Neural correlates of categories and concepts
Earl K Miller*, Andreas Nieder, David J Freedman and Jonathan D Wallis

The ability to readily adapt to novel situations requires something beyond storing specific stimulus–response associations. Instead, many animals can detect basic characteristics of events and store them as generalized classes. Because these representations are abstracted beyond specific details of sensory inputs and motor outputs, they can be easily generalized and adapted to new circumstances. Explorations of neural mechanisms of sensory processing and motor output have progressed to the point where studies can begin to address the neural basis of abstract, categorical representations. Recent studies have revealed their neural correlates in various cortical areas of the non-human primate brain.

Addresses
Picower Center for Learning and Memory, RIKEN-MIT Neuroscience Research Center, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA
*e-mail: ekm@ai.mit.edu

Abbreviations
ITC   inferior temporal cortex
PFC   prefrontal cortex
SEFs   supplementary eye fields
S–R   stimulus–response

Introduction
In the first half of the 20th century, scientific investigations of the mind were dominated by BF Skinner and the radical behaviorist philosophy that all mental life could be explained by stimulus–response (S–R) associations and, therefore, that virtually every mental event could be investigated by examining details of sensory input and motor output. The failure of this philosophy (and the advent of modern cognitive science) came with the realization that it cannot account for all behavior. For one thing, the capacity of humans and other animals to readily adapt to novel events with no prior S–R associations must depend on 'internal' representations — those that encode general characteristics of events into classes abstracted beyond specific details of sensory inputs and motor outputs. This is an efficient and flexible way to deal with a complex world. These abstracted representations allow the navigation of many different situations with a minimal amount of storage. Thus, they can be easily generalized to novel situations that share basic characteristics but that may, on the surface, appear very different.

As in experimental psychology, investigations in systems neuroscience have followed an outward-in trajectory. We needed to understand something about processing at the 'periphery' of the brain — the input/output details — before we could tackle aspects of more 'internal' processing. The wealth of data about sensory and motor processing collected over the past several decades provides a backdrop to address questions about the neural basis of abstract categories and concepts in non-human primates.

Perceptual categories
Categorical representations do not faithfully track exact sensory input. They provide useful groupings and divisions not present in the external world. Consider a simple example: crickets sharply divide a certain range of pure tones into 'mate' versus 'bat' (a predator) [1]. Even though the input varies along a continuum, behavior does not. At the low end of the range, crickets approach sounds equally; however, when the tone of the sound reaches 16 kHz, their behavior suddenly flips to avoidance, and then remains equivalent across another wide range. This allows the crickets to maximize reproduction while minimizing disaster. Another example is humans' perception of the facial expressions of emotion [2], which also flip at a discrete point (e.g., from 'happy' to 'sad'). Thus, the representation of perceptual categories must involve something distinct from the neural tuning that encodes physical attributes — the gradual changes in neural activity as attributes gradually change (e.g. shape, orientation, direction). Perceptual categories have sharp boundaries (not gradual transitions) and members of the same category are treated as similar even though their physical appearance may vary widely.

By learning to take multiple dimensions into account in order to make sharp distinctions and groupings, advanced animals such as monkeys can acquire higher-level perceptual categories such as 'animal', [3] 'food', [4] 'tree', and 'fish'. [5] The search for neural correlates of such high-level categories has naturally focused on brain regions at the final stages of visual processing, such as the inferior temporal cortex (ITC), a brain region critical for visual recognition [6–9], and the prefrontal cortex (PFC), which receives highly processed visual information from the ITC and orchestrates voluntary, goal-directed behaviors [10] (Figure 1).

Figure 1
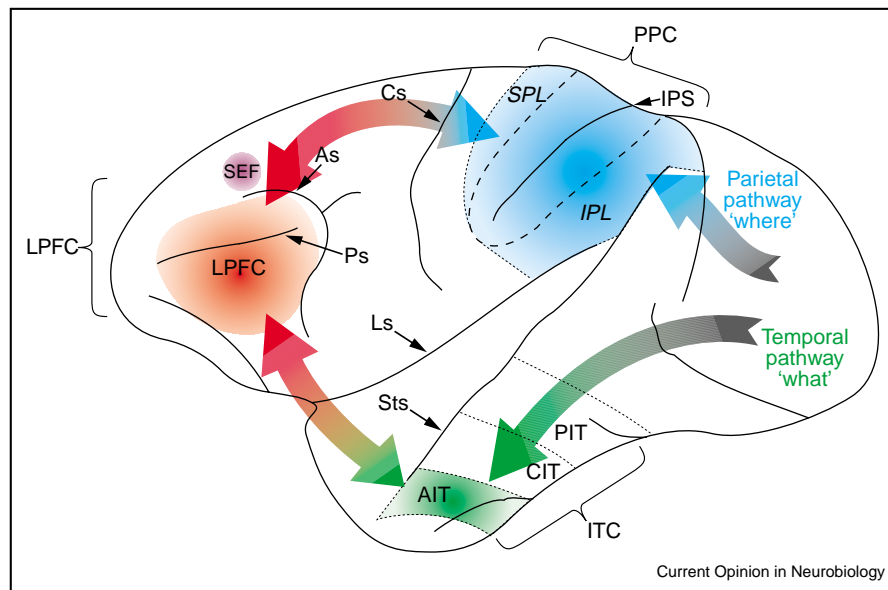
Regions of the macaque neocortex discussed in this review. Illustrated are the two general visual system pathways thought to mediate form and color vision ('what') and visuospatial functions ('where'). The differences between these pathways are relative, however. There is a substantial intermixing and interchange of this information between the pathways. Abbreviations: AIT, anterior inferior temporal; As, arcuate sulcus; CIT, center inferior temporal; Cs, central sulcus; IPL, inferior parietal lobule; IPS, interparietal sulcus (unfolded); LPFC, lateral prefrontal cortex; Ls, lateral sulcus; PIT, posterior inferior temporal; Ps, principal sulcus; SPL, superior parietal lobule; Sts, superior temporal sulcus.

ITC neurons with tuning properties suggestive of categories have been known since the seminal work of Gross and co-workers [7]. They described a small population of 'face cells' that were strongly activated by the gestalt of a face, but not by individual features of faces or by other stimuli. Face cells have also been reported in the lateral PFC [11]. More recently, Vogels [12] reported ITC neurons in trained monkeys that were selectively activated by 'trees' or 'fish' and that showed relatively little differentiation between diverse examples of those categories. Interpretation of such work rests on the assumption that similar mechanisms operate in the human brain; indeed, analogous results have been found in human experiments. Kreiman *et al.* [13] recorded from the hippocampus, amygdala, and adjacent cortex of human epileptic patients and found neurons in these areas that were selectively activated by certain classes of stimuli. Functional imaging of blood flow signals suggests differential distributions of neurons in human visual cortex activated by the categories 'face' and 'place' [14•].

This use of natural images can help identify neurons that ultimately contribute to category judgments. However, unless neurons are tested for the identifying characteristics of perceptual categories (sharp boundaries, equivalence within a category), it is possible that their activity reflects physical similarity rather than category membership *per se*; trees and faces, after all, look more like one another than like other stimuli. So, Freedman *et al.* [15••]

trained monkeys to categorize stimuli along a morphing continuum of different blends of 'cats' and 'dogs'. This revealed lateral PFC neurons with hallmarks of category representations: sharp differences in activity to similar-looking stimuli across a discrete category boundary, yet similar activity to different-looking members of the same category.

But how exactly are perceptual categories acquired? Some theories posit that we build up prototypes of each category; others suggest that we construct a 'list' of defining features. Insight into the learning process comes from a recent study by Sigala and Logothetis [16••]. They found that stimulus features relevant for category judgements relative to irrelevant features were enhanced in ITC activity. This suggests a process that weights features according to their relevance for category membership, and perhaps supports more feature-based models of categorization. In general, there is ample evidence to show that neural circuitry in the ITC and PFC has the plasticity to acquire new categories: ITC neurons are preferentially activated by trained stimuli [17] and show selectivity for learned groupings of pictures [18]. A modest amount of experience causes the formation of local clusters of neurons with similar properties in the perirhinal cortex [19] and modifies PFC stimulus representations to make them much smaller, but more robust and efficient [20]. Baker *et al.* [21••] recently trained monkeys that certain features of complex objects 'go together'; this resulted in ITC

neural selectivity for those feature combinations in the trained animals. Mechanisms that piece together unified representations of related features have obvious advantages for learning categories.

Numerosity: quantitative categories

A number is a most abstract category: 'two' could mean two beers or two countries. Human neuropsychological and imaging studies point to the involvement of the parietal lobe and frontal lobe in numerical judgments [22]; recent studies in monkeys (which have basic numerical abilities [23–25]) have searched for their neural correlates in these areas. For practical purposes, the first experiments in this line of work started with smaller numbers that can be handled by monkeys and other animals [26]. Of course, animals cannot derive numerical information by verbally and serially 'counting' items (like humans); rather they encode numerosities in a non-verbal, analogue magnitude format, as do pre-verbal human infants [27,28]. Such approximate representations of numerical values may be regarded as biological precursors of adult humans' counting abilities.

Sawamura *et al.* [29**] trained monkeys to alternate between five arm movements of one type and five of another. They found neurons in a somatosensory-responsive region of the superior parietal lobule (SPL) that kept track of the movement number. Relatively few such neurons were found in the same lateral PFC regions in which other perceptual categories and face cells had been found. By contrast, Nieder *et al.* [30**] trained monkeys to judge the number of items (between one and five) in a visual display and found ample number-tuned neurons in this lateral PFC region. One possibility for the difference between these studies may be modality employed (touch versus vision), but another significant factor may be the level of abstraction. Most movement–number representations found by Sawamura *et al.* [29**] (85%) were not abstract: number-selective activity depended on whether the monkey's movement was 'push' or 'turn'. In contrast, the visual number representations found in the lateral PFC were abstract and generalized. Changes in the physical appearance of the displays had little effect on the activity of the majority of the number-tuned neurons (Figure 2; [30**]).

Such neurophysiological studies can tell us where and how category information might be represented in the brain. For example, they show that categories are reflected in the activity of single neurons, just like physical attributes such as color or shape. Ingraining categories at this low level (rather than having them emerge as interactions between neurons) may explain why objects can be categorized so rapidly [4]. Even so, these neurons are not 'grandmother cells' (i.e., single neurons that represent an entire percept or concept). Neither categories nor any other type of information seems to be

Figure 2
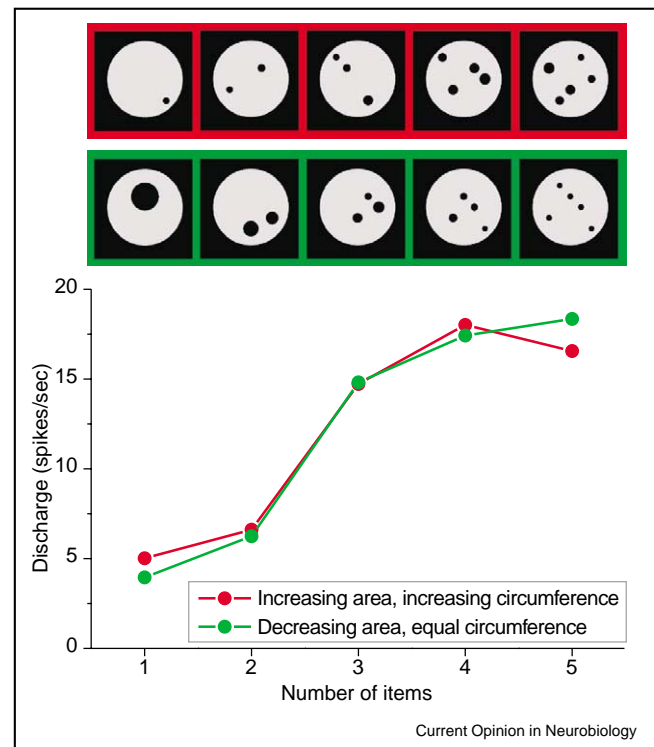
A neuron from the lateral PFC tuned to the number of displayed items. Monkeys were required to judge the number of items in visual displays. This neuron is tuned to number but virtually unaffected by changes in the exact appearance of the displays. The red and green lines show activity in response to two different stimulus sets. This neuron was unaffected by changes in the position and size of the items. Adapted from Nieder *et al.* [30**].

ultimately encapsulated in an individual neuron. Just as physical attributes, such as color, ultimately depend on a pattern of activity across a neural population, so do categories. Single neurons are rarely reliable detectors (but see [31]) and neural regions biased towards face or place notwithstanding, there is little evidence that such specificity is a general coding scheme. Finally, neurophysiological studies can also place constraints on theory. For example, both behavioral and physiological data [30**] argue for parallel extraction of numerosities, supporting one type of network model of numerosity [32].

Rules: forming general plans

Meaningful groupings of information do not only occur in the sensory domain. We can lump together sets of events and actions into general guidelines, principles, or rules for behavior. Consider restaurants: we have knowledge of generic rules such as 'wait to be seated', 'order', and 'pay the bill' that are long decoupled from the specific experiences in which they were learned. We then have a notion of what is expected of us the first time we walk into a new restaurant.

As learning and applying rules are central to complex, goal-directed behavior, several investigators have recently argued that their acquisition and representation may be a cardinal PFC function; their lack could explain the 'goal neglect' that follows PFC damage [10,33•,34]. Indeed, neural correlates of rules are evident in the PFC of both monkeys and rodents [35–37,38••]. For example, in monkeys trained to alternate between a spatial rule (go to the location of a cue) and a conditional rule (go to the location associated with the cue), the sensory and motor-related activities of many PFC neurons are gated or modulated according to which rule the animal is currently using [35,36]. In other words, it seems that PFC neurons do not simply reflect stimuli and actions, they also encode their behavioral context.

When rules involve familiar cues and responses (such as 'stop at red'), they are 'concrete' and can be represented as a set of specific S–R associations. With varied experience, rules can be abstracted beyond such specificity. Wallis *et al.* [38••] trained monkeys to alternate between applying either a 'match' or 'non-match' rule to pairs of

Figure 3
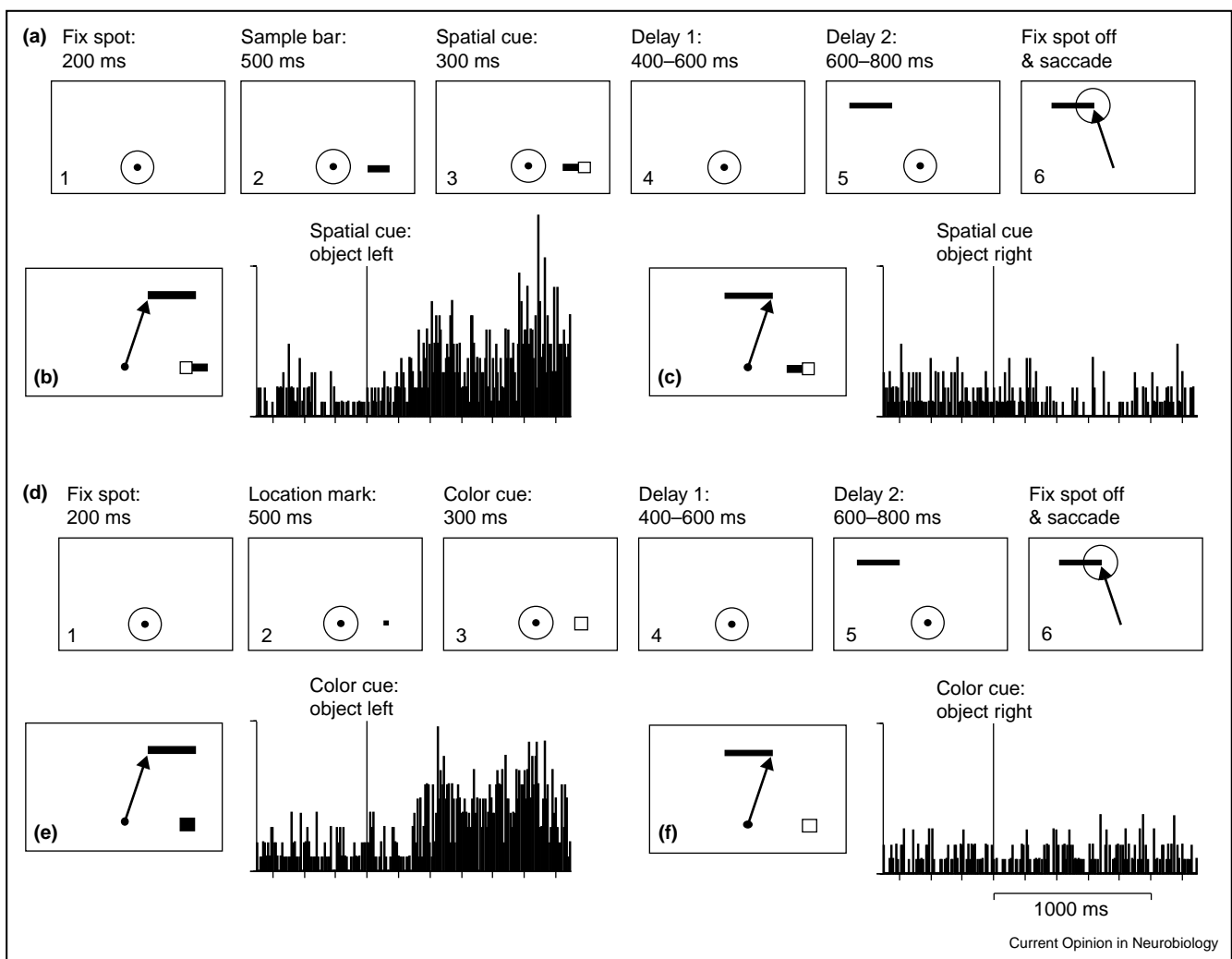
Abstract encoding of left versus right in the SEF. Data were obtained from a supplementary eye field neuron exhibiting categorical selectivity for object-centred location (left end versus right end of object). **(a)** Object-centred task with spatial cue. While the monkey maintained central fixation (panel 1), a sample bar appeared (panel 2) and a cue spot flashed on either its left or right end (panel 3). After a delay (panel 4), a target bar appeared at an unpredictable location in the upper visual field (panel 5). The monkey had to execute a saccadic eye movement to the end of the target bar corresponding to the cued end of the sample bar (panel 6). The circle in each panel indicates the direction of gaze at the corresponding stage of the trial. Strong neuronal activity occurred when the instruction was to select **(b)** the bar's left end as target but not when the instruction was to **(c)** select the bar's right end. **(d)** Object-centred task with color cue. Events were identical to those in (a) with the exception that the color of the cue (blue or yellow) instructed the monkey whether to select the left or right end of the target bar. The neuron still fired strongly on **(e)** bar-left trials but not on **(f)** bar-right trials. Thus, it expressed categorical selectivity for object-centred location, we regardless of the physical attributes of the cue conveying the instruction. Figure courtesy of Carl Olson, after [39].

pictures. After training with a wide range of pictures, monkeys could apply these rules successfully upon seeing a novel picture for the first time (and therefore had no prior S–R association). The modal group of task-related lateral and orbital PFC neurons reflected whichever rule was currently being used. This neural activity was independent of which specific cue signalled the rule and was not linked to the behavioral response.

Action-related categories

Categorical representations are also evident in the motor domain. Olson and co-workers [39,40••] explored correlates of representations of 'right' and 'left' by training monkeys to make an eye movement to either the right or left side of a small stimulus. Neurons in the supplementary eye fields (SEFs) — a frontal cortical area involved in eye movement control (Figure 1) — exhibited activity that varied depending on whether monkeys were cued to saccade to the left or right side, but did not vary with the exact position of the bar (and hence with the details of the eye movements), nor with the sensory conditions or rule used to cue the side (Figure 3).

Conclusions and future directions

Even in this short review, we present examples of abstract categorical representations across different levels of processing, brain regions, and species. The great usefulness of storing groupings of events and actions as meaningful classes that transcend physicality no doubt explains this ubiquity. However, this means that one future challenge for cognitive neuroscience will involve sorting out the roles that different areas play in representing, storing, and recalling different types and levels of abstraction.

For example, take the PFC, ITC and the visual categories described above. Where are the categories stored? The temporal lobe is a good bet: category-specific agnosias follow its damage. Is category information, then, merely copied to the PFC from the ITC when it is brought 'online' to guide behavior? Or does the PFC further abstract and elaborate the information as it retrieves it? One model of object recognition suggests the latter [41••]. Objects can belong to multiple categories that vary according to current context and purpose. So, it may be more efficient to save their explicit representation until the highest level of the perception–action cycle (the PFC), where the greatest amount of flexibility is needed. Indeed, a whole host of questions remain to be asked. How are superordinate and subordinate categories and concepts represented? How are multiple category memberships of a stimulus stored and gated? The work described here is a first stab at understanding how our brains acquire, store, and use the knowledge that makes us intelligent.

Acknowledgements

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. Wyttenbach RA, May ML, Hoy RR: **Categorical perception of sound frequency by crickets**. *Science* 1996, **273**:1542-1544.

2. Beale JM, Keil FC: **Categorical effects in the perception of faces**. *Cognition* 1995, **57**:217-239.

3. Roberts WA, Mazmanian DS: **Concept at different levels of abstraction by pigeons, monkeys and people**. *J Exp Psychol Anim Behav Proc* 1988, **14**:247-260.

4. Fabre-Thorpe M, Richard G, Thorpe SJ: **Rapid categorization of natural images by rhesus monkeys**. *Neuroreport* 1998, **9**:303-308.

5. Vogels R: **Categorization of complex visual images by rhesus monkeys. Part 1: behavioural study**. *Eur J Neurosci* 1999, **11**:1223-1238.

6. Gross CG: **Visual functions of inferotemporal cortex**. In *Handbook of Sensory Physiology*. Edited by Jung R. Berlin: Springer-Verlag; 1973.

7. Desimone R, Albright TD, Gross CG, Bruce C: **Stimulus-selective properties of inferior temporal neurons in the macaque**. *J Neurosci* 1984, **4**:2051-2062.

8. Logothetis NK, Sheinberg DL: **Visual object recognition**. *Annu Rev Neurosci* 1996, **19**:577-621.

9. Tanaka K: **Inferotemporal cortex and object vision**. *Annu Rev Neurosci* 1996, **19**:109-139.

10. Miller EK, Cohen JD: **An integrative theory of prefrontal function**. *Annu Rev Neurosci* 2001, **24**:167-202.

11. O Scalaidhe SP, Wilson FA, Goldman-Rakic PS: **Areal segregation of face-processing neurons in prefrontal cortex**. *Science* 1997, **278**:1135-1138.

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

13. Kreiman G, Koch C, Fried I: **Category-specific visual responses of single neurons in the human medial temporal lobe**. *Nat Neurosci* 2000, **3**:946-953.

14. Spiridon M, Kanwisher N: **How distributed is visual category**
•   **information in human occipito-temporal cortex? An fMRI study**. *Neuron* 2002, **12**:1157-1165.
Experiments in humans using brain imaging suggest that certain visual cortical regions are primarily involved in the analysis of a single class of stimulus. More specifically, there is a region in ventral visual cortex that exhibits stronger blood-oxygen-level-dependent (BOLD) signals to faces than to other stimuli; another region shows stronger signals for familiar places (plazas and intersections, etc.) than for other stimuli.

15. Freedman DJ, Riesenhuber M, Poggio T, Miller EK: **Categorical**
•• **representation of visual stimuli in the primate prefrontal cortex**. *Science* 2001, **291**:312-316.
Using morphing techniques, the authors systematically varied the shape of a set of computer-generated images and then taught monkeys to categorize the stimuli into two classes, 'cats' and 'dogs'. PFC neural activity was found to reflect category membership, rather than exact physical appearance. The category representations seemed to result from training; when the stimuli were assigned to new categories and a monkey was retrained, neural activity shifted to reflect the new category assignments.

16. Sigala N, Logothetis NK: **Visual categorization shapes feature**
•• **selectivity in the primate temporal cortex**. *Nature* 2002, **415**:318-320.
In this study, monkeys performed a categorization task with two sets of parametric stimuli. Each stimulus set contained four varying features, two

of which were relevant for the categorization task (the diagnostic features). The authors report enhanced neural representation of the diagnostic features relative to the non-diagnostic ones.

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

18. Tomita H, Ohbayashi M, Nakahara K, Hasegawa I, Miyashita Y: **Top-down signal from prefrontal cortex in executive control of memory retrieval**. *Nature* 1999, **401**:699-703.

19. Erickson CA, Jagadeesh B, Desimone R: **Clustering of perirhinal neurons with similar properties following visual experience in adult monkeys**. *Nat Neurosci* 2000, **3**:1143-1148.

20. Rainer G, Miller EK: **Effects of visual experience on the representation of objects in the prefrontal cortex**. *Neuron* 2000, **27**:179-189.

21. Baker CI, Behrmann M, Olson CR: **Impact of learning on
•• representation of parts and wholes in monkey inferotemporal cortex**. *Nat Neurosci* 2002, **5**:1210-1216.
Here, monkeys were trained to discriminate elongated stimuli with discrete tops and bottoms that were joined by a vertical stem. After training, the authors compared ITC neural activity in response to learned stimuli (those with familiar combinations of tops and bottoms) versus unlearned stimuli. ITC neural selectivity was enhanced for both learned individual parts and whole stimuli. This suggests the existence of neural mechanisms for piecing together parts into a holistic representation.

22. Dehaene S, Spelke E, Pinel P, Stanescu R, Tsivkin S: **Sources of mathematical thinking: behavioral and brain-imaging evidence**. *Science* 1999, **284**:970-974.

23. Brannon EM, Terrace HS: **Ordering of the numerosities 1 to 9 by monkeys**. *Science* 1998, **282**:746-749.

24. Hauser MD, Carey S, Hauser LB: **Spontaneous number representation in semi-free-ranging rhesus monkeys**. *Proc R Soc Lond Ser B* 2000, **267**:829-833.

25. Hauser MD, Dehaene S, Dehaene-Lambertz G, Patalano AL: **Spontaneous number discrimination of multi-format auditory stimuli in cotton-top tamarins (*Saguinus oedipus*)**. *Cognition* 2002, **86**:B23-B32.

26. Davis H, Pérusse R: **Numerical competence in animals: definitional issues, current evidence, and a new research agenda**. *Behav Brain Sci* 1988, **11**:561-615.

27. Xu F, Spelke ES: **Large number discrimination in 6-month-old infants**. *Cognition* 2000, **74**:B1-B11.

28. Wynn K, Bloom P, Chiang WC: **Enumeration of collective entities by 5-month-old infants**. *Cognition* 2002, **83**:55-62.

29. Sawamura H, Shima K, Tanji J: **Numerical representation for
•• action in the parietal cortex of the monkey**. *Nature* 2002, **415**:918-922.
Sawamura *et al*. taught monkeys to alternate between five arms movements of one type (push) and five of another type (turn). Neurons in the parietal cortex kept track of the number of each specific movement. That is, neurons were selectively activated by the conjunction of a specific movement and its number in the current sequence.

30. Nieder A, Freedman DJ, Miller EK: **Representation of the quantity
•• of visual items in the primate prefrontal cortex**. *Science* 2002, **297**:1708-1711.
In this paper, the authors taught monkeys to judge whether two successively presented visual displays contained the same number of items (between one and five). A wide range of stimuli were used during training

so that monkeys would learn to make number judgements independent of the specific stimulus being judged. This seemed to result in abstracted representations of numbers in the lateral PFC. Neurons in the PFC were maximally activated by a specific number of items, but were unaffected by changes in the exact appearance of the stimuli. Also, changes in low-level visual features, such as density, total area, and circumference, did not account for the neural tuning for numbers.

31. Parker AJ, Newsome WT: **Sense and the single neuron: probing the physiology of perception**. *Annu Rev Neurosci* 1998, **21**:227-277.

32. Dehaene S, Changeux JP: **Development of elementary numerical abilities: a neural model**. *J Cogn Neurosci* 1993, **5**:390-407.

33. Duncan J: **An adaptive coding model of neural function in
• prefrontal cortex**. *Nat Rev Neurosci* 2001, **2**:820-829.
A review of recent work in the PFC of humans and monkeys. Duncan proposes that PFC neurons adapt their properties to carry specific information that is relevant to current concerns, producing a dense, distributed representation of related inputs, actions, rewards and other information.

34. Fuster JM: **The prefrontal cortex — an update: time is of the essence**. *Neuron* 2001, **30**:319-333.

35. White IM, Wise SP: **Rule-dependent neuronal activity in the prefrontal cortex**. *Exp Brain Res* 1999, **126**:315-335.

36. Asaad WF, Rainer G, Miller EK: **Task-specific activity in the primate prefrontal cortex**. *J Neurophysiol* 2000, **84**:451-459.

37. Schoenbaum G, Setlow B: **Integrating orbitofrontal cortex into prefrontal theory: common processing themes across species and subdivisions**. *Learn Mem* 2001, **8**:134-147.

38. Wallis JD, Anderson KC, Miller EK: **Single neurons in the
•• prefrontal cortex encode abstract rules**. *Nature* 2001, **411**:953-956.
Here, the authors provide a demonstration that neurons in the PFC can encode abstract concepts of 'match' and 'non-match'. Monkeys were trained to switch between performing a match or non-match judgement until they could do so with novel pictures. Neurons in the PFC encoded match or non-match, yet were unaffected by which cues were used to signal the match or non-match judgement or by which pictures needed to be judged.

39. Olson CR, Gettner SN: **Macaque SEF neurons encode object-centered directions of eye movements regardless of the visual attributes of instructional cues**. *J Neurophysiol* 1999, **81**:2340-2346.

40. Tremblay L, Gettner SN, Olson CR: **Neurons with object-centered
•• spatial selectivity in macaque SEF: do they represent locations or rules?** *J Neurophysiol* 2002, **87**:333-350.
To determine whether SEF object-centred spatial selectivity represents location information or the rules used to select them, the authors recorded neural activity while monkeys selected targets by either an object-centred spatial rule or a color rule. On some trials, monkeys had to select a target on the same side as the cue (right or left), whereas on other trials, monkeys selected a target that was the same color as the cue (red or green). Object-centred neural selectivity for right versus left side of a stimulus was only weakly affected by the type of rule the monkey was following (object-centred-location or color) or by the color of the cue and target (red or green).

41. Riesenhuber M, Poggio T: **Neural mechanisms of object
•• recognition**. *Curr Opin Neurobiol* 2002, **12**:162-168.
Riesenhuber and Poggio review the empirical and modeling work on mechanisms of object recognition. They argue that high-level categories are not explicitly encoded until the final stages of visual processing.