Neuronal codes for visual perception and memory

Rodrigo Quian Quiroga

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

ARTICLE INFO

Article history:
Received 12 June 2015
Received in revised form 8 December 2015
Accepted 17 December 2015

Keywords:
Perception
Memory
Concept cells
Neuronal coding
Memory models

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 memory based on partially overlapping assemblies.

© 2015 Elsevier Ltd. All rights reserved.

1. 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 it will eventually lead to the formation of memories about the new encounter.

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 and Malach, 2004; Logothetis and Sheinberg, 1996; Roelfsema, 2006; Tanaka, 1996; Tsao and Livingstone, 2008). From high level visual areas in the VVP, there are strong connections to the medial temporal lobe (MTL) (see Fig. 1) (Lavenex and Amaral, 2000; Saleem and Tanaka, 1996; Suzuki, 1996), where evidence from patient H.M. (Corkin, 2002; Milner et al., 1968; Scoville and Milner, 1957; Squire, 2009), other patients with lesions in the MTL (Moscovitch et al., 2005; Rosenbaum et al., 2005; Squire et al., 2004) and animal studies (Mishkin, 1978; Squire and Zola-Morgan, 1991; Squire et al., 2004) have demonstrated the key role of this area in declarative memory (Squire and Zola-Morgan, 1991; 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 et al. (2013), Clarke et al. (2013), Cowell et al. (2010), Miyashita (1993), Naya et al. (2001), and 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.

2. 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 and 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

---

Please cite this article as: Quian Quiroga, R., Neuronal codes for visual perception and memory. Neuropsychologia (2016), http://dx.doi.org/10.1016/j.neuropsychologia.2015.12.016
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 and Sheinberg, 1996; Rosch et al., 1976) – depending on the knowledge and expertise of the subject (Bukach et al., 2006; Palmeri and Gauthier, 2004; Tanaka and Taylor, 1991) – mainly extracting prototypes and leaving aside a number of details and individual differences (Fabre-Thorpe, 2003; Logothetis and Sheinberg, 1996; Palmer, 1999; Palmeri and 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 et al. (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 parallel 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 and 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 and McDermott, 1995), or when, among a set of faces, subjects incorrectly recognize the prototype formed with these faces as being previously presented (Gao and 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 et al. (2009), Loftus and Palmer (1974), Schacter (2001), Schacter and Slotnick (2004)).

3. Perception in the ventral visual pathway

Given the vast evidence about visual perception along the ventral visual pathway reviewed elsewhere (Grill-Spector and Malach, 2004; Logothetis and Sheinberg, 1996; Roelfsema, 2006; Tanaka, 1996; Tsao and Livingstone, 2008), in this section I will just summarize evidence from electrophysiology recordings in monkeys and humans about the 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).

3.1. 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) (Fig. 1A) (Goodale and Milner, 1992; Ungerleider and Mishkin, 1982) (though it has been argued that the VVP should be extended further anteriorly to include the prefrontal cortex (Kraivitz et al., 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 and Sheinberg, 1996; Roelfsema, 2006; Tanaka, 1996; Tsao and Livingstone, 2008). In fact, neurons in IT have been shown to respond to experimentally defined (Sigala and Logothetis, 2002) and natural category of stimuli, such as faces, objects or body parts (Gross, 1992, 2008; Logothetis and 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 and Tsao, 2010; Tsao et al., 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 et al. (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) and Quian Quiroga and 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 and Tovee, 1995; Tamura and Tanaka, 2001; Tsao et al., 2006). Moreover, as it has been shown with single cell (Hung et al., 2005; Young and 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 and 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 perceptions – most notably with binocular rivalry (Blake and 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 et al., 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 et al., 2007; Kriegeskorte et al., 2008; Rolls and 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 et al., 2003; Freedman et al., 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 (Quian Quiroga, 2012c)) and also comparing perpendicular and oblique electrode penetrations (spanning neurons from the same and different putative columns, respectively) (Fujita et al., 1992). Further evidence along this line was given with optical imaging (Wang and 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 et al., 2009; Kanwisher, 2010; Konkle and Oliva, 2012; Tsumoda et al., 2001) as well as by the presence of discrete face (Freiwald and Tsaö, 2010; Tsaö et al., 2006) and body patches (Popivanov et al., 2014) studied with combined fMRI and electrophysiology recordings.

3.2. 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 et al., 2005) and, in order to make a clinical decision about the surgical procedure, they are monitored 24/7 for about 1–2 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 et al., 1994, 1999; Halgren et al., 1994; McCarthy et al., 1999; Puče et al., 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 et al., 1992) and fMRI (Kanwisher et al., 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, 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 (Quian Quiroga and 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 et al., 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 and Tsao, 2010; 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).

3.3. 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 et al., 2001). These activations are very sensitive to variations 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 et al., 1996; Rossion, 2014a). Among many studies, Bentin and colleagues used ambiguous schematic shapes and showed that these stimuli triggered N170 responses only after the subjects learned to associate them with faces (Bentin et al., 2002). These findings support the above-mentioned 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 et al., 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., 2014)).
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 et al., 2007), the temporal separation to a mask (Bacon-Mace et al., 2005; Harris et al., 2011), the phase coherence of the images (Philatides and Sajda, 2006) or the amount and type of noise added to them (Jemel et al., 2003; Tanskanen et al., 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 et al., 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 Fig. 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 and Changeux, 2011; Del Cul et al., 2007). However, as we have previously argued (Navajas et al., 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 et al., 2014; Clarke et al., 2013, 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.

4. 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 (Fig. 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 and Amaral, 2000; Quian Quiroga et al., 2009; Saleem and 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; Squire and Zola-Morgan, 1991; 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 et al., 1973; Engel et al., 2005; Fried et al., 1997; Rey et al., 2015) (see Fig. 3). Studies using this technique showed single neuron responses to individual words and faces (Heit et al., 1988), to oddball stimuli (Heit et al., 1990), conjuction of stimulus features, like gender and facial expression (Fried et al., 1997), associated word pairs (Cameron et al., 2001), and the degree of novelty and familiarity of the stimuli (Rutishauser et al., 2006; Viskontas et al., 2006). They also described neurons in the MTL that were active during visual imagery (Kreiman et al., 2000) and recall (Gelbard-Sagiv et al., 2008).
4.1. 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 Fig. 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 et al., 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 Fig. 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 and Nadel, 1997; Moscovitch et al., 2005; Rosenbaum et al., 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 et al., 2007; Hassabis and Maguire, 2007). Conversely, patients with semantic dementia, due to a focal atrophy in the anterolateral temporal lobe, have a severe deficit in retrieving semantic information but a much largely preserved episodic memory (Graham et al., 2000; Hodges and Graham, 2001; Snowden et al., 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

Fig. 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).
100 per session) (Ison et al., 2011; Mormann et al., 2008; Quian Quiroga et al., 2007, 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 and Kreiman, 2010; Quian Quiroga et al., 2008) and (iii) we have recently showed how these neurons can expand their tuning to encode new associations (Ison et al., 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 Fig. 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 and 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 et al., 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 et al., 2008). The modulation of single cell MTL responses with conscious perception has been also demonstrated using flash suppression (Kreiman et al., 2002), change blindness (Reddy et al., 2006) and more recently, using ambiguous morphed images (Quian Quiroga et al., 2014). In this latter study, the images to which the neurons initially fired (e.g., Jodie Foster; see Fig. 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 et al., 2005) – we triggered different percepts (recognized Jodie Foster or Nicole Kidman) from the same ambiguous image. As shown in Fig. 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 and 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 Fig. 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 (Json et al., 2015). The fact that the neurons encoded arbitrary associations (i.e. without a previous relationship between the items) 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 et al., 1995) (see last section for a more detailed discussion).

Sixth, concept cells tend to fire to personally relevant images (Viskontas et al., 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 (Squire et al., 2004) or not, in the latter case being always dependent on the hippocampus (Moscovitch and 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 relationships (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 hospitalward), 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. 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 et al., 1987; Redish et al., 2001), where a random connectivity between neurons has been described in CA3 (Li et al. 1994), the hippocampal area that has been proposed to be involved in forming associations (Tieves and 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 Fig. 2) in the human hippocampus (Quian Quiroga et al., 2009, 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 and Thorpe, 2006; Liu et al., 2009; Thorpe et al., 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 et al., 2014). In fact, when briefly flashing pictures at the threshold of conscious perception (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 and 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.

5. Concept cells in memory formation and recall

5.1. Partially overlapping assemblies

In line with the large bulk of evidence demonstrating the role of the MTL in declarative memory (Moscovitch et al., 2005; Squire and Zola-Morgan, 1991; 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 (Quian 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, we 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 et al., 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 and Kreiman, 2010).

Taking these two pieces of evidence together, namely, that
concepts are encoded by cell assemblies and that these neurons sometimes 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 Fig. 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 Fig. 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 et al., 1999; Hasselmo, 2012; Levy, 1996; Lisman, 1999). Compared to these works, the main difference is that the partially overlapping assemblies proposed here do 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, this model 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.

5.2. 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.

5.3. Do we have enough neurons?

In the following 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 and Rolls, 1994) – which in humans has only a few million neurons (West et al., 1991). A possible solution is given by the standard consolidation model, which postulates that memories first stored in the MTL consolidate in cortex (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 and Nadel, 1997; Moscovitch et al., 2005; Steinworth 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.

5.4. 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 the environment becomes familiar, has been proposed to explain the formation of place cells in the rodent hippocampus (Karlsson and Frank, 2008; Wilson and 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.

6. 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 subserve 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 and 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,
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 and 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 and O’Reilly, 2003; O’Reilly and 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).

7. 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 sketchy 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 I 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 (Squire and Zola-Morgan, 1991), the rodent hippocampus has been linked to spatial representations and navigation (O’Keefe and Dostrovsky, 1971; O’Keefe and 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 and Buffalo, 2010; Rolls et al., 1993; Rolls et al., 2005; Sliva et al., 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 versions of the manuscript and the Human Frontiers Science Program (grant nr. RGP0015/2013) 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.

References