

- Carpenter A F, Georgopoulos A P, Pellizzer G 1999 Motor cortical encoding of serial order in a context-recall task. *Science* **283**: 1752–7
- Carter C S, Macdonald A M, Botvinick M, Ross L L, Stenger V A, Noll D, Cohen J D 2000 Parsing executive processes: strategic vs. evaluative functions of the anterior cingulate cortex. *Proceedings of the National Academy of Science USA* **97**: 1944–8
- Connolly J D, Goodale M A, Desouza J F, Menon R S, Vilis T A 2000 Comparison of frontoparietal fMRI activation during anti-saccades and anti-pointing. *Journal of Neurophysiology* **84**: 1645–55
- Fu Q G, Suarez J I, Ebner T J 1993 Neuronal specification of direction and distance during reaching movements in the superior precentral premotor area and primary motor cortex of monkeys. *Journal of Neurophysiology* **70**: 2097–116
- Geyer S, Matelli M, Luppino G, Zilles K 2000 Functional neuroanatomy of the primate isocortical motor system. *Anatomy and Embryology* **202**: 443–74
- Graziano M S, Gross C G 1998 Spatial maps for the control of movement. *Current Opinion in Neurobiology* **8**: 195–201
- Hikosaka O, Sakai K, Miyauchi S, Takino R, Sasaki Y, Putz B 1996 Activation of human presupplementary motor area in learning sequential procedures: a functional MRI study. *Journal of Neurophysiology* **76**: 617–21
- Kalaska J F, Scott S H, Cisek P, Sergio L E 1997 Cortical control of reaching movements. *Current Opinion in Neurobiology* **7**: 849–59
- Picard N, Strick P L 1996 Motor areas of the medial wall: a review of their location and functional activation. *Cerebral Cortex* **6**: 342–53
- Sanes J N 2000 The relation between human brain activity and hand movements. *Neuroimage* **11**: 370–4
- Sanes J N, Donoghue J P 1997 Static and dynamic organization of motor cortex. *Advances in Neurology* **73**: 277–96
- Shima K, Tanji J 2000 Neuronal activity in the supplementary and presupplementary motor areas for temporal organization of multiple movements. *Journal of Neurophysiology* **84**: 2148–60
- Strick P L, Dum R P, Picard N 1998 Motor areas on the medial wall of the hemisphere. *Novartis Foundation Symposium* **218**: 64–75
- Todorov E 2000 Direct cortical control of muscle activation in voluntary arm movements: a model. *Nature Neuroscience* **3**: 391–8
- Wise S P, Boussaoud D, Johnson P B 1997 Caminiti Premotor and parietal cortex: corticocortical connectivity and combinatorial computations. *Annual Review of Neuroscience* **20**: 25–42R

J. Donoghue

Neural Representations of Objects

1. How Small Numbers of Neurons Code Information About Objects

One of the most commonly measured units of representation in the brain is the neuron. Somewhat surprisingly, functional information about objects

seems to be coded at this incredibly fine scale. Single-unit recordings of neurons in the visual cortex reveal activity that is often tuned to individual objects or classes of objects. That is, a particular stimulus object presented in the receptive field of a given neuron will produce significantly greater activity (more spikes per second) in that neuron as compared to any other tested stimulus (typically a varied assortment of natural and artificial objects). For example, neurons in many cortical extrastriate areas (the part of the primate brain that is thought to support high-level vision, including object recognition) in the monkey brain are highly selective to primate faces and, sometimes, even individual faces, and recent evidence has revealed similar selectivity for many classes of familiar objects (Sheinberg and Logothetis 2001). There is similar evidence for cortical neurons specifically tuned to individual nonface objects, at least when the stimuli are highly familiar to the monkey (Logothetis and Sheinberg 1996). At the same time, few researchers believe that single neurons actually code for individual objects. Rather, it is commonly held that large populations of neurons represent objects or object classes in a distributed fashion. Some evidence for this hypothesis is provided by single-unit studies that systematically decompose the response of a given neuron by exploring which particular features of an object are actually driving neural activity. It is often the case that the maximum response is maintained when only schematic elements of the object are presented, suggesting that single neurons must work in concert to represent an object (Tanaka 1996).

To the extent that individual neurons participate in the representation of a given object, there remains the question of the format of this representation (Marr 1982). Object representations can be more or less invariant over changes in the appearance of the object in the image (see *Object Recognition: Theories*). Ideally, perception and recognition should remain robust when an object is moved, rotated, or illuminated differently. Most object-tuned neurons exhibit some invariance: the level of response to a preferred object is not dramatically affected by changes in the size or the spatial position of an object for which a given neuron is selective. On the other hand, these same neurons often show great sensitivity to changes in orientation, viewpoint, and illumination direction. For example, a large percentage of object-tuned neurons are also 'view-tuned' in that they show the highest level of activity to specific objects in specific viewpoints. Face-selective neurons may respond most strongly when presented with a frontal view of a face, but show a progressively diminished response as the face rotates away from this viewpoint. Similarly, in the visual cortex of monkeys taught to recognize novel objects, neurons that become selective for these objects are typically highly tuned to the particular viewpoints used during training. These findings are consistent with much of the behavioral data on object recogni-

tion, where recognition performance is often strongly dependent on the familiarity of a given viewpoint (Tarr and Bülthoff 1998). Thus, at the finest scale of analysis, single neurons participate in the representation of objects in a manner that is invariant over size and location, but specific to viewpoint and illumination.

1.1 Hierarchical Processing in the Visual Cortex

How can single neurons or populations of neurons code the complex patterns and feature conjunctions present in most objects? Many models of object recognition assume that the answer is that object- or feature-tuned neurons are the end-point of a hierarchy of progressively more and more complex visual codes (see *Visual System in the Brain*). Supporting this approach, neurons in the earliest visual areas (see Occipital lobe) respond to very simple generic properties of objects, for instance, oriented edges or blobs of color. As one moves ‘up’ the visual pathway (see Dorsal and ventral visual systems) the coding becomes more and more elaborate, with neurons at one level having responses that generally code more complex properties of objects than the level(s) below them (Van Essen and Maunsell 1983). The principle behind this hierarchy is straightforward: from one level to the next, neurons respond only if they receive input from multiple neurons, thereby creating a code that is necessarily more complex than the individual inputs. This many-to-one principle not only pools information in the straightforward sense, but because of the pattern of connectivity from one layer to the next, assembles new and more complex features as one moves up the hierarchy. Given the strong evidence for this sort of feed-forward processing in the visual system, many computational models, even those that are quite dissimilar in other respects, have adopted a hierarchical processing architecture to assemble complex object features (Hummel and Biederman 1992). Although the hierarchical approach is appealing, some computational theorists have pointed out that constructing complex features in this fashion may lead to overly specific object representations. For instance, imagine using the features present in the image to derive a very precise description of an object. If the object is then rotated in depth how does one combine the new very precise description with its predecessor? To address this need for invariance over changes in the image, some computational models have adopted a hierarchical architecture in which there is an alternation between conjunctions of features—the standard hierarchical model—and disjunctions of features (Riesenhuber and Poggio 1999). For example, one layer might consist of neurons that respond only if they receive input from a *collection* of neurons coding for simpler features, while the next layer might consist

of neurons that respond if they receive input from *any* of several neurons. This sort of hierarchical coding is exemplified by a single neuron that codes for a three-dimensional object regardless of its orientation, but only by virtue of the fact that this view-invariant neuron is driven by any one of several view-tuned neurons, each coding for a different viewpoint of the same object (Logothetis and Sheinberg 1996).

1.2 The Building Blocks of Object Representation

The hierarchical organization of the visual cortex suggests that objects are represented in terms of a ‘vocabulary’ of features constructed from simpler features (Marr 1982). This assumption implies that somewhere between orientation-tuned neurons in early vision and object-tuned neurons at the highest levels of the visual system, are neurons representing the critical features that distinguish one object from another. Under one view, these critical features should look something like object parts; intuitively the first level of decomposition when we look at objects (Marr 1982). There is, however, little evidence to indicate that neurons coding for parts as we know them are the precursors to object-tuned neurons. When the responses of neurons in intermediate and high-level visual areas are systematically decomposed, they show consistent responses to patterns that are more complex than edges, but less complex than objects. However, these ‘features’ appear arbitrary, coding for odd, almost random shape properties. For example, one neuron might respond most strongly to a lollipop shape, while another neuron might respond to a eight-point star. While such features may appear to be random or, at best, schematic versions of certain objects, they are not arrayed haphazardly across the visual cortex. Rather, as with most visual areas, there is a columnar organization in which neurons within a column all code for the same feature and adjacent columns tend to code for similar features, albeit with columnar boundaries sometimes representing dramatic changes in feature selectivity (Tanaka 1996). It is worth noting that although such features might provide the ‘building blocks’ for representing objects, there is, at present, no plausible account that can explain how this occurs or why these particular features are used.

2. How Large Numbers of Neurons Code Information About Objects

As discussed above, despite neurophysiological results that seem to indicate that objects are represented by single neurons, there is little fondness for this stance. Many researchers have observed that single-unit methods limit the direct measurement of simultaneous



Figure 1
A novel object known as a 'Greeble'

neural activity to small numbers of units and it is only by considering single neuron responses in the context of population codes that we can understand the true nature of object representation (Booth and Rolls 1998). Thus, to the extent that a given neuron plays a larger role in the representation of one object over all other objects (or at least tested objects), a single-unit approach may bias the experimenter to interpret the neural response as 'tuned' for one object. The critical point is that this preference is not equivalent to 'representation.' Even if a neuron does not respond maximally to a given object, its partial (or even low) response may be part of the code by which a complex object is represented. Thus, individual objects and classes of objects may be represented by ensembles of 1,000s or 1,000,000s of neurons—something impossible to assess completely using neurophysiological measurements. In contrast, neuroimaging (Positron Emission Tomography—'PET'; and functional Magnetic Resonance Imaging—'fMRI') indirectly measures the conjoint activity of large numbers of neurons; the smallest unit of measurement being about a 3mm^3 'voxel' which encompasses approximately 1–2 million neurons.

2.1 Category Selectivity in the Visual Cortex

The neural representation of objects at this larger scale, as measured by PET or fMRI, reveals a degree of organization in extrastriate areas that mirrors the selectivity observed at the single-neuron level. That is, much as individual neurons respond preferentially to individual objects, localized regions of the visual

cortex respond preferentially to classes of objects. Moreover, as with neurophysiological methods, the clearest preferences are obtained with highly-familiar object classes. A region of the visual cortex known as mid-fusiform gyrus (midFG) shows significantly higher activity when observers view faces as compared to when they view common objects (Kanwisher 2000). Similar selectivity for small regions of the cortex near the midFG has been found for letters, places, houses, and chairs. Thus, object classes appear to be represented minimally by large numbers of neurons in localized regions of the extrastriate cortex.

However, even this level of analysis may be misleading. Evidence for localized category-selective cortical regions comes from neuroimaging methods that compare the activation observed for one object class, e.g., faces, to a second object class, e.g., flowers. In truth, viewing any class of objects produces a pattern of activity across much of the ventral temporal cortex that is different from the activation pattern obtained for any other class. These differences, however, are often subtle and relatively small compared to the large differences seen between highly-overlearned categories such as faces and less familiar objects. Such differences may be critical elements in the complete neural code for objects; if so, objects and classes of objects may be represented as large-scale networks of neurons distributed over much of the visual cortex.

2.2 The Origins of Category Selectivity

If category selectivity is not a marker for the representation of objects *per se*, how can we explain the preferential responses obtained in localized regions of the visual cortex? One interpretation is that the visual cortex is organized into dedicated *modules* that are prewired to be selective for the geometries of particular object classes (Kanwisher 2000). Alternatively, category selectivity may reflect particular computations that, with experience, become automatically executed on distributed representations of objects and classes (Tarr and Gauthier 2000)—the 'Process Map' hypothesis. Again, the most salient example is faces: we are trained from birth to recognize faces at the individual level. The result of this lifetime experience is that, as adults, we are experts at processing faces and, by default, apply a degree of perceptual analysis that is not necessary for recognizing objects at more categorical levels. This interpretation suggests that the same computational resources should be applied to any object class (not just faces) for which observers are experts at recognizing individuals. Several studies bear out this prediction. Neuroimaging reveals that both car and bird experts show selectively higher activation in midFG for their domain of expertise. Similarly, for novel classes of objects (e.g., 'Greebles'; see Fig. 1), trained experts, but not untrained novices, exhibit increased midFG activation for the trained object class (Tarr and Gauthier 2000). Apparently, visual

experience with many similar objects ‘tunes’ midFG, as well as other category selective extrastriate areas, to respond automatically when those or other objects from the class are seen again (see *Neural Plasticity*). Consequently, category selectivity is perhaps best understood as a consequence of how experience reorganizes the visual recognition process, rather than an indicator of how objects themselves are neurally represented.

3. The Functional Role of Localized Brain Regions in Object Representation

Neuroimaging methods are useful for localizing *where* in the brain particular computations occur, but are less informative regarding the different roles these neural substrates play in producing behavior. In contrast, neuropsychological methods allow us to explore deficits in behavior as a direct consequence of brain injuries to particular brain regions. Given the neuroimaging results discussed in the previous section, damage to category-selective areas such as midFG should impair face recognition, and, perhaps, performance in any difficult individual-level recognition task. Such neuropsychological cases have been documented: Patients who suffer damage to extrastriate visual areas often lose the ability to recognize many or all object classes—a syndrome known as *agnosia* (see *Agnosia*). Although these patients are grouped under a single label, agnosia includes damage to many different cortical regions and many different recognition deficits.

Accounts of category selectivity as revealed by neuroimaging argue for either modular subsystems organized by geometrically-defined object classes or for a flexible process map of distributed representations and processing mechanisms that are recruited by expertise and individual-level recognition. If the modularity argument is correct, then damage to the module specialized for faces should only result in deficits at face recognition but no deficits in nonface object recognition. Conversely, damage to the module specialized for objects or classes of objects should only result in deficits in object recognition, but not face recognition. The logic of the double dissociation method is clear: if two distinct brain regions perform independent functions, then lesions to one region should not affect the functioning of the other region and vice versa. Alternatively, if the process map model is correct and processing mechanisms can be recruited for the recognition of any object class, there should be common deficits in face and object recognition when there is damage to the ventral temporal cortex. It is critical, however, that the recognition tasks be equated in such comparisons: spurious differences may be found simply because different computational components of the recognition process are being recruited for different tasks.

3.1 Patterns of Sparing and Loss in Visual Object Recognition

Given knowledge about the location of the brain damage in an individual, a clear pattern of sparing and loss may help reveal the functional role of intact and damaged brain regions (see *Cognitive Neuropsychology, Methodology of*). Patients with injuries to their ventral temporal cortex often suffer a specific form of agnosia referred to as *prosopagnosia* (see *Prosopagnosia*) in which face recognition appears to be disproportionately impaired relative to object recognition (Farah 1990). At the same time, there is one case study in which the brain-injured patient appears to have intact face recognition, but severely impaired object recognition. Although this pattern is consistent with the modularity account, further examination of prosopagnosia does not support this conclusion. When patients with apparent face recognition deficits perform difficult individual-level discriminations with nonface objects (e.g., snowflakes), they show equal impairment with faces and nonfaces (Gauthier et al. 1999). Thus, neuropsychology yields a mixed bag of results supporting both modular and nonmodular accounts.

See also: Agnosia; Binding Problem, Neural Basis of; Face Recognition Models; Face Recognition: Psychological and Neural Aspects; Neural Synchrony as a Binding Mechanism

Bibliography

- Booth M C A, Rolls E T 1998 View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. *Cerebral Cortex* **8**: 510–523
- Farah M J 1990 *Visual Agnosia: Disorders of Object Recognition and What They Tell Us About Normal Vision*. MIT Press, Cambridge, MA
- Gauthier I, Behrmann M, Tarr M J 1999 Can face recognition really be dissociated from object recognition? *Journal of Cognitive Neuroscience* **11**: 349–370
- Hubel D H 1995 *Eye, Brain, and Vision*. Freeman, San Francisco, CA
- Hummel J E, Biederman I 1992 Dynamic binding in a neural network for shape recognition. *Psychological Review* **99**: 480–517
- Kanwisher N 2000 Domain specificity in face perception. *Nature Neuroscience* **3**: 759–63
- Logothetis N K, Sheinberg D L 1996 Visual object recognition. *Annual Review of Neuroscience*. Annual Reviews, Palo Alto, CA
- Marr D 1982 *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Freeman, San Francisco, CA
- Palmer S E 1999 *Vision Science: Photons to Phenomenology*. MIT Press, Cambridge, MA
- Riesenhuber M, Poggio T 1999 Hierarchical models of object recognition in cortex. *Nature Neuroscience* **2**: 1019–1025
- Sheinberg D L, Logothetis N K 2001 Finding familiar objects in real-world scenes: The role temporal cortical neurons in natural vision. *Journal of Neuroscience* **21**: 1340–1350

- Tanaka K 1996 Inferotemporal cortex and object vision. *Annual Review of Neuroscience*. Annual Reviews, Palo Alto, CA
- Tarr M J, Bülthoff H H 1998 *Object Recognition in Man, Monkey, and Machine*. MIT Press, Cambridge, MA
- Tarr M J, Gauthier I 2000 FFA: A flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nature Neuroscience* 3: 764–69
- Van Essen D C, Maunsell J H R 1983 Hierarchical organization and functional streams in the visual cortex. *Trends in Neuroscience* 6: 370–375

M. J. Tarr and Q. C. Vuong

Neuronal Synchrony as a Binding Mechanism

Neuronal systems have to solve immensely complex combinatorial problems and require efficient binding mechanisms in order to generate representations of perceptual objects and movements. In the context of cognitive functions, combinatorial problems arise from the fact that perceptual objects are defined by a unique constellation of features, the diversity of possible constellations being virtually unlimited. Combinatorial problems of similar magnitude have to be solved for the acquisition and execution of motor acts. Although the elementary components of motor acts—the movements of individual muscle fibres—are limited in number, the spatial and temporal diversity of movements that can be composed by combining the elementary components in ever-changing constellations is again virtually infinite. In order to establish neuronal representations of perceptual objects and movements the manifold relations among elementary sensory features and movement components have to be encoded in neural responses. This requires binding mechanisms that can cope efficiently with combinatorial complexity. Evolved brains have acquired an extraordinary competence to solve such combinatorial problems and it appears that this competence is a result of the evolution of the cerebral cortex.

1. Two Complementary Binding Strategies

In the primary visual cortex of mammals, relations among the responses of retinal ganglion cells are evaluated and represented by having the output of selected arrays of ganglion cells converge in diverse combinations onto individual cortical neurons. Distributed signals are bound together by selective convergence of feed forward connections (Hubel and Wiesel 1962). Through iteration of this strategy neurons in higher cortical areas acquire increasingly sophisticated response properties, representing complex constellations of features and, in the special case

of faces, even whole perceptual objects (Gross 1992, Tanaka 1997).

However, this strategy—of binding features together by recombining input connections in ever-changing variations and representing relations explicitly by responses of specialized binding cells—results in a combinatorial explosion of the number of required binding units. This problem is further accentuated by the fact that perceptual objects are often defined by conjunctions of features encoded in different sensory systems; they have specific visual, haptic, or acoustic properties. Moreover, if binding is achieved solely by the formation of conjunction-specific neurons, difficulties arise if novel objects with unfamiliar feature constellations need to be represented. It has been proposed, therefore, that the cerebral cortex uses a second, complementary strategy, which permits utilization of the same neurons for the representation of different objects (Hebb 1949). Here, the particular constellation of features characterizing perceptual objects is thought to be represented by the joint and coordinated activity of a dynamically associated ensemble of cells, whereby each of the participating neurons represents explicitly only one of the elementary features that characterize a particular perceptual object. Different objects can then be represented by recombining in various constellations neurons tuned to elementary features, whereby a particular cell can participate at different times in different assemblies. This sharing of neurons economizes on neuron numbers and copes effectively with the huge combinatorial space occupied by real-world objects.

For assembly coding two constraints need to be met, however. First, a selection mechanism is required that permits dynamic, context-dependent association of neurons into distinct, functionally coherent assemblies. Second, grouped responses must get labeled so that they can be distinguished by subsequent processing stages as components of one coherent representation and do not get confounded with other unrelated responses.

Such labeling can be achieved by jointly raising the saliency of the selected neurons, because this ensures that they are processed and evaluated together at the subsequent processing stage. Neuronal systems have three options to increase the relative saliency of selected responses: first, non-grouped responses can be inhibited; second, the amplitude of the selected responses can be enhanced; and third, the selected cells can be made to discharge in precise temporal synchrony. Inhibition simply eliminates potentially confounding signals, while enhanced firing raises the impact of the grouped responses through temporal summation and synchronization through spatial summation of synaptic potentials (Singer 1999). Selection of responses by enhancing discharge rates is common in attentional gating (Luck et al. 1997) but it has certain disadvantages when used as the only mech-

10567