Selectivity and invariance for visual object perception

Matias J. Ison, Rodrigo Quian Quiroga

Department of Engineering, University of Leicester, LE1 7RH, Leicester, United Kingdom

TABLE OF CONTENTS

1. ABSTRACT

The sight of an object triggers a complex set of processes in the brain. Although it is already well established that object perception is performed by a hierarchical network, the so-called ventral visual pathway, we are only starting to understand how neurons along this pathway encode visual information at each processing stage. In this review, we discuss basic principles of neural coding for object perception and describe evidence showing that it mainly relies on two principles: selectivity and invariance.

2. INTRODUCTION

Without effort we can recognize the face of, for example, Robert de Niro playing a key role in the film 'Heat'. This seemingly simple process, however, involves an extraordinary complex cascade of neural activations in the brain transforming pixel information in the retina to a conscious percept in higher-level visual areas. To have a gist of how complex the process of object recognition is, think of creating a robot, as the ones developed by researchers working on machine vision, that can reliably identify faces. The algorithm for this robot should at least

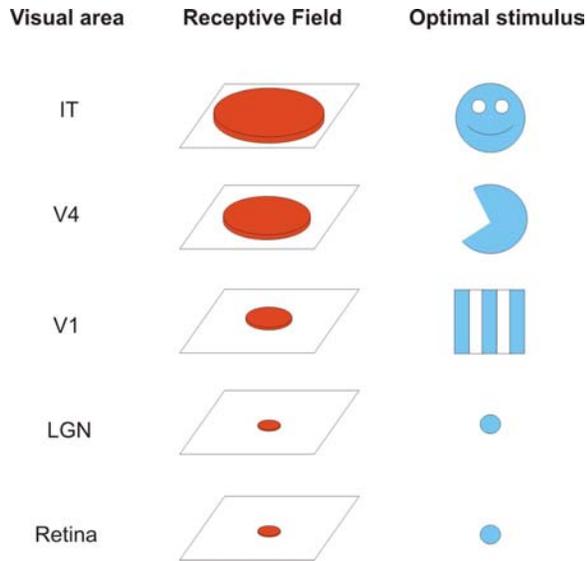| Visual area	Receptive Field	Optimal stimulus

Figure 1. Organization of the ventral visual pathway. Along the ventral stream, the receptive fields of cells increase (pictures not in scale) and the optimal stimulus becomes progressively complex.

have: i) some degree of invariance, to recognize Robert de Niro as the same person all over the film, ii) selectivity, to avoid confusing Robert de Niro with Al Pacino, and iii) related to invariance, the robot should have the ability to generalize, since very likely it have never seen before the particular views of de Niro in the film.

Converging evidence from monkey recordings shows that object recognition is performed by neurons in the ventral visual pathway, extending from the primary visual area (V1) to the inferotemporal cortex (IT). Although much (but far from all) is known about how different areas across the ventral visual pathway process different aspects of a given visual stimulus, we are only starting to understand how neurons in each area encode this information. As one of the central issues of Neuroscience, there exist a number of reviews in the literature covering several aspects of object perception: The type of representations of an object that are formed during the process of visual recognition is extensively discussed in (1). The neural correlates of perception along the ventral visual pathway is described in (2-4). In (5, 6), several models of object recognition are presented and discussed. The neural organization of visual object recognition in the inferotemporal cortex have been described in detail in (7).

Here we discuss the basic principles of neural encoding for object perception and review major findings in animals and humans at this respect. In our discussion, we include evidence from the ventral visual pathway, the inferotemporal cortex, and the medial temporal lobe, since perception is intrinsically related to memory formation. We stress the idea that in higher-level visual areas object perception is achieved using two features: selectivity and invariance.

3. THE VENTRAL VISUAL PATHWAY

It is already well established that the processing of visual information in the cerebral cortex has a hierarchical organization along two different pathways (8, 9). The dorsal or 'action' pathway, going from V1 through the parietal cortex to motor cortex, performs visuo-motor transformations, for example, to translate the sight of a cup of coffee into a reach towards it (10). It is also called the 'where' pathway, since it determines where the object is, as required for performing an action. The ventral pathway performs the actual recognition of the object (11). This organization should not be conceived as a strictly serial scheme or a strictly feedforward model involving unidirectional information flow, but as a hierarchical organization in the broad sense, where extensive feedback as well as parallel processing is also present (12). Moreover, there should be communication between both pathways, since we have to recognize the object we are reaching to (e.g. the cup of coffee and not the sugar).

In this section, we will focus on the ventral visual pathway, including a brief description of its hierarchical organization, landmark imaging studies on visual processing, the correlation between neural activity along this pathway and subjective perception, and a succinct overview on computational models of object recognition. We do not review the extensive work on other aspects of visual perception such as the neural mechanisms of perceptual learning (the modification of perception following sensory experience), perceptual grouping (the ability to group an object's elements segregating them from the background), or attention. We refer the reader to (13-15) for some recent experimental findings, (16-19), for works with remarkable theoretical contributions, and (20-23) for extensive reviews on these topics.

3.1. From retina to inferotemporal cortex

As sketched in Figure 1, the processing of visual information starts in the retina, which receives a projected inverted view of the object whose image enters through the lens of the eye. The receptive field (RF) of photoreceptors, cells in the deepest layer of the retina, is very small. Therefore, these cells are selective to pixel intensity at a given location and almost entirely insensitive to the spatial structure of a stimulus. Most retinal cells project to the lateral geniculate nucleus (LGN), where there is a topographical map of visual space in which spatial relations between neighboring neurons are maintained. In turn, most LGN cells project to the primary visual cortex (V1) at the very back of the occipital lobe. In this region, a large proportion of cortical cells are selective for local orientation. As shown by Hubel and Wiesel, a typical V1 cell, which they named 'simple cell', fires mainly when elongated stimuli (e.g. a bar) appears at a particular location with a particular orientation (6). This is in sharp contrast to cells in the retina and LGN, which are not selective for orientation. Hubel and Wiesel also reported the finding of 'complex cells', for which the precise position of the stimulus within the RF is less critical, i.e. they show some degree of invariance to the exact stimulus location (24). The increase in complexity and selectivity further continues
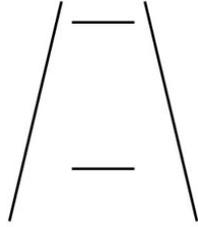
Figure 2. The Ponzo illusion. The two horizontal lines are of equal size, but the upper one appears to be longer.

along the ventral visual pathway: From V1 the information is conveyed to V2 and then to V4, where neurons exhibit larger RFs than the ones in V1. Moreover, many neurons in V4 are selective for contour feature orientation, and respond to angles and curves pointing in a particular direction (25). At the top of the visual hierarchy, neurons in the inferotemporal cortex have RFs that are often very large and can respond to stimuli as complex as faces (26, 27).

3.2. Imaging studies on visual processing

Besides the vast amount of new information that neurophysiology has provided about the ventral visual pathway, its understanding has also profited in the last two decades from major technological developments, such as functional magnetic resonance imaging (fMRI) and positron emission tomography (PET). In the context of our discussion, we will mention a number of landmark studies and bring up some recent specific findings.

A parallel between the human and non-human primate organization of the visual pathways using non-invasive studies was first demonstrated by pioneering findings using PET (28, 29), and soon after extended with a variety of fMRI studies (30-32). In particular, a seminal work by Malach and colleagues (32), found that the lateral occipital complex (LO) showed preferential activation for objects, reminiscent of what was known from neurophysiology in monkey area IT (33). Since then, a number of studies have successfully identified many domain-specific areas: i) a region of the extrastriate cortex called the fusiform face area (FFA), which responds strongly to faces, as found by Kanwisher and colleagues (34, 35); ii) the parahippocampal place area (PPA), which is highly selective to pictures of places and scenes (36, 37); iii) the extrastriate body area (EBA), which responds selectively to images of the human body (38, 39).

Two alternative models have challenged the view of domain specificity in perception. Gauthier and collaborators proposed that the selectivity described in the above-mentioned studies is strongly driven by experience (40-42). These authors suggested that the specialization of areas such as the FFA is determined by the level of categorization and expertise. In other words, the claim is that FFA is responsive to faces because of our every day's experience at recognizing them. Another view on the subject (43, 44) proposes that the ventral visual pathway is topographically organized to represent information about object forms in a continuous way. Interestingly, Haxby and

co-workers found that a given object category could be inferred from the information of the pattern of responses along other regions, excluding the region that is maximally activated (43) . These different hypothesis are still a ripe subject of debate (5, 45-48).

Arguably, one of the most productive recent contributions to this debate was found by Tsao and colleagues (49, 50). To overcome the resolution limitations of fMRI, they combined neurophysiology with fMRI studies in macaque monkeys using a very appealing paradigm: they targeted single-unit recordings to already identified fMRI face-selective regions. Interestingly, they found that the selectivity of individual neurons was strongly biased towards faces, since across the population of 405 neurons recorded, 310 turned out to be visually selective cells, out of which 97% were face selective.

3.3. Perception along the ventral stream

What we perceive is not necessarily what we actually see. This can be easily exemplified with visual illusions, as the one shown in Figure 2. For more complex and fascinating illusions, see (51). The two horizontal bars of Figure 2 are of equal size. However, due to the illusion of perspective by the oblique lines we perceive the upper one to be longer than the bottom one. Given that both bars create a similar input in the retina (they only differ in their location), where along the visual pathway this information is integrated with its context (i.e. the oblique lines) and the perception of the upper one to be longer is created?

Even though it is experimentally difficult to address this issue, paradigms have been designed to produce multistable phenomena, in which perception changes between alternative views (2, 3, 52). Two of the most popular of such paradigms are binocular rivalry and flash suppression. Binocular rivalry is produced by presenting two distinct images to the two eyes. Instead of a fusion of the two images, the brain alternates between the two views every few seconds. At each time there is only one dominant image that correlates with perception, the other one being perceptually suppressed. In flash suppression, the perceptual suppression of a constant monocular stimulus is obtained by flashing a different stimulus to the other eye. Although two stimuli are presented, subjects are only aware of the flashed stimulus. Then, the experimental design to study the neural correlates of conscious perception is very simple. We only need to present to one eye a picture that elicits the activation of a neuron of a given area and to the other eye a picture that does not. For example, with binocular rivalry, we could see if the neuron keeps its firing constant --since the visual stimulus does not change-- or if it modulates its firing according to the subjective perception.

In a series of studies aiming at understanding how neural activity correlates with perception along the ventral visual pathway, Logothetis and coworkers trained macaque monkeys to report their perceptual responses (2, 3). By recording from single neurons, they found that the correlation between neural activity and the animal's perception increased at successive stages of early visual

cortical areas (3). In V1, the firing of the majority of the neurons remained unaffected by the conscious percept reported by the animal, since the activity of only 20% of them followed perception (53). In area V4, their proportion increased to approximately 40% (2). The striking result was found at higher visual areas, like IT and the superior temporal sulcus (STS), where approximately 90% of the recorded cells followed reliably the perceptual state of the animal (53). In a very recent study Afraz and colleagues (54), applied electrical microstimulation to activate small clusters of IT neurons recorded from monkeys performing a categorization task.   Interestingly, they found that microstimulation of face-selective sites strongly affected the judgment of a monkey towards the face category. Microstimulation biased choices significantly in favor of face choices for 19 of 31 face selective sites. Its effect on perceptual decision was even much larger when microstimulation was applied to clusters of face-selective neurons, demonstrating a causal relationship between the activity of these neurons and face perception.

In humans, the possibility of performing invasive recordings are very rare and limited, in the sense that areas from which recordings are performed are exclusively determined by clinical criteria (55). For this reason, most studies of visual perception in humans have used non-invasive techniques, such as fMRI to identify brain regions that correlate with perception, as inferred indirectly from the blood oxygen level-dependent (BOLD) signal. In an early work, Lumer and colleagues (56) studied the neural activity associated with perceptual transitions during binocular rivalry, finding neural activity correlated with rivalrous transitions at multiple levels of the ventral visual pathway, but not in V1. Grill-Spector and collaborators (57), found a significant correlation between recognition performance and fMRI signal in an area preferably activated by objects when subjects were presented with pictures in condition of difficult recognition (stimuli were presented from 200 to 500 milliseconds). Tong and Kanwisher (58) studied fMRI responses of the fusiform face area (FFA) and the parahippocampal place area (PPA), two regions of human visual cortex highly selective to specific stimulus classes (faces and houses, respectively). In line with monkey studies in IT, they found that activity in the FFA and the PPA --at a high stage in the ventral visual pathway-- reflects the perceived, rather than the actual stimulus.

Undoubtedly, Imaging studies in humans have provided very valuable information about visual processing in different areas along the ventral visual pathway. The main limitation of these studies is that they measure the neuronal activity indirectly and therefore they cannot deal with the question of how neurons encode visual information.

3.4. Models of object recognition

As stated in the Introduction, the main problems that an algorithm for object recognition must face are selectivity (or specificity), invariance, and generalization. When building models to understand how object recognition can be achieved by the brain given the

architectural and timing constraints, additional problems appear (1, 5, 6).

The first model of visual recognition we discuss is the one by David Marr, originally introduced by Marr and Nishihara in 1978 (59, 60). Focusing on vision from an information-processing point of view, he proposed that objects are represented in a view-invariant manner, where an object is decomposed into parts and it is the important features of an object and their spatial relations what underlies recognition (23).

A well-known successor of Marr's model is the 'recognition by components' (RBC) theory developed by Biederman (61-63). Its central idea is that any object can be segmented into an arrangement of a small set of simple geometric components, called 'geons'. This theory is also view-invariant but, in contraposition to Marr's theory, it received some empirical support. However, the RBC theory has a number of limitations. One of the main criticisms is based on the fact that humans can easily recognize objects with rather similar structural descriptions, such as faces, which present similar or identical geon-based representations (64).

Despite the economical advantages of a visual system using a small set of components, psychophysical (65, 66) and physiological (1, 67) evidence support a viewpoint-dependent representation. In such a scheme, parts of objects, or entire images of them are encoded in a view-specific manner and object recognition processes are based on the similarity between encoded and perceived images (65).

Among the numerous models that have been proposed to achieve the goals of invariance and specificity with biological plausibility, a few of them mimic the hierarchical organization of the visual pathway described in Section 3.1.

One of these pioneering hierarchical models is the one developed by Fukushima in 1980, a multilayer neural network which was dubbed "neocognitron" (68). This model successfully accounted for translational invariance, since its responses are not affected by the object's position (68). However, a model in agreement with physiological and neural data remained elusive for many years. An important step in this direction was given by the HMAX model of Riesenhuber and Poggio (69), a simple hierarchical model that mimics the architecture of the ventral visual pathway. HMAX, being related to previous approaches (70, 71), is based on a feedforward architecture and is capable of dealing with both invariance and selectivity by means of (to some extent) biologically plausible mechanisms. Numerous extensions and alternatives have been presented since then (6, 72-74). In particular, Deco and Rolls' approach (75) extended a previous model of object recognition (71) incorporating top-down backprojections to account for attentional effects.

Whereas it has recently been shown that feedforward architecture can successfully account for rapid

categorization (76), many other features related to object recognition, including our every day's visual experience with natural images, are known to be highly influenced by feedback and lateral interactions. An interesting proposition (16) trying to reconcile the impressive speed of certain visual tasks (as determining whether a picture contains a face or not), with slower detailed perception (as understanding ambiguous images) states that processing along the feedforward hierarchy is automatic and implicit, whereas explicit perception (conscious perception), begins at the high-level cortex, and gradually returns downward as needed. For a recent discussion on the subject see (77).

4. THE MEDIAL TEMPORAL LOBE

The inferotemporal cortex, the last exclusively visual area in the ventral pathway, includes massive direct projections to the Medial Temporal Lobe (MTL) (7, 78, 79). The MTL comprises multiple interconnected areas including the hippocampus, amygdala, parahippocampal gyrus, entorhinal cortex, and perirhinal cortex.

The entorhinal cortex provides the main input to the hippocampus, which in turn receives projections from the perirhinal cortex, the parahippocampal cortex, as well as other direct inputs from orbital frontal cortex, cingulate cortex, insular cortex, and superior temporal gyrus (80).

4.1. Medial Temporal Lobe and memory

Based mainly on evidence from neurological patients and studies with animal models with selective lesions of the MTL, it is widely accepted that the MTL is crucial for declarative memory, the conscious or explicit memory for facts and events (80-83).

Our current knowledge about the role of MTL in memory has largely benefited from the study of patient H.M., whose severe epilepsy was treated by an experimental bilateral MTL resection in 1953 (84-86); for a current overview, see (87). Although after surgery H.M. had normal intelligence --as assessed by neuropsychogical tests-- and short-term memory (i.e. memory lasting from seconds to minutes), he developed severe anterograde amnesia; i.e. he could not later remember new information. Interestingly, this amnesia was limited to declarative memory, since he could still acquire non-declarative memories, such as learning a motor skill. Due to the fact that H.M could still recognize people and objects, it has been postulated that the MTL is not involved in recognition. Since both his short-term and long-term memory systems seemed not compromised by the surgery, what appeared to be disrupted was his ability to transform new short-term memories into long-term memories.

Studies from other patients support this view. One of these cases is patient R.B., who developed anterograde amnesia after an ischemic episode involving both hippocampi, as assessed by post-mortem histological analysis (88). Converging results were also obtained with patient K.C., who suffered a severe head injury -- compromising the MTL-- following a motorcycle accident (89). Subsequently, K.C. developed a severe anterograde and retrograde amnesia involving episodic information.

However, he had an average IQ and normal short-term memory. His semantic memory for knowledge acquired prior to the accident as well as his performance in perceptual priming, a form of non-declarative memory, were intact. However, he had no episodic memory of having seen the cues that had been used in previous sessions. These clinical data are corroborated by animal studies. Surgical lesions in different areas of MTL in monkeys affect memory performance, as assessed by delay non-matching to sample tasks (80, 90, 91). Imaging studies in healthy human subjects also support this view (92-95).

The above-mentioned interpretation follows the 'standard model of consolidation', which states that MTL is necessary for the consolidation of new semantic and episodic memories in neocortex. This view has been challenged by more recent evidence, suggesting that the hippocampus is always necessary for the retrieval of contextual aspects of episodic memory (96). A re-examination of H.M. episodic memory seems to support this 'memory trace' view (97). Note that both models are in line with the finding of abstract MTL neurons that follow conscious perception (see section 8). This could be interpreted as the consolidation of long-term semantic and episodic memories, according to the standard model of consolidation, or also for the retrieval of episodic memories, according to the memory trace theory.

5. FROM FULLY DISTRIBUTED TO SPARSE REPRESENTATIONS

Among the possible strategies that the brain could use to encode visual information, two opposite views have been discussed in the literature: a distributed and a sparse representation.

In a distributed representation, the information about a particular stimulus is represented by the orchestrated activity of a large ensemble of neurons. One of the advantages of such a scheme is its robustness against neuronal variability (98). Another advantage is its large memory capacity and ability to represent a number of different stimuli attributes simultaneously.

In a sparse representation, only a small number of neurons are simultaneously activated by a given stimulus. The theoretical limit of a sparse coding, where the representation of an object is reduced to a single neuron, is often referred to as a grandmother cell (see Figure 3 for a sketch of possible neural coding strategies). Sparse codes present many advantages for an organism. In particular, they are convenient in terms of energy consumption because only a relatively small number of neurons is needed to represent a concept, thus saving energy that is metabolically expensive (99, 100). Moreover, sparse codes represent information in an explicit way, which makes it easy to read-out information for decoding algorithms or neurons in other areas (101) .

In a distributed representation, the number of stimuli that can be represented increases exponentially with the number of components (64), i.e. a set of 2 binary
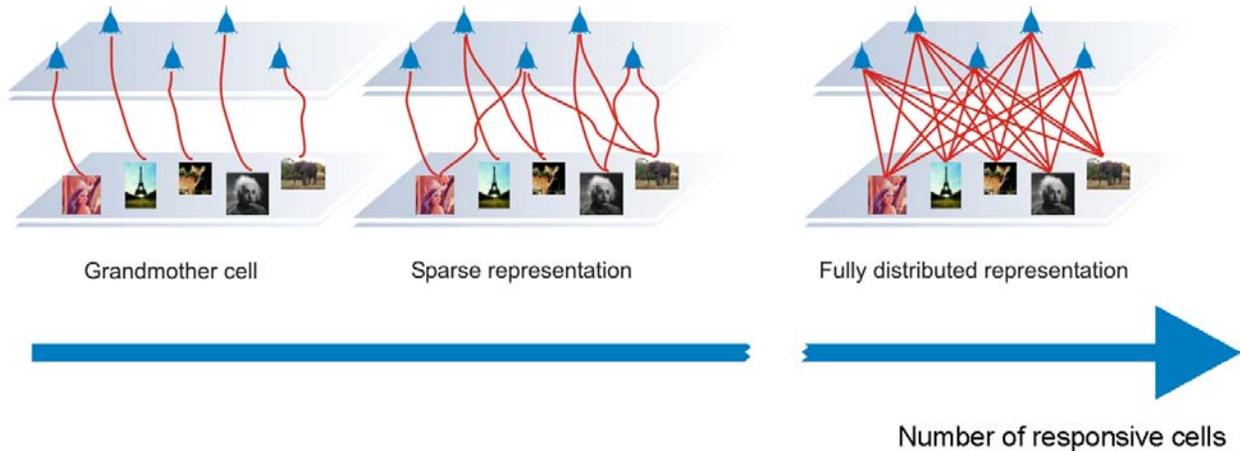
Figure 3. Sketch of neural coding strategies in order of increasing sparseness (see text). Left. A Grandmother Cell type of coding, where each cell codifies one stimulus. Middle. sparse representation, where only a few neurons are active in response to a stimulus. Right. Fully distributed representation. Any stimulus elicits activity of a large population of cells.

neurons can be in any of 4 states (on-on, on-off, off-on, off-off), and therefore can code for any of 4 stimuli seen; 3 neurons can code for 8 stimuli, and so forth. In contrast, for a grandmother cell coding –where each neuron encodes one and only one stimulus- the representational capacity increases linearly with the number of components.

As we have described in Section 3, cells of the visual pathway are hierarchically organized, with cells in higher levels having a preferred firing towards more complex stimuli. A possible interpretation of selectivity in terms of neural coding implies that each area of the hierarchy elaborates on the representation based on the processing in earlier areas, representing the stimulus in a more specific way, from local representations that are closely linked to the retinal image to abstract representations that are extensively connected to perception (46) .

In order to quantify the type of representation used to encode information, some measures of "selectivity" or "sparseness" have been proposed (102-104). These measures usually assess "lifetime sparseness"; i.e. the number of stimuli to which a neuron responds to. Lifetime sparseness is related to 'population sparseness', which is the number of neurons that respond to one stimulus. In particular, one expects that if neurons in a given area respond to very few stimuli, then each of these stimuli will elicit activations in a relatively low number of neurons -- although technically this does not need to be the case, since very few neurons can respond to a large number of stimuli and vice versa--. Ideally, one would like to measure selectivity values at the different stages of the ventral visual pathway and see whether there is an increase in selectivity as one goes from V1 to higher visual areas. However, it is difficult to compare selectivity values in different areas, since they depend on the stimuli set (102, 104) and different types of stimuli are usually used in different areas (e.g. oriented bars in V1 and faces in IT). On top of this, a recent work in which cells were recorded from areas V1, V2, and V4 of awake macaques that were presented with simple (bars and sinusoids)

and relatively complex stimuli (angles, intersections, arcs, and non-Cartesian gratings), revealed that shape representation along recorded areas do not parallel directly the stepwise organization of the anatomical hierarchy (105).

6. GRANDMOTHER CELLS

At the extreme of a sparse representation, a neuron would respond to one and only one person or concept, such as a grandmother, following the popular term "Grandmother Cell", coined by Jerry Lettvin (106-108).

As early as in 1890, the psychologist and philosopher William James in his Principles of Psychology first introduced the related idea of "pontifical cells". James said: "There is, however among the cells one central or pontifical one to which our consciousness is attached" (109) (p.179). About 50 years later, Sir Charles Sherrington discussed James´s idea of pontifical cells and favored the view of the mind as a "million-fold democracy whose each units is a cell" (110).

In 1967, the neuropsychologist Jerzy Konorski went a step further and predicted the existence of single neurons sensitive to complex stimuli such as faces, hands, emotional expressions, etc. (111). He named them Gnostic units, after "gnosis", the Greek term for recognition. He not only suggested their existence, but also predicted that certain specific brain areas were devoted to the representation of different categories, the impairment of which would cause category-specific agnosias (like prosopagnosia, an impairment in the ability to recognize familiar faces). This discussion had a significant boost after 1967, when Jerome Lettvin introduced -- and basically disregarded-- the idea of 'Grandmother Cells' in a course at MIT, with the fictitious and by now very famous story of the Neurosurgeon Akakhi Akakhievitch (67). Akakhievitch, seeking for long awaited recognition, ablated all the mother cells in Portnoy, one of his patients, and as a result, Portnoy lost the concept of his mother. Akakhievitch then decided to go for the grandmother cells…

Soon after Konorski and Lettvin, Horace Barlow proposed in a seminal work five dogmas to account for the relationship between single-unit activity and sensation (112). Barlow's central proposition was that "our perceptions are caused by the activity of a rather small number of neurons selected from a very large population of predominantly silent cells. The activity of each single cell is thus an important perceptual event and it is thought to be related quite simply to our subjective experience. The subtlety and sensitivity of perception results from the mechanisms determining when a single cell becomes active, rather than from complex combinatorial rules of usage of nerve cells" (112). Gathering information from his own experiments on the frog's retina (where he found retinal neurons that act as "bug detectors" and respond invariantly to the same external patterns despite changes of average luminance), the hierarchical organization of selectivity from Hubel and Wiesel experiments on the visual cortex in cats (24), and the interpretation of multisensory experiments linking sensation and patterns of neurophysiological activity in terms of the brain as an information processing system (112), he built a framework to link single neuron activity and perception. He revisited Sherrington´s idea of pontifical cells but favored the view of a relatively low number of "cardinal cells" encoding each percept. "Among the many cardinals only a few speak at once; each makes a complicated statement, but not, of course, as complicated as that of the pontiff if he were to express the whole of perception in one utterance" (112).

One of the main criticisms to grandmother cell coding is that it leads to the so-called "combinatorial explosion" in the number of neurons necessary to represent a given percept (113, 114). As mentioned in Section 5, the number of stimuli that can be encoded using a grandmother cell coding increases linearly with the number of cells. Therefore, a huge amount of neurons would be required to represent the total number of patterns that can be seen. Moreover, this scheme is not robust against cell death. Supposing our grandmother is represented by only one neuron, the death of this particular cell would imply to lose this concept, as achieved by Akakhievitch in Lettvin's parable.

7. INVARIANCE

As discussed in the previous section, probably the strongest argument against grandmother cells, or extremely sparse representations, comes from the fact that, if only one neuron were to represent a given object, then there are just not enough neurons in the brain to represent all possible things we can store (113). For example, one would need one neuron for our own grandmother in frontal view, another one of her in profile, another one for her dressing her favorite dress, etc. However, the selectivity of a cell can be related to changes in its invariance against stimuli, e.g. some cells in the frog's retina respond invariantly to the same external patterns despite changes of average luminance (112). In the primary visual cortex, neurons display a variety of responses to motion, color, orientation, etc. irrespective of all other stimulus features. As we will see in the following, the key to avoid combinatorial explosion with very sparse representations is given by invariance and an abstract representation.

7.1. Invariance in IT

The first visually evoked responses in monkey IT were obtained by Gross and collaborators (26, 115). These seminal findings were later confirmed and expanded by many studies (27, 33, 116-119), setting IT as the last exclusively visual area in the ventral pathway. More than 85% of the neurons in this area were found to be responsive to different simple or complex visual stimuli (33). Interestingly, the anterior part of IT (anterior IT or TE) is the earliest area in which selectivity for complex shapes, such as faces, can extensively be found (26, 33, 116). Although selective responses to complex features have been traced to V4 and the posterior part of IT (120), which are earlier stages of the visual pathway, most responsive cells in these areas fire to simple features and the size of their receptive fields are much smaller than the ones of neurons in anterior IT. This last property turns out to be crucial for achieving invariance. IT neurons show various degrees of invariance to image transformations. Indeed, more than half of the IT neurons exhibit size and position invariance (67, 121). The response of the rest of the neurons in IT indicate some degree of size specificity, suggesting that at least certain images of objects might be stored in a size-specific manner (122).

Even though cells in IT respond to complex features of a stimulus, the specificity is not enough as to represent an object through the activity of a single cell (1, 7). By recording simultaneously from different units, Tanaka and collaborators found that a large part of IT is composed of columnar modules in which cells within the same column respond to similar features (7). Following this result, Tanaka proposed that a given feature of an object is represented by the activity of many cells within a single columnar module (7). He argued that such a representation would help to solve two of the main problems in visual recognition: invariance among subtle changes in input images and selectivity. However, it is not clear how the information from different columns is combined to represent a whole image.

Other investigators have studied the tuning of individual IT neurons by using parametrical variations of the stimuli (123-125). For instance, monkeys were presented with faces varying in its degree of characterization, from a neutral identity-ambiguous face to a caricature where some distinguishing features have been exaggerated (125, 126). In one of such experiments, Leopold and colleagues found that most individual face-responsive neurons in monkey IT exhibited a roughly linear increase in their activity as a function of face identity, and are therefore tuned around an 'average', identity-ambiguous face[1]. In (125), it is suggested that a norm-based mechanism in which an incoming facial expression would be compared to the stored average face could be advantageous for face recognition in terms of time saving and storage capacity. Further evidence of norm-based coding strategies have also been reported following different paradigms (124, 128).
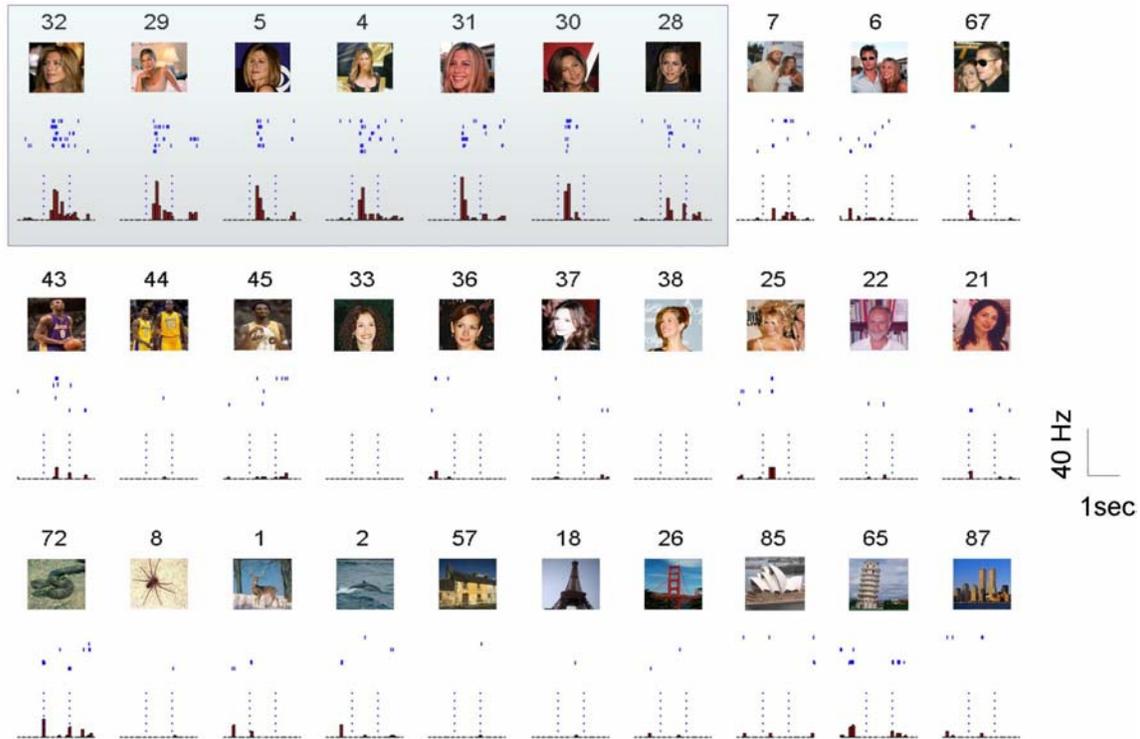
Figure 4. A single unit in the right posterior hippocampus that responded selectively to different pictures of the actress Jennifer Aniston, but not to other pictures (including pictures of the same actress with Brad Pitt). For each picture, the raster plots and the post-stimulus time histograms are depicted. The vertical dotted lines (1 sec apart) mark the onset and offset of picture presentations. Adapted with permission from (136).

Current evidence indicates that individual IT neurons do not explicitly encode information about the stimulus identity and that this information is encoded at the population level. Indeed, Hung and collaborators found that very small IT neuronal populations (around 100 randomly selected cells) convey scale- and position-invariant object category and identity information (129). Notably, the same study found no difference in the time course of the information conveyed for "categorization" and "identification" tasks, arguing against previous findings claiming that single neurons conveyed two different scales of facial information in the latencies of their spike trains (130). Similar results were recently found by Kiani and colleagues (131), by following a different experimental paradigm in which monkeys were presented with a large (>1000) set of natural object images. Taken together, these studies suggest that category and identity information in IT is largely distributed over a rather small cell population, corresponding to a sparse coding strategy.

In sum, extensive studies have shown that many neurons in IT are highly selective for complex stimuli. Among these neurons, some of them show scale invariance, i.e. their firing is independent from the size of the object at the retina, and some others show location invariance, i.e. their responses do not depend on the object position in the visual field. However, this type of invariance in IT is very basic. In particular, the tolerance of individual neurons for rotations is very limited (1). A word of caution is in order at this point. A limited invariance in IT is not necessarily a drawback in terms of coding strategies. Indeed, it is also possible that a population code with highly selective and non-invariant neurons, as well as non-selective neurons showing invariance to image transformation, encodes different aspect of a given percept, as recently proposed (132, 133).

7.2. Invariant representation in the MTL

Under exceptional circumstances, epilepsy patients who do not improve with medication are implanted with intracranial electrodes to identify the seizure focus for potential resective surgery. In many of these cases, recordings are done in the medial temporal lobe, including the hippocampus, amygdala, and entorhinal and parahippocampal cortices. These recordings offer the unique opportunity to study single cell activity in the human brain. These studies led to the discovery of "category-specific" neurons in the MTL that responded selectively to visual stimuli from different categories, such as faces, landmarks, and animals (134, 135). More recently, a remarkable type of neuron that responded selectively to the identity of given individuals or objects was found (136). For example, Figure 4 shows one of such neurons that responded strongly in an invariant manner to different pictures of the actress Jennifer Aniston, but not to other persons, objects, etc. This particular neuron dramatically increased its firing rate from 0.03 spikes/s to 16 spikes/s during the presentation of pictures of Jennifer Aniston, being almost silent to the other 80 pictures shown.

These results suggest that the MTL is encoding an abstract representation of the identity of the individual shown, rather than a complex set of features of the particular picture.

Are these true Grandmother cells? In other words, is this the only cell in the brain firing to Jennifer Aniston and absolutely nothing else? Very simple arguments make this interpretation extremely unlikely. In particular, with the current recording and data processing technology, using advance mathematical methods to identify the activity of different neurons from each recording electrode --i.e. spike sorting algorithms-- (137), it is possible to record from up to 100 neurons simultaneously. Given that MTL has of the order of 10^9 neurons (138), the chance of hitting the one and only cell responding to Jennifer Aniston is negligible. Therefore, if one of these cells was found, there should definitely be more[2]. Furthermore, the fact that the neuron responded selectively to Jennifer Aniston within the dataset of about 100 pictures shown, does not imply that it will not respond to anything else if more pictures had been shown. In fact, this same cell also responded to a picture of Lisa Kudrow in a follow-up experiment, as we will describe in the next section.

Given the total number of neurons in MTL, the subset of neurons from which we record from, the number of pictures shown, and the probability of getting responses, using Bayesian inference it was estimated that about 2 to 5 million neurons are activated by a typical stimulus, and that each neuron fires in response to 50-150 distinct representations (138). This is however an upper bound, because: i) for these experiments pictures of things very familiar to the patient were used, thus increasing the probability of getting responses and ii) it is possible that more selective cells remained silent during the experiments since none of the very few pictures to which they respond were shown. Still, an activation of the order of million of cells seems far away from a grandmother cell representation but given that there are about 10^9 neurons in MTL, this supports the idea of ultra-sparse coding, along with the view of 'Cardinal cells' by Barlow (see Section 6). Indeed, abstract neurons that are essentially silent during the presentation of most pictures and only respond to a small subset of images (like pictures of Jennifer Aniston in the example) are certainly far away from a distributed representation and suggest a sparse and invariant coding. It is interesting to notice that silent neurons with baseline firing rates smaller than 1 spikes/s are in agreement with energy consumption estimations for pyramidal neurons in human neocortex (100). Moreover, increasing evidence from studies using diverse recording methods seem to support a scenario where many neurons are silent most of the time (139).

We remark that given the well-established evidence described in section 4.1., it is very unlikely that these cells are involved in perception. Based on these findings, it is very plausible that these MTL neurons play a key role in the transformation of visual percepts into long-term and abstract memories (136). In other words, the neuron responding to Jennifer Aniston is not necessary for recognizing her, but it may be necessary later to remember seeing her again. Alternatively, following the memory trace view (see section 4.1) these neurons may be critical for the recollection of memory traces related to a concept (96). Either of these interpretations is in agreement with the idea of invariance and abstract representation, since we tend to remember abstract concepts rather than details.

In spite of the fact that these cells are not directly involved in perception (but see (140)), they are giving us a hint about the final representation of visual information. No such an extreme sparse and invariant coding has been reported in monkey IT, but unless this representation only arises in hippocampus, it is in principle possible that such a scheme is present in IT as well. Alternatively, as mentioned above it has been proposed that the limited invariance found in IT is important to encode different aspects of a given stimulus (133). We already mentioned that MTL cells are likely involved in storing percepts into long-term memory (something like the RAM memory in a PC). Then, this representation can be dynamic (with neurons changing its selectivity on a day-by-day basis according to the relevant items that have to be stored) and highly redundant. In other words, in MTL we may have many more cells coding for concepts we want to store into memory in comparison to IT. Therefore, it would be more likely to find such neurons there. Future experiments carefully testing for invariance in monkey IT and MTL may shed more light to this issue.

8.   ASSOCIATIONS   AND   ABSTRACT REPRESENTATIONS

The existence of invariant responses in the MTL presents clear benefits, given the role of this area in the formation of new memories. In fact, one does not usually remember unnecessary details about a known person or object.

Invariance can directly be linked to association and categorization, which are important concepts in the fields of object recognition and memory. Indeed, humans can easily recognize objects regardless of large variations in viewpoint, illumination conditions, etc. As seen in Section 2, to accomplish this, a vision system needs to capture the invariant properties of objects. For a memory system, associations are essential. For example, we associate the different pictures of Jennifer Aniston in Figure 4 to the same individual. Similarly, we may associate different individuals or objects to the same concept, e.g. mother and father are associated to the concept of parents. Largely, learning and memory can be though of as the creation of associations and categories. For example, one can link a group of people as a single category (e.g. the people we met at the party). Of course, when having the chance to know them better one may individualize them and each subject will be a category that could also be linked to other concepts, for example, to form episodic memories.

To illustrate this discussion, in Figure 5 we show the responses of a cell in the right posterior hippocampus of
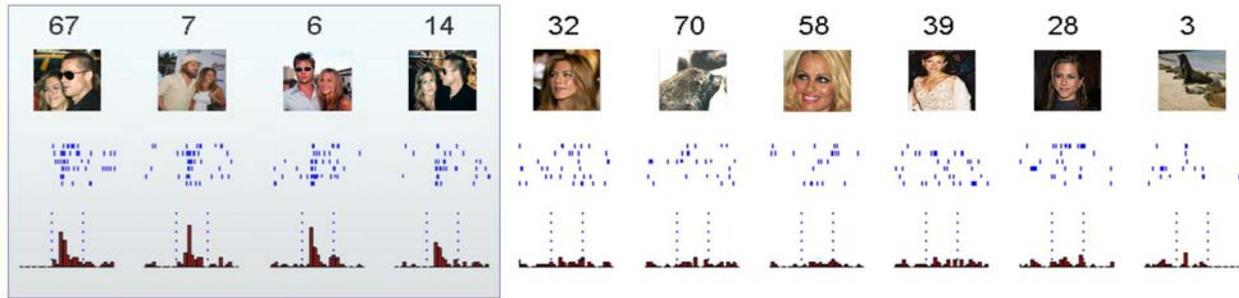
Figure 5. A unit in the right posterior hippocampus that responded selectively to different pictures of the Pisa tower and a picture of the Eiffel tower. Adapted with permission from (136).
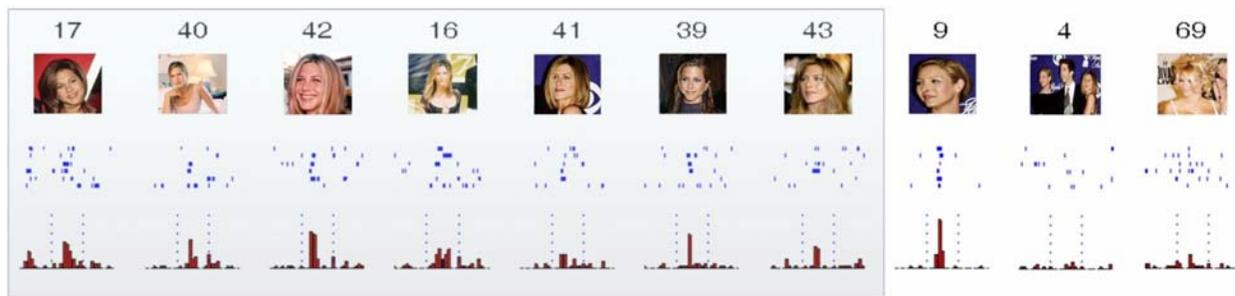

Figure 6. A single unit in the right posterior hippocampus that responded selectively to different pictures of Jennifer Aniston and also to a picture of the actress Lisa Kudrow, a co-star of the TV series "Friends". Adapted with permission from (136).

a patient. This particular cell responded selectively to different pictures of the Pisa tower and also very strongly to a picture of the Eiffel tower[3].

It is evident that such an invariant pattern of activation cannot be explained in terms of common visual features of the images shown. What seems to be important for these cells is the meaning of the stimuli, which is very likely linked to a single concept for the patient (e.g. landmarks of Europe).

Figure 6 shows another example of associated concepts. This cell responded to both pictures of Jennifer Aniston and Lisa Kudrow, her co-star on the television show Friends. Therefore, it may be showing the association of these two linked persons into a single concept, which fits well with the fact that the same neuron on a previous day (Figure 4) did not respond to Jennifer Aniston with Brad Pitt since it may have evoked a different concept of association, even if involving the same person. Interestingly, single neurons recorded from the temporal cortex in monkeys, have been shown to acquire selectivity for visual patterns through associative learning (141, 142).

In terms of coding strategies, the encoding of associated concepts in similar (but not identical) neural ensembles would allow a big representational capacity with only a few active neurons. As mentioned in Section 5, energy consumption imposes big constraints on the number of active neurons engaged in a certain task (100). A sparse and invariant abstract representation of visual stimuli would contain the necessary ingredients for achieving an optimal coding strategy with enough representational capacity and the economic benefit of having a few active neurons.

9. PERSPECTIVE

The understanding of how our brains perceive and recognize objects so fast and accurately (143) is still one of the major challenges of neuroscience. The precise role of neural ensembles in information encoding and the strategies for neural coding are likely to continue attracting enormous interest in the forthcoming years. Since the pioneering work of Hubel and Wiesel describing the neural representation in primary visual areas, much progress has been made. It is clear that object perception is performed by neurons in a hierarchical network, whose basic properties we are starting to understand.

We have described a representation in the human MTL that is abstract and selective. Animal studies may elucidate whether a representation with such degree of invariance and abstraction is characteristic for humans or can be also found in other species. In particular, the Jennifer Aniston type of cells has striking similarities with place cells in rats (144). Interestingly, for a rat it is crucial to navigate in an environment, whereas for us it is crucial to recognize each other. Monkeys have social interactions and can recognize each other in a colony, so it is very likely that such neurons will be also eventually found with monkey recordings. Maybe these studies should also look at MTL neurons, since due to its function in memory storage the information in MTL may be more redundant than in IT, thus dramatically increasing the likelihood of finding them.

10. ACKNOWLEDGEMENTS

11. REFERENCES

1. N. K. Logothetis and D. L. Sheinberg: Visual object recognition. *Annu Rev Neurosci*, 19, 577-621 (1996)

2. D. A. Leopold and N. K. Logothetis: Multistable phenomena: Changing views in perception. *Trends in Cognitive Sciences*, 3(7), 254-264 (1999)

3. R. Blake and N. K. Logothetis: Visual competition. *Nature Reviews Neuroscience*, 3(1), 13-23 (2002)

4. N. Kanwisher: Neural events and perceptual awareness. *Cognition*, 79(1-2), 89-113 (2001)

5. J. J. Peissig and M. J. Tarr: Visual object recognition: Do we know more now than we did 20 Years ago? *Annual Review of Psychology*, 58, 75-96 (2007)

6. M. Riesenhuber and T. Poggio: Models of object recognition. *Nature Neuroscience*, 3(SUPPL.), 1199-1204 (2000)

7. K. Tanaka: Inferotemporal Cortex and Object Vision. *Annual Review of Neuroscience*, 19(1), 109-139 (1996)

8. D. C. Van Essen and J. H. Maunsell: Hierarchical organization and functional streams in the visual cortex. *Trends in Neuroscience*, 6, 370-375 (1983)

9. L. G. Ungerleider and M. Mishkin: Two cortical visual systems. In: *Analysis of Visual Behavior*. Ed D. J. Ingle, M. A. Goodale&R. J. W. Mansfield. MIT Press, Cambridge, MA (1982)

10. R. A. Andersen and C. A. Buneo: Intentional maps in posterior parietal cortex. *Ann. Review Neuroscience*, 25, 189-220 (2002)

11. M. A. Goodale and A. D. Milner: Separate visual pathways for perception and action. *Trends in Neuroscience*, 15, 20-25 (1992)

12. D. C. Van Essen, C. H. Anderson and D. J. Felleman: Information processing in the primate visual system: An integrated systems perspective. *Science*, 255(5043), 419-423 (1992)

13. R. Cabeza and L. Nyberg: Imaging cognition II: An empirical review of 275 PET and fMRI studies. *Journal of Cognitive Neuroscience*, 12(1), 1-47 (2000)

14. W. Li, V. Pie?ch and C. D. Gilbert: Perceptual learning and top-down influences in primary visual cortex. *Nature Neuroscience*, 7(6), 651-657 (2004)

15. J. W. Bisley and M. E. Goldberg: Neuronal Activity in the Lateral Intraparietal Area and Spatial Attention. *Science*, 299(5603), 81-86 (2003) doi:10.1126/science.1077395

16. S. Hochstein and M. Ahissar: View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, 36(5), 791-804 (2002)

17. S. Grossberg and J. R. Williamson: A Neural Model of how Horizontal and Interlaminar Connections of Visual Cortex Develop into Adult Circuits that Carry Out Perceptual Grouping and Learning. *Cereb. Cortex*, 11(1), 37-58 (2001) doi:10.1093/cercor/11.1.37

18. L. Itti, C. Koch and E. Niebur: A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11), 1254-1259 (1998)

19. P. R. Roelfsema, V. A. F. Lamme and H. Spekreijse: The implementation of visual routines. *Vision Research*, 40(10-12), 1385-1411 (2000)

20. L. Itti and C. Koch: Computational modelling of visual attention. *Nature Reviews Neuroscience*, 2(3), 194-203 (2001)

21. C. D. Gilbert, M. Sigman and R. E. Crist: The neural basis of perceptual learning. *Neuron*, 31(5), 681-697 (2001)

22. P. R. Roelfsema: Cortical algorithms for perceptual grouping. (2006)

23. V. A. F. Lamme and P. R. Roelfsema: The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, 23(11), 571-579 (2000)

24. D. H. Hubel and T. N. Wiesel: Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol*, 160, 106-154 (1962)

25. A. Pasupathy and C. E. Connor: Responses to contour features in macaque area V4. *Journal of Neurophysiology*, 82(5), 2490-2502 (1999)

26. C. G. Gross, C. E. Rocha-Miranda and D. B. Bender: Visual properties of neurons in inferotemporal cortex of the Macaque. *J Neurophysiol*, 35(1), 96-111 (1972)

27. R. Desimone and C. G. Gross: Visual areas in the temporal cortex of the macaque. *Brain Research*, 178(2-3), 363-380 (1979)

28. S. Zeki, J. D. Watson, C. J. Lueck, K. J. Friston, C. Kennard and R. S. Frackowiak: A direct demonstration of functional specialization in human visual cortex. *J Neurosci*, 11(3), 641-9. (1991)

29. M. Corbetta, F. M. Miezin, S. Dobmeyer, G. L. Shulman and S. E. Petersen: Selective and divided attention during visual discriminations of shape, color, and speed: Functional anatomy by positron emission tomography. *Journal of Neuroscience*, 11(8), 2383-2402 (1991)

30. M. I. Sereno, A. M. Dale, J. B. Reppas, K. K. Kwong, J. W. Belliveau, T. J. Brady, B. R. Rosen and R. B. Tootell: Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science*, 268(5212), 889-93. (1995)

31. R. B. Tootell, J. B. Reppas, A. M. Dale, R. B. Look, M. I. Sereno, R. Malach, T. J. Brady and B. R. Rosen: Visual motion aftereffect in human cortical area MT revealed by functional magnetic resonance imaging. *Nature*, 375(6527), 139-41. (1995)

32. R. Malach, J. B. Reppas, R. R. Benson, K. K. Kwong, H. Jiang, W. A. Kennedy, P. J. Ledden, T. J. Brady, B. R. Rosen and R. B. H. Tootell: Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 92(18), 8135-8139 (1995)

33. R. Desimone, T. D. Albright, C. G. Gross and C. Bruce: Stimulus-selective properties of inferior temporal neurons in the macaque. *Journal of Neuroscience*, 4(8), 2051-2062 (1984)

34. N. Kanwisher, J. McDermott and M. M. Chun: The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, 17(11), 4302-4311 (1997)

35. N. Kanwisher: Domain specificity in face perception. *Nature Neuroscience*, 3(8), 759-763 (2000)

36. R. Epstein and N. Kanwisher: A cortical representation of the local visual environment. *Nature*, 392(6676), 598-601 (1998)

37. R. Epstein, A. Harris, D. Stanley and N. Kanwisher: The parahippocampal place area: Recognition, navigation, or encoding? *Neuron*, 23(1), 115-125 (1999)

38. P. E. Downing, Y. Jiang, M. Shuman and N. Kanwisher: A cortical area selective for visual processing of the human body. *Science*, 293(5539), 2470-2473 (2001)

39. M. V. Peelen and P. E. Downing: The neural basis of visual body perception. *Nat Rev Neurosci*, 8(8), 636-648 (2007)

40. I. Gauthier, M. J. Tarr, A. W. Anderson, P. Skudlarski and J. C. Gore: Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nature Neuroscience*, 2(6), 568-573 (1999)

41. I. Gauthier, M. J. Tarr, J. Moylan, P. Skudlarski, J. C. Gore and A. W. Anderson: The fusiform 'face area' is part of a network that processes faces at the individual level. *Journal of Cognitive Neuroscience*, 12(3), 495-504 (2000)

42. M. J. Tarr and I. Gauthier: FFA: A flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nature Neuroscience*, 3(8), 764-769 (2000)

43. J. V. Haxby, M. I. Gobbini, M. L. Furey, A. Ishai, J. L. Schouten and P. Pietrini: Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293(5539), 2425-2430 (2001)

44. A. Ishai, L. G. Ungerleider, A. Martin, J. L. Schouten and J. V. Haxby: Distributed representation of objects in the human ventral visual pathway. *Proceedings of the National Academy of Sciences of the United States of America*, 96(16), 9379-9384 (1999)

45. K. Grill-Spector, N. Knouf and N. Kanwisher: The fusiform face area subserves face perception, not generic within-category identification. *Nature Neuroscience*, 7(5), 555-562 (2004)

46. K. Grill-Spector and R. Malach: The human visual cortex. *Annual Review of Neuroscience*, 27, 649-677 (2004)

47. C. M. Bukach, I. Gauthier and M. J. Tarr: Beyond faces and modularity: the power of an expertise framework. *Trends in Cognitive Sciences*, 10(4), 159-166 (2006)

48. I. Gauthier and C. Bukach: Should we reject the expertise hypothesis? *Cognition*, 103(2), 322-330 (2007)

49. D. Y. Tsao, W. A. Freiwald, R. B. H. Tootell and M. S. Livingstone: A cortical region consisting entirely of face-selective cells. *Science*, 311(5761), 670-674 (2006)

50. D. Y. Tsao, W. A. Freiwald, T. A. Knutsen, J. B. Mandeville and R. B. H. Tootell: Faces and objects in macaque cerebral cortex. *Nature Neuroscience*, 6(9), 989-995 (2003)

51. A. Seckel: The Great Book of Optical Illusions. Firefly Books, (2002)

52. N. K. Logothetis and J. D. Schall: Neuronal correlates of subjective visual perception. *Science*, 245(4919), 761-3. (1989)

53. D. L. Sheinberg and N. K. Logothetis: The role of temporal cortical areas in perceptual organization. *Proceedings of the National Academy of Sciences of the United States of America*, 94(7), 3408-3413 (1997)

54. S. R. Afraz, R. Kiani and H. Esteky: Microstimulation of inferotemporal cortex influences face categorization. *Nature*, 442(7103), 692-695 (2006)

55. A. K. Engel, C. K. E. Moll, I. Fried and G. A. Ojemann: Invasive recordings from the human brain: Clinical insights and beyond. *Nature Reviews Neuroscience*, 6(1), 35-47 (2005)

56. E. D. Lumer, K. J. Friston and G. Rees: Neural Correlates of Perceptual Rivalry in the Human Brain. *Science*, 280(5371), 1930-1934 (1998) doi:10.1126/science.280.5371.1930

57. K. Grill-Spector, T. Kushnir, T. Hendler and R. Malach: The dynamics of object-selective activation correlate with recognition performance in humans, 3(8), 837-843 (2000)

58. F. Tong, K. Nakayama, J. T. Vaughan and N. Kanwisher: Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron*, 21(4), 753-759 (1998)

59. D. Marr: Vision: a computational investigation into the human representation and processing of visual information. W. H. Freeman, San Francisco (1982)

60. D. Marr and H. K. Nishihara: Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London. Series B. Biological sciences*, 200(1140), 269-294 (1978)

61. I. Biederman: Recognition-by-components: a theory of human image understanding. *Psychological review*, 94(2), 115-147 (1987)

62. I. Biederman and E. E. Cooper: Evidence for complete translational and reflectional invariance in visual object priming. *Perception*, 20(5), 585-593 (1991)

63. J. E. Hummel and I. Biederman: Dynamic binding in a neural network for shape recognition. *Psychological Review*, 99(3), 480-517 (1992)

64. E. Rolls and G. Deco: Computational neuroscience of vision. Oxford University Press, (2002)

65. M. J. Tarr, P. Williams, W. G. Hayward and I. Gauthier: Three-dimensional object recognition is viewpoint dependent. *Nature Neuroscience*, 1(4), 275-277 (1998)

66. M. J. Tarr and S. Pinker: Mental rotation and orientation-dependence in shape recognition. *Cognitive psychology*, 21(2), 233-282 (1989)

67. N. K. Logothetis and J. Pauls: Psychophysical and Physiological Evidence for Viewer-centered Object Representations in the Primate. *Cereb. Cortex*, 5(3), 270-288 (1995)

68. K. Fukushima: Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36(4), 193-202 (1980)

69. M. Riesenhuber and T. Poggio: Hierarchical models of object recognition in cortex, 2(11), 1019-1025 (1999)

70. D. I. Perrett and M. W. Oram: Neurophysiology of shape processing. *Image and Vision Computing*, 11(6), 317-333 (1993)

71. G. Wallis and E. T. Rolls: Invariant face and object recognition in the visual system. *Progress in Neurobiology*, 51(2), 167-194 (1997)

72. T. Serre, L. Wolf and T. Poggio: Object recognition with features inspired by visual cortex. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. (2005)

73. T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber and T. Poggio: Robust object recognition with cortex-like

mechanisms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(3), 411-426 (2007)

74. T. Masquelier and S. J. Thorpe: Unsupervised learning of visual features through spike timing dependent plasticity. *PLoS Computational Biology*, 3(2), 0247-0257 (2007)

75. G. Deco and E. T. Rolls: A Neurodynamical cortical model of visual attention and invariant object recognition. *Vision Research*, 44(6), 621-642 (2004)

76. T. Serre, A. Oliva and T. Poggio: A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences of the United States of America*, 104(15), 6424-6429 (2007)

77. J. Hegde and D. J. Felleman: Reappraising the Functional Implications of the Primate Visual Anatomical Hierarchy. *Neuroscientist*, 13(5), 416-421 (2007) doi:10.1177/1073858407305201

78. K. S. Saleem and K. Tanaka: Divergent Projections from the Anterior Inferotemporal Area TE to the Perirhinal and Entorhinal Cortices in the Macaque Monkey. *J. Neurosci.*, 16(15), 4757-4775 (1996)

79. W. A. Suzuki: Neuroanatomy of the monkey entorhinal, perirhinal and parahippocampal cortices: Organization of cortical inputs and interconnections with amygdala and striatum. *Seminars in Neuroscience*, 8(1), 3-12 (1996)

80. L. Squire and S. Zola-Morgan: The medial temporal lobe memory system. *Science*, 253(5026), 1380-1386 (1991)

81. H. Eichenbaum, G. Schoenbaum, B. Young and M. Bunsey: Functional organization of the hippocampal memory system. *PNAS*, 93(24), 13500-13507 (1996)

82. H. Eichenbaum: A cortical-hippocampal system for declarative memory. *Nature Reviews Neuroscience*, 1(1), 41-50 (2000)

83. L. R. Squire, C. E. L. Stark and R. E. Clark: THE MEDIAL TEMPORAL LOBE. *Annual Review of Neuroscience*, 27(1), 279-306 (2004)

84. W. PENFIELD and B. MILNER: Memory deficit produced by bilateral lesions in the hippocampal zone. *AMA Arch Neurol Psychiatry*, 79(5), 475-97 (1958)

85. W. B. Scoville and B. Milner: Loss of recent memory after bilateral hippocampal lesions. 1957. *The Journal of neuropsychiatry and clinical neurosciences*, 12(1), 103-113 (2000)

86. B. Milner, S. Corkin and H. Teuber: Further analysis of the hippocampal amnesic syndrome. *Neuropsychologia*, 6, 215-234 (1968)

87. S. Corkin: What's new with the amnesic patient H.M.? *Nature Reviews Neuroscience*, 3(2), 153-160 (2002)

88. S. Zola-Morgan, L. R. Squire and D. G. Amaral: Human amnesia and the medial temporal region: enduring memory impairment following a bilateral lesion limited to field CA1 of the hippocampus. *J. Neurosci.*, 6(10), 2950-2967 (1986)

89. E. Tulving, C. A. Hayman and C. A. Macdonald: Long-lasting perceptual priming and semantic learning in amnesia: a case experiment. *Journal of experimental psychology. Learning, memory, and cognition*, 17(4), 595-617 (1991)

90. M. Mishkin: Memory in monkeys severely impaired by combined but not by separate removal of amygdala and hippocampus. *Nature*, 273(5660), 297-298 (1978)

91. S. Zola-Morgan, L. R. Squire, R. P. Clower and N. L. Rempel: Damage to the perirhinal cortex exacerbates memory impairment following lesions to the hippocampal formation. *Journal of Neuroscience*, 13(1), 251-265 (1993)

92. J. V. Haxby, L. G. Ungerleider, B. Horwitz, J. M. Maisog, S. I. Rapoport and C. L. Grady: Face encoding and recognition in the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, 93(2), 922-927 (1996)

93. L. Nyberg, A. R. McIntosh, R. Cabeza, R. Habib, S. Houle and E. Tulving: General and specific brain regions involved in encoding and retrieval of events: What, where, and when. *Proceedings of the National Academy of Sciences of the United States of America*, 93(20), 11280-11285 (1996)

94. A. D. Wagner, D. L. Schacter, M. Rotte, W. Koutstaal, A. Maril, A. M. Dale, B. R. Rosen and R. L. Buckner: Building memories: Remembering and forgetting of verbal experiences as predicted by brain activity. *Science*, 281(5380), 1188-1191 (1998)

95. J. B. Brewer, Z. Zhao, J. E. Desmond, G. H. Glover and J. D. E. Gabrieli: Making memories: Brain activity that predicts how well visual experience will be remembered. *Science*, 281(5380), 1185-1187 (1998)

96. M. Moscovitch, R. S. Rosenbaum, A. Gilboa, D. R. Addis, R. Westmacott, C. Grady, M. P. McAndrews, B. Levine, S. Black, G. Winocur and L. Nadel: Functional neuroanatomy of remote episodic, semantic and spatial memory: A unified account based on multiple trace theory. *Journal of Anatomy*, 207(1), 35-66 (2005)

97. S. Steinvorth, B. Levine and S. Corkin: Medial temporal lobe structures are needed to re-experience remote autobiographical memories: Evidence from H.M. and W.R. *Neuropsychologia*, 43(4), 479-496 (2005)

98. A. Pouget, P. Dayan and R. Zemel: Information processing with population codes. *Nature Reviews Neuroscience*, 1(2), 125-132 (2000)

99. D. Attwell and S. B. Laughlin: An energy budget for signaling in the grey matter of the brain. *Journal of Cerebral Blood Flow and Metabolism*, 21(10), 1133-1145 (2001)

100. P. Lennie: The Cost of Cortical Computation. *Current Biology*, 13(6), 493-497 (2003)

101. R. Q. Quiroga, L. Reddy, C. Koch and I. Fried: Decoding visual inputs from multiple neurons in the human temporal lobe. *Journal of Neurophysiology*, 98(4), 1997-2007 (2007)

102. B. A. Olshausen and D. J. Field: Sparse coding of sensory inputs. *Current Opinion in Neurobiology*, 14(4), 481-487 (2004)

103. E. T. Rolls and M. J. Tovee: Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *Journal of Neurophysiology*, 73(2), 713-726 (1995)

104. R. Quian Quiroga, L. Reddy, C. Koch and I. Fried: Decoding visual inputs from multiple neurons in the human temporal lobe. *J Neurophysiol*, 00125.2007 (2007)

105. J. Hegde and D. C. Van Essen: A Comparative Study of Shape Representation in Macaque Visual Areas V2 and V4. *Cereb. Cortex*, 17(5), 1100-1116 (2007)

106. C. G. Gross: Genealogy of the "grandmother cell". *Neuroscientist*, 8(5), 512-518 (2002)

107. D. Rose: Some reflections on (or by?) grandmother cells. *Perception*, 25(8), 881-886 (1996)

108. K. A. C. Martin: A brief history of the "feature detector". *Cerebral Cortex*, 4(1), 1-7 (1994)

109. W. James: The Principles of Psychology. H. Holt, (1890)

110. C. Sherrington: Man on His Nature. Cambridge University Press, (1951)

111. J. Konorski: Integrative activity of the brain. University of Chicago Press Chicago, (1967)

112. H. B. Barlow: Single units and sensation: A neuron doctrine for perceptual psychology? *Perception*, 1(4), 371-394 (1972)

113. C. von der Malsburg: The Correlation Theory of Brain Function (1981)

114. W. Singer and C. M. Gray: Visual Feature Integration and the Temporal Correlation Hypothesis. *Annual Review of Neuroscience*, 18(1), 555-586 (1995)

115. C. Gross, D. Bender and C. Rocha-Miranda: Visual Receptive Fields of Neurons in Inferotemporal Cortex of the Monkey. *Science*, 166(3910), 1303 (1969)

116. I. Fujita, K. Tanaka, M. Ito and K. Cheng: Columns for visual features of objects in monkey inferotemporal cortex. *Nature*, 360(6402), 343-6. (1992)

117. K. Tanaka, H.-A. Saito, Y. Fukada and M. Moriya: Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *Journal of Neurophysiology*, 66(1), 170-189 (1991)

118. M. Young and S. Yamane: Sparse population coding of faces in the inferotemporal cortex. *Science*, 256(5061), 1327-1331 (1992)

119. N. K. Logothetis, J. Pauls and T. Poggio: Shape representation in the inferior temporal cortex of monkeys. *Current Biology*, 5(5), 552-563 (1995)

120. E. Kobatake and K. Tanaka: Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *Journal of Neurophysiology*, 71(3), 856-867 (1994)

121. E. Rolls and G. Baylis: Size and contrast have only small effects on the responses to faces of neurons in the cortex of the superior temporal sulcus of the monkey. *Experimental Brain Research*, 65(1), 38-48 (1986)

122. M. Ito, H. Tamura, I. Fujita and K. Tanaka: Size and position invariance of neuronal responses in monkey inferotemporal cortex. *J Neurophysiol*, 73(1), 218-226 (1995)

123. N. Sigala, F. Gabbiani and N. K. Logothetis: Visual categorization and object representation in monkeys and humans. *Journal of Cognitive Neuroscience*, 14(2), 187-198 (2002)

124. G. Kayaert, I. Biederman, H. P. Op De Beeck and R. Vogels: Tuning for shape dimensions in macaque inferior temporal cortex. *European Journal of Neuroscience*, 22(1), 212-224 (2005)

125. D. A. Leopold, I. V. Bondar and M. A. Giese: Norm-based face encoding by single neurons in the monkey inferotemporal cortex. *Nature*, 442(7102), 572-575 (2006)

126. G. Rhodes, S. Brennan and S. Carey: Identification and ratings of caricatures: Implications for mental representations of faces. *Cognitive Psychology*, 19(4), 473-497 (1987)

127. G. Loffler, G. Yourganov, F. Wilkinson and H. R. Wilson: fMRI evidence for the neural representation of faces. *Nature Neuroscience*, 8(10), 1386-1390 (2005)

128. W. De Baene, E. Premereur and R. Vogels: Properties of shape tuning of macaque inferior temporal neurons examined using rapid serial visual presentation. *Journal of Neurophysiology*, 97(4), 2900-2916 (2007)

129. C. P. Hung, G. Kreiman, T. Poggio and J. J. DiCarlo: Fast Readout of Object Identity from Macaque Inferior Temporal Cortex. *Science*, 310(5749), 863-866 (2005)

130. Y. Sugase, S. Yamane, S. Ueno and K. Kawano: Global and fine information coded by single neurons in the temporal visual cortex. *Nature*, 400(6747), 869-873 (1999)

131. R. Kiani, H. Esteky, K. Mirpour and K. Tanaka: Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *Journal of Neurophysiology*, 97(6), 4296-4309 (2007)

132. D. Zoccolan, M. Kouh, T. Poggio and J. J. DiCarlo: Trade-Off between Object Selectivity and Tolerance in Monkey Inferotemporal Cortex. *J. Neurosci.*, 27(45), 12292-12307 (2007) doi:10.1523/jneurosci.1897-07.2007

133. J. J. DiCarlo and D. D. Cox: Untangling invariant object recognition. *Trends in Cognitive Sciences*, 11(8), 333-341 (2007)

134. I. Fried, K. A. MacDonald and C. L. Wilson: Single neuron activity in human hippocampus and amygdala during recognition of faces and objects. *Neuron*, 18(5), 753-65 (1997)

135. G. Kreiman, C. Koch and I. Fried: Category-specific visual responses of single neurons in the human medial temporal lobe, 3(9), 946-953 (2000)

136. R. Q. Quiroga, L. Reddy, G. Kreiman, C. Koch and I. Fried: Invariant visual representation by single neurons in the human brain. *Nature*, 435(7045), 1102-7 (2005)

137. R. Q. Quiroga, Z. Nadasdy and Y. Ben-Shaul: Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural Comput*, 16(8), 1661-87 (2004)

138. S. Waydo, A. Kraskov, R. Quian Quiroga, I. Fried and C. Koch: Sparse Representation in the Human Medial Temporal Lobe. *J. Neurosci.*, 26(40), 10232-10234 (2006)

139. S. Shoham, D. H. O'Connor and R. Segev: How silent is the brain: Is there a "dark matter" problem in neuroscience? *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, 192(8), 777-784 (2006)

140. M. J. Buckley and D. Gaffan: Perirhinal cortical contributions to object perception. *Trends in Cognitive Sciences*, 10(3), 100-107 (2006)

141. S. Wirth, M. Yanike, L. M. Frank, A. C. Smith, E. N. Brown and W. A. Suzuki: Single Neurons in the Monkey Hippocampus and Learning of New Associations. *Science*, 300(5625), 1578-1581 (2003)

142. K. Sakai and Y. Miyashita: Neural organization for the long-term memory of paired associates. *Nature*, 354(6349), 152-155 (1991)

143. S. J. Thorpe and M. Fabre-Thorpe: NEUROSCIENCE: Seeking Categories in the Brain. *Science*, 291(5502), 260-263 (2001)

144. M. A. Wilson and B. L. McNaughton: Dynamics of the Hippocampal Ensemble Code for Space. *Science*, 261(5124), 1055-1058 (1993)

Footnotes: [1] It is interesting that similar results have recently been reported in humans by means of fMRI experiments in the fusiform face area (127). [2] Otherwise one of us, RQQ, would be by now living out of his amazing luck in Las Vegas. [3] This is actually another strong blow to the view of these neurons as grandmother cells.

Abbreviations: IT: inferotemporal cortex, LGN: lateral geniculate nucleus, V1: primary visual area, MTL: medial temporal lobe, PET: positron emission tomography, fMRI: functional magnetic resonance imaging, BOLD: blood oxygen level-dependent, STS: superior temporal sulcus, RF: receptive field, FFA: fusiform face area, PPA: parahippocampal place area, EBA: extrastriate body area.

Key Words: Object Recognition, Memory, Perception, Sparse Coding, Grandmother Cells, Invariance, Visual Representation, Ventral Visual Pathway, Medial Temporal Lobe, Inferotemporal Cortex, Review

Send correspondence to: Dr. Matias J. Ison, Department of Engineering, University of Leicester, University Road, LE1 7RH, Leicester, United Kingdom, Tel: 44 116 2522823, Fax: 44116 2522619, E-mail: mji3@le.ac.uk