

Neuronal codes for visual perception and memory

Rodrigo Quian Quiroga

Centre for Systems Neuroscience
University of Leicester.
9 Salisbury Rd, LE1 7QR, Leicester, UK.

Corresponding author:
Rodrigo Quian Quiroga
Tel: +44 116 252 5200
Fax: +44 116 252 5043
Email: rqqg1@le.ac.uk

Keywords: perception; memory; concept cells; neuronal coding; memory model

Abstract

In this review, I describe and contrast the representation of stimuli in visual cortical areas and in the medial temporal lobe (MTL). While cortex is characterized by a distributed and implicit coding that is optimal for recognition and storage of semantic information, the MTL shows a much sparser and explicit coding of specific concepts that is ideal for episodic memory. I will describe the main characteristics of the coding in the MTL by the so-called concept cells and will then propose a model of the formation and recall of episodic information based on partially overlapping assemblies.

Introduction

One of the most important and at the same time fascinating functions of the brain is to construct meaning from relatively little, noisy and ambiguous sensory information. We indeed see, and are aware of seeing, a dog, a laptop, a famous actor or our mother, rather than the minute features that characterize these images or the combination of wavelengths that impinge our retina. The notion that our sensations and thoughts are based on constructions we make of the external world goes all the way back to the beginnings of scientific reasoning in the thoughts of Aristotle (Aristotle, reprinted 2004) and has been particularly substantiated by Hermann von Helmholtz, who argued that perception is given by representations made from unconscious inferences (Helmholtz, 1878). However, perception is just the beginning of elaborate brain processes that are triggered by seeing, for example, the face of an old friend in a cafe. In particular, recognizing the face of our friend will elicit the recall of related information and this will eventually lead to the formation of new memories about the encounter in the cafe.

In the following, I will focus on visual perception - given that this is the dominant sensory modality in humans and many other animals, and also given the vast number of studies dealing with vision compared to other sensory modalities - although I would argue that the principles described below apply not only to vision but to sensory processing in general. A myriad of studies have characterized the neural substrates of visual perception along the so-called ventral visual pathway (VVP) (Grill-Spector & Malach, 2004; Logothetis & Sheinberg, 1996; Roelfsema, 2006; Tanaka, 1996; D.Y Tsao & Livingstone, 2008). From high level visual areas in the VVP, there are strong connections to the medial temporal lobe (MTL) (see Figure 1) (Lavenex & Amaral, 2000; Saleem & Tanaka, 1996; Suzuki, 1996), where evidence from patient H.M. (Corkin, 2002; Milner, Corkin, & Teuber, 1968; Scoville & Milner, 1957; L. Squire, 2009), other patients with lesions in the MTL (Moscovitch, et al., 2005; Rosenbaum, et al., 2005; L. R. Squire, Stark, & Clark, 2004) and animal studies (Mishkin, 1978; L. Squire & Zola-Morgan, 1991; L. R. Squire, et al., 2004) have demonstrated the key role of this area in declarative memory (L. Squire & Zola-Morgan, 1991; L. R. Squire, et al., 2004). In spite of the strong relationship between perception and

memory, both behaviorally (considering that perception triggers memory formation) and anatomically (given that high level VVP areas project to the MTL), these two processes have been mostly studied independently (for previous studies linking perception and memory see (Carlson, Simmons, Kriegeskorte, & Slevc, 2013; Clarke, Taylor, Debereux, Randall, & Tyler, 2013; Cowell, Bussey, & Saksida, 2010; Miyashita, 1993; Naya, Yoshida, & Miyashita, 2001; Quian Quiroga, 2012b)), without focusing on the processes and the subjacent neuronal representations that determine how external stimuli elicits the creation of the conceptual representations that we use, for example, to form new episodic memories. (In the following, by 'concept' I refer to the meaning of a stimulus – e.g. a given person, place or object - as opposed to a representation of sensory features.) These are precisely the two processes – visual perception in the VVP and memory in the MTL – that I will discuss and contrast in this short review. In particular, I will argue that both perception and memory involve a construction of meaning by selecting little information and using inferences and abstractions. But while perception involves distributed representations, with category responses and a hierarchical organization of information that is ideal for the robust identification of a stimulus and its related semantic information, (episodic) memory involves much sparser, invariant and not organized representations that are better suited for the fast creation of associations.

The review is structured as follows: I will first outline basic similarities between perception and memory processes, then I will briefly describe perception along the VVP – just highlighting a few selected studies, given that this topic has already been comprehensively reviewed elsewhere (see references above) – and in the final two sections I will describe more in detail the findings with single cell recordings in the MTL and, based on this type of representation, I will then discuss a very simple model of memory formation based on partial overlapping assemblies.

Perception and memory

Although perception and memory are two different processes, they are intrinsically related because the perception of a person or an object requires contrasting sensory features with stored representations. In fact, cases of visual

associative agnosia provide clear evidence of this link considering that, due to specific lesions, these patients cannot associate a percept with its meaning (i.e. they cannot access the memory of the percept), in spite of having normal vision (Farah & Feinberg, 2000).

More generally, it has been argued that both perception and memory are constructions by the brain based on the meaning we attribute to percepts or events (for alternative views, particularly the view of perception as a direct processing of information, see (Gibson, 1972)). In particular, in the late XIX century Helmholtz argued that the information we get through our senses is very limited and that perception is largely given by unconscious inferences based on previous experiences (Gregory, 1997; Helmholtz, 1878). Then, according to Helmholtz we construct rich sensory representations based on these assumptions and we see prototypes or *signs* of reality rather than the reality itself. The attribution of meaning in perception is supported by redundant, parallel processing (Barlow, 1981) of relatively little visual information (Koch, et al., 2006), and it involves categorizations to deal with the millions of bits of information that enter our senses (Fabre-Thorpe, 2003). These categorizations are partially subjective and can be done at different levels (Logothetis & Sheinberg, 1996; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976) – depending on the knowledge and expertise of the subject (Bukach, Gauthier, & Tarr, 2006; Palmeri & Gauthier, 2004; Tanaka & Taylor, 1991) – mainly extracting prototypes and leaving aside a number of details and individual differences (Fabre-Thorpe, 2003; Logothetis & Sheinberg, 1996; Palmer, 1999; Palmeri & Gauthier, 2004).

With respect to memory, as in the case of perception it has long been recognized that we remember the subjective meaning we attribute to episodes using abstractions and the consequent loss of details (James, 1890) (for a review of experimental evidence see e.g. (Koriat, Goldsmith, & Pansky, 2000)). For example, in a series of experiments where subjects were asked to remember unfamiliar folk tales, Frederic Bartlett showed that memory retrieval is a constructive process that is far from a passive replay of stored information. In fact, the recalled stories tended to be shorter, more coherent and fitting closely the particular viewpoint of the subjects (Bartlett, 1932). In other words, subjects

actively looked for the meaning of the stories they were told and created a *schema*, a sense of the story they stored and later recalled.

The similarity between perceptual and memory illusions stresses further the link between perception and memory (Roediger, 1996). In perception, the fact that we constantly make inferences about what we see is manifest in visual illusions, where unconscious, unavoidable assumptions lead to false interpretations and astonishing effects (Eagleman, 2001; Gregory, 1997). In memory, analogous type of assumptions elicit the creation of false memories, for example, when subjects are asked to memorize a list of words (e.g.: thread, pin, eye, sewing, sharp, etc.) that have a close associate (needle) which tends to be incorrectly recognized as part of the list (Deese, 1959; Roediger & McDermott, 1995), or when, among a set of faces, subjects incorrectly recognize the prototype formed with these faces as being previously presented (Gao & Wilson, 2014). Along this line, there is large experimental evidence about memory distortions and the creation of false memories described in the literature (see e.g. (Baddeley, Eysenck, & Anderson, 2009; Loftus & Palmer, 1974; Schacter, 2001; Schacter & Slotnick, 2004)).

Perception in the ventral visual pathway

Given the vast evidence about visual perception along the ventral visual pathway reviewed elsewhere (Grill-Spector & Malach, 2004; Logothetis & Sheinberg, 1996; Roelfsema, 2006; Tanaka, 1996; D.Y Tsao & Livingstone, 2008), in this section I will just summarize evidence from electrophysiology recordings in monkeys and humans about encoding of stimuli in visual cortex, which I will later contrast with evidence from recordings in the human MTL. In particular, I will argue that in visual cortical areas there is a topographically organized and distributed representation of visual stimuli, a representation that is ideal for the subjective extraction of meaning and semantic categorizations. I will also show that response onsets in high level visual areas are at about 100 ms, which are much earlier than the responses obtained in the MTL, at about 300 ms. A critical limitation when comparing VVP and MTL responses is that it is in principle not possible to record single neuron activity in human visual cortex (the placing of depth electrodes in these areas is not clinically justified) and, in this case,

electrophysiology recordings are limited to electrocorticograms (i.e. grids of electrodes placed on the surface of the brain) or scalp EEGs. These recordings offer complementary information to the one given by single cell recordings, but lack specificity because they reflect the common activity of large neuronal populations. However, given the similarity between intracranial LFP responses in both species, it seems plausible to assume that similar principles may apply (particularly with respect to response latencies in high level visual areas).

Electrophysiology recordings in monkeys

The recognition of visual stimuli is carried out in the ventral visual pathway, extending from the primary cortical visual area (V1) to the inferotemporal cortex (IT) (Figure 1A) (Goodale & Milner, 1992; Ungerleider & Mishkin, 1982) (though it has been argued that the VVP should be extended further anteriorly to include the prefrontal cortex (Kravitz, Saleem, Baker, Ungerleider, & Mishkin, 2013)). Converging evidence from monkey electrophysiology recordings in different areas along this pathway have shown that, while neurons in V1 respond to local orientations, thus representing the minute details that compose an image, neurons in IT have larger receptive fields and respond, at about 100 ms after stimulus onset (see Table 1 in (Mormann, et al., 2008)), to a high level representation of the visual inputs, showing selectivity to complex visual features, together with a larger degree of visual invariance – i.e. a robustness to simple image transformations, such as changes of the object size, position or 2D rotations – (Gross, 1992, 2008; Logothetis & Sheinberg, 1996; Roelfsema, 2006; Tanaka, 1996; D.Y Tsao & Livingstone, 2008). In fact, neurons in IT have been shown to respond to experimentally defined (Sigala & Logothetis, 2002) and natural category of stimuli, such as faces, objects or body parts (Gross, 1992, 2008; Logothetis & Sheinberg, 1996; Tanaka, 1996). Along this line, a set of remarkable studies, using fMRI activations to determine the areas of interest for single cell recordings, described 6 cortical patches with a large majority of neurons that responded exclusively to faces (Freiwald & Tsao, 2010; D.Y. Tsao, Freiwald, Tootell, & Livingstone, 2006).

Given the representation of minute visual details in early visual areas that gives rise to a more complex representation of information in higher areas, it has

been proposed that such hierarchical processing may end up in *gnostic* or *grandmother cells* that represent individual stimuli in a specific and invariant way – that means, neurons that fire to different views of a specific individual or object but not to other individuals or objects (Barlow, 1994; Konoski, 1967)(for a discussion see (Gross, 2002; Quian Quiroga, Fried, & Koch, 2013)). However, the degree of selectivity and invariance found in IT cortex is relatively limited and very far from a grandmother cell representation (see, however, (Bowers, 2009; Quian Quiroga & Kreiman, 2010) for a debate of whether these and other neurons should be considered grandmother cells or not). In fact, neurons in the macaque temporal cortex show a distributed representation, given that they tend to respond to a relatively large number of stimuli (typically more than 20% in average) mainly within the neuron’s preferred category (Rolls & Tovee, 1995; Tamura & Tanaka, 2001; D.Y. Tsao, et al., 2006). Moreover, as it has been shown with single cell (Hung, Kreiman, Poggio, & DiCarlo, 2005; Young & Yamane, 1992) and local field potential recordings (Kreiman, et al., 2006), the information about the specific stimulus identity is given by modulations in the response strength to the different stimuli when considering a *population* of neurons. In other words, while the information about the stimulus category is explicit at the single cell level (i.e. from the firing of a neuron we can tell whether the stimulus was a face or not), the information about the specific stimulus identity within a category is implicit (i.e. from the firing of a neuron we cannot tell which face it is) and distributed across a neural population (Quian Quiroga & Kreiman, 2010).

Altogether, these results support the notion that neurons along the ventral visual pathway (and the prefrontal cortex) are involved in extracting the subjective meaning of the presented stimulus. In line with this view, using images eliciting ambiguous percepts – most notably with binocular rivalry (Blake & Logothetis, 2002) and variants like flash suppression (Wolfe, 1984) – it has been shown that while about 20% of the neurons are modulated by the subjective percept in V1, about 40% are modulated in V4 and 90% in IT (Kanwisher, 2001; Logothetis, 1998; Panagiotaropoulos, Kapoor, & Logothetis, 2014).

The representation of information in the monkey temporal cortex is topographically organized. In particular, it has been argued that from the activity

of neurons in the temporal cortex (IT and the superior temporal sulcus) it is possible to extract information about the hierarchical structure of semantic categories (Kiani, Esteky, Mirpour, & Tanaka, 2007; Kriegeskorte, et al., 2008; Rolls & Tovee, 1995). Although other authors have argued that this reflects a categorization of visual features (Baldassi, et al., 2013), in line with studies showing semantic categorization in neurons downstream from the VVP, in the prefrontal cortex (Freedman, Riesenhuber, Poggio, & Miller, 2003; Freedman, Riessenhuber, Poggio, & Miller, 2001). Neurons in IT cortex have been also reported to have columnar organization, in the sense that nearby neurons have a similar (though slightly different) tuning (Tanaka, 1996, 2003). Such spatial clustering is supported by the fact that responses to similar features were found in nearby neurons recorded from the same electrode (separated after spike sorting (Quiñero Quiroga, 2012c)) and also comparing perpendicular and oblique electrode penetrations (spanning neurons from the same and different putative columns, respectively) (Fujita, Tanaka, Ito, & Cheng, 1992). Further evidence along this line was given with optical imaging (Wang & Tanaka, 1996) and large-scale multielectrode recordings (Sato, et al., 2013). At a broader scale, a spatial organization of information in temporal cortex was supported by imaging studies (Bell, Hadj-Bouziane, Frihauf, Tootell, & Ungerleider, 2009; Kanwisher, 2010; Konkle & Oliva, 2012; Tsunoda, Yamane, Nishizaki, & Tanifuji, 2001) as well as by the presence of discrete face (Freiwald & Tsao, 2010; D.Y. Tsao, et al., 2006) and body patches (Popivanov, Jastorff, Vanduffel, & Vogels, 2014) studied with combined fMRI and electrophysiology recordings.

Intracranial recordings in humans

Invasive electrophysiology recordings are performed very rarely in humans due to obvious ethical reasons. One of these few exceptional cases is when studying patients suffering from epilepsy refractory to medication, in order to identify the area originating the seizures and evaluate the possibility of a surgical resection (Rey, et al., 2015). These patients are implanted with depth intracranial electrodes or with subdural strips or grids (Engel, Moll, Fried, & Ojermann, 2005) and, in order to make a clinical decision about the surgical procedure, they are monitored 24/7 for about 1 to 2 weeks, until a minimum

number of seizures is recorded. During this time, researchers have the unique opportunity to obtain intracranial recordings from awake and behaving human subjects that can perform a variety of tasks. As with monkey studies, this allows investigating visual responses in different cortical areas, but with the caveat that in humans there is typically no access to the activity of single neurons (but see next section) and only intracranial EEG data, reflecting the common activity of large neural populations, is studied.

Initial works analyzing intracranial EEG responses to picture presentations described spatially segregated areas in the occipito-temporal cortex with specific evoked potentials triggered by different category of stimuli (objects, body parts, words, numbers) and particularly by faces (Allison, McCarthy, Nobre, Puce, & Belger, 1994; Allison, Puce, Spencer, & McCarthy, 1999; Halgren, et al., 1994; McCarthy, Puce, Belger, & Allison, 1999; Puce, Allison, & McCarthy, 1999). This predominance of responses to faces is not surprising given the importance that face recognition has for normal behavior and social interactions. In line with these findings, selective responses to faces have been also described in the fusiform gyrus with positron emission tomography (Sergent, Ohta, & MacDonald, 1992) and fMRI (Kanwisher, McDermott, & Chun, 1997). Moreover, the involvement of the occipitotemporal cortex in the processing of faces is supported by cases of prosopagnosia – i.e. a deficit in face recognition following a brain injury (Rossion, 2014b) – and by studies showing that electrical stimulation in these areas impairs face recognition (Jonas, et al., 2012; Jonas, et al., 2014).

More recent studies focused on the timing of responses in visual cortical areas, showing first activations selective to faces at about 100 ms after stimulus onset (Bansal, et al., 2012; Davidesco, et al., 2014; Jacques, et al., (in press)). Moreover, using a decoding approach – i.e. predicting the stimulus category based on the neural responses (Quiñero Quiroga & Panzeri, 2009) – another study showed that already at 100 ms after picture presentation there was enough information in the evoked responses to tell apart the category of the stimulus (Liu, Agam, Madsen, & Kreiman, 2009); a finding in line with the category information obtained from single neuron (Hung, et al., 2005) and local field potential (Kreiman, et al., 2006) recordings in high level visual areas in monkeys.

Furthermore, as it was the case with monkey recordings, intracranial evoked responses in human cortical areas were modulated by the recognition of the stimulus (Fisch, et al., 2009). Another point of similarity between human and monkey cortical visual processing is the fact that responses were triggered by object categories and were not selective to specific exemplars within each category (Bansal, et al., 2012). Furthermore, as in the case of monkey recordings (Freiwald & Tsao, 2010; D.Y. Tsao, et al., 2006), it has been shown with human intracranial recordings that face selective neurons are anatomically grouped according to perceptual similarity (Davidesco, et al., 2014).

Scalp EEG recordings

With scalp EEG recordings, responses to visual stimuli are observed in early visual cortical areas within 100 ms after stimulus onset (Di Russo, Martinez, Sereno, Pitzalis, & Hillyard, 2001). These activations are very sensitive to changes in the stimulus parameters, like changes in contrast or spatial frequency (Luck, 2005). Following the finding of selective responses to faces with intracranial recordings (Allison, et al., 1994), an analogous negative potential that was larger for faces compared to other category of stimuli (e.g. cars) was observed with scalp EEG in the occipito-temporal cortex, peaking at about 170 ms after stimulus onset (Bentin, Allison, Puce, Perez, & McCarthy, 1996; Rossion, 2014a). Among many studies, Bentin and colleagues used ambiguous schematic stimuli and showed that these stimuli triggered N170 responses when the subjects learned to associate them with faces (Bentin, Sagiv, Mecklinger, Friederici, & von Cramon, 2002). These findings support the abovementioned view of perception as a constructive process that attributes meaning to the stimulus, rather than being just a direct representation of stimulus features. Although the N170 is sensitive to faces but not to the specific identity of the face (Rossion, 2014a) (though, adaptation effects of the N170 responses have been shown to dissociate face identity (Jacques, d'Arripe, & Rossion, 2007; Rossion, 2014a)), this does not necessarily rule out an involvement of the areas generating the N170 in the recognition of individual faces. In fact, the N170 represents the common activation of large neuronal populations and the coding of face identity may be given by modulations in the

firing of these neurons (as it was the case with spike and LFP recordings in monkey IT (Hung, et al., 2005; Kreiman, et al., 2006)), which may cancel out when considering the N170 response.

The N170 has been also used as a marker of conscious perception, contrasting responses to perceived versus not perceived faces (Rossion, 2014a). In particular, to elicit perceived and non-perceived conditions several works have varied the physical properties of the stimulus, like the presentation time (Tanskanen, Nasanen, Ojanpaa, & Hari, 2007), the temporal separation to a mask (Bacon-Mace, Mace, Fabre-Thorpe, & Thorpe, 2005; Harris, Wu, & Woldorff, 2011), the phase coherence of the images (Philiastides & Sajda, 2006) or the amount and type of noise added to them (Jemel, et al., 2003; Tanskanen, Nasanen, Montez, Paallysaho, & Hari, 2005). These studies reported a correlation of the N170 with face perception. Yet, in all these cases the unseen condition was obtained by changing the stimulus parameters to make the recognition of the faces more difficult. Therefore, the N170 modulations could be attributed to the processing of different low-level features, irrespective of face recognition. To rule out this confound, in a recent study (Navajas, Ahmadi, & Quian Quiroga, 2013) we used short presentation times coupled with backward masking, and varied the noise added to the images using a double staircase procedure in order to establish the amount of noise that gave a 50/50 recognition performance (i.e. the stimulus recognized as a face in half of the trials – note that we did not ask subjects to identify individual faces, but to assess whether there was a face or not in each trial). We then compared the N170 responses to the same stimuli, but separating trials according to the subjects' perception. As shown in Figure 2, the recognized faces elicited a larger N170 compared to the non-recognized ones. Interestingly, from the single-trial N170 peak amplitudes (but not from the earlier P1 or the later P2) it was possible to decode whether the subjects recognized the stimulus as a face or not, thus showing that, using identical stimuli, conscious perception led to a larger single-trial activation in occipitotemporal areas (Navajas, et al., 2013). This result seems in principle at odds with previous claims that consciousness emerges later on, at about 300 ms after stimulus onset (Dehaene & Changeux, 2011; Del Cul, Baillet, & Dehaene, 2007). However, as we have previously argued (Navajas, Rey, & Quian Quiroga,

2014), this apparent disagreement may be due to different uses of the term “consciousness”. In particular, it may be useful to distinguish between the perceptual awareness involved in recognizing a person, which is linked to the process of retrieving related semantic information – taking place in the VVP at about 100-300 ms, as shown with magnetoencephalography data (Cichy, Pantazis, & Oliva, 2014; Clarke, et al., 2013; Clarke, Taylor, & Tyler, 2011) – and the awareness of episodic information given by the retrieval of experiences related to this specific person. In the next section, I will argue that the latter process takes place after 300 ms and relies on the activation of the MTL (and further activations in cortical areas). So, while first perceptual awareness processes may take place at about 100 ms, a richer awareness, combining information from different sensory modalities and experiences related to a given concept, will only develop after 300 ms.

Single cell recordings in the human medial temporal lobe

The hippocampus and its surrounding cortex, what is known as the medial temporal lobe, is involved in certain forms of epilepsy and it is therefore a typical area targeted for the implantation of intracranial electrodes (Engel, et al., 2005; Niedermeyer, 1993; Rey, et al., 2015). The MTL comprises several interconnected areas organized in a hierarchical structure (Figure 1B): the perirhinal and parahippocampal cortex (PHC), at the bottom of this hierarchy, receive inputs from cortical sensory areas and send information to the entorhinal cortex, which in turn project to the hippocampus (Lavenex & Amaral, 2000; Quiñones Quiroga, Kraskov, Koch, & Fried, 2009; Saleem & Tanaka, 1996; Suzuki, 1996). The amygdala, also part of the MTL, has direct connections to sensory areas and the other MTL structures. As we have discussed above, the MTL is critically involved in declarative (i.e. episodic and semantic) memory (Moscovitch, et al., 2005; L. Squire & Zola-Morgan, 1991; L. R. Squire, et al., 2004), which gives researchers the unique opportunity to study memory processes with intracranial recordings in humans.

Since the early 1970's, a new type of electrodes has been developed at UCLA that allows recording not only intracranial EEG but also the activity of multiple single neurons, via microwires that pass through the depth electrode

and protrude a few millimeters from the electrode tip (Babb, Carr, & Crandall, 1973; Engel, et al., 2005; Fried, MacDonald, & Wilson, 1997; Rey, et al., 2015) (see Figure 3). Studies using this technique showed single neuron responses to individual words and faces (Heit, Smith, & Halgren, 1988), to oddball stimuli (Heit, Smith, & Halgren, 1990), conjunction of stimulus features, like gender and facial expression (Fried, et al., 1997), associated word pairs (Cameron, Yashar, Wilson, & Fried, 2001), and the degree of novelty and familiarity of the stimuli (Rutishauser, Mamelak, & Schuman, 2006; Viskontas, Knowlton, Steinmetz, & Fried, 2006). They also described neurons in the MTL that were active during visual imagery (Kreiman, Koch, & Fried, 2000) and recall (Gelbard-Sagiv, Mukamel, Harel, Malach, & Fried, 2008).

Concept cells

Several advances – in particular, the use of: i) a stimulus set optimized for each subject according to his own preferences and background, ii) screening sessions to determine which pictures elicit responses in any of the recorded neurons, and iii) optimal processing of the data, most notably, the use of optimal spike detection and sorting (Quian Quiroga, 2012c)(see Figure 3) – allowed us finding sparsely firing neurons in the human MTL, with very selective responses to various pictures of specific persons or objects (Quian Quiroga, Reddy, Kreiman, Koch, & Fried, 2005). For example, the first of these neurons fired to 7 different pictures of Jennifer Aniston (thus, they were dubbed “Jennifer Aniston neurons”) but not to 80 pictures of other actors, famous people, places, animals, etc. In a series of studies (reviewed in (Quian Quiroga, 2012b)) we have shown how these neurons fire to specific concepts, like different views of a person, and even to the person’s written or spoken name (i.e. shown in a computer screen or pronounced by a synthesized voice) (Quian Quiroga, et al., 2009) (see Figure 4). Given the role of the MTL in declarative memory and given, among other factors discussed below, the relatively late onset of MTL single neuron responses compared to responses in visual perception areas, we have postulated that these neurons have a representation of concepts, in order to create links between them and form (or recall) episodic memories (Quian Quiroga, 2012b). In other words, the activation of concept cells brings the particular concept into

awareness to embed it within its related circumstances and to enable the creation of associations, memories and the flow of consciousness. At the same time, the activation of concept cells points towards to, and links, related and more detailed and semantic representations in different cortical areas. These concepts are the subjective meaning we attribute to external stimuli, depending on how we may want to remember them. They are the *signs* of Helmholtz (Helmholtz, 1878) that we use to construct the *schemas* of Bartlett (Bartlett, 1932).

This interpretation is in line with evidence from patients with lesions in the MTL, who have a devastating deficit in assessing and combining contextual information and in forming or recalling episodic memories (Moscovitch & Nadel, 1997; Moscovitch, et al., 2005; Rosenbaum, Gilboa, Levine, Winocur, & Moscovitch, 2009; Steinvorth, Levine, & Corkin, 2005; Vargha-Khadem, et al., 1997). Furthermore, these patients are impaired at imagining new experiences, as they are able to imagine only fragmented events without any environmental context (Hassabis, Kumaran, Vann, & Maguire, 2007; Hassabis & Maguire, 2007). Conversely, patients with semantic dementia, due to a focal atrophy in the antero-lateral temporal lobe, have a severe deficit in retrieving semantic information but a much largely preserved episodic memory (Graham, Simons, Pratt, Patterson, & Hodges, 2000; Hodges & Graham, 2001; Snowden, Griffiths, & Neary, 1996). In the following I revise the main characteristics of concept cells, which supports their proposed role in declarative memory.

First, concept cells show a *very sparse tuning* to specific concepts, as they tend to fire to around 2-3% of images shown (about 100 per session) (Ison, et al., 2011; Mormann, et al., 2008; Quian Quiroga, Reddy, Koch, & Fried, 2007; Quian Quiroga, et al., 2005). Furthermore, this degree of selectivity increases along the hierarchical structure of the MTL, going from about 5% in PHC to less than 2% in the hippocampus (Quian Quiroga, 2012b). Related to this finding, the number of visually responsive neurons is more than double in PHC (18%) compared to the hippocampus (8%). Such high selectivity, with responses to specific individuals rather than to broader categories (e.g. faces, males, etc.), is in line with the proposed role of these neurons in forming and recalling episodic memories, as we tend to remember episodes involving specific persons and places: we will

remember seeing Arnold Schwarzenegger at Venice Beach, rather than seeing a famous, athletic Austrian actor at a hot place by the sea. But it is important to highlight that in spite of such very high selectivity, these neurons should not be considered grandmother cells because: i) it is in principle impossible to demonstrate that a neuron fires to one and only one concept (as we cannot show all possible concepts in an experiment), ii) many of these neurons do actually fire to more than one concept (Quian Quiroga, et al., 2009; Quian Quiroga & Kreiman, 2010; Quian Quiroga, Kreiman, Koch, & Fried, 2008) and iii) we have recently showed how these neurons can expand their tuning to encode new associations (Ison, Quian Quiroga, & Fried, 2015) (see below).

Second, concept cells show a very high level of visual and *multimodal invariance*, as they fire selectively to different pictures of specific persons, places or objects and even to their written and spoken names (Quian Quiroga, et al., 2009) (see Figure 4). The degree of invariance also increases along the hierarchical structure of the MTL: 52% of the neurons show visual invariance in PHC versus 85% in the hippocampus, and while no neuron in PHC showed responses to the text or sound presentations, about half the neurons did so in the hippocampus (Quian Quiroga, 2012b; Quian Quiroga, et al., 2009). Note that such level of invariance implies that these neurons do not fire to specific visual features, because the responses are similar to completely different pictures of the same person (and even to the person's name). This is in agreement with the fact that we tend to remember relationships between concepts and forget specific details (Bartlett, 1932; Koriat, et al., 2000; Quian Quiroga, 2010; Quian Quiroga, et al., 2008): we will remember seeing Arnold at Venice Beach but we will likely forget how exactly he looked like that day or the color of the shirt he was wearing. Note also that invariance is critical to avoid “combinatorial explosion” (i.e. having too many possible combinations to be stored), in the sense that we only need neurons firing to Arnold in this area and not separate sets of cells for Arnold giving a speech, lifting weights, or in a press conference releasing a movie. It is just “Arnold” and the specific details will not even be encoded in the MTL – although some of these details (like a very colorful shirt) could potentially be remembered if we pay particular attention to them.

Third, concept cells have an *explicit representation of the meaning* of the stimulus. The representation is explicit in the sense that, using a decoding algorithm (Quian Quiroga & Panzeri, 2009), we showed that from the firing of these neurons it was possible to tell way above chance which of many specific concepts was being seen (Quian Quiroga, et al., 2007) or even thought (Cerf, et al., 2010) by the subjects. But in line with the invariant representation described above, in most cases it was not possible to predict which specific picture of a given person was presented (Quian Quiroga, et al., 2007). The increase in decoding performance was linear with the number of neurons, thus showing that each neuron contributed explicit information about a specific concept (Quian Quiroga, et al., 2007). This contrasts with the implicit representation found in visual cortical areas, where it is not possible to infer which stimulus is being presented from the activity of a single neuron (from IT neurons we could potentially tell it is a face but in general we cannot tell which face) and there is a nonlinear increase of decoding performance with the number of neurons considered (Abbott, Rolls, & Tovee, 1996; Hung, et al., 2005; Kreiman, et al., 2006).

Fourth, using short presentation times coupled with backward masking, we showed that these neurons *follow conscious perception*, mostly with “all or none” responses. That means, concept cells fired whenever the pictures eliciting responses were recognized and remained at baseline levels when they were not, although the pictures (shown at the threshold of conscious perception) were exactly the same (Quian Quiroga, Mukamel, Isham, Malach, & Fried, 2008). The modulation of single cell MTL responses with conscious perception has been also demonstrated using flash suppression (Kreiman, Fried, & Koch, 2002), change blindness (Reddy, Quian Quiroga, Wilken, Koch, & Fried, 2006) and more recently, using ambiguous morphed images (Quian Quiroga, Kraskov, Mormann, Fried, & Koch, 2014). In this latter study, the images to which the neurons initially fired (e.g. Jodie Foster; see Figure 5) were morphed with other images to which they did not (Nicole Kidman). Using a face adaptation paradigm – i.e. presenting the picture of one or the other person before the morphed image (Leopold, Rhodes, Mueller, & Jeffery, 2005) – we triggered different percepts (recognized Jodie Foster or Nicole Kidman) from the same ambiguous image. As

shown in Figure 5, the neurons dramatically changed their firing according to the subjects' perception, in spite of the fact that the morphed images were the same in both conditions (Quian Quiroga, et al., 2014). Moreover, the response to the original picture, for example, Jodie Foster, was the same as the one to the morphed picture, as long as it was recognized as Jodie Foster. This stresses the fact that these neurons respond to the subjective percept (what the subject believed he saw), irrespective of the features of the stimulus triggering it.

Fifth, we already mentioned that concept cells may fire to more than one concept, but if this is the case, these *concepts tend to be associated* (Quian Quiroga, et al., 2009; Quian Quiroga & Kreiman, 2010). For example, the neuron firing to Jennifer Aniston also responded to Lisa Kudrow (a costar in the TV series "Friends") in a recording performed next day; a neuron that fired to Seinfeld also responded to Kramer (another character from the same TV show); a neuron firing to Luke Skywalker also fired to Yoda (both characters of the film "Star Wars"; see Figure 4), and so on. In other words, these neurons encode associations between specific concepts. Given these results, we wondered whether the tuning of concept cells could be changed by having subjects learning new associations (specific persons at specific places) and found that neurons initially firing to a given person (place) started firing to the associated place (person) from the moment the subject learnt the association (Ison, et al., 2015). The fact that the neurons encoded arbitrary associations (i.e. without a previous relationship between them) speaks against a category/semantic organization, as found in cortical areas. In agreement with this finding, modeling studies have shown that sparse representations, as the ones we found in the MTL, are critical to rapidly form associations, whereas the distributed (and organized) representations in cortex are optimal for robust perception but cannot support rapid learning (McClelland, McNaughton, & O'Reilly, 1995) (see last section for a more detailed discussion).

Sixth, concept cells tend to *fire to personally relevant images* (Viskontas, Quian Quiroga, & Fried, 2009), namely, to the concepts that are relevant enough to form new memories. In particular, the largest probability of responses was to pictures of the experimenters performing the recordings (which the patients got to know quite well after daily interactions) and family members, followed by

pictures of celebrities, and then by pictures of not famous people (Viskontas, et al., 2009). This is actually the reason for tuning the stimulus set for each patient (mentioned above), considering their different backgrounds and interests, in order to maximize the chances of getting responses. The fact that the pictures of experimenters triggered the largest proportion of responses is interesting for 2 reasons: i) the experimenters were unknown to the patients about a couple of days before the recordings took place, thus showing that the explicit representation of concepts can be created relatively rapidly, within a day or two or even less; ii) there were more neurons firing to experimenters compared to family members (and celebrities), thus showing a preferential encoding of novel and salient persons and events. Below and in the next section we discuss further about the plasticity and stability of the representations by concept cells.

Seventh, the responses of concept cells show *repetition suppression* but they remain way above baseline even after several (up to 30) presentations of the stimulus eliciting responses (Pedreira, et al., 2010; Rey, et al., 2015). Moreover, responses were present from the very first trial in the very first experiment, thus meaning that the neuron was already encoding the specific concept before the experiment took place (Pedreira, et al., 2010). In the literature there is a long-standing debate about whether episodic memories consolidate in cortex (L. R. Squire, et al., 2004) or not, in the latter case being always dependent on the hippocampus (Moscovitch & Nadel, 1997). With our recordings we cannot assess whether there is a consolidation of episodic memories in cortex, but these results show that there is, at least partly, a relatively stable representation of concepts and their relations (the skeleton of episodic memories) in the MTL. By relatively stable we mean that the representation by concept cells in the MTL is not created *de novo* for a given task or context (seeing pictures in the hospital ward), as it was there before the experiments took place and seemed to outlast their duration - i.e. following an initial decay, after about 5 trials the response strength reached an asymptote largely above baseline. This does not imply that a given neuron will always encode the same concept. For example, it is likely that the neurons that responded to the experimenters several years ago do not do so in present times, since the patients had no interaction with the experimenters after the electrodes

were deplanted. This lack of interaction implies that these concepts have likely not been revisited and consolidated and, therefore, the corresponding neurons may be encoding something else and the episodic memories involving the experimenters are eventually forgotten.

Eighth, concept cells have a *non-topographic representation* – i.e. nearby neurons do not tend to fire to similar (or related) concepts. This is reminiscent of the non-topographic organization of place cells – cells encoding specific locations in the environment – in the rat hippocampus (Muller, Kubie, & Ranck, 1987; Redish, et al., 2001), where a random connectivity between neurons has been described in CA3 (Li, Somogyi, Ylinen, & Buzsaki, 1994), the hippocampal area that has been proposed to be involved in forming associations (Treves & Rolls, 1994). In line with this finding, it is common to find close-by neurons separated after spike sorting that respond to completely unrelated concepts (like Vladimir Putin and the Taj Majal, see Figure 2) in the human hippocampus (Quian Quiroga, et al., 2009; Quian Quiroga, et al., 2007). Noteworthy, such sparse and non-topographic representation is ideal for a fast encoding of new associations between disparate items (Marr, 1971).

Ninth, concept cells have a relatively *late response latency*, at about 300 ms after stimulus onset (Mormann, et al., 2008; Quian Quiroga, et al., 2009), much later than what would be expected for neurons involved in visual perception (Davidesco, et al., 2014; Kirchner & Thorpe, 2006; Liu, et al., 2009; Thorpe, Fize, & Marlot, 1996). Such late responses suggest that MTL neurons are activated after relatively long processing in cortical areas, likely to extract the meaning of the stimulus that will be conveyed to the MTL for memory functions. The fact that these neurons are beyond sensory processing is also supported by the fact that they fire when the subjects imagine the concept following a specific cue (Kreiman, et al., 2000), when they think of the concept to solve a task (Cerf, et al., 2010), or even when they recall it, in the absence of any external stimulation (Gelbard-Sagiv, et al., 2008).

Tenth, the spiking *response onset of concept cells is shortly preceded by an LFP deflection* upon recognition of the stimulus (Rey, Fried, & Quian Quiroga, 2014). In fact, when briefly flashing pictures at the threshold of conscious perception (R. Quian Quiroga, et al., 2008) we found a global (across all MTL

recording sites) LFP response in the theta band that was present only when the pictures were recognized. The response was non-selective, in the sense that it was present for any picture (as long as it was recognized), likely generated by projections from cortical areas –corresponding to the visual responses observed in high level visual cortex (Bansal, et al., 2012; Davidesco, et al., 2014; Jacques, et al., (in press); Liu, et al., 2009). Based on this evidence, we have postulated that the theta LFP responses, generated in cortex, open a time window for concept cells in the MTL to fire consistently at a relatively late latency – much later than what would be expected from direct feedforward activations from visual cortex (Thorpe & Fabre-Thorpe, 2001) – to synchronize information from different sensory modalities (Rey, et al., 2014). At about the same latency of the spiking responses and only for recognized trials, there was also a more localized and selective single trial power increase in the high gamma band, which likely reflects the activation of the assembly encoding the particular concept presented.

Summarizing, concept cells provide an invariant, explicit and sparse representation of the meaning of consciously perceived stimuli, which, as argued in the next section, is critical for memory functions. The role of concept cells in memory is further supported by the fact that: i) these neurons have a relatively late response onset, in line with cognitive and not perceptual functions, ii) they encode personally relevant stimuli, that means, the ones that are likely to be remembered and iii) they can rapidly encode new associations (which is the basis of memory formation), supported by a very sparse representation that lacks topographic organization.

Concept cells in memory formation and recall

Partially overlapping assemblies

In line with the large bulk of evidence demonstrating the role of the MTL in declarative memory (Moscovitch, et al., 2005; L. Squire & Zola-Morgan, 1991; L. R. Squire, et al., 2004), we have postulated that concept cells are critically involved in the formation and recall of episodic memories by encoding associations between concepts (Quiñones Quiroga, 2012b). The representation by concept cells seems indeed optimal for this function, as we tend to remember

concepts and forget myriads of irrelevant details, and also considering that their sparse coding is ideal for the fast formation of associations (Marr, 1971). In the following, I will present a very simple mechanistic model of how associations can be encoded by these neurons based on the findings described in the previous section.

We have already mentioned that neurons in IT cortex should not be considered grandmother cells. Concept cells neither. First, if we do find a neuron firing to a particular concept, there have to be more, because the chance of hitting the one and only neuron encoding the concept among about billion neurons in the MTL (besides many other neurons firing to the concept in a more implicit way in cortex) is way too small. So, in the MTL each concept is encoded by an assembly of concept cells (Waydo, Kraskov, Quian Quiroga, Fried, & Koch, 2006) that fire together whenever the subject is aware of the specific concept. Note that we can talk about an assembly encoding a concept because the representation is sparse. In a distributed representation, we cannot identify a group of neurons firing to a specific concept because each neuron fires to many different things. Second, although it is in principle possible that an assembly of concept cells could fire to one and only one concept, this is impossible to prove (we cannot test every possible concept) and in many cases we have found concept cells that responded to more than one person or object – though when this was the case, the concepts to which the neuron fired were relatively very few and they tended to be related (Quian Quiroga & Kreiman, 2010).

Taking these two pieces of evidence together, namely, that concepts are encoded by cell assemblies and that these neurons fire to related concepts, we can postulate that the encoding of associations in the MTL is given by partially overlapping assemblies. This idea is illustrated in Figure 6, where there is a cell assembly encoding the concept of Luke Skywalker and another one encoding Yoda. However, since these two concepts are related, the assemblies are partially overlapping – i.e. some neurons fire to both concepts (as it was the case for the neuron in Figure 4). The establishment of associations between concepts can be very rapid, even in one trial (Ison, et al., 2015), through Hebbian synaptic plasticity (Hebb, 1949). In fact, there are many instances in which the two

related concepts – Luke and Yoda – appear (or are recalled) together and both assemblies coactivate, thus generating the overlap.

The model of partially overlapping assemblies gives a simple and effective mechanism to rapidly encode associations. It resembles models of sequence coding in the rodent hippocampus – neurons encoding the sequence of places the rodent visits in a trajectory – which has actually been proposed as a model of episodic memory (Buzsaki, 2005, 2006; Eichenbaum, Dudchenko, Wood, Shapiro, & Tanila, 1999; Hasselmo, 2012; Levy, 1996; Lisman, 1999). Compared to these works, the main difference is that the model of partially overlapping assemblies proposed here does not necessarily code all the aspects and sequence of events that determine a memory trace (as a sequence of place cell activations can encode all the places visited during a run). Instead, it relies on the representation of relatively few and highly specific concepts – those that are relevant enough to be remembered – and the associations made between them. This constitutes only the skeleton of episodic memories (e.g. remembering seeing *Arnold* at *Venice Beach*), which is enriched by pointers to cortical areas encoding salient details of the memory trace, as well as related semantic information based on previous abstractions and categorizations (e.g. the fact that *Arnold* is an Austrian actor, that *Venice Beach* is in California, etc).

A similar mechanism of associations via partially overlapping assemblies can explain how we can quickly retrieve related episodic information and the flow of consciousness, namely, going from one concept to the other. For example, seeing a picture of Luke Skywalker will elicit the firing of Luke’s MTL assembly, which will also activate part and eventually the whole of Yoda’s assembly through pattern completion, and so on. Again, details related to Luke Skywalker and Yoda (e.g. the features of their faces, the sound of their voices, the fact that the first is human and the second is not, etc.) are not encoded in the hippocampus but in cortex, so the flow of consciousness involves activating sequences of hippocampal concept cell assemblies and their linked cortical representations.

Associations and unitization

It is important to stress that the overlap between two different concept cell assemblies should be partial, because otherwise the concepts would not be separable anymore. In other words, there would be no way to distinguish between Luke and Yoda from the firing of MTL neurons if these neurons fire equally to both concepts. Much larger overlaps may give a mechanism for integrating different versions of a stimulus into a unified concept, for example, to link a face with a name. As described before, MTL neurons show visual and multimodal invariance, firing to completely different pictures of the same person and to its spoken and written name. One could then postulate that, in a first instance, the different pictures of Luke may trigger the firing of an assembly of MTL neurons and that his name may trigger the firing of a different set. As these two events, seeing his face and hearing (or recalling) his name, will tend to occur together, there will be a large enough overlap for them to fuse into a single concept, which can then be evoked by the picture or the name through pattern completion. In the case of two different but associated concepts, their co-occurrence will not be as high (when we see Luke, we may see or think of Yoda, but also Darth Vader, Hans Solo, etc.), and the degree of the assembly overlap will be smaller. This way, when a concept is triggered, different concepts associated to it will be only partially active and will compete to reach awareness.

Do we have enough neurons?

In the next sections we discuss some obvious challenges to the model. We can first ask if it is possible to store all possible concepts and situations we may be aware of in the MTL. This problem is aggravated if we consider that associations may be only established in the hippocampus and particularly in area CA3 – as proposed by modeling works (Treves & Rolls, 1994) – which in humans has only a few million neurons (West, Slomianka, & Gundersen, 1991). A possible solution is given by the standard consolidation model, which postulates that memories first stored in the MTL consolidate in cortex (L. R. Squire, et al., 2004). This way, hippocampal neurons will be flexible enough to change their tuning to encode new memories (or reconsolidate previous ones) and a stable representation will be only present in cortex. However, we have seen that

concept cells fire selectively to their preferred stimulus from the very first trial of the very first experiment and they reach an asymptotic response strength that remains way above baseline levels even after 30 presentations (Pedreira, et al., 2010; Rey, et al., 2015). Furthermore, several studies have shown that the MTL is always necessary for the encoding of episodic memories (Moscovitch & Nadel, 1997; Moscovitch, et al., 2005; Steinvorth, et al., 2005), thus suggesting that there should be a more stable representation of episodic memories in the MTL. I argue that this representation is given by concept cells, perhaps even with only a few million neurons. This is feasible because: First, concept cells show visual invariance, in the sense that they respond to different pictures of the same person or object – i.e. they do not represent differentially all possible views of the same concept or all possible instances in which a concept may be involved. Second, concept cells just code the sketch of an episodic memory and point to more detailed representations in cortex. Third, concept cells fire to personally relevant things. So, we do not need to encode all possible concepts we may be able to recognize but only those that are relevant enough to potentially create new memories. Fourth, as discussed in the first sections, we actually remember relatively little information and the rest is constructed based on assumptions from previous experiences.

How can concept cell assemblies be formed?

We have shown that concept cell representations can be created relatively quickly: in many cases we found that these neurons fired to different pictures and even the written or spoken names of researchers performing experiments (who were previously unknown to the patients). Furthermore, there was a larger probability of getting responses to familiar persons, but the largest proportion of responses was to the pictures of the experimenters, even larger than the one to pictures of well-known family members (Viskontas, et al., 2009). In line with this observation, other studies have described that the presentation of unknown pictures trigger non-selective responses in about 20% of MTL neurons in the first trials (Rutishauser, et al., 2006; Viskontas, et al., 2006). Therefore, we can postulate that, on the one hand, well known concepts trigger the firing of specific cell assemblies, even from the first presentation

(Pedreira, et al., 2010), which recruit interneurons that inhibit the firing of other assemblies (Ison, et al., 2011). On the other hand, new concepts (e.g. a new person that we meet) would initially trigger the firing of large number of neurons, which do not yet form a specific assembly recruiting interneurons to compete against other representations. These initial and relatively large assemblies will be non-selective and different neurons of these assemblies will gradually narrow their tuning to encode one (or a few related) concept/s that are consolidated through sensory stimulation or recall. Of course, this idea should be tested experimentally in future works, but it is worth noting that a similar model of an initial non-selective activation with relatively little inhibition, followed by a more narrower tuning as an environment becomes familiar, has been proposed to explain the formation of place cells in the rodent hippocampus (Karlsson & Frank, 2008; Wilson & McNaughton, 1993).

We could further argue that after the initial generation of an assembly representing a concept, new associations will be created as the concept is revisited, thus enlarging and further consolidating it, and, at the same time, slowly establishing related semantic information in cortex. This mechanism leads to the creation of relatively stable representations, but which are always dynamic because the assemblies will shrink and eventually disappear if their corresponding concepts are not reconsolidated –i.e. the information related to concepts that lose relevance and are not revisited will be forgotten.

Contrasting cortical and MTL neuronal representations

In the previous sections we have argued that along the areas involved in processing sensory stimuli there is an increase of invariance and selectivity to complex features. High level visual areas project to the MTL, where this process continues further and reaches its pinnacle in the hippocampus, with neurons firing selectively to different pictures and even the names of specific individuals or objects. Although the general process of increasing invariance and selectivity to complex representations is common in the cortex and the MTL, there are major differences in the type of encoding by these areas to subservise different functions. In particular, cortical visual areas have a topographically organized representation, whereas the hippocampus contains area CA3, which has a

random connectivity between neurons and acts as an autoassociative network that is able to rapidly create links between any given representations (Treves & Rolls, 1994). Furthermore, high-level areas along the VVP have a relatively distributed coding of the stimuli that is ideal for robust recognition and the storage of semantic information, whereas the representation in the MTL is much sparser and invariant, which is ideal for episodic memory (see below). So, while the high level visual cortex in humans show initial differential activations separating different stimuli at about 100 ms (Bansal, et al., 2012; Davidesco, et al., 2014; Jacques, et al., (in press); Liu, et al., 2009), the process of extracting the meaning attributed to the stimuli occurs within less than 300 ms (Cichy, et al., 2014; Clarke, et al., 2013; Clarke, et al., 2011), and after about 300 ms these activations are conveyed to MTL neurons, which have an explicit representation of concepts to form and recall associations between them (and also link to cortical representations). The cortical and MTL representations are clearly related and we have suggested a mechanism by which LFPs generated by activations upon conscious perception in cortex may open a time window for the processing of information in the MTL (Rey, et al., 2014). Once the assemblies of concept cells are activated, this leads to the formation and recall of associations with two processes taking place in parallel: on the one hand, the sequential activation of concept cell assemblies, which gives a sketch of the flow of consciousness between associated concepts and, on the other hand, pointers from these assemblies to related cortical representations, which enrich these memories with details (e.g. the features of a face) and related semantic information.

The distinction between distributed representations in cortex for recognition and sparse representations in MTL for memory fits well with evidence from modeling studies. On the one hand, the processing of visual information is distributed in nature, thus allowing parallel processing, generalizations and robustness to different viewing conditions (McClelland & Rogers, 2003). On the other hand, the information for episodic memory should be encoded in a much sparser and explicit manner, to allow a simple access for recall and a fast creation of associations (Marr, 1971; McClelland, et al., 1995; Norman & O'Reilly, 2003; O'Reilly & Norman, 2002). Besides technical details,

the main reason is very simple. Imagine having a distributed (and implicit) representation of episodic memories. Then, meeting Arnold at Venice Beach would imply activating a relatively large number of neurons coding for minute features of Arnold and another large set of neurons coding for features of Venice Beach. This way, establishing an association would require creating links between the many different details of the two concepts but without mixing them up with others, as Arnold may look like another athletic actor and Venice Beach may look like another seaside place we know. Creating these links with distributed networks is very slow and leads to the mixing of memories. In contrast, making these links with sparse and explicit networks is fast and easy, as it just requires having a few neurons that start firing to both concepts, thus creating a partially overlapping representation. Moreover, adding new concepts with a sparse representation is relatively straightforward as it doesn't impact much on the rest of the network, whereas it involves shifting boundaries for the entire network with a distributed representation, and thus the need of slow consolidation of new memories in cortex to avoid catastrophic interferences (McClelland, et al., 1995).

Conclusions and open questions

In this review I have described and contrasted basic principles of neural coding in cortex and the MTL, arguing that while distributed and organized representations in cortex are ideal for perception and storage of semantic information, the non-organized and sparse representations in the MTL are ideal for the formation and recall of episodic memories. Moreover, I have described a model of how associations can be rapidly created in the MTL based on partially overlapping assemblies. Such associations constitute the skeleton of episodic memories, originally described as the recall of personal experiences or, more colloquially, a "mental time travel" (Tulving, 2002). Within this context, it is tempting to speculate to what extent the model of partially overlapping assemblies can explain episodic memory. One possibility is that associations in the MTL point to more detailed representations in cortex and the recall of personal experiences would then be given by an interplay between sketched activations in MTL and richer cortical representations in different sensory areas.

Another non-exclusive possibility is to consider that, as with visual perception, episodic memory is basically a construction that relies only in few associations: we just remember meeting a person in a given place, talking about a particular subject, and the rest we infer based on assumptions. For example, it has been argued that we actually don't remember time as much as we may think we do, and that the idea of time, which is key in episodic memory, can be based on assumptions (Friedman, 2004). My personal view is that the answer might be a combination of both possibilities, namely, that what we call episodic memory relies in few associations between concepts in MTL and a few pointers to details stored in cortex, and the rest we make it up, thus the occurrence of false memories. Within this interplay between the MTL and cortex, it would be interesting to establish the type of associations that underlie both episodic and semantic memory, perhaps in the first case relying more on arbitrary and quickly learned associations in the MTL (as shown in (Ison, et al., 2015)), and in the latter case relying more on slowly changing, topographic and semantically organized relationships in cortex. These are, of course, very speculative statements that should be addressed with further experiments, but a main confounding factor is the fact that there is not always a clear-cut distinction between episodic and semantic relationships, as these can be intermixed (e.g. I may know that Luke and Yoda are two Jedis but may also remember the event of seeing them in Star Wars).

Another intriguing issue is the apparent dissociation between the function of the human and the rodent hippocampus: while the human hippocampus has been mainly related to declarative memory (L. Squire & Zola-Morgan, 1991), the rodent hippocampus has been linked to spatial representations and navigation (O'Keefe & Dostrovsky, 1971; O'Keefe & Nadel, 1978). A possible link between both theories might be to realize that, on the one hand, specific places (encoded by place cells) can be considered concepts that are behaviorally critical for rodents, whereas persons and objects are more salient for humans. On the other hand, spatial representations play an important role in episodic memory, especially to avoid interferences between different events: for example, I may remember where I was standing in a conference when having a conversation with a colleague and this spatial layout may help me

distinguish this conversation from another one that I had later on with another colleague in a different place. Within the model of partially overlapping assemblies, spatial layouts may be considered concepts that are associated with specific events, and which, associated with other related concepts, may create the “context” in which episodic memories are embedded. In other words, the context of memories might be just given by a set of associations. As before, these are speculations that cry out for experimental validation or refutation, on the one hand, studying the influence of spatial context in the encoding of concepts in the human hippocampus and, on the other hand, seeking for memory-related activations in the rodent hippocampus that go beyond spatial representations.

Comparative experiments may lay the ground to understand commonalities of the function of the hippocampus and eventual differences across species. We may then assess quantitatively if the representation by concept cells is uniquely human or if it is shared, and to what extent, by other animals. A final interesting observation is that while the latency of activations in the ventral visual pathway seem to be comparable in monkeys and humans (at about 100 ms), there is a large difference between the latency of activations observed in the monkey hippocampus (at about 150-200 ms; something that could in principle be explained by direct feedforward activations from cortex) (Jutras & Buffalo, 2010; Rolls, Cahusac, Feigenbaum, & Miyashita, 1993; Rolls, Xiang, & Franco, 2005; Sliwa, Plante, Duhamel, & Wirth, in press), and the ones observed in the human hippocampus (at about 300 ms), a delay that might result from further cortical processing in humans to select information and extract a more elaborated meaning, combining different sensory modalities.

Acknowledgements

I thank Hernan Rey, Joaquin Navajas and Theofanis Panagiotaropoulos for useful comments on earlier version of the manuscript and the Human Frontiers Science Program for funding. I also thank Bruno Rossion (the action editor) and an anonymous reviewer for very useful and detailed feedback that helped improving this contribution.

Figure captions

Figure 1: A) Sketch of the (ventral) visual perception and memory pathways. Neurons in V1, the first cortical visual processing area, respond to local orientations (in this case, a vertical bar). This information is further processed along the ventral visual pathway, ending in the inferotemporal cortex (IT), where neurons were found to respond selectively to faces. IT has numerous connections to the medial temporal lobe (MTL) - the hippocampus and surrounding cortex - where neurons were found to respond selectively to specific persons or objects, in this case, football player Diego Maradona. B) Hierarchical processing of information in the MTL. The connectivity within MTL areas and with visual and auditory cortex is marked with grey and orange arrows, respectively. Right inset in Fig. 1A adapted from (Hubel & Wiesel, 1959), bottom left inset from (Gross, 2008) and center inset from (Quian Quiroga, 2012a). Figure 1B adapted from (Quian Quiroga, et al., 2009)

Figure 2: A) Data from one subject showing the double staircase procedure implemented to determine the noise level that gave a 50% probability of seeing the face. Starting from a 0% and 100% noise level respectively, the noise level was increased after a 'seen' trial and decreased after an 'unseen' one. Trials corresponding to each staircase were interleaved. B) Grand average responses (N =22 subjects) for the 'seen' (blue) and 'unseen' (red) faces from a scalp recording in the right occipitotemporal cortex (electrode PO8). Note the larger response upon seeing the faces in spite of the fact that the visual stimuli was the same in both conditions. Shaded areas indicate SEM. Adapted from (Navajas, et al., 2013).

Figure 3: A) Electrodes used for single cell recordings in humans. The macroelectrodes record the intracranial EEG used to localize the epileptic activity. The microelectrodes protruding from the tip of the depth electrode allow the recording of LFPs and spiking activity. B) Recording from a microelectrode in the hippocampus of one patient. The upper plot shows the continuous (high pass filtered) data and the threshold for spike detection. The

lower plots show all detected spikes (leftmost panel) and the spikes separated according to their shapes into 5 different clusters (corresponding to 1 multiunit –in blue– and 4 putative single units). C) Responses from all detected spikes (top) and from clusters 3 and 5 after spike sorting. Note that the very selective responses to president Putin (cluster 3) and the Taj Majal (cluster 5) cannot be observed with the detected spikes, before optimal spike sorting. In all rasters the first trial is plotted on top and time zero corresponds to stimulus onset. Adapted from (Rey, et al., 2015).

Figure 4: A) Responses of a single unit in the entorhinal cortex that fired selectively to different pictures of Luke Skywalker, to his name written in the computer screen (stimulus 58) and pronounced by a male (stimulus 71) and a female voice (stimulus 72). The neuron also fired to a picture of Yoda (stimulus 63; only 1 picture of Yoda was shown), another character from the movie ‘Star Wars’. For space reasons, only 20 out of 76 responses are shown, but they were no significant responses to any of the other stimuli. B) Median number of spikes for all stimuli presented in the recorded session. Bars in red correspond to the presentations of Luke Skywalker and the horizontal bar denotes 5 s.d. above baseline firing of the neuron. Adapted from (Quiñ Quiroga, et al., 2009).

Figure 5: A) A single unit in the hippocampus that responded to Jodie Foster (100% B) and not to Nicole Kidman (100% A). The middle panels show the responses to the presentation of an ambiguous morph between Foster and Kidman, separated according to the subject’s recognition. The response to the morphed picture was larger when the subject recognized it as Foster (Decision B) compared to when he recognized it as Kidman (Decision A). In each case (Decision A and Decision B) the responses to the morph picture were similar to the ones obtained with the original pictures without morphing (100% A and 100% B). Adapted from (Quiñ Quiroga, et al., 2014)

Figure 6: Partially overlapping assemblies. On the left, a hypothetical cell assembly encoding the concept “Luke Skywalker” (in orange) and, on the right, another one encoding the concept “Yoda” (in blue). Some of these neurons

(marked in orange and blue) fire to Luke and Yoda, thus encoding the association between the two concepts through a partial overlap of the corresponding assemblies.

References

- Abbott, L. F., Rolls, E. T., & Tovee, M. J. (1996). Representational capacity of face coding in monkeys. *Cerebral Cortex*, 6, 498-505.
- Allison, T., McCarthy, G., Nobre, A., Puce, A., & Belger, A. (1994). Human extrastriate visual cortex and the perception of faces, words, numbers, and colors. *Cerebral Cortex*, 5, 544-554.
- Allison, T., Puce, A., Spencer, D., & McCarthy, G. (1999). Electrophysiological studies of human face perception. I: Potentials generated in occipitotemporal cortex by face and non-face stimuli. *Cerebral Cortex*, 9, 415-430.
- Aristotle. (reprinted 2004). *De anima*. London: Penguin, (Chapter Chapter).
- Babb, T. L., Carr, E., & Crandall, P. H. (1973). Analysis of extracellular firing patterns of deep temporal lobe structures in man. *Electroencephalogr Clin Neurophysiol*, 34, 247-257.
- Bacon-Mace, N., Mace, M., Fabre-Thorpe, M., & Thorpe, S. (2005). The time course of visual processing: backward masking and natural scene categorisation. *Vision Research*, 45, 1459-1469.
- Baddeley, A., Eysenck, M. W., & Anderson, M. C. (2009). *Memory*. New York: Psychology Press, (Chapter Chapter).
- Baldassi, C., Alemi-Neissi, A., Pagan, M., Di Carlo, J. J., Zecchina, R., & Zoccolan, D. (2013). Shape similarity, better than semantic membership, accounts for the structure of visual object representations in a population of monkey inferotemporal neurons. *PLOS Computational Biology*, 9, 1-21.
- Bansal, A., Singer, J., Anderson, W., Golby, A., Madsen, J., & Kreiman, G. (2012). Temporal stability of visually selective responses in intracranial field potentials recorded from human occipital and temporal lobes. *Journal of Neurophysiology*, 108, 3073-3086.
- Barlow, H. B. (1981). The Ferrier lecture 1980: Critical limiting factors in the design of the eye and visual cortex. *Proceedings of the Royal Society of London B*, 212.
- Barlow, H. B. (1994). The neuron doctrine in perception. In M. Gazzaniga (Ed.), *The Cognitive Neurosciences*. Boston: MIT Press.
- Bartlett, F. C. (1932). *Remembering*. Cambridge: Cambridge University Press, (Chapter Chapter).
- Bell, A., Hadj-Bouziane, F., Frihauf, J., Tootell, R. B., & Ungerleider, L. G. (2009). Object representation in the temporal cortex of monkeys and humans as revealed by functional magnetic resonance imaging. *Journal of Neurophysiology*, 101, 688-700.
- Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience*, 8, 551-565.
- Bentin, S., Sagiv, N., Mecklinger, A., Friederici, A., & von Cramon, D. Y. (2002). Priming visual face-processing mechanisms: Electrophysiological evidence. *Psychological Science*, 13, 190-193.
- Blake, R., & Logothetis, N. K. (2002). Visual competition. *Nature Reviews Neuroscience*, 3, 13-21.
- Bowers, J. (2009). On the biological plausibility of grandmother cells: Implications for neural network theories in psychology and neuroscience. *Psychological Review*, 116., 220-251.

- Bukach, C. M., Gauthier, I., & Tarr, M. J. (2006). Beyond faces and modularity: the power of an expertise framework. *Trends in Cognitive Sciences*, *10*, 159-166.
- Buzsaki, G. (2005). Theta rhythm of navigation: Link between path integration and landmark navigation, episodic and semantic memory. *Hippocampus*, *15*, 827-840.
- Buzsaki, G. (2006). *Rhythms of the brain*. Oxford: Oxford University Press, (Chapter Chapter).
- Cameron, K. A., Yashar, S., Wilson, C. L., & Fried, I. (2001). Human hippocampal neurons predict how well word pairs will be remembered. *Neuron*, *30*, 289-298.
- Carlson, T., Simmons, R., Kriegeskorte, N., & Slevc, L. (2013). The emergence of semantic meaning in the ventral temporal pathway. *Journal of Cognitive Neuroscience*, *26*, 120-131.
- Cerf, M., Thiruvengadam, N., Mormann, F., Kraskov, A., Quiñero, R., Koch, C., & Fried, I. (2010). On-line, voluntary control of human temporal lobe neurons. *Nature*, *467*, 1104-1108.
- Cichy, R., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. *Nature Neuroscience*, *17*, 455-462.
- Clarke, A., Taylor, K., Debereux, B., Randall, B., & Tyler, L. (2013). From perception to conception: How meaningful objects are processed over time. *Cerebral Cortex*, *23*, 187-197.
- Clarke, A., Taylor, K., & Tyler, L. (2011). The evolution of meaning: Spatio-temporal dynamics of visual object recognition. *Journal of Cognitive Neuroscience*, *23*, 1887-1899.
- Corkin, S. (2002). What's new with the amnesic patient H.M.? *Nature Reviews Neuroscience*, *3*, 153-160.
- Cowell, R. A., Bussey, T. J., & Saksida, L. M. (2010). Functional dissociations within the ventral object processing pathway: cognitive modules or a hierarchical continuum? *Journal of Cognitive Neuroscience*, *22*, 2460-2479.
- Davidesco, I., Zion-Golombic, E., Bickel, S., Harel, M., Groppe, D., Keller, C. J., Schevon, C., McKhann, G., Goodman, R., Goelman, G., Schroeder, C. E., Mehta, A. D., & Malach, R. (2014). Exemplar selectivity reflects perceptual similarities in the human fusiform cortex. *Cerebral Cortex*, *24*, 1879-1893.
- Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, *58*, 17-22.
- Dehaene, S., & Changeux, J. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, *70*, 200-227.
- Del Cul, A., Baillet, S., & Dehaene, S. (2007). Brain dynamics underlying the nonlinear threshold for access to consciousness. *PLoS Biology*, *5*, e260.
- Di Russo, F., Martinez, A., Sereno, M., Pitzalis, S., & Hillyard, S. A. (2001). Cortical sources of the early components of the visual evoked potential. *Human Brain Mapping*, *15*, 95-111.
- Eagleman, D. (2001). Visual illusions and neurobiology. *Nature Reviews Neuroscience*, *2*, 920-926.
- Eichenbaum, H., Dudchenko, P., Wood, E., Shapiro, M., & Tanila, H. (1999). The hippocampus, memory, and place cells: Is it spatial memory or a memory space? *Neuron*, *23*, 209-226.

- Engel, A. K., Moll, C. K. E., Fried, I., & Ojermann, G. A. (2005). Invasive recordings from the human brain: clinical insights and beyond. *Nature Reviews Neuroscience*, *6*, 35-47.
- Fabre-Thorpe, M. (2003). Visual categorization: accessing abstraction in non-human primates. *Phil. Trans. R. Soc. Lond. B*, *358*, 1215-1223.
- Farah, M., & Feinberg, T. (2000). Visual object agnosia. In M. F. a. T. Feinberg (Ed.), *Patient-based approaches to cognitive neuroscience*. Cambridge: MIT Press.
- Fisch, L., Privman, E., Ramot, M., Harel, M., Nir, Y., Kipervasser, S., Andelman, F., Neureld, M., Kramer, U., Fried, I., & Malach, R. (2009). Neural "ignition": Enhanced activation linked to perceptual awareness in human ventral stream visual cortex. *Neuron*, *64*, 562-574.
- Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *The Journal of Neuroscience*, *23*, 5235-5246.
- Freedman, D. J., Riessenhuber, M., Poggio, T., & Miller, E. K. (2001). Categorical Representation of Visual Stimuli in the Primate Prefrontal Cortex. *Science*, *291*, 312-316.
- Freiwald, F. A., & Tsao, D. Y. (2010). Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science*, *330*, 845-851.
- Fried, I., MacDonald, K. A., & Wilson, C. L. (1997). Single neuron activity in human hippocampus and amygdala during recognition of faces and objects. *Neuron*, *18*, 753-765.
- Friedman, W. J. (2004). Time in autobiographical memory. *Social Cognition*, *22*, 591-605.
- Fujita, I., Tanaka, K., Ito, M., & Cheng, K. (1992). Columns for visual features of objects in monkey inferotemporal cortex. *Nature*, *360*, 343-346.
- Gao, X., & Wilson, H. R. (2014). Implicit learning of geometric eigenfaces. *Vision Research*, *99*, 12-18.
- Gelbard-Sagiv, H., Mukamel, R., Harel, M., Malach, R., & Fried, I. (2008). Internally generated reactivation of single neurons in human hippocampus during free recall. *Science*, *322*, 96-101.
- Gibson, J. J. (1972). A theory of direct visual perception. In J. R. a. W. Rozenboom (Ed.), *The psychology of knowing*. New York: Gordon & Breach.
- Goodale, M., & Milner, D. (1992). Separate visual pathways for perception and action. *Trends in Neuroscience*, *15*, 20-25.
- Graham, K. S., Simons, J. S., Pratt, K. H., Patterson, K., & Hodges, J. R. (2000). Insights from semantic dementia on the relationship between episodic and semantic memory. *Neuropsychologia*, *38*, 313-324.
- Gregory, R. L. (1997). Knowledge in perception and illusion. *Phil. Trans. R. Soc. Lond. B*, *352*, 1121-1128.
- Grill-Spector, K., & Malach, R. (2004). The Human Visual Cortex. *Annual Review of Neuroscience*, *27*, 649-677.
- Gross, C. (1992). Representation of visual stimuli in inferior temporal cortex. *Phil. Trans. R. Soc. Lond. B*, *335*, 3-10.
- Gross, C. (2002). Genealogy of the "Grandmother Cell". *The Neuroscientist*, *8*, 512-518.

- Gross, C. (2008). Single neuron studies of inferior temporal cortex. *Neuropsychologia*, *46*, 841-852.
- Halgren, E., Baudena, P., Heit, G., Clarke, J. M., Marinkovic, K., & Clarke, M. (1994). Spatio-temporal stages in face and word processing. I. Depth-recorded potentials in the human occipital, temporal and parietal lobes. *Journal of Physiology*, *88*, 1-50.
- Harris, J. A., Wu, C., & Woldorff, M. (2011). Sandwich masking eliminates both visual awareness of faces and face-specific brain activity through a feed-forward mechanism. *Journal of Vision*, *11*, 1-12.
- Hassabis, D., Kumaran, D., Vann, S., & Maguire, E. A. (2007). Patients with hippocampal amnesia cannot imagine new experiences. *Proceedings of the National Academy of Sciences*, *104*, 1726-1731.
- Hassabis, D., & Maguire, E. A. (2007). Deconstructing episodic memory with construction. *Trends in Cognitive Science*, *11*, 299-306.
- Hasselmo, M. E. (2012). *How do we remember - Brain mechanism of episodic memory*. Cambridge, MA: MIT Press, (Chapter Chapter).
- Hebb, D. O. (1949). *The organization of behavior*. NY: John Wiley & Sons, (Chapter Chapter).
- Heit, G., Smith, M. E., & Halgren, E. (1988). Neural encoding of individual words and faces by the human hippocampus and amygdala. *Nature*, *333*, 773-775.
- Heit, G., Smith, M. E., & Halgren, E. (1990). Neuronal activity in the human medial temporal lobe during recognition memory. *Brain*, *113*, 1093-1112.
- Helmholtz, H. (1878). The facts of perception. In R. Kahl (Ed.), *Selected writings of Hermann von Helmholtz*. Middletown Conn.: Wesleyan University Press.
- Hodges, J. R., & Graham, K. S. (2001). Episodic memory: insights from semantic dementia. *Phil. Trans. R. Soc. Lond. B*, *356*, 1423-1434.
- Hubel, D., & Wiesel, T. N. (1959). Receptive Fields of Single Neurones in the Cat's Striate Cortex. *Journal of Physiology*, *148*, 574-591.
- Hung, C., Kreiman, G., Poggio, T., & DiCarlo, J. (2005). Fast read-out of object information in inferior temporal cortex. *Science*, *310*, 863-866.
- Ison, M., Mormann, F., Cerf, M., Koch, C., Fried, I., & Quiñan Quiroga, R. (2011). Selectivity of pyramidal cells and interneurons in the Human Medial Temporal Lobe. *Journal of Neurophysiology*, (in press).
- Ison, M., Quiñan Quiroga, R., & Fried, I. (2015). Rapid encoding of new memories by individual neurons in the human brain. *Neuron*, *87*, 220-230.
- Jacques, C., d'Arripe, O., & Rossion, B. (2007). The time course of the inversion effect during individual face discrimination. *Journal of Vision*, *7*, 1-9.
- Jacques, C., Witthoft, N., Weiner, K. S., Foster, B. L., Rangarajan, V., Hermes, D., Miller, K. J., Parvizi, J., & Grill-Spector, K. ((in press)). Corresponding ECoG and fMRI category-selective signals in human ventral temporal cortex. *Neuroimage*.
- James, W. (1890). *The principles of psychology*. New York: Cosmo classics, (Chapter Chapter).
- Jemel, B., Schuller, A., Cheref-Khan, Y., Goffaux, V., Crommelinck, M., & Bruyer, R. (2003). Stepwise emergence of the face-sensitive N170 event-related potential component. *NeuroReport*, *14*, 2035-2039.
- Jonas, J., Descoins, M., Koessler, L., Colnat-Coulbois, S., Sauvee, M., Guye, M., Vignal, J.-P., Vespignani, H., Rossion, B., & Maillard, L. (2012). Focal

- electrical intracerebral stimulation of a face-sensitive area causes transient prosopagnosia. *Neuroscience*, 222, 281-288.
- Jonas, J., Rossion, B., Krieg, J., Koessler, L., Colnat-Coulbois, S., Verpignani, H., Jacques, C., Vignal, J.-P., Brissart, H., & Maillard, L. (2014). Intracerebral electrical stimulation of a face-selective area in the right inferior occipital cortex impairs individual face discrimination. *Neuroimage*, 99, 487-497.
- Jutras, M., & Buffalo, E. A. (2010). Recognition memory signals in the macaque hippocampus. *Proceedings of the National Academy of Sciences*, 107, 401-406.
- Kanwisher, N. (2001). Neural events and perceptual awareness. *Cognition*, 79, 89-113.
- Kanwisher, N. (2010). Functional specificity in the human brain: A window into the functional architecture of the mind. *Proceedings of the National Academy of Sciences*, 107, 11163-11170.
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci*, 17, 4302-4311.
- Karlsson, M., & Frank, L. M. (2008). Network dynamics underlying the formation of sparse, informative representations in the hippocampus. *The Journal of Neuroscience*, 24, 14271-14281.
- Kiani, R., Esteky, H., Mirpour, K., & Tanaka, K. (2007). Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *Journal of Neurophysiology*, 97, 4296-4309.
- Kirchner, H., & Thorpe, S. (2006). Ultra-rapid object detection with saccadic eye-movements: visual processing speed revisited. *Vision Research*, 46, 1762-1776.
- Koch, K., McLean, J., Segev, R., Freed, M., Berry, M., Balasubramanian, V., & Sterling, P. (2006). How Much the Eye Tells the Brain. *Current Biology*, 16, 1428-1434.
- Konkle, T., & Oliva, A. (2012). A real-world size organization of object responses in occipitotemporal cortex. *Neuron*, 74, 1114-1124.
- Konoski, J. (1967). *Integrative activity of the brain: An interdisciplinary approach.*: University of Chicago Press, (Chapter Chapter).
- Koriat, A., Goldsmith, M., & Pansky, A. (2000). Toward a psychology of memory accuracy. *Annual Review of Psychology*, 51, 481-537.
- Kravitz, D., Saleem, K., Baker, C., Ungerleider, L., & Mishkin, M. (2013). The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends in Cognitive Science*, 17, 26-49.
- Kreiman, G., Fried, I., & Koch, C. (2002). Single-neuron correlates of subjective vision in the human medial temporal lobe. *Proceedings of the National Academy of Sciences*, 99, 8378-8383.
- Kreiman, G., Hung, C. P., Kraskov, A., Quiñones Quiroga, R., Poggio, T., & DiCarlo, J. J. (2006). Object selectivity of local field potentials and spikes in the macaque inferior temporal cortex. *Neuron*, 49, 433-445.
- Kreiman, G., Koch, C., & Fried, I. (2000). Imagery neurons in the human brain. *Nature*, 408, 357-361.
- Kriegeskorte, N., Mur, M., Ruff, D., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., & Bandettini, P. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, 60, 1126-1141.

- Lavenex, P., & Amaral, D. G. (2000). Hippocampal-Neocortical interaction: a hierarchy of associativity. *Hippocampus*, *10*, 420-430.
- Leopold, D. A., Rhodes, G., Mueller, K.-M., & Jeffery, L. (2005). The dynamics of visual adaptation to faces. *Proceedings of the Royal Society B*, *272*, 897-904.
- Levy, W. B. (1996). A sequence predicting CA3 is a flexible associator that learns and uses context to solve hippocampal-like tasks. *Hippocampus*, *6*, 579-590.
- Li, X.-G., Somogyi, P., Ylinen, A., & Buzsaki, G. (1994). The hippocampal CA3 network: an in vivo intracellular labeling study. *The Journal of Comparative Neurology*, *339*, 181-208.
- Lisman, J. E. (1999). Relating hippocampal circuitry to function: recall of memory sequences by reciprocal dentate-CA3 interactions. *Neuron*, *22*, 233-242.
- Liu, H., Agam, Y., Madsen, J., & Kreiman, G. (2009). Timing, timing, timing: Fast decoding of object information from intracranial field potentials in human visual cortex. *Neuron*, *62*, 281-290.
- Loftus, E., & Palmer, J. (1974). Reconstruction of automobile destruction: an example of interaction between language and memory. *Journal of verbal learning and verbal behavior*, *13*, 585-589.
- Logothetis, N. K. (1998). Single units and conscious vision. *Phil. Trans. R. Soc. Lond. B*, *353*, 1801-1818.
- Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annu Rev Neurosci*, *19*, 577-621.
- Luck, S. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press, (Chapter Chapter).
- Marr, D. (1971). Simple memory: A theory for archicortex. *Proc. Royal Soc. London, B*, *262*, 23-81.
- McCarthy, G., Puce, A., Belger, A., & Allison, T. (1999). Electrophysiological studies of human face perception. II: Response properties of face-specific potentials generated in occipitotemporal cortex. *Cerebral Cortex*, *9*, 431-444.
- McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*, 419-457.
- McClelland, J. L., & Rogers, T. T. (2003). The parallel distributed processing approach to semantic cognition. *Nature Reviews Neuroscience*, *4*, 310-322.
- Milner, B., Corkin, S., & Teuber, H. (1968). Further analysis of the hippocampal amnesic syndrome: 14-years follow-up study of H.M. *Neuropsychologia*, *6*, 215-234.
- Mishkin, M. (1978). Memory in monkeys severely impaired by combined but not separate removal of the amygdala and hippocampus. *Nature*, *273*, 297-298.
- Miyashita, Y. (1993). Inferior temporal cortex: Where visual perception meets memory. *Annual Review of Neuroscience*, *16*, 245-263.
- Mormann, F., Kornblith, S., Quiñero, R., Kraskov, A., Cerf, M., Fried, I., & Koch, C. (2008). Latency and Selectivity of Single Neurons Indicate Hierarchical Processing in the Human Medial Temporal Lobe. *Journal of Neuroscience*, *28*, 8865-8872.

- Moscovitch, M., & Nadel, L. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Current Opinion in Neurobiology*, 7, 217-227.
- Moscovitch, M., Rosenbaum, R. S., Gillboa, A., Addis, D. R., Westmacott, R., Grady, C., McAndrews, M. P., Levine, B., Black, S., Winocur, G., & Nadel, L. (2005). Functional neuroanatomy of remote episodic, semantic and spatial memory: a unified account based on multiple trace theory. *Journal of Anatomy*, 207, 35-66.
- Muller, R. U., Kubie, J. L., & Ranck, J. B. (1987). Spatial firing patterns of hippocampal complex-spike cells in a fixed environment. *The Journal of Neuroscience*, 7, 1935-1950.
- Navajas, J., Ahmadi, M., & Quian Quiroga, R. (2013). Uncovering the Mechanisms of Conscious Face Perception: A Single-Trial Study of the N170 Responses. *Journal of Neuroscience*, 33, 1337-1343.
- Navajas, J., Rey, H., & Quian Quiroga, R. (2014). Perceptual and contextual awareness: methodological considerations in the search for the neural correlates of consciousness. *Frontiers in Psychology*, 5, article 959.
- Naya, Y., Yoshida, M., & Miyashita, Y. (2001). Backward spreading of memory-retrieval signal in the primate temporal cortex. *Science*, 291, 661-664.
- Niedermeyer, E. (1993). Epileptic seizure disorders. In F. Lopes da Silva & E. Niedermeyer (Eds.), *Electroencephalography*. Baltimore: Williams and Wilkins.
- Norman, K., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. *Psychological Review*, 110, 611-646.
- O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map: preliminary evidence from unit activity in freely moving rats. *Brain Research*, 34, 171-175.
- O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*.: Oxford University Press, (Chapter Chapter).
- O'Reilly, R. C., & Norman, K. A. (2002). Hippocampal and neocortical contributions to memory: advances in the complementary learning systems framework. *Trends in Cognitive Science*, 6, 505-510.
- Palmer, S. E. (1999). *Vision Science*. Cambridge, Massachusetts: MIT Press, (Chapter Chapter).
- Palmeri, T. J., & Gauthier, I. (2004). Visual object understanding. *Nature Reviews Neuroscience*, 5, 291-303.
- Panagiotaropoulos, T., Kapoor, V., & Logothetis, N. K. (2014). Subjective visual perception: from local processing to emergent phenomena of brain activity. *Proceedings of the Royal Society B*, 369, 20130534.
- Pedreira, C., Mormann, F., Kraskov, A., Cerf, M., Fried, I., Koch, C., & Quian Quiroga, R. (2010). Responses of human medial temporal lobe neurons are modulated by stimulus repetition. *Journal of Neurophysiology*, 103, 97-107.
- Philiastides, M., & Sajda, P. (2006). Temporal characterization of the neural correlates of perceptual decision making in the human brain. *Cerebral Cortex*, 16, 509-518.

- Popivanov, I., Jastorff, J., Vanduffel, W., & Vogels, R. (2014). Heterogeneous Single-Unit Selectivity in an fMRI-Defined Body-Selective Patch. *Journal of Neuroscience*, *34*, 95-111.
- Puce, A., Allison, T., & McCarthy, G. (1999). Electrophysiological studies of human face perception. III: Effects of top-down processing on face-specific potentials. *Cerebral Cortex*, *9*, 445-458.
- Quian Quiroga, R. (2010). In retrospect: Funes the memorious. *Nature*, *463*, 611.
- Quian Quiroga, R. (2012a). *Borges and memory: encounters with the human brain*. Cambridge, MA: MIT Press, (Chapter Chapter).
- Quian Quiroga, R. (2012b). Concept cells: The building blocks of declarative memory functions. *Nature Reviews Neuroscience*, *13*, 587-597.
- Quian Quiroga, R. (2012c). Quick guide: Spike sorting. *Current Biology*, *22*, R45-R46.
- Quian Quiroga, R., Fried, I., & Koch, C. (2013). Brain cells for grandmother. *Scientific American*, *308*, 30-35.
- Quian Quiroga, R., Kraskov, A., Koch, C., & Fried, I. (2009). Explicit encoding of multimodal percepts by single neurons in the human brain. *Current Biology*, *19*, 1308-1313.
- Quian Quiroga, R., Kraskov, A., Mormann, F., Fried, I., & Koch, C. (2014). Single-Cell Responses to Face Adaptation in the Human Medial Temporal Lobe. *Neuron*, *84*, 363-369.
- Quian Quiroga, R., & Kreiman, G. (2010). Measuring sparseness in the brain: comment on Bowers (2009). *Psychological Review*, *117*, 291-299.
- Quian Quiroga, R., Kreiman, G., Koch, C., & Fried, I. (2008). Sparse but not 'Grandmother-cell' coding in the medial temporal lobe. *Trends Cogn Sci*, *12*, 87-91.
- Quian Quiroga, R., Mukamel, R., Isham, E. A., Malach, R., & Fried, I. (2008). Human single-neuron responses at the threshold of conscious recognition. *Proceedings of the National Academy of Sciences*, *105*, 3599-3604.
- Quian Quiroga, R., & Panzeri, S. (2009). Extracting information from neural populations: Information theory and decoding approaches. *Nature Reviews Neuroscience*, *10*, 173-185.
- Quian Quiroga, R., Reddy, L., Koch, C., & Fried, I. (2007). Decoding Visual Inputs From Multiple Neurons in the Human Temporal Lobe. *J Neurophysiol*, *98*, 1997-2007.
- Quian Quiroga, R., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*, *435*, 1102-1107.
- Reddy, L., Quian Quiroga, R., Wilken, P., Koch, C., & Fried, I. (2006). A single-neuron correlate of change detection and change blindness in the human medial temporal lobe. *Curr Biol*, *16*, 2066-2072.
- Redish, A. D., Battaglia, F. P., Chawla, M. K., Ekstrom, A. D., Gerrard, J. L., Lipa, P., Rosenzweig, E. S., Worley, P. F., Guzowski, J. F., McNaughton, B. L., & Barnes, C. A. (2001). Independence of firing correlates of anatomically proximate hippocampal pyramidal cells. *The Journal of Neuroscience*, *21*, 1-6.
- Rey, H., Fried, I., & Quian Quiroga, R. (2014). Timing of single neuron and local field potential responses in the human medial temporal lobe. *Current Biology*, *24*, 299-304.

- Rey, H., Ison, M., Pedreira, C., Valentin, A., Alarcon, G., Selway, R., Richardson, M., & Quiñero, R. (2015). Single cell recordings in the human medial temporal lobe. *Journal of Anatomy*, *227*, 394-408
- Roediger, H. L. (1996). Memory illusions. *Journal of memory and language*, *35*, 76-100.
- Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: remembering words not presented in lists. *Journal of Experimental Psychology*, *21*, 803-814.
- Roelfsema, P. R. (2006). Cortical algorithms for perceptual grouping. *Annual Review of Neuroscience*, *29*, 203-227.
- Rolls, E. T., Cahusac, P. M. B., Feigenbaum, J. D., & Miyashita, Y. (1993). Responses of single neurons in the hippocampus of the macaque related to recognition memory. *Experimental Brain Research*, *93*, 299-306.
- Rolls, E. T., & Tovee, M. J. (1995). Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *Journal of Neurophysiology*, *73*, 713-726.
- Rolls, E. T., Xiang, J., & Franco, L. (2005). Object, Space, and Object-Space Representations in the Primate Hippocampus. *Journal of Neurophysiology*, *94*, 833-844.
- Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382-439.
- Rosenbaum, R. S., Gilboa, A., Levine, B., Winocur, G., & Moscovitch, M. (2009). Amnesia as an impairment of detail generation and binding: Evidence from personal, fictional, and semantic narratives in K.C. *Neuropsychologia*, *47*, 2181-2187.
- Rosenbaum, R. S., Koeler, S., Schacter, D., Moscovitch, M., Westmacott, R., Black, S., Gao, F., & Tulving, E. (2005). The case of K.C.: contributions of a memory-impaired person to memory theory. *Neuropsychologia*, *43*, 989-1021.
- Rossion, B. (2014a). Understanding face perception by means of human electrophysiology. *Trends in Cognitive Science*, *18*, 310-318.
- Rossion, B. (2014b). Understanding face perception by means of prosopagnosia and neuroimaging. *Frontiers in Bioscience*, *6*, 308-317.
- Rutishauser, U., Mamelak, A. N., & Schuman, E. M. (2006). Single-trial learning of novel stimuli by individual neurons of the human hippocampus-amygdala complex. *Neuron*, *49*, 805-813.
- Saleem, K. S., & Tanaka, K. (1996). Divergent projections from the anterior inferotemporal area TE to the perirhinal and entorhinal cortices in the macaque monkey. *The Journal of Neuroscience*, *16*, 4757-4775.
- Sato, T., Uchida, G., Lescroart, M., Kitazono, J., Okada, M., & Tanifuji, M. (2013). Object representation in inferior temporal cortex is organized hierarchically in a mosaic-like structure. *The Journal of Neuroscience*, *33*, 16642-16656.
- Schacter, D. (2001). *The seven sins of memory: How the mind forgets and remembers*. Boston: Houghton Mifflin, (Chapter Chapter).
- Schacter, D., & Slotnick, S. (2004). The cognitive neuroscience of memory distortion. *Neuron*, *44*, 149-160.

- Scoville, W., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesion. *Journal of Neurology, Neurosurgery and Psychiatry*, *20*, 11-21.
- Sergent, J., Ohta, S., & MacDonald, B. (1992). Functional neuroanatomy of face and object processing: a positron emission tomography study. *Brain*, *115*, 15-36.
- Sigala, N., & Logothetis, N. K. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature*, *415*, 318-320.
- Sliwa, J., Plante, A., Duhamel, J. R., & Wirth, S. (in press). Independent Neuronal Representation of Facial and Vocal Identity in the Monkey Hippocampus and Inferotemporal Cortex. *Cerebral Cortex*.
- Snowden, J. S., Griffiths, H. L., & Neary, D. (1996). Semantic-Episodic memory interactions in semantic dementia: Implications for retrograde memory function. *Cognitive Neuropsychology*, *13*, 1101-1137.
- Squire, L. (2009). The legacy of patient H.M. for neuroscience. *Neuron*, *61*, 6-9.
- Squire, L., & Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science*, *253*, 1380-1386.
- Squire, L. R., Stark, C. E. L., & Clark, R. E. (2004). The medial temporal lobe. *Annual Review of Neuroscience*, *27*, 279-306.
- Steinvorth, S., Levine, B., & Corkin, S. (2005). Medial temporal lobe structures are needed to re-experience remote autobiographical memories: evidence from H.M. and W.R. *Neuropsychologia*, *43*, 479-496.
- Suzuki, W. A. (1996). Neuroanatomy of the monkey entorhinal, perirhinal and parahippocampal cortices: organization of cortical inputs and interconnections with amygdala and striatum. *Seminar Neurosciences*, *8*, 3-12.
- Tamura, H., & Tanaka, K. (2001). Visual response properties of cells in the ventral and dorsal parts of the macaque inferotemporal cortex. *Cerebral Cortex*, *11*, 384-399.
- Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review on Neuroscience*, *19*, 109-139.
- Tanaka, K. (2003). Columns for complex visual object features in the inferotemporal cortex: Clustering of cells with similar but slightly different stimulus selectivities. *Cerebral Cortex*, *13*, 90-99.
- Tanaka, K., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology*, *23*, 457-482.
- Tanskanen, T., Nasanen, R., Montez, T., Paallysaho, J., & Hari, R. (2005). Face recognition and cortical responses show similar sensitivity to noise spatial frequency. *Cerebral Cortex*, *15*, 526-534.
- Tanskanen, T., Nasanen, R., Ojanpaa, H., & Hari, R. (2007). Face recognition and cortical responses: Effect of stimulus duration. *Neuroimage*, *35*, 1636-1644.
- Thorpe, S., & Fabre-Thorpe, M. (2001). Seeking Categories in the Brain. *Science*, *291*, 260-263.
- Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*, 520-522.
- Treves, A., & Rolls, E. T. (1994). Computational analysis of the role of the hippocampus in memory. *Hippocampus*, *4*, 374-391.

- Tsao, D. Y., Freiwald, W. A., Tootell, R. B., & Livingstone, M. (2006). A cortical region consisting entirely of face-selective cells. *Science*, *311*, 670-674.
- Tsao, D. Y., & Livingstone, M. (2008). Mechanisms of face perception. *Annual Review of Neuroscience*, *31*, 411-437.
- Tsunoda, K., Yamane, Y., Nishizaki, M., & Tanifuji, M. (2001). Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. *Nature Neuroscience*, *4*, 832-838.
- Tulving, E. (2002). Episodic memory: from mind to brain. *Annual Review of Psychology*, *53*, 1-25.
- Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In M. A. G. D. J. Ingle, and R. J. W. Mansfield. (Ed.), *Analysis of Visual Behavior*. Cambridge: MIT Press.
- Vargha-Khadem, F., Gadian, D. G., Watkins, K. E., Connelly, A., van Paesschen, W., & Mishkin, M. (1997). Differential effects of early hippocampal pathology on episodic and semantic memory. *Science*, *277*, 376-380.
- Viskontas, I., Knowlton, B. J., Steinmetz, P. N., & Fried, I. (2006). Differences in mnemonic processing by neurons in the human hippocampus and parahippocampal regions. *Journal of Cognitive Neuroscience*, *18*, 1654-1662.
- Viskontas, I., Quian Quiroga, R., & Fried, I. (2009). Human medial temporal lobe neurons respond preferentially to personally-relevant images. *Proceedings of the National Academy of Sciences USA*, *106*, 21329-21334.
- Wang, G., & Tanaka, K. (1996). Optical imaging of functional organization in the monkey inferotemporal cortex. *Science*, *272*, 1665-1668.
- Waydo, S., Kraskov, A., Quian Quiroga, R., Fried, I., & Koch, C. (2006). Sparse Representation in the Human Medial Temporal Lobe. *Journal of Neuroscience*, *26*, 10232-10234.
- West, M. J., Slomianka, L., & Gundersen, H. J. G. (1991). Unbiased stereological estimation of the total number of neurons in the subdivisions of the rat hippocampus using the optical fractionator. *Anat. Rec.*, *231*, 482-497.
- Wilson, M. A., & McNaughton, B. L. (1993). Dynamics of the Hippocampal Ensemble Code for Space. *Science*, *261*, 1055-1058.
- Wolfe, J. M. (1984). Reversing ocular dominance and suppression in a single flash. *Vision Research*, *24*, 471-478.
- Young, M., & Yamane, S. (1992). Sparse population coding of faces in the inferotemporal cortex. *Science*, *256*, 1327-1331.