Cell PRESS

# Sparse but not 'Grandmother-cell' coding in the medial temporal lobe

R. Quian Quiroga[1,2,3], G. Kreiman[4,5], C. Koch[2] and I. Fried[3,6]

[1] Department of Engineering, University of Leicester, LE1 7RH, Leicester, UK
[2] Computation and Neural Systems, California Institute of Technology, 91125 Pasadena, CA, USA
[3] Division of Neurosurgery, David Geffen School of Medicine and Semel Institute for Neuroscience and Human Behavior, University of California Los Angeles, 90095 Los Angeles, CA, USA
[4] Division of Neuroscience and Ophthalmology, Children's Hospital Boston, Harvard Medical School, 02115 Boston, MA, USA
[5] Center for Brain Science, Harvard University, 02138 Cambridge, MA, USA
[6] Functional Neurosurgery Unit, Tel-Aviv Medical Center and Sackler Faculty of Medicine, Tel-Aviv University, 69978 Tel-Aviv, Israel

**Although a large number of neuropsychological and imaging studies have demonstrated that the medial temporal lobe (MTL) plays an important role in human memory, there are few data regarding the activity of neurons involved in this process. The MTL receives massive inputs from visual cortical areas, and evidence over the last decade has consistently shown that MTL neurons respond selectively to complex visual stimuli. Here, we focus on how the activity patterns of these cells might reflect the transformation of visual percepts into long-term memories. Given the very sparse and abstract representation of visual information by these neurons, they could in principle be considered as 'grandmother cells'. However, we give several arguments that make such an extreme interpretation unlikely.**

## Introduction

The question of how visual information is transformed across different brain areas, and how it is finally represented, has occupied neuroscientists for decades. Consider, for example, the variety of processes triggered by the single sight of a face – from neurons in the retina responding to intensity and wavelength, to several areas in the cortex responding to the direction of gaze of the eyes, the identity of the face, its emotional expression and so on – culminating in a conscious percept. Evidence from electrophysiology and lesion studies in monkeys [1] supports the existence of a hierarchical organization along the 'ventral visual pathway', extending from primary visual cortex (V1) to the inferior temporal (IT) cortex. Neurons in V1 code for local orientations, wavelength and other basic visual features, whereas neurons in IT show selectivity for more complex shapes and even for faces [2–4].

Although it is widely accepted that visual objects are processed along the ventral pathway, the question of how this information is represented in the upper stages of the hierarchy – and made accessible to perceptual, cognitive and mnemonic processes – remains unclear [3–5]. At least two alternatives have been proposed. The 'distributed population coding' view [6–8] assumes that a given percept is represented by the activity of very large neuronal ensembles, in which each neuron is broadly tuned to particular metric features. Thus, for any one object, a large fraction of the population will fire. Alternatively, the 'sparse coding' view [5,9] holds that the same percept is represented by much smaller neuronal ensembles, the members of which respond in an explicit manner to specific features, objects or concepts. In such a sparse representation, the majority of neurons are silent for any one object. Taking this argument to the limit, neuroscientists considered the hypothesis that a single cell might respond to one and only one object or person, independently of, for example, its angle of gaze, location on the retina or facial expression. These hypothetical cells are popularly known as 'grandmother cells', as named by Jerry Lettvin [10,11]. Various versions of this view have been named 'pontificial cells' by Sherrington [12], 'gnostic cells' by Konorski [13] and 'cardinal cells' by Barlow [5]. Here, we discuss the evidence for such a representation. We argue that, although neurons in the human medial temporal lobe have recently been shown to display a very sparse and abstract representation [14], this representation is still a far cry from an extreme Grandmother-cell-like coding.

## Are there 'grandmother' cells in the medial temporal lobe?

The IT cortex conveys visual information to the medial temporal lobe (MTL) [15–17]. The MTL consists of hierarchical interconnected areas including the amygdala, hippocampus, entorhinal cortex, parahippocampal cortex and perirhinal cortex. The hippocampus receives direct input from the entorhinal cortex, which in turn receives its major inputs from the perirhinal cortex, the parahippocampal cortex and to a lesser degree directly from the IT cortex [17]. Clearly, the MTL should not be seen as a homogeneous structure with a single function. For example, the differential combination of various regions in the MTL to declarative memory processes is still a subject of intense investigation in studies using both neuropsychological observations and neuroimaging techniques in humans, as well as lesion and single cell recordings in animals (for a review of the hippocampal–rhinal system, see Ref. [18]). Moreover, consistent evidence

*Corresponding author:* Quiroga, R.Q. (rodri@vis.caltech.edu).

1364-6613/$ – see front matter  doi:10.1016/j.tics.2007.12.003 Available online 11 February 2008

has demonstrated that the amygdala is implicated in fear-relevant emotional processing (for reviews see Refs [19,20]).

Over the last few years, several studies have been conducted within a clinical setting where electrodes are implanted in patients with epilepsy to localize the seizure focus for possible curative resection [21]. The location of the electrodes is based exclusively on clinical criteria and the majority of them are usually placed in the MTL, given the prevalence of epileptic foci there.

Neurons in the human MTL were found to be selectively activated by conjunctions of gender and facial expression [21], by pictures of particular categories of objects, such as animals, faces and houses [22], as well as by the degree of novelty and familiarity of the images [21,23,24].

Figure 1 shows a single unit in the right anterior hippocampus that is selective to pictures of the American actor Steve Carell. The neuron fires with up to 20 Hz in response to pictures of Carell and is nearly silent during baseline – with an average activity of 0.04 Hz – or during presentations of other faces. About 40% of the responsive units recorded in the MTL had such a selective and invariant representation [14]. This combination of selectivity and invariance leads to an explicit representation [25], in which a single cell can indicate whether the picture of a particular person is being shown. In fact, a simple linear classifier applied to the spiking activity of a handful of simultaneously recorded units predicted which picture the patient was seeing in each trial far above chance [26].

Although these cells bear some similarities to 'grandmother cells', several arguments make this interpretation unlikely. First, it is implausible that there is one and only one cell responding to a person or concept because the probability of finding this cell, out of a few hundred million neurons in the MTL, would be very small. From the number of responsive units in a recording session, the number of stimuli presented and the total number of recorded neurons, a Bayesian probabilistic argument was used to estimate that out of approximately one billion MTL neurons, fewer than two million represent a given percept [27]. This is a far cry from a grandmother-cell-like representation. This number could be much lower because:

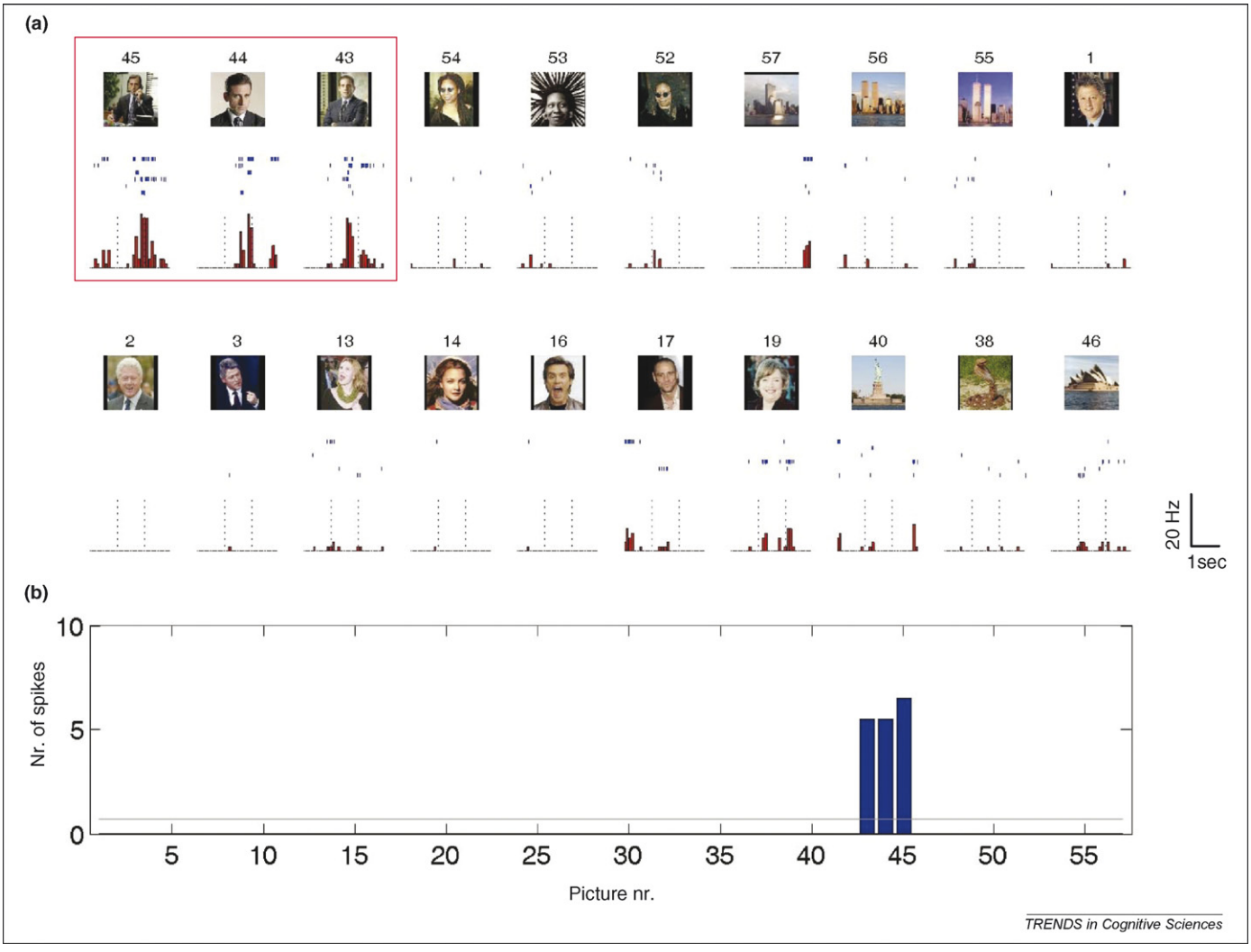


**Figure 1.** **(a)** Responses of a single unit in the right anterior hippocampus that is activated by pictures of the actor Steve Carell. For space reasons we display only 20 out of 57 pictures shown to the patient. There were no statistically significant responses to the other 37 pictures. For each picture, the corresponding raster plots and post-stimulus time histograms are given. The vertical dashed lines indicate image onset and offset. Pictures were shown for 1 s. **(b)** The median responses to all pictures. Note that using the median, instead of the mean, decreases the effect of outliers, such as the burst of spikes in the fourth trial of picture 19. The image numbers correspond to those displayed in (a). The horizontal line show the mean baseline activity plus five standard deviations, used for defining significant responses. The only significant responses are to the three pictures of Carell.

(i) images known to the subjects – which are more likely to elicit responses than unfamiliar stimuli – were used, and (ii) neurons with a higher degree of sparseness are very difficult to detect in our recording sessions lasting, on average, only ~30 min.

Second, although it was found that the cell of Figure 1 responds only to a particular person, Steve Carell, one should not discard the possibility that a response to other persons or objects would have been found if more pictures had been presented. Indeed, some units fired to more than one individual in an invariant manner. For example, one neuron in the hippocampus of another patient was activated by pictures of Jennifer Aniston and Lisa Kudrow, both actresses in the TV series 'Friends', whereas another neuron in the parahippocampal cortex fired to pictures of the Tower of Pisa and the Eiffel Tower, but not to other landmarks (see Figs S6 and S7 in Ref. [14]). Note that in these examples the stimuli eliciting responses were clearly related. Finally, theoretical considerations estimate that each cell most probably responds to between 50 and 150 distinct individuals or objects [27]. Thus, although these and other examples [28,29] suggest the existence of a very sparse representation of information (for a review see Ref. [9]), this should not be taken to the extreme of assuming that there is a one-to-one mapping between cells and objects or concepts.

It will be important to study whether an abstract representation such as the one hinted at in Figure 1 is also present in other species or systems, as the recently reported neurons in rats responding to the concept of a 'nest' [30]. In particular, one would expect that such sparse and abstract cells would be present in neocortical areas for object recognition, such as the IT cortex. However, it has been argued that the limited invariance found in the IT cortex serves to differentiate different views of a given person or object [31]. It should also be noted that it is difficult to detect sparse firing neurons. This is particularly the case when using single electrode recordings with movable probes that tend to miss sparsely firing cells – which might be quiet when the electrode passes in their vicinity unless their 'trigger' stimulus is shown – and are more likely to report neurons with high spontaneous rates and broadly tuned responses [9,32].

How can such specific and invariant cellular responses arise in the MTL? Computational work gives one compelling answer. We showed that an unsupervised learning principle that induces sparse representations [33] naturally leads to the development of units that respond only to a specific individual or object [34]. In this study, a two-layered network was exposed many times to 40 different face images of ten individuals – obtained from a standard machine-vision database of images [35] – and the synaptic weights were changed, without any external supervision, to maximize the sparseness of the representation. That is, each of the output units was constrained to fire to the smallest possible number of inputs and, consequently, the smallest number of units represented each image. After the network had stabilized, it was tested with a different set of 40 novel images of the same individuals, and it was found that most of the units responded uniquely to a single individual. This finding demonstrates how, in principle, a very sparse and invariant neuronal representation could emerge in the MTL using unsupervised learning.

## What are these MTL cells doing?

On the basis of findings in neurological patients with lesions or animal models with resections of the hippocampus and other parts of the temporal lobe, it appears that the MTL is not necessary for visual recognition. In particular, the hippocampal–rhinal system is involved in long-term declarative memory processes, a view supported by a large set of studies in both humans and animals (see for example [18,36].

In line with this evidence, it is very likely that the responses of human MTL cells described in the previous section link visual (or other forms of) perception to memory. For example, the cell responding to pictures of Steve Carell might not be involved in recognizing him (something that might require processing in the IT cortex and other, more posterior regions), but might be crucial for the storage of new long-term memories related to him and to the fact that the patient viewed his pictures in the clinic. This interpretation is in agreement with the latency of these responses, ~250–350 ms after stimulus onset [14]. These latencies are much longer than those found in neurons in the IT cortex of the macaque – the final purely visual processing area – at ~130 ms [37] and also long after rapid recognition occurs in the human brain, at ~150 ms [38]. Given the direct synaptic connections between the IT cortex and MTL in the monkey [17], response latencies of about 150 ms would have been expected for MTL neurons. This is clearly not the case for our human data and it is plausible that the additional delay of ~100–200 ms is due to high-level processing needed to transform visual percepts into memories to be stored.

It is common to remember abstract concepts and not details, unless attention is explicitly paid to them (see Box 1). It is also common to link abstract concepts, like persons and places, to form associations that can be retrieved in years to come. Consequently, if the MTL is involved in consolidation into long-term memory storage, it is plausible that these neurons might not differentiate all visual details. Clearly, neurons in other parts of the brain could represent details, as it is the case for view-sensitive neurons in the IT cortex [39]. However, abstraction might indeed be a crucial feature for declarative memory processes. As Borges says (see Box 1), without such abstraction it is not possible to generalize, to learn or even to think.

The existence of category cells [22], or cells responding to single individuals [14], is compatible with the view that they encode aspects of the *meaning* of any one stimulus that we might wish to remember. Indeed, a given object can be relevant as a category (e.g. a dog) or as an individual (my dog). A similar argument can be made for other types of category including faces.

These cells might also be involved in learning associations and relational encoding, in line with previous findings in monkeys [40–44] and humans [45]. In particular, we mentioned a cell firing to pictures of two actresses appearing on the same popular TV series, and another one firing to pictures of both the Tower of Pisa and the

**Box 1. Funes, the memorious**

In 1944, Jorge Luis Borges (1899–1986) published a compilation of short stories called 'Ficciones'. One of these, 'Funes, the memorious', recounts the fate of Irineo Funes, who after hitting his head when falling down from his horse, acquired the amazing talent of being able to remember absolutely everything. In this story, Borges gives us an extraordinary perspective of concepts like memory, invariance and learning. He writes:

'He was, let us not forget, almost incapable of general, platonic ideas. It was not only difficult for him to understand that the generic term *dog* embraced so many unlike specimens of differing sizes and different forms; he was disturbed by the fact that a dog at three-fourteen (seen in profile) should have the same name as the dog at three-fifteen (seen from the front). His own face in the mirror, his own hands, surprised him on every occasion.'

'Without effort, he had learned English, French, Portuguese, Latin. I suspect, nevertheless, that he was not very capable of thought. To think is to forget a difference, to generalize, to abstract. In the overly replete world of Funes there were nothing but details, almost contiguous details.'

As compellingly illustrated by Borges' short story, it might be detrimental to remember every single detail. This fits nicely with the finding of 'abstract' neurons in the human MTL, given the role of this area in memory consolidation [18]. There are several other advantages of such a sparse and invariant coding. First, it gives an explicit representation [25,26], in which the firing of these neurons carries a large amount of information because they represent the end product of many previous computations. This facilitates the readout by simple decoding algorithms [26] or by neurons in other areas. Second, sparse codes are energy efficient [46,47], because a relatively small number of neurons should be active to signal a percept. Third, sparse codes allow a large storage capacity in associative neural networks by avoiding interference between different memories [48–50]. Such a reduced interference due to sparseness has also been shown to be important for rapid learning in a network model of birdsong [51]. Finally, in contrast to extreme 'Grandmother cell' coding schemes, sparse representations have relatively large storage capacity, as well as tolerance to degradations of the network and noise in the inputs [49].

Eiffel Tower. This raises the possibility that MTL cells provide the substrate for such high-level associations.

## Conclusions

MTL neurons are situated at the juncture of transformation of percepts into constructs that can be consciously recollected. These cells respond to percepts rather than to the detailed information falling on the retina. Thus, their activity reflects the full transformation that visual information undergoes through the ventral pathway. A crucial aspect of this transformation is the complementary development of both selectivity and invariance. The evidence presented here, obtained from recordings of single-neuron activity in humans, suggests that a subset of MTL neurons possesses a striking invariant representation for consciously perceived objects, responding to abstract concepts rather than more basic metric details. This representation is sparse, in the sense that responsive neurons fire only to very few stimuli (and are mostly silent except for their preferred stimuli), but it is far from a Grandmother-cell representation. The fact that the MTL represents conscious abstract information in such a sparse and invariant way is consistent with its prominent role in the consolidation of long-term semantic memories. Furthermore, it is possible that these cells are involved in learning associations, a subject ripe for further investigation (see Box 2).

**Box 2. Open questions**

- Is the tuning of MTL cells stable or is there a continuous rearranging of the preferred stimuli of the cell?
- How many cells encode any one percept and, conversely, to how many different objects or individuals does a cell respond?
- Do MTL cells have the same type of response for other sensory modalities?
- How are MTL cells involved in learning associations?
- How are MTL cells involved in free recall or the spontaneous emergence of recollection in the human mind?
- Although the MTL receives direct inputs from the IT cortex, there is a very long delay between the neuronal responses in the IT cortex (at ~130 ms) and those in the MTL (at ~300 ms). What is happening between 130 ms and 300 ms?
- Are the sparse and invariant human MTL cells generalized versions of rodent place cells?

## Acknowledgements

We thank all patients for their participation, E. Behnke, T. Fields, E. Ho, E. Isham, A. Kraskov and C. Wilson for technical assistance and S. Waydo and F. Mormann for comments on this manuscript. This work was supported by grants from the NINDS, NIMH, NSF, DARPA, Epilepsy Foundation, EPSRC and the Life Science Interface Programme, the Office of Naval Research, the W. M. Keck Foundation Fund for Discovery in Basic Medical Research, the McGovern Institute, Children's Hospital Ophthalmology Foundation, the Gordon Moore Foundation, the Sloan Foundation and the Swartz Foundation for Computational Neuroscience.

## References

1 Mishkin, M. *et al.* (1983) Object vision and spatial vision: two cortical pathways. *Trends Neurosci.* 6, 414–417
2 Gross, C.G. *et al.* (1969) Visual receptive fields of neurons in inferotemporal cortex of the monkey. *Science* 166, 1303–1306
3 Tanaka, K. (1996) Inferotemporal cortex and object vision. *Annu. Rev. Neurosci.* 19, 109–139
4 Logothetis, N.K. and Sheinberg, D.L. (1996) Visual object recognition. *Annu. Rev. Neurosci.* 19, 577–621
5 Barlow, H.B. (1972) Single units and sensation: a neuron doctrine for perception. *Perception* 1, 371–394
6 Georgopoulos, A.P. *et al.* (1986) Neuronal population coding of movement direction. *Science* 233, 1416–1419
7 Abbott, L.F. (1994) Decoding neuronal firing and modeling neural networks. *Q. Rev. Biophys.* 27, 291–331
8 deCharms, R.C. and Zador, A. (2000) Neural representation and the cortical code. *Annu. Rev. Neurosci.* 23, 613–647
9 Olshausen, B.A. and Field, D.J. (2004) Sparse coding of sensory inputs. *Curr. Opin. Neurobiol.* 14, 481–487
10 Gross, C.G. (2002) Genealogy of the "Grandmother cell". *Neuroscientist* 8, 512–518
11 Rose, D. (1996) Some reflections on (or by?) grandmother cells. *Perception* 25, 881–886
12 Sherrington, C.S. (1940) *Man on His Nature,* Cambridge University Press
13 Konorski, J. (1967) *Integrative Activity of the Brain,* University of Chicago Press
14 Quian Quiroga, R. *et al.* (2005) Invariant visual representation by single-neurons in the human brain. *Nature* 435, 1102–1107
15 Lavenex, P. and Amaral, D.G. (2000) Hippocampal–neocortical interaction: a hierarchy of associativity. *Hippocampus* 10, 420–430
16 Suzuki, W.A. (1996) Neuroanatomy of the monkey entorhinal, perirhinal and parahippocampal cortices: Organization of cortical inputs and interconnections with amygdala and striatum. *Semin. Neurosci.* 8, 3–12
17 Saleem, K.S. and Tanaka, K. (1996) Divergent projections from the anterior inferotemporal area TE to the perirhinal and entorhinal cortices in the macaque monkey. *J. Neurosci.* 16, 4757–4775
18 Squire, L.R. *et al.* (2004) The medial temporal lobe. *Annu. Rev. Neurosci.* 27, 279–306
19 LeDoux, J. (2003) The emotional brain, fear, and the amygdala. *Cell. Mol. Neurobiol.* 23, 727–738

20 Zald, D.H. (2003) The human amygdala and the emotional evaluation of sensory stimuli. *Brain Res. Brain Res. Rev.* 41, 88–123
21 Fried, I. *et al.* (1997) Single neuron activity in human hippocampus and amygdala during recognition of faces and objects. *Neuron* 18, 753–765
22 Kreiman, G. *et al.* (2000) Category-specific visual responses of single neurons in the human medial temporal lobe. *Nat. Neurosci.* 3, 946–953
23 Rutishauser, U. *et al.* (2006) Single-trial learning of novel stimuli by individual neurons of the human hippocampus–amygdala complex. *Neuron* 49, 805–813
24 Viskontas, I.V. *et al.* (2006) Differences in mnemonic processing by neurons in the human hippocampus and parahippocampal regions. *J. Cogn. Neurosci.* 18, 1654–1662
25 Koch, C. (2004) *The Quest for Consciousness: A Neurobiological Approach,* Roberts and Company
26 Quian Quiroga, R. *et al.* (2007) Decoding visual inputs from multiple neurons in the human temporal lobe. *J. Neurophysiol.* 98, 1997–2007
27 Waydo, S. *et al.* (2006) Sparse representation in the human medial temporal lobe. *J. Neurosci.* 26, 10232–10234
28 Perez-Orive, J. *et al.* (2002) Oscillations and sparsening of odor representations in the mushroom body. *Science* 297, 359–365
29 Hahnloser, R.H. *et al.* (2002) An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature* 419, 65–70
30 Lin, L. *et al.* (2007) Neural encoding of the concept of nest in the mouse brain. *Proc. Natl Acad. Sci. USA* 104, 6066–6071
31 DiCarlo, J.J. and Cox, D.D. (2007) Untangling invariant object recognition. *Trends Cogn. Sci.* 11, 333–340
32 Shoham, S. *et al.* (2006) How silent is the brain: is there a "dark matter" problem in neuroscience? *J. Comp. Physiol. A Neuroethol. Sens. Neural. Behav. Physiol.* 192, 777–784
33 Olshausen, B.A. and Field, D.J. (1996) Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 381, 607–609
34 Waydo, S. and Koch, C. (2008) Unsupervised learning of individuals and categories from images. *Neural Comput.* 20, 1–14
35 Griffin, G. *et al.* (2006) *The Caltech-256*, Caltech Technical Report
36 Gazzaniga, M.S. *et al.* (2002) *Cognitive Neuroscience,* W. W. Norton and Company
37 Hung, C.P. *et al.* (2005) Fast read-out of object information in inferior temporal cortex. *Science* 310, 863–866
38 Thorpe, S.J. and Fabre-Thorpe, M. (2001) Neuroscience. Seeking categories in the brain. *Science* 291, 260–263
39 Logothetis, N.K. *et al.* (1995) Shape representation in the inferior temporal cortex of monkeys. *Curr. Biol.* 5, 552–563
40 Miyashita, Y. (1988) Neuronal correlate of visual associative long-term memory in the primate temporal cortex. *Nature* 335, 817–820
41 Miyashita, Y. and Chang, H.S. (1988) Neuronal correlate of pictorial short-term memory in the primate temporal cortex. *Nature* 331, 68–71
42 Miyashita, Y. *et al.* (1989) Activity of hippocampal formation neurons in the monkey related to a conditional spatial response task. *J. Neurophysiol.* 61, 669–678
43 Rolls, E.T. *et al.* (1989) Hippocampal neurons in the monkey with activity related to the place in which a stimulus is shown. *J. Neurosci.* 9, 1835–1845
44 Wirth, S. *et al.* (2003) Single neurons in the monkey hippocampus and learning of new associations. *Science* 300, 1578–1581
45 Cameron, K.A. *et al.* (2001) Human hippocampal neurons predict how well word pairs will be remembered. *Neuron* 30, 289–298
46 Attwell, D. and Laughlin, S.B. (2001) An energy budget for signaling in the grey matter of the brain. *J. Cereb. Blood Flow Metab.* 21, 1133–1145
47 Lennie, P. (2003) The costs of cortical computation. *Curr. Biol.* 13, 493–497
48 Willshaw, D.J. *et al.* (1969) Non-holographic associative memory. *Nature* 222, 960–962
49 Rolls, E.T. and Treves, A. (1990) The relative advantages of sparse versus distributed encoding for associative neuronal networks in the brain. *Network: Comp. Neural Syst.* 1, 407–421
50 Norman, K.A. and O'Reilly, R.C. (2003) Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. *Psychol. Rev.* 110, 611–646
51 Fiete, I.R. *et al.* (2004) Temporal sparseness of the premotor drive is important for rapid learning in a neural network model of birdsong. *J. Neurophysiol.* 92, 2274–2282