# What constrains the organization of the ventral temporal cortex?

## Isabel Gauthier

Ishai and colleagues[1] recently challenged the current concepts of the organization of the ventral temporal cortex with an intriguing interpretation of their neuroimaging work in object perception. Using fMRI, Ishai *et al.* compared activation in the ventral temporal lobe of normal subjects during observation of images of three object categories: faces, houses and chairs. The results concur with previous imaging studies that compared these or analogous categories in a pairwise fashion[2–5]. Specifically, faces activated part of the fusiform gyrus more than did houses or chairs; houses activated a medial temporal region more than did faces and chairs, and chairs activated a more lateral temporal region more than did faces or houses. Such results are typically interpreted as evidence for a modular organization of extrastriate cortex, with more or less independent processing systems or modules dedicated to different object categories (where 'module' describes a domain-specific entity implemented in a specific neural substrate which may or may not be innately specified).

The most important aspect of this paper is not so much the data as the authors' treatment and interpretation. In particular, Ishai and her colleagues distanced themselves from the classic modular interpretation by considering more than simply the strongest response elicited in each area by the stimuli. They emphasized the distributed nature of the activation patterns: activation for any of the categories extended into regions that were maximally responsive to other categories. In other words, the specialization is not very 'clean'. Ishai *et al.* also noted that activation for non-preferred categories (e.g. for objects and faces in the 'house-selective' region) showed replicable differences. This is a very important point because it suggests that if there is a 'house-processing module', it can also tell you if it's looking at a face or at a chair.

Ishai *et al.* propose that there is a topologically organized continuous representation of object attributes in extrastriate cortex, rather than a collection of modules each dedicated to a single category. This is a surprising conclusion because it is based on results that would, at first glance, rejoice any modularist. However, as I outline below, this study does not present convincing evidence either for or against modular organiz-ation. This can be illustrated by considering four different theories of ventral temporal cortical organization, presented here in order from most modular to less modular.

### 'Faces are special'

According to one model[6,7], faces are processed in a separate module, reasonably enough, given their sociobiological relevance, and all other objects (perhaps with the exception of letters, which also often get their own module) are processed by a 'general' system. Support for this model is obtained from neuro-imaging experiments in which activation to faces is compared with another condition using a variety of non-face objects[2,6]. However, such a design cannot reveal possible differences between categories of non-face objects. Of course, this model can be easily refuted: if different object categories can be dissociated to any significant degree (as was demonstrated by Ishai *et al.*), then why assume that faces are special and everything else is processed in a 'hodgepodge' module? This paper, then, offers one example from the fMRI literature that refutes this model. Another example comes from the ERP literature, where amplitude differences of the N170 component have also been observed between different categories of non-face objects (e.g. cars and shoes) that are as large as those between faces and these categories[8].

### Category-specific model

In this model (the 'strawman' put up by Ishai *et al.*), it is proposed that there could be a module for every category. A category-specific model receives support from rare patients with selectively impaired (or preserved) recognition of a category, such as cows in one case[9]. Ishai and colleagues believe that such an interpretation leads to a fundamental storage problem: that there is not enough cortex to support all of the possible categories. I would add that it is not even clear what defines a category (and we know that most do not have sharp definitions)[10], so it is far from clear how a 'one category–one module' organization could be implemented in the brain.

### Feature-map model

This is the model favored by Ishai *et al.*, which proposes that object-recognition cortex may be 'featurotopic' in the same way that primary visual cortex is retino-topic and somato/motor cortex is somato-topic. The idea is simple and elegant: if extrastriate cortex includes a map in which similar visual features are found near one another, then visually similar objects should activate similar regions of this map. Keiji Tanaka, has championed a similar model on the basis of single-cell recording evidence[11]. The cortical organization that Tanaka and colleagues have observed in inferior temporal cortex (IT) consists of very small regions of cortex in which cells within a tangential distance of 0.2–0.7 mm respond to similar stimuli. Beyond this distance, cells no longer respond to those stimuli, although another small patch of neurons may be found that responds to similar features a distance of 0.4–1.0 mm away. While there is some degree of clustering according to visual features in IT, there is no evidence for a larger scale map of features whereby two of Tanaka's modules near to each other are more likely to respond to similar features than two other modules further apart. It is difficult to map the results of Ishai *et al.* onto the organization described by Tanaka given that the fMRI category-selective areas reported are at least 7 voxels in size, and each 3.125 × 3.125 × 5.0 mm voxel would therefore include a large number of Tanaka's modules with very different selectivity.

One issue that remains problematic for a feature-map organization is that Ishai *et al.* did not include a manipulation that could refute their hypothesis: for example, what region of IT would be activated by an object that we know is a chair but which shares more visual features with the typical house than the typical chair? Epstein and Kanwisher[3] found that the house-selective region (which they call the 'parahippocampal place area' or PPA) responded strongly to a picture of a house, and to a furnished room or the same room when empty, but did not respond to an image of the furniture from the same room placed in a rearranged configuration that did not reflect spatial information about the environment. This result is difficult to explain by the feature-map model because it suggests that there is not a set of features common to all objects that activate the PPA, which are not present in objects to which the PPA does not respond.

### Process-map model

According to this model, extrastriate cortex contains areas that are best suited

*I. Gauthier is at the Department of Psychology, Vanderbilt University, 502 Wilson Hall, Nashville, Tennessee, TN 37240, USA.*

tel: +1 615 322 1778
fax: +1 615 343 8449
e-mail:
isabel.gauthier@
vanderbilt.edu

for different computations. These might be encoding subtle differences between visually similar objects or, conversely, finding underlying similarities between visually dissimilar objects, or encoding objects in terms of spatial layout. If this were the case, faces, houses and chairs would be expected to activate these areas to different extents as we process these objects in different ways. But how can a process-map model account for category differences when subjects are asked to perform the same task with different categories (e.g. the delayed-matching task in Ishai et al.)[2]? The explanation is that different recognition goals are associated with different categories of objects through experience, and this leads to automatic processing biases. For example, we learn to recognize faces at the individual level while very young and this is crucial to most of our social interactions. By contrast, for many objects such as chairs, plants and blenders, we are rarely placed in the situation where we need to discriminate two instances of the class and instead learn to recognize such objects at a higher 'entry-level' ('chair' rather than 'Windsor'). Similarly, when we look at houses, we may have learned that spatial information is of prime importance. These recognition habits might, in time, tune the process map: thus, an area originally best suited for any spatial-layout computation might learn to respond more and more automatically to houses.

In this framework, although neurons in extrastriate cortex display preferences for certain visual features, such preferences are not fixed or intrinsic to their functional role. Imagine, for example, moving to a new planet where shelters look entirely different from what we know as houses. After a while, all of the neurons that previously responded strongly to images of houses might now respond to images of the new habitations. According to the process-map model, this would not reflect a change in the functional role of these neurons (or the neuronal assemblies they constitute): they would be performing very similar computations, only using different inputs.

The strongest support for this model has been obtained in the so-called 'face area'. In some studies, pictures of non-face objects produced more activation in the 'face area' when subject matched them with very specific labels rather than with more categorical ones (e.g. pelican versus bird)[12,13]. Recently, the activation seen in the 'face area' for novel objects was found to increase with intensive training on objects from the same novel category ('Greebles')[14]. Interestingly, the 'face area' was more activated in 'experts' then in novices for Greebles even when the experts were passively looking at them. This supports the idea that experience may lead to automatic processing biases. Finally, autistic adults who have paid little attention to people and their faces – and thus may never have acquired face expertise – also show less

cortical specialization for faces than normal subjects[15].

Recently, a similar explanation – in terms of processes rather than visual features – was offered for the PPA area that responds to houses, landmarks, scenes of different sorts and, in general, any image that carries spatial information about the environment[3].

## Assessment of the models
Do the experiments reported by Ishai et al. allow us to interrogate these different theories? Certainly, Ishai et al.'s results are inconsistent with the 'faces (only) are special' model. However, it is not clear that their results can address the value of the category-specific model. The main reason for this is the choice of categories in their experiment. Houses activate a region that is not considered by some researchers to reflect object recognition per se, but rather the encoding of spatial layout information[3]. Advocates of a category-specific model already consider that faces and chairs should be processed separately and they rarely specify how exclusive the activation must be to be considered as evidence for category-specific modules. I suspect that even a convincing demonstration that activation is fairly distributed might not be perceived as sufficient grounds for abandoning the model. Nonetheless, the emphasis on the distributed nature of activation patterns in Ishai et al.'s paper highlights the necessity for a very clear definition of what should be considered a category-specific module.

The experimental results presented in this paper do not constitute strong support for the feature-map model. In order to do so, it would be necessary to establish that objects perceived as coming from different categories but sharing visual features can lead to activations that cluster together to a greater degree than is the case for objects perceived as coming from the same category but which do not look alike. Ishai et al. did not address this question specifically.

Finally, the choice of stimulus categories also means that this study cannot definitively provide evidence against the process-map model. Faces and houses are each associated with a processing hypothesis, the 'face area' thought to be involved in expert discrimination of visually similar objects[14] and the 'place area' thought to process spatial information, which is associated with houses[3]. In addition, for both categories, experiments have found that the face or place 'areas' can be activated with stimuli that do not look like houses or faces[3,12–14]. This is crucial: the same voxels (neurons) may be recruited by a car, a bird, a Greeble or a face, to the extent that they share some processing requirements, even though the visual features are very different. This leaves us with the area that is selective for chairs, which could also process all other objects that are processed like chairs (that is, mainly at what is often called the 'basic' level).

It would be possible to pose a serious challenge to the process-map model. For example, could evidence be provided for a robust topological arrangement for categories such as birds, fishes and insects, or perhaps for cars, bikes and boats, categories that most of us treat in a similar manner but that are visually dissimilar? Ishai et al. offer a new interpretation of the organization in the visual system. They may have paved the way for studies in which the visual and functional aspects of objects are not confounded, which will allow us to determine what constrains the organization of the ventral temporal cortex.

...............................................................

## References

**1** Ishai, A. et al. Distributed representation of objects in the human ventral visual pathway. Proc. Natl. Acad. Sci. U. S. A. 96, 9379–9384

**2** Kanwisher, N. et al. (1997). The fusiform face area: a module in human extrastriate specialized for face perception. J. Neurosci. 17, 4302–4311

**3** Epstein, R. and Kanwisher, N. (1998) A cortical representation of the local visual environment. Nature 392, 598–601

**4** Aguirre, G.K. et al. (1998). An area within human ventral cortex sensitive to 'building' stimuli: evidence and implications. Neuron 21, 373–383

**5** Haxby J.V. et al. (1999). The effect of face inversion on activity in human neural systems for face and object perception. Neuron 22, 189–199

**6** McCarthy, G. et al. (1997) Face-specific processing in the human fusiform gyrus. J. Cogn. Neurosci. 9, 605–610

**7** Moscovitch, M. et al. (1997). What is special in face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. J. Cogn. Neurosci. 9, 555–604

**8** Rossion, B. et al. The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: an electrophysiological account of face-specific processes in the human brain NeuroReport (in press)

**9** Assal, G. et al. (1984). Nonrecognition of familiar animals by a farmer: zooagnosia or prosopagnosia for animals. Rev. Neurol. (Paris) 140, 580–584

**10** Armstrong, S.L. et al. (1983) What some concepts might not be. Cognition 13, 263–308

**11** Tanaka, K. (1996). Inferotemporal cortex and object vision. Annu. Rev. Neurosci. 19, 109–139

**12** Gauthier, I. et al. (1997). Levels of categorization in visual recognition studied using functional magnetic resonance imaging. Curr. Biol. 7, 645–651

**13** Gauthier, I. et al. Does subordinate-level categorization engage the functionally-defined face area? J. Cogn. Neurosci. (in press)

**14** Gauthier, I. et al. (1999). Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. Nat. Neurosci. 2, 568–573

**15** Schultz, R.T. et al. Abnormal ventral temporal cortical activity among individuals with autism and asperger syndrome during face discrimination. Arch. Gen. Psychiatry (in press)