

# Visual Neurons: Categorization-Based Selectivity Dispatch

Isabel Gauthier and Thomas J. Palmeri

**A computational framework that can account for object categorization and identification has gained empirical support from recent studies of monkey behavior and neurophysiology.**

Primates are avid categorizers, devoting considerable energy to grouping objects, labeling them and interacting with them. These are critical skills for survival, for example when distinguishing predator from prey, friend from enemy or edible object from inedible object. Categorization can also be a source of enjoyment, as exemplified by collectors who create novel categories such as ‘stamps with waterfalls on them’ or ‘impaled garden gnomes.’ The occipito-temporal visual pathway is necessary for such skills — damage to this pathway can lead to a variety of deficits in visual object categorization [1]. There is a strong consensus that the occipito-temporal pathway is responsible for categorizing objects on the basis of visual information (but see [2]). Neuroscientists are eager to explain *how* the neural networks in the occipito-temporal pathway represent and process information for this purpose. Two recent papers by Sigala and colleagues [3,4] tackle this issue, drawing on methods from traditionally different fields, such as formal modeling of categorization and single-cell recording in animals.

The distributed nature of object representations in the occipito-temporal pathway has recently been revealed by functional magnetic resonance imaging (fMRI) [5]. Various object categories were found each to elicit a stable and distributed pattern of activity across the occipito-temporal pathway. Rejecting the popular approach of focusing on small areas highly selective for a particular category [6], Haxby *et al.* [5] uncovered a unique signature for each category across the *entire* occipito-temporal pathway, containing sufficient information to support recognition at a general level, for example distinguishing cats from chairs, houses and faces. Although this distributed representation may be capable of supporting categorization at a general level — chair, face, house and so on — the work did not reveal how discrimination of exemplars within the same category is performed. How, for example, does the occipito-temporal pathway code for the distinction between office and dining chairs or male and female faces, and how does it uniquely code for one’s own house?

On the face of it, grouping objects into the same category and discriminating between objects within the same category may seem different problems.

Including all chairs in one category requires ignoring many differences between them, whereas identifying unique chairs requires attending to those very differences. Indeed, some argue [7] that these problems are so different that the left hemisphere is specialized for making generalizations, as in categorization, whereas the right hemisphere is specialized for making fine distinctions, as in identification. There are, however, computational approaches that can account for both categorization and identification within the same system. This is perhaps best illustrated in one exemplar-based model of categorization and identification, the generalized context model [8].

According to this model, categorization and identification are based on similarity to memories for previously encountered exemplars. Evidence for a particular categorization is based on the summed similarity to category exemplars, whereas evidence for a particular identification is based on the similarity to a specific exemplar; a categorization or identification decision is based on the relative evidence across alternatives. Exemplars are represented as points in a multidimensional psychological space, where dimensions are qualities such as color, shape, spatial frequency content and the like. Similarity between a stimulus and an exemplar is an exponentially decreasing function distance in that space.

One central aspect of the generalized context model is that dimensions are attended according to their diagnosticity for categorization (or identification). This selective attention process acts to stretch (or shrink) the multidimensional space so that some dimensions become more (or less) important than others in determining similarity between items (Figure 1). This assumption of dimension-selective attention is critical for the generalized context model accounting for both categorization and identification [8–10] — stimulus identification typically requires attention to multiple dimensions, whereas many categorization problems may require attention to a limited number of diagnostic dimensions.

The generalized context model has accounted for a variety of fundamental findings in categorization, identification, and other domains. The basic model has been extended by incorporating connectionist learning mechanisms [11], stochastic elements necessary for time-course predictions [12,13], and additional attentional elements for stimulus and response selection [14]. Until recently, however, tests of the generalized context model and related models have been confined to human behavior.

Sigala *et al.* [3] have now shown that the generalized context model can account for categorization by both humans *and* monkeys. They found that the generalized context model accounts much better for the available data than a number of competing models, including those based on prototype abstraction or conditional probabilities. Monkey and human

observers learned to classify exemplars of faces or fish varying along four dimensions, for example, in the case of faces: eye separation, eye height, nose length and mouth height. Only two of the dimensions, eye separation and eye height, were diagnostic, in that both had to be attended for perfect categorization; the other two dimensions were irrelevant. During categorization, monkeys and humans placed more weight on diagnostic dimensions than non-diagnostic dimensions in computing similarities. Furthermore, monkeys and humans showed increased perceptual sensitivity along diagnostic dimensions when making similarity judgments after categorization experience. Essentially, observers attended to dimensions that were diagnostic for categorization — ignoring non-diagnostic dimensions — even outside of the context of the categorization task.

Neurons in the occipito-temporal visual pathway, particularly within the inferior temporal cortex in the monkey, can develop selectivity for objects after categorization experience, appearing to represent those objects in a low-dimensional psychological space [15]. What remains unclear, however, is whether all dimensions of an object are equally represented by inferior temporal cells, or whether task relevance influences selectivity to particular dimensions. In the work of Sigala *et al.* [3], all feature values along all dimensions were presented with equal frequency: this could presumably lead to equivalent representations of diagnostic and non-diagnostic dimensions, as frequency can be an important determinant of visual responses [16].

In a companion paper, Sigala and Logothetis [4] recorded from inferior temporal neurons of monkeys while they categorized faces or fish learned in the behavioral study [3]. The authors did not focus on cells highly selective for one category, such as ‘face cells’ or ‘fish cells’. Indeed, some of the cells apparently discriminated within both face and fish dimensions (see also [15]). After recording the activity of all visually responsive cells, Sigala and Logothetis [4] found that a large proportion showed selectivity along at least one stimulus dimension. For example, one cell might respond more to wide eyes than close eyes and another cell (quite possibly the same cell) might respond more to a round fish tail than a curved fish tail.

Strikingly, although neural responses were not specifically responsive to faces versus fish, neurons showing dimensional selectivity were limited to those coding for dimensions that were diagnostic for categorization (but see [15]). These findings [4] thus demonstrate a tight relationship between categorization behavior and the distributed information represented in the responses of inferior temporal cells in the occipito-temporal visual pathway. The computations described by the generalized context model framework could be supported by such as network.

What still remains rather unclear is how best to characterize the difference in firing rate for diagnostic versus non-diagnostic dimensions. The generalized context model has typically characterized ‘attentional weighting’ of diagnostic dimensions as a flexible process, with dimension weights shifting in an optimal

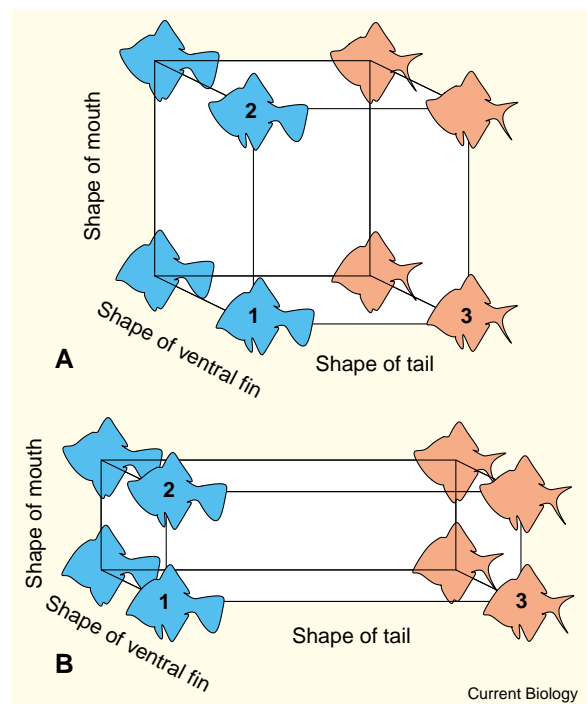


Figure 1. Selective attention to a dimension that is diagnostic for categorization stretches the psychological space.

In this simplified version of the task used by Logothetis and colleagues [3,4], as subjects learn to categorize fish into two groups the multidimensional space becomes relatively stretched along a diagnostic dimension (here shape of the tail) relative to the non-diagnostic dimensions. Before categorization (A), object 1 is equally similar to objects 2 and 3. During categorization (B), object 1 becomes more similar to object 2 than to object 3 through selective attention to the shape of the tail.

fashion depending on task demands [10]. In addition to such an adaptive process, dimensional weighting could also reflect a form of perceptual learning in which diagnostic dimensions become perceptually more discriminable, perhaps because of enhanced receptive field properties or an increased number of neurons coding a dimension. Finally, this weighting could also reflect perceptual learning of another kind: the creation of entirely new dimensions within the visual system [17]. In the case of the fish stimuli, presumably novel to the monkeys, it is not unreasonable to imagine that new feature dimensions for the shape of the dorsal fin, ventral fin and so forth may have been ‘created’ during the categorization task. Unfortunately, Sigala and Logothetis [4] did not report recordings of inferior temporal cell activity before and after categorization, so the true nature of this dimensional attending remains unclear.

It also seems important to measure neural activity in different kinds of task that do not emphasize the same diagnostic features. Is the dimensional selectivity expressed in the activity of single neurons ‘hard-coded’ from perceptual learning, or is it ‘soft-coded’ in the service of a particular categorization task? In humans, expertise identifying objects from a category, birds or cars for example, leads to changes

in selectivity in the fusiform face area — a face-selective part of the human occipito-temporal pathway — even in the context of a location task where identity is completely irrelevant [18]. This type of perceptual learning, which generalizes both to new exemplars of a category and to other tasks where diagnostic dimensions are irrelevant, has been studied mainly for acquired expertise at identifying objects at an individual level. Whether the categorization task used by Sigala and Logothetis [4] can lead to similar generalizations remains to be investigated.

These new studies of categorization in monkeys [3,4] demonstrate once again the power of the exemplar-based framework for understanding visual object recognition. This should not come as a surprise: image-based, viewpoint-dependent theories of object recognition [19] are essentially exemplar-based accounts, and have received support from neurophysiology [20]. More than evoked potentials or fMRI, single-cell recordings may provide the right level of analysis for testing models such as the generalized context model, and the neurophysiology appears particularly well positioned to bridge the large gap currently existing between human imaging studies and computational models of object categorization.

#### References

1. Farah, M.J. (1990). *Visual Agnosia: Disorders of Object Recognition and What They Tell Us About Normal Vision*. (The MIT Press, Cambridge, MA).
2. Freedman, D.J., Riesenhuber, M., Poggio, T. and Miller, E.K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science* 291, 312–316.
3. Sigala N, Gabbiani F and Logothetis NK. (2002). Visual categorization and object representation in monkeys and humans. *J. Cogn. Neurosci.*
4. Sigala, N. and Logothetis, N.K. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature* 415, 318–320.
5. Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L. and Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293, 2425–2430.
6. Kanwisher, N., McDermott, J. and Chun, M.M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *J. Neurosci.* 17, 4302–4311.
7. Marsolek, C. (1999). Dissociable neural subsystems underlie abstract and specific object recognition. *Psychol. Sci.* 107, 111–118.
8. Nosofsky, R.M. (1986). Attention, similarity and the identification-categorization relationship. *J. Exp. Psychol. Gen.* 115, 39–61.
9. Nosofsky, R.M. (1984). Choice, similarity and the context theory of classification. *J. Exp. Psychol. Learn. Mem. Cogn.* 10, 104–114.
10. Nosofsky, R.M. (1998). Optimal performance and exemplar models of classification. In *Rational Models Of Cognition*, M Oaksford, N Chater, eds. (Oxford University Press).
11. Kruschke, J.K. (1992). ALCOVE: an exemplar-based connectionist model of category learning. *Psychol. Rev.* 99, 22–44.
12. Lamberts, K. (2000). Information-accumulation theory of speeded categorization. *Psych Review* 107, 227–260.
13. Nosofsky, R.M. and Palmeri, T.J. (1997). An exemplar-based random walk model of speeded classification. *Psychol. Rev.* 104, 266–300.
14. Logan, G.D. An instance theory of attention and memory. *Psych. Rev.* in press.
15. Op de Beeck, H., Wagemans, J. and Vogels, R. (2001). Inferotemporal neurons represent low-dimensional configurations of parameterized shapes. *Nat. Neurosci.* 4, 1244–1252.
16. Watanabe, T., Nanez, J.E. and Sasaki, Y. (2001). Perceptual learning without perception. *Nature* 413, 844–848.
17. Schyns, P.G., Goldstone, R.L. and Thibaut, J.-P. (1998). The development of features in object concepts. *Behav. Brain Sci.* 21, 1–54.
18. Gauthier, I., Skudlarski, P., Gore, J.C. and Anderson, A.W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nat. Neurosci.* 3, 191–197.
19. Riesenhuber, M. and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nat. Neurosci.* 2, 1019–1025.
20. Logothetis, N.K. and Pauls, J. (1995). Psychophysical and physiological evidence for viewer-centered object representations in the primate. *Cereb. Cortex* 5, 270–288.