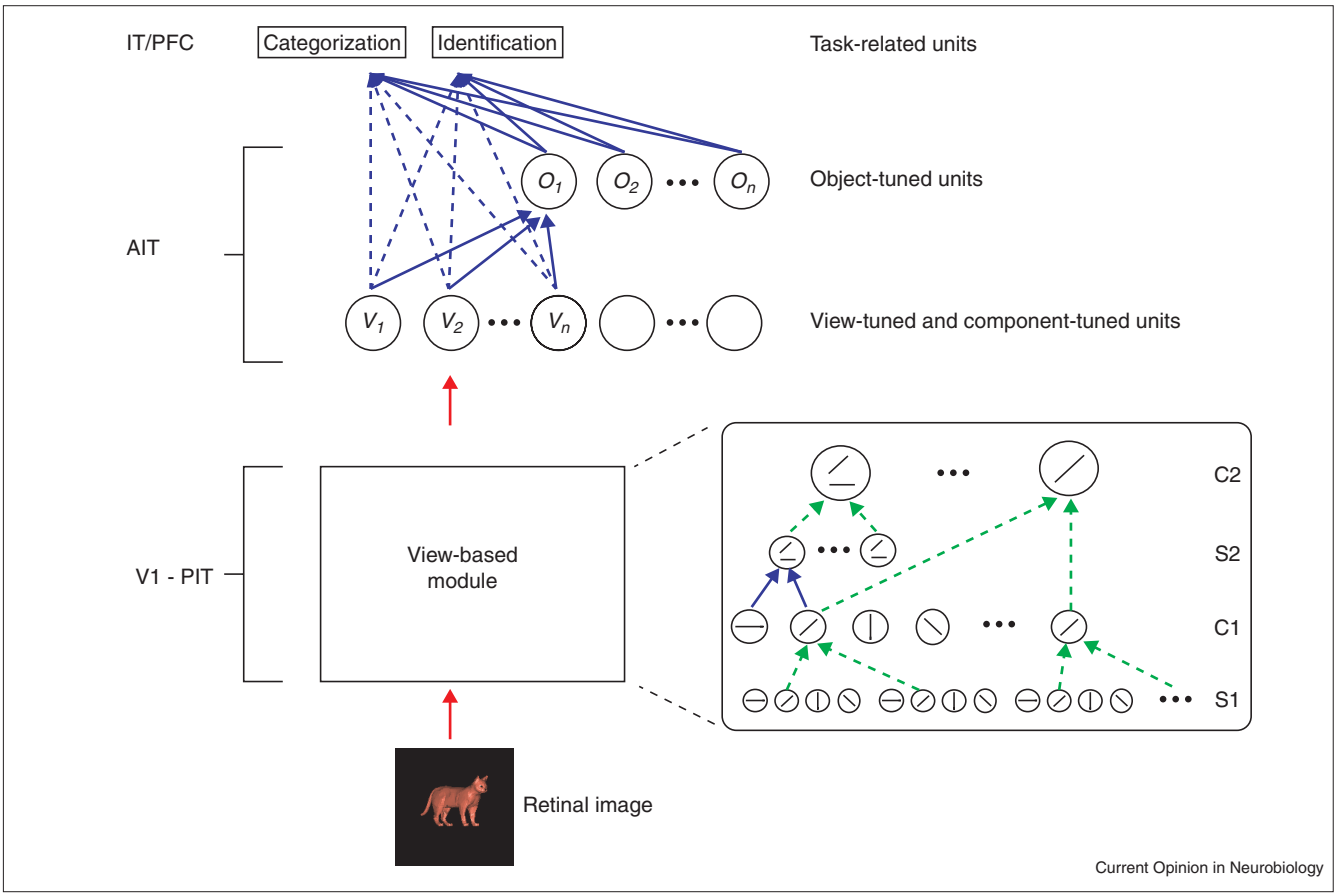
13. SUPPLEMENTARY NOTES
The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**		**7**	

Figure 1



Model of the architecture of recognition in the cortex [18•]. The model combines and extends several recent models [9,13,14,55,56] and effectively summarizes many experimental findings. A view-based module [14], consisting of a hierarchical extension of the classical paradigm of building complex cells from simple cells [57]. The hierarchy of layers have two different types of pooling mechanisms. The first layer in V1 represents linear oriented filters similar to simple cells; each unit in the next layer pools the outputs of simple cells of the same orientation but at slightly different positions (scales). Each of these units is still orientation-selective but more invariant to position (scale), similarly to some complex cells. In the next stage, signals from complex cells with different orientations but similar positions are combined to create neurons (S2) tuned to a small dictionary of more complex features. The next layer is equivalent to complex cells in V1: by pooling together signals from S2, cells of the same type but at slightly different positions, the C2 units become more invariant to position (and scale) but preserve feature selectivity. They may correspond roughly to V4 cells. In the model, the C2 cells feed into view-tuned cells (V_n), with connection weights that are learned from exposure to a view of an object. There may be more levels in this hierarchy, after the C2 layer. The key idea in the view-tuned module alternates two types of pooling: the first to provide increasing pattern selectivity (blue lines in the inset) and the second (founded on the Max operation; dashed green lines in the inset) to provide invariance. Invariance to translation is achieved by pooling over afferents tuned to different positions, and invariance to scale (not shown) is accomplished by pooling over afferents tuned to different scales. The output of the view-based module is represented by view-tuned model units that exhibit tight tuning to rotation in depth (and other object-dependent transformations, such as illumination and facial expression) but are tolerant to scaling and translation of their preferred object view. Notice that the cells labeled here as view-tuned units, encompass, between the anterior IT (AIT) and posterior IT (PIT), a spectrum of tuning from views to complex features: depending on the synaptic weights determined during learning, each view-tuned cell becomes effectively connected to all or only a few of the units activated by the object view [20]. The second part of the model starts with the view-tuned cells. Invariance to rotation in depth is obtained by combining, in a learning module, several view-tuned units tuned to different views of the same object [13], creating view-invariant units (O_n). These, as well as the view-tuned units, can then serve as inputs to task modules that learn to perform different visual tasks such as identification/discrimination or object categorization. They consist of same generic learning circuitry (similar to an RBF network [13]) but are trained with appropriate sets of examples to perform specific tasks. In addition to the feed-forward processing, there are likely feedback pathways for top-down modulation of neuronal responses throughout the processing hierarchy and to support the learning phase. All the units in the model represent single cells modeled as simplified neurons with modifiable synapses.

organization of two stages of processing. The first, extending from V1 to IT, is comprised of units showing the same scale and position invariance properties as the view-tuned IT neurons described by Logothetis *et al.* [8] using the same stimuli. Computationally, this is accomplished by a scheme best explained by taking striate complex cells as an example: invariance to changes in the position of an optimal stimulus (within a range) is obtained by means of a maximum (Max) operation performed on the simple cell inputs to the complex cells. Both simple and complex cells

are assumed to have the same optimal orientation but at different positions. The key idea is that the two steps — filtering followed by a Max operation — are equivalent to a simple but powerful signal processing technique: select the peak of the correlation between the signal and a given matched filter (here the correlation is over either position or scale). The model alternates layers of units combining simple filters into more complex ones with layers using the Max operation, in order to build invariance to position and scale while increasing pattern selectivity. In the second part of the architecture, learning from multiple examples, represented by view-tuned units, leads to view-invariant units, as well as neural circuits performing specific tasks. The key idea here is that interpolation and generalization can be obtained by simple networks that learn to combine the output of cells, each broadly tuned to the features of an example image [8,13]. Simple learning networks of this type can learn to identify an object across different viewpoints [13] and illuminations, as well as categorizing objects across exemplars of a class [19].

The model described above predicts several experimental results and provides interesting perspectives on still other data and claims. For instance, the model accounts (see [14,18•,19,20]) for the response of tuned IT cells to multiple objects in the receptive field [21], scrambled objects [22], cluttered [23] and mirror views [2]. It also shows a degree of performance roughly in agreement with physiological and psychophysical data obtained from specific tasks. These include the cat versus dog categorization task described by Freedman *et al.* [24•], object identification, gender classification and possibly the face habituation effect of Leopold *et al.* [25], as well as the effects of contrast, mirror and figure-ground reversal described by Baylis and Driver [26]. Preliminary data [27] support a specific prediction of the model — the existence of a Max-like pooling operation to increase invariance (see Figure 1).

A key function of models is to clarify basic issues and the interpretation of relevant data. In the following, we will use the model shown in Figure 1 to discuss three focal topics of recent research in object recognition: the feature tuning of neurons in higher visual areas, the nature and organization of object representation, and the relationship between identification and categorization tasks.

Selectivity

Invariance is one requirement for object recognition, the other one being selectivity. Several studies have established that IT neurons can become tuned to task-relevant objects and their views [8,28,29•,30•] or to objects in the monkey's environment [10], suggesting that the activity of these neurons may be part of the representation of objects occurring in an animal's environment. The preferred stimuli of neurons in intermediate stages of the ventral stream are less clear, possibly because of the difficulty of knowing which stimuli to use to probe their neural selectivity. Reports of preferred features of neurons in V4, the visual area preceding IT in the ventral pathway, vary depending on the set of stimuli used to probe responses, including cartesian gratings [31], polar and hyperbolic sinusoidal gratings [32], and contour features [33]. In the secondary visual cortex (V2), a recent study [34] has reported neuronal preferences to complex stimuli such as arcs, intersecting lines and non-cartesian gratings.

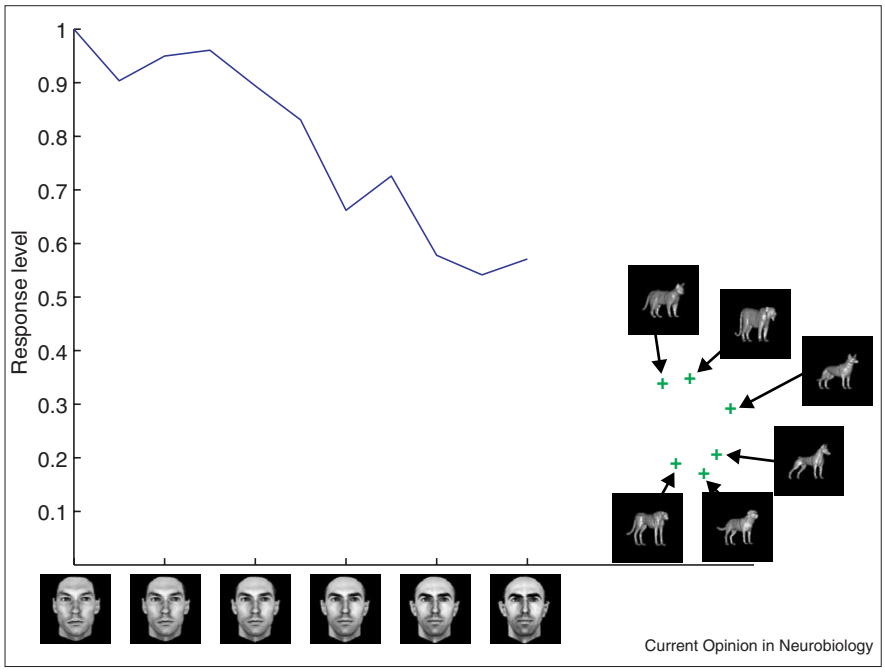
Instead of probing neuronal tuning with a fixed set of stimuli, another set of studies [1,3,35–37] has employed a 'simplification procedure' in an effort to define the features crucial to activate a neuron. In this approach, a complex natural stimulus (such as a face) to which the neuron under study responds, is progressively 'simplified' (e.g. by removing color or texture, or simplifying complex shapes into simpler geometric primitives) such that the magnitude of the response remains the same as that elicited by the original, unsimplified object. The stimulus that cannot be 'simplified' further without decreasing the firing rate is then defined as the effective stimulus for that cell. A study using this paradigm [3] has reported an increase in feature complexity from area V2 to anterior IT. However, a recent IT optical imaging study [37], supported by single cells recordings, demonstrates the fundamental difficulty of determining a neuron's preferred feature in higher visual areas. These authors report that, in fact, in the majority of cases, 'simplifying' a stimulus led to the activation of additional IT neurons relative to the original 'complex' stimulus. Interestingly, the model described in Figure 1 does actually qualitatively predict what is observed — neurons tuned to a dictionary of features at different levels of complexity. Moreover, preliminary simulations suggest that, for IT model units, the effect of the 'simplification' procedure may well lead to the observations reported by Tsunoda *et al.* [37].

Representation

Related to the issue of neuronal tuning is the question of the precise nature of object representation in cortex. It has recently been put forward, on the basis of a set of human fMRI studies, that some object classes — faces [38], places [39] and body parts [40•] — are processed by distinct modules in cortex. Another fMRI study [41•] has shown that objects of a certain class (e.g. faces) evoke a distributed pattern of activity that is not confined to the aforementioned specialized modules (e.g. the fusiform face area [FFA] [38]), and that activation patterns outside a specific module are sufficient for object categorization. Some data, therefore, appear to argue for a 'modular' framework of object representation in cortex, where specific brain areas are posited to perform computations unique to the object class at hand. Other data, however, support a model in which objects from different classes are represented in a distributed fashion, and their recognition is founded on the same computations.

The model represented in Figure 1 supports the latter claim. Figure 2 helps to reconcile the two sets of data.

Figure 2



Tuning of a model face unit. The unit is a view-tuned unit as the V_n shown in Figure 1, tuned to the leftmost face on the bottom axis. The blue line shows the unit's response changes as the stimulus is gradually morphed away from the preferred stimulus to another face (along the axis). The unit's response changes gradually with changes in the stimulus, permitting subordinate level discrimination (especially when using a population code consisting of several units tuned to different representatives of the class [46]). The same unit also responds to the animal stimuli shown on the right (green crosses), but at a lower level than to the faces. This permits a coarse categorization of a stimulus as an animal stimulus, on the basis of the face unit's firing [41•]. Units such as these can form the basis of a categorization circuit [19]. Face images courtesy of T Vetter [58].

Model IT units have preferred afferent activation patterns that can represent a full or partial view of an object [20]. At the highest levels of the model, the tuning ranges from the one described by Tanaka [1] to the view-tuning described for paperclips and faces. Contrary to some claims [17], both the model and the experimental data [42••,43] suggest that faces are not special. In fact, the model predicts the existence of neurons tuned to moderately complex features and cells tuned specifically to views of objects, depending on task training and difficulty [44]. In doing so, it suggests that the distinction between 'complex features' and 'objects' is largely semantic: during training, a cell may well become tuned to a feature that is diagnostic for the object rather than to a full view, depending on the specificity and number of its afferents. What is relevant for object recognition is that the objects to be discriminated produce distinct activation patterns.

From a computational point of view, groups of neurons responding to representatives from different object classes do not have to be segregated, but are likely to be interdigitated. Moreover, the same neuron can respond to objects from different classes, depending on their visual similarity (Figure 2). Because the activity of one fMRI voxel is typically the average of hundreds of thousands of neurons, a strong activation of the FFA for faces would argue for a higher density of face neurons in that part of cortex, perhaps owing to the great cognitive importance of faces. (For cautionary notes about the interpretation of fMRI images see [45••].) However, subjects with substantial expertise for other object classes could be expected to have a greater number of neurons tuned to objects from their field of expertise [46], and correspondingly might show significant activation of the same cortical regions for these objects. Indeed, in bird and car experts, brain areas overlapping with the FFA have been found [47] to be specifically activated by birds and cars, and subjects trained to recognize objects from a novel class of objects ('greebles') showed activation of the FFA by the training objects [48•].

What is the mechanism that permits the usage of similar neural circuits for representing objects as diverse as faces, birds and cars? The architecture and operational principles of our model offer one putative mechanism (for detailed computational simulations, see [19,46]). A particular object, say a specific face, will elicit different activities in the view-specific V_n and object-specific O_n cells of Figure 1 (an example of which is shown in Figure 2). Thus, the memory of the particular face is represented in an implicit way, by a sparse population code through the activation pattern over the coarsely tuned V_n and O_n cells. Discrimination, or memorization of specific objects, can then proceed by comparing activation patterns over the strongly activated object-tuned or view-tuned units [46] tuned to a small number of 'prototypical' faces [49]. For a certain level of specificity, only the activations of a small number of units have to be stored, forming a sparse code. This is in contrast to activation patterns at lower levels, where units are less specific and hence activation patterns tend to involve more neurons. In a similar fashion, neural circuitry for categorization, located putatively in the PFC [24•], can be trained [19] to receive input from relevant object-tuned units. For instance, a unit that categorizes

cats versus dogs would receive input from units responding to some individual cats and dogs. This is in line with a very recent finding that PFC units show more category tuning than IT neurons, in a macaque trained to categorize cats and dogs [50].

In conclusion, the model depicted in Figure 1 suggests that the same basic circuitry, replicated many times in IT, can learn from visual experience to represent and recognize different types of objects. Computations do not need to be fully hardwired by genes and specialized for specific classes of objects.

Levels of recognition

An object can be recognized at different levels — a face can be recognized as a face, but also more specifically as 'a male face', 'Tommy Poggio's face' or 'Tommy Poggio's smiling face'. It has been common in cognitive science to assume that recognition of an object at different levels relies on different computational mechanisms [51,52]. In particular, it has been proposed that 'subordinate level' recognition (identification) is derived from 'configurational' judgements, whereas 'basic level' categorization (a face? a dog? a car?) relies on a qualitative representation formulated on the presence or absence of features.

However, as Figure 1 makes clear and as we have pointed out earlier [18•], all supervised recognition tasks — in which the subject is trained with labeled examples — are identical from a computational point of view: they all involve a classification established on positive and negative exemplars. Indeed, it is not clear why different computations should be required to recognize a face at the subordinate level or, for example, to determine its gender. In fact, it is worth noticing that the basic radial basis function (RBF) network [8,13] replicated at different levels in Figure 1 (e.g. from view-tuned to view-invariant units), can learn to perform different tasks from the same set of training images. For instance, units tuned to distinct expressions of a face can feed into an identification unit that responds to a specific face; the same units can also be used with different synaptic weights by an expression unit that has learned to respond, say, to smiling. In line with the model in Figure 1, recent findings indicate that the FFA is involved not just in subordinate level face recognition but also in face detection [53], arguing against a specialization of brain areas for recognition tasks, such as subordinate level recognition independent of object class. The problem with many experiments investigating the relationship between categorization and identification that claim an advantage of basic level recognition over subordinate level recognition, is that the tasks used for the different recognition levels are of differing difficulties. Discriminating a face from a chair (categorization) is a much easier task than discriminating between two faces (identification), as the latter are more similar to each other. Assuming that physically similar stimuli produce similar neuronal activation patterns, and that the ability to discriminate between two stimuli requires a certain level of evidence (in the form of firing rate differences), a finer discrimination would require the accumulation of evidence over a longer time period. A prediction would be that if categorization and identification tasks were equalized in terms of difficulty, they would take a similar amount of time. From the point of view of the model represented in Figure 1, the two tasks are computationally equivalent and can be learned with equal ease.

Conclusions and future directions

Most of the old and new data on object recognition in cortex can be summarized and interpreted in a quantitative and consistent way, by a simple hierarchical, mostly feed-forward architecture as shown in Figure 1. Of course, many aspects of how object recognition is performed are left open by simple models of this kind. Furthermore, future experiments may require modifications and extensions of this model and still others may falsify significant parts of it. For instance, data on the neural correlates of border ownership in V2 [54•] are hard to incorporate in feed-forward models, especially if they hold true for natural scenes. In any case, the road ahead will require close interactions between experimental and computational work.

Acknowledgements

Supported by grants from Office of Naval Research, National Science Foundation and Honda. M Riesenhuber is supported by a McDonnell-Pew Award in Cognitive Neuroscience. T Poggio is supported by the Uncas and Helen Whitaker Chair at the Whitaker College, Massachusetts Institute of Technology. With special thanks to Nikos Logothetis for his substantial help in the preparation of this review.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Tanaka K: **Inferotemporal cortex and object vision.** *Annu Rev Neurosci* 1996, **19**:109-139.
2. Logothetis NK, Sheinberg DL: **Visual object recognition.** *Annu Rev Neurosci* 1996, **19**:577-621.
3. Kobatake E, Tanaka K: **Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex.** *J Neurophysiol* 1994, **71**:856-867.
4. Gross CG, Rocha-Miranda CE, Bender DB: **Visual properties of neurons in inferotemporal cortex of the macaque.** *J Neurophysiol* 1972, **35**:96-111.
5. Desimone R, Albright TD, Gross CG, Bruce C: **Stimulus-selective properties of inferior temporal neurons in the macaque.** *J Neurosci* 1984, **4**:2051-2062.
6. Desimone R: **Face-selective cells in the temporal cortex of monkeys.** *J Cogn Neurosci* 1991, **3**:1-8.
7. Perrett DI, Hietanen JK, Oram MW, Benson PJ: **Organization and functions of cells responsive to faces in the temporal cortex.** *Philos Trans R Soc B* 1992, **335**:23-30.
8. Logothetis NK, Pauls J, Poggio P: **Shape representation in the inferior temporal cortex of monkeys.** *Curr Biol* 1995, **5**:552-563.
9. Perrett D, Oram M: **Neurophysiology of shape processing.** *Imaging Vis Comput* 1993, **11**:317-333.
10. Booth MC, Rolls ET: **View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex.** *Cereb Cortex* 1998, **8**:510-523.

11. Hietanen JK, Perrett DI, Benson PJ, Dittrich WH: **The effects of lighting conditions on responses of cells selective for face views in the macaque temporal cortex.** *Exp Brain Res* 1992, **89**:157-171.

12. Wang G, Tanaka K, Tanifuji M: **Optical imaging of functional organization in the monkey inferotemporal cortex.** *Science* 1996, **272**:1665-1668.

13. Poggio T, Edelman S: **A network that learns to recognize 3D objects.** *Nature* 1990, **343**:263-266.

14. Riesenhuber M, Poggio T: **Hierarchical models of object recognition in cortex.** *Nat Neurosci* 1999, **2**:1019-1025.

15. Grill-Spector K, Kourtzi Z, Kanwisher N: **The lateral occipital complex and its role in object recognition.** *Vis Res* 2001, **41**:1409-1422.

16. Malach R, Reppas JB, Benson RR, Kwong KK, Jiang HH, Kennedy WA, Ledden PJ, Brady TJ, Rosen BR, Tootell RB: **Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex.** *Proc Natl Acad Sci USA* 1995, **92**:8135-8139.

17. Tanaka K: **Mechanisms of visual object recognition: monkey and human studies.** *Curr Opin Neurobiol* 1997, **7**:523-529.

18. • Riesenhuber M, Poggio T: **Models of object recognition.** *Nat Neurosci* 2000, **3**:1199-1204.

A review of computational models of object recognition slanted towards the authors' own work but with many relevant references. It contains more details related to the present review.

19. Riesenhuber M, Poggio T: **A note on object class representation and categorical perception.** AI Memo 1679, CBCL Paper 183, MIT AI Lab and CBCL,Cambridge, MA 1999. URL: http://www.ai.mit.edu/research/publications/publications.shtml

20. Riesenhuber M, Poggio T: **Are cortical models really bound by the 'binding problem'?** *Neuron* 1999, **24**:87-93.

21. Sato T, Kawamura T, Iwai E: **Responsiveness of inferotemporal single units to visual pattern stimuli in monkeys performing discrimination.** *Exp Brain Res* 1980, **38**:313-319.

22. Vogels R: **Categorization of complex visual images by rhesus monkeys. Part 2: single-cell study.** *Eur J Neurosci* 1999, **11**:1239-1255.

23. Missal M, Vogels R, Orban GA: **Responses of macaque inferior temporal neurons to overlapping shapes.** *Cereb Cortex* 1997, **7**:758-767.

24. • Freedman D, Riesenhuber M, Poggio T, Miller E: **Categorical representation of visual stimuli in the primate prefrontal cortex.** *Science* 2001, **291**:312-316.

The authors describe neural correlates of a supervised categorization behavior – a monkey was trained to categorize cats versus dogs – in PFC.

25. Leopold DA, O'Toole AJ, Vetter T, Blanz V: **Prototype-references shape encoding revealed by high-level aftereffects.** *Nat Neurosci* 2001, **4**:3-5.

26. Baylis GC, Driver J: **Shape-coding in IT cells generalizes over contrast and mirror reversal, but not figure-ground reversal.** *Nat Neurosci* 2001, **4**:937-942.

27. Lampl I, Riesenhuber M, Poggio T, Ferster D: **The Max operation in cells in the cat visual cortex.** *Soc Neurosci Abs* 2001, **619**.30.

28. Kobatake E, Wang G, Tanaka K: **Effects of shape-discrimination training on the selectivity of inferotemporal cells in adult monkeys.** *J Neurophys* 1998, **80**:324-330.

29. • Sheinberg DL, Logothetis NK: **Noticing familiar objects in real world scenes: the role of temporal cortical neurons in natural vision.** *J Neurosci* 2001, **21**:1340-1350.

This study of IT selectivity during free viewing of natural scenes shows similar selectivity of neurons for familiar stimuli presented in isolation or encountered during exploration. During exploration, neural activation may start before effective targets are fixated.

30. • DiCarlo JJ, Maunsell JHR: **Form representation in monkey inferotemporal cortex is virtually unaltered by free viewing.** *Nat Neurosci* 2000, **3**:814-821.

Whether targets are flashed at the endpoints of saccades or under the usual controlled viewing condition – in which monkeys hold their direction of gaze fixed while isolated stimuli are presented – the stimulus selectivity of IT neurons remains the same.

31. Desimone R, Schein SJ: **Visual properties of neurons in area V4 of the macaque: sensitivity to stimulus form.** *J Neurophys* 1987, **57**:835-868.

32. Gallant JL, Connor CE, Rakshit S, Lewis JW, van Essen DC: **Neural responses to polar, hyperbolic, and cartesian gratings in area V4 of the macaque monkey.** *J Neurophysiol* 1996, **76**:2718-2739.

33. Pasupathy A, Connor CE: **Responses to contour features in macaque area V4.** *J Neurophysiol* 1999, **82**:2490-2502.

34. Hegde J, van Essen DC: **Selectivity for complex shapes in primate visual area V2.** *J Neurosci* 2000, **20**:R61:1-6.

35. Tanaka K: **Neuronal mechanisms of object recognition.** *Science* 1993, **262**:685-688.

36. Fujita I, Tanaka K, Ito M, Cheng K: **Columns for visual features of objects in monkey inferotemporal cortex.** *Nature* 1992, **360**:343-346.

37. Tsunoda K, Yamane Y, Nishizaki M, Tanifuji M: **Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns.** *Nat Neurosci* 2001, **4**:832-838.

38. Kanwisher N, McDermott J, Chun MM: **The fusiform face area: a module in human extrastriate cortex specialized for face perception.** *J Neurosci* 1997, **17**:4302-4311.

39. Epstein R, Kanwisher N: **A cortical representation of the local visual environment.** *Nature* 1998, **392**:598-601.

40. • Downing PE, Jiang Y, Shuman M, Kanwisher N: **A cortical area selective for visual processing of the human body.** *Science* 2001, **293**:2470-2473.

This fMRI study reports a region in human visual cortex activated maximally by images of the human body and some of its parts. It remains an open issue whether the activated voxels correspond to a computationally distinct module, perhaps involved in the recognition of biological motion.

41. • Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P: **Distributed and overlapping representations of faces and objects in the ventral temporal cortex.** *Science* 2001, **293**:2425-2430.

Using fMRI of human ventral temporal cortex, these authors found a distinctive pattern of response for each of several categories of objects. The distinctiveness of response was not due to the region that responded maximally to the specific category.

42. •• Gauthier I, Logothetis N: **Is face recognition not so unique after all?** *Cognit Neuropsych* 2000, **17**:125-142.

A unique comparative review of fMRI studies in humans and physiological findings in monkeys. fMRI data have well known counterparts in monkey physiology. In terms of neural encoding, faces are not 'special': other objects can be represented in a similar way for comparable levels of expertise and classification difficulty.

43. Tarr M, Gauthier I: **FFA: a flexible fusiform area for subordinate-level visual processing automatized by experience.** *Nat Neurosci* 2000, **3**:764-769.

44. Pauls J: **The representation of 3-dimensional objects in the primate visual system [PhD Thesis].** Houston: Baylor College of Medicine; 1997.

45. •• Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A: **Neurophysiological investigation of the basis of the fMRI signal.** *Nature* 2001, **412**:150-157.

Here, an experimental tour de force that provides the first comprehensive analysis of the relation between the fMRI signal and neural activity is described. The BOLD (blood oxygenation level dependence) signal used in fMRI is correlated with local field potentials probably originating in dendrites. Activation in human fMRI experiments is very often underestimated owing to the variability of the vascular response.

46. Riesenhuber M, Poggio T: **The individual is nothing, the class everything: psychophysics and modeling of recognition in object classes.** AI Memo 1682, CBCL Paper 185, MIT AI Lab and CBCL, Cambridge, MA 2000. URL: http://www.ai.mit.edu/research/publications/publications.shtml

47. Gauthier I, Skudlarski P, Gore JC, Anderson AW: **Expertise for cars and birds recruits brain areas involved in face recognition.** *Nat Neurosci* 2000, **3**:191-197.

48. • Gauthier I, Tarr MJ, Anderson AW, Skudlarski P, Gore JC: **Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects.** *Nat Neurosci* 1999, **2**:568-573.

Gauthier *et al.* show that expertise with novel objects ('greebles') led to increased activation – as measured by fMRI – of the right hemisphere face areas in humans.

49. Young MP, Yamane S: **Sparse population coding of faces in the inferotemporal cortex.** *Science* 1992, **256**:1327-1331.

50. Freedman D, Riesenhuber M, Poggio T, Miller EK: **Comparison of primate prefrontal and anterior temporal cortex activity during visual categorization.** *Soc Neurosci Abs* 2001, **852**.14.

51. Tversky B, Hemenway K: **Objects, parts, and categories.** *J Exp Psychol Gen* 1984, **113**:169-197.

52. Murphy GL, Brownell HH: **Category differentiation in object recognition: typicality constraints on the basic category advantage.** *J Exp Psychol Learn Mem Cogn* 1985, **11**:70-84.

53. Grill-Spector K, Kanwisher NG: **The functional organization of human ventral temporal cortex is based on stimulus selectivity, not recognition task.** *Soc Neurosci Abs* 2001, **122**.10.

54. • Zhou H, Friedman HS, von der Heydt R: **Coding of border ownership in monkey visual cortex.** *J Neurosci* 2000, **20**:6594-6611.
The authors recorded single cell activity in areas V1, V2 and V4 of awake behaving monkeys. Displays were used in which the same border – typically from a square against uniform lighter or darker background – could be presented as part of different figures. Coding of border ownership modulated the response of a significant percentage of neurons, especially in V2 and V4, even for figure sizes up to 21°. The border ownership-related difference emerged within 25 msec after the response onset.

55. Fukushima K: **Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position.** *Biol Cybern* 1980, **36**:193-202.

56. Wallis G, Rolls ET: **A model of invariant object recognition in the visual system.** *Prog Neurobiol* 1997, **51**:167-194.

57. Hubel DH, Wiesel TN: **Receptive fields, binocular interaction and functional architecture in the cat's visual cortex.** *J Physiol* 1962, **160**:106-154.

58. Blanz V, Vetter T: **A morphable model for the synthesis of 3D faces.** In *Proceedings of SIGGRAPH '99*: 1999 Aug 8–13; Los Angeles. New York: ACM Computer Soc Press; 1999:187-194.