

LONG-TERM CODING OF ASSOCIATIONS IN THE HUMAN MEDIAL TEMPORAL LOBE

Thesis submitted for the degree of
Doctor of Philosophy
at the University of Leicester

by

EMANUELA DE FALCO

Centre for Systems Neuroscience

University of Leicester



UNIVERSITY OF
LEICESTER

June 26, 2017

Abstract

Long-term Coding of Associations in the Human Medial Temporal Lobe

EMANUELA DE FALCO

This PhD thesis aims to investigate the role of human medial temporal lobe (MTL) neurons in the encoding of associations between items and the characteristics of single neuron activity in response to related items. The invaluable opportunity to record single neuron activity occurs when patients suffering from refractory epilepsy have to be implanted with intracranial depth electrodes to treat their clinical condition.

It has been long recognised that the MTL plays a critical role for declarative memory functions. MTL neurons have been shown to change their tuning during associative learning tasks. However, it is still not clear whether their involvement is confined to the task execution or goes beyond it. To address this issue, we studied the responses of MTL neurons in neurosurgical patients to known concepts (people and places), in conjunction with two different metrics measuring the degree of association between items (one metric based on the patients' evaluations and the other based on web searches). We found that whenever MTL neurons responded to more than one concept, these concepts were typically related, therefore providing evidence for a long-term involvement of MTL neurons in the representation of durable associations, which is essential to declarative memory functions.

We also analysed the differences between spiking responses elicited by different stimuli in a single neuron, and how these differences related to the degree of association between stimuli. We found that, in general, MTL neurons exhibit a similar neural activity in response to different stimuli, and that eventual differences in the responses are smaller the more two stimuli are associated to each other. Our results support the idea that a process of unitisation of neural responses occurs in the MTL, and that information about the stimulus identity is not encoded in individual neurons, but rather at the neural assembly level.

Acknowledgements

I would like to start by thanking my supervisor Rodrigo Quian Quiroga for his invaluable guidance during my PhD and for placing his trust in me from the very beginning.

Thanks to Matias J. Ison for co-supervising my work and always offering his support and an interesting point of view while working together.

To Hernan G. Rey for co-supervising my work, for sharing his experience with me while working together, and for being the most meticulous co-worker ever.

To all my colleagues at the CSN and friends in Leicester, who made this experience so rich and full of happy moments: Nishad, Alex, Julieta, Joaco, Hernan, Alex, Matias, Ceci, Leòn, Manolo, Fernando, Vitor, Natalia, Fanis, Sandra, Manolo, Emanuele, Alex, Hugo, Sonia, Trisha, Munisha, Maurizio, Luca.

To all my friends from Italy who always encouraged me during these years: Mario, Marco, Robertina, Giappo, Nicola, Giuseppe, Ciro.

To my parents, Patrizia and Raffaele, and to my grandparent Raffaele, for their continuous support no matter where my goals were taking me.

To my sisters, Francesca and Alessia, for always being there for me despite the distance.

To Enrico, for filling my life with love and happiness.

©Emanuela De Falco, 2017.

This thesis is copyright material and no quotation from it may be published without proper acknowledgement.

Contents

Abstract	2
Acknowledgements	3
Introduction	13
1 Literature review: The medial temporal lobe memory system	16
1.1 The visual pathways	17
1.2 Medial temporal lobe structure and connections	21
1.3 Memory functions	25
1.3.1 Classification of memory	25
1.3.2 The case of H.M.	27
1.3.3 MTL and declarative memory	28
1.3.4 Spatial memory	32
1.3.5 Consolidation models	33
1.4 Single cell recordings in human MTL	35
1.4.1 Concept cells	36
1.4.2 Encoding of associations	39
2 Experimental Methods	41
2.1 Intracranial recordings	41
2.1.1 Spike detection and sorting	45
2.2 Experimental paradigms	47

Contents	5
-----------------	----------

2.2.1	Responsiveness criteria	48
2.3	Association metrics	50
2.3.1	Personal association metric	50
2.3.2	Web-based association metric	51
2.3.3	Correlation personal and web-based metrics	54
2.3.4	Definition of mean association scores	55

3 Long-term coding of personal and universal associations in the human MTL **58**

3.1	Subjects and recordings	59
3.2	Personal association metric	60
3.3	Web-based association metric	64
3.3.1	Effect of relative number of hits	67
3.3.2	Visual similarity analysis	68
3.3.3	Effect of semantic categories	69
3.3.4	Response latency	71
3.4	Location analysis	72
3.5	Non-topographic organisation	75
3.6	Probability of pair responses	76
3.7	Cell-by-cell analysis	78
3.7.1	Correlation between neural responses and association	79
3.7.2	Decoding analysis	85
3.8	Firing rate distributions and sparseness	86
3.8.1	Sparseness of the representation	87
3.9	Summary on the distribution of responses	90
3.10	Conclusions to Chapter 3	93

4 Firing properties of pluri-responsive neurons in the human MTL **98**

4.1	Subjects and recordings	99
4.1.1	Spike-shape surrogate test	101
4.2	Quantification of spiking activity	102
4.2.1	Pairwise response strength comparison	106
4.2.2	Strength difference between responsive pairs	108
4.2.3	Decoding analysis of response strength	109
4.2.4	Response strength difference vs number of trials	111
4.3	Spiking response latency	113
4.3.1	Latency estimation with number of trials	113
4.3.2	Pairwise latency comparison	115
4.3.3	Latency difference on responsive pairs	116
4.4	Association metric results	118
4.4.1	Association scores	118
4.4.2	Differences in neural responses and association	120
4.5	Response characteristics with number of responses	123
4.6	Conclusions to Chapter 4	125
5	Conclusions and future directions	130
5.1	MTL neurons encode well-established associations	131
5.2	Overlapping assemblies encode associations	132
5.3	Unitisation of the neural responses	133
5.4	Future directions	135

List of Figures

1	Drawing of the rodent hippocampal formation	13
1.1	The structure of human neocortex	18
1.2	Dorsal and ventral visual pathways	20
1.3	Hippocampal system structures	21
1.4	Human hippocampal formation	22
1.5	Structure and connectivity of the human MTL	24
1.6	Long-term memory classification	27
2.1	Electrode probe for single cell recordings	42
2.2	Processing of extracellular recordings	44
2.3	Screening session paradigm	49
2.4	Personal association matrix	51
2.5	Web-based association matrix	53
2.6	Correlation between personal and web-based association scores .	56
3.1	Personal association scores for a single neuron in theHIPP . . .	61
3.2	Personal association scores	62
3.3	Numbers of responses (personal association metric)	64
3.4	Web-based association scores	65
3.5	Numbers of responses (web-based association metric)	67
3.6	Mean association scores per MTL areas	73
3.7	Location analysis	74

3.8	Non-topographic organization	76
3.9	Probability of pair responses	77
3.10	Correlation analysis for single neuron in the HIPP	81
3.11	Correlation analysis for single neuron in the PHC	82
3.12	Correlation analysis for a single neuron in the EC	83
3.13	Joint-neural activation map	84
3.14	Spontaneous and responsive firing rates	87
3.15	Sparseness distributions	89
3.16	Distribution of responsive neurons	91
3.17	Proportion of responses per category	92
4.1	Numbers of responses across recorded units	100
4.2	Spontaneous and responsive firing rates	103
4.3	Normalised strength of response activity	104
4.4	Cumulative distribution function for response strength	105
4.5	Strength difference for a pluri-responsive neuron in the HIPP	107
4.6	Surrogate test on response strength	108
4.7	Response strength for responsive and random pairs	109
4.8	Decoding analysis for exemplary unit in HIPP	110
4.9	Decoding analysis on response strength	111
4.10	Strength difference estimation with number of trials	112
4.11	Latency estimation with number of trials	114
4.12	Latency estimation with a sliding window of 6 trials	115
4.13	Latency difference for a single unit in HIPP	116
4.14	Surrogate test on response latency	117
4.15	Latency difference for responsive and random pairs	117
4.16	Mean association scores	118
4.17	Response differences vs association strength	121
4.18	Percentages of different pairs separated by AS	122

4.19 Response strength with number of responsive stimuli 123

4.20 Response latencies with number of responsive stimuli 124

List of Tables

3.1	Number of units for personal association metric	63
3.2	Number of units for web-based association metric	66
3.3	Summary on the numbers of units per analysis	90
4.1	Number of units for web-based association metric	119

List of Abbreviations

AMY	Amygdala
ANOVA	Analysis of Variance
AS	Association Score
CA	Cornu Ammonis Fields
CDF	Cumulative Distribution Function
CT	Computed Tomography
DG	Dentate Gyrus
EC	Entorhinal Cortex
EEG	Electroencephalography
HF	Hippocampal Formation
HIPP	Hippocampus
iEEG	Intracranial Electroencephalogram
IFR	Instantaneous Firing Rate
ISI	Interspike Interval
IT	Inferior Temporal Cortex

LFP	Local Field Potential
LGN	Lateral Geniculate Nucleus
MRI	Magnetic Resonance Imaging
MST	Medial Superior Temporal Area
MT	Middle Temporal Area
MTL	Medial Temporal Lobe
MU	Multi Unit
PER	Perirhinal Cortex
PHC	Parahippocampal Cortex
PMI	Pointwise Mutual Information
R-NR	Response-Non-Response
RR	Response-Response
SD	Standard Deviation
sdf	spike density function
SEM	Standard Error of the Mean
SPC	Superparamagnetic Clustering
SU	Single Unit
V1	Primary Visual Cortex
VS	Visual Similarity Score

Introduction

Some of the pioneering and most famous studies about the nervous system structure were done by Santiago Ramón y Cajal, considered the father of modern neuroscience. The Spanish neuroscientist widely used Golgi's dark staining technique to analyse the microscopic structure of the brain, producing incredibly detailed illustrations of brain cells and structures, and laying the basis for single neuron theory [Ramón y Cajal, 1904]. Figure 1 shows his famous drawing of the basic circuit of the hippocampus, the main component of the medial temporal lobe.

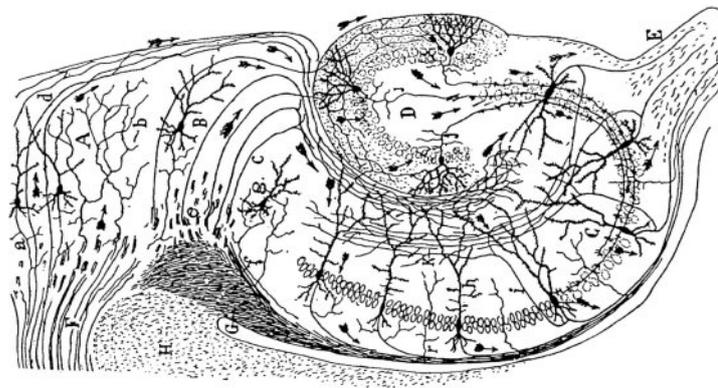


Figure 1 – Drawing of the rodent hippocampal formation by Santiago Ramón y Cajal. Main hippocampal areas: B, subiculum; C, Ammon's horn; D, dentate gyrus [Ramón y Cajal, 1909]

The essential role of the medial temporal lobe in explicit memory processing was firmly established in 1957 when the devastating effects of MTL resection were systematically studied by William Beecher Scoville and Brenda Milner in a patient known as H.M. [Scoville and Milner, 1957]. In the following decades, many behavioural and lesion studies in animals, together with indirect

measurements from human subjects, have deepened our knowledge on the anatomical components of the medial temporal lobe and their roles in memory functions (see for a review: Squire and Zola-Morgan [1991], Squire et al. [2004], Eichenbaum et al. [2007]).

The development of single cell recording techniques in human MTL [Halgren et al., 1978] has allowed us to study, directly, the activation of single neurons in behaving patients, leading to several studies exploring the role of MTL in language, memory and perception processing (e.g. Ojemann et al. [1988], Heit et al. [1988], Fried et al. [1997], Kreiman et al. [2000b]). This technique brought to the recent discovery, in human MTL, of the so-called concept cells [Quian Quiroga et al., 2005]. Concept cells are neurons that respond selectively to different representations of a single concept (e.g. a specific person or place) and have been postulated to form the basis of MTL declarative memory processing [Quian Quiroga, 2012].

Moreover, MTL neurons have been reported to change their tuning during associative learning tasks in several animal studies (e.g. Miyashita et al. [1989], Cahusac et al. [1993], Erickson and Desimone [1999], Wirth et al. [2003], Yanike et al. [2004]) and, very recently, also in humans [Ison et al., 2015]. However, in all of these reports, recordings were done during associative learning tasks, leaving it unclear whether the MTL provides only a transient encoding of associations during learning or whether more stable representations persist after the execution of the task.

Following previous observations of MTL neurons responding to presumably associated concepts (e.g. two co-stars in a television show, Quian Quiroga et al. [2005, 2009]), we addressed the open question of whether there is a consistent tendency for MTL neurons to encode meaningful associations, independent from the performing of associative learning tasks, and how similar (or dissimilar) the activation of a single neuron firing in response to different stimuli was.

Organisation of the thesis

This thesis describes, in detail the motivation, techniques and results of my PhD work, and it is organised as follows.

Chapter 1 contains a literature review about the MTL and its role in long-term declarative memory processing. In particular, the anatomical and connectivity structures of the MTL will be described here, followed by evidences and theories about the MTL role in memory functions, and by the latest findings based on human intracranial recordings.

A description of all the methodological tools employed in this work will be presented in Chapter 2. Here I will report the technical details for data recording and processing, extraction of single neuron activity, and the experimental paradigms used. Moreover, this chapter also contains detailed descriptions of the two metrics used to estimate the degree of association between concepts, together with the validation of the web-based method.

Chapter 3 will describe my main PhD project, aimed to prove the encoding of long-term associations in the MTL. Here I will illustrate the details of the different analysis performed and the results supporting our main claim, together with several other remarks from the study.

Chapter 4 contains the details and results of our follow up study, aimed at analysing the response features of single neurons firing to more than one stimulus. I will describe here how the findings from this study support the idea that abstract representations are encoded in the MTL through the activation of sparse neural assemblies, rather than in the activity of single neurons.

Finally, in Chapter 5 I will summarise the main contributions of the present work to the current discussion about MTL and associative coding, highlighting the novelty and implications of our findings, and suggesting possible future directions to follow.

Chapter 1

Literature review: The medial temporal lobe memory system

The medial temporal lobe (MTL) is located, as its name suggests, in the medial (internal) part of the temporal lobe. It comprises the hippocampus and surrounding structures, known as the parahippocampal region. The hippocampal and parahippocampal areas lie at the end of the hierarchy for processing visual stimuli perception and recognition [Logothetis and Sheinberg, 1996]. The first clear evidence of the MTL's essential role in explicit memory storage was the case of patient H.M. His profound and selective memory impairment, following a bilateral resection of medial temporal lobes, was systematically studied and reported by Milner and Scoville in the 1957 [Scoville and Milner, 1957]. This evidence gave rise to several behavioural and lesion studies in rats, monkeys, and other animals aimed to better understand the anatomical components of the medial temporal lobe and their roles in memory functions [Squire et al., 2004, Squire and Zola-Morgan, 1991]. In addition, over recent decades, the development of intracranial recording techniques in epileptic patients [Halgren et al., 1978] has allowed to directly study the activity of single neurons in the human MTL, leading to several major findings about the role of MTL neurons

in processing external stimuli (e.g. Fried et al. [1997], Kreiman et al. [2000b], Quiñan Quiroga et al. [2005, 2009]).

This chapter is organized as follows. Firstly, I briefly review the visual pathways that carry visual information from the retina to the MTL. Next, I describe the anatomical and connectivity structure of the MTL followed by evidences and interpretations about its fundamental role in declarative memory. Finally, I review the latest literature based on single cell recordings in the human MTL, focusing in particular on the so-called concept cells and the encoding of associations in the MTL.

1.1 The visual pathways

Visual perception is a constructive process that goes from the discrimination of low level features (colour, orientation, contrast) up to the conscious perception of the visual scene. This process involves multiple pathways and structures of the brain.

The *primary visual pathway* constitutes the first stage of the visual perception process; it starts from the retina and reaches the primary visual cortex. The visual information from retinal photoreceptors is passed to retinal ganglion cells. Ganglion cell axons bundle together to form the optic nerve, run through the optic chiasm and reach their major target: the lateral geniculate nucleus (LGN) of the thalamus. The LGN neurons analyse low level features as contrast and colour and send projection to the primary visual cortex, also referred as V1 or Brodmann's area 17 (see Figure 1.1 for a schematic of Brodmann's areas), located in the posterior portion of the occipital cortex. Neurons in the primary visual cortex are able to identify orientation, lines and boundaries (the contours in the image), binocular disparity, and direction of movement. From area V1 cortical projections are sent forward along two parallel major

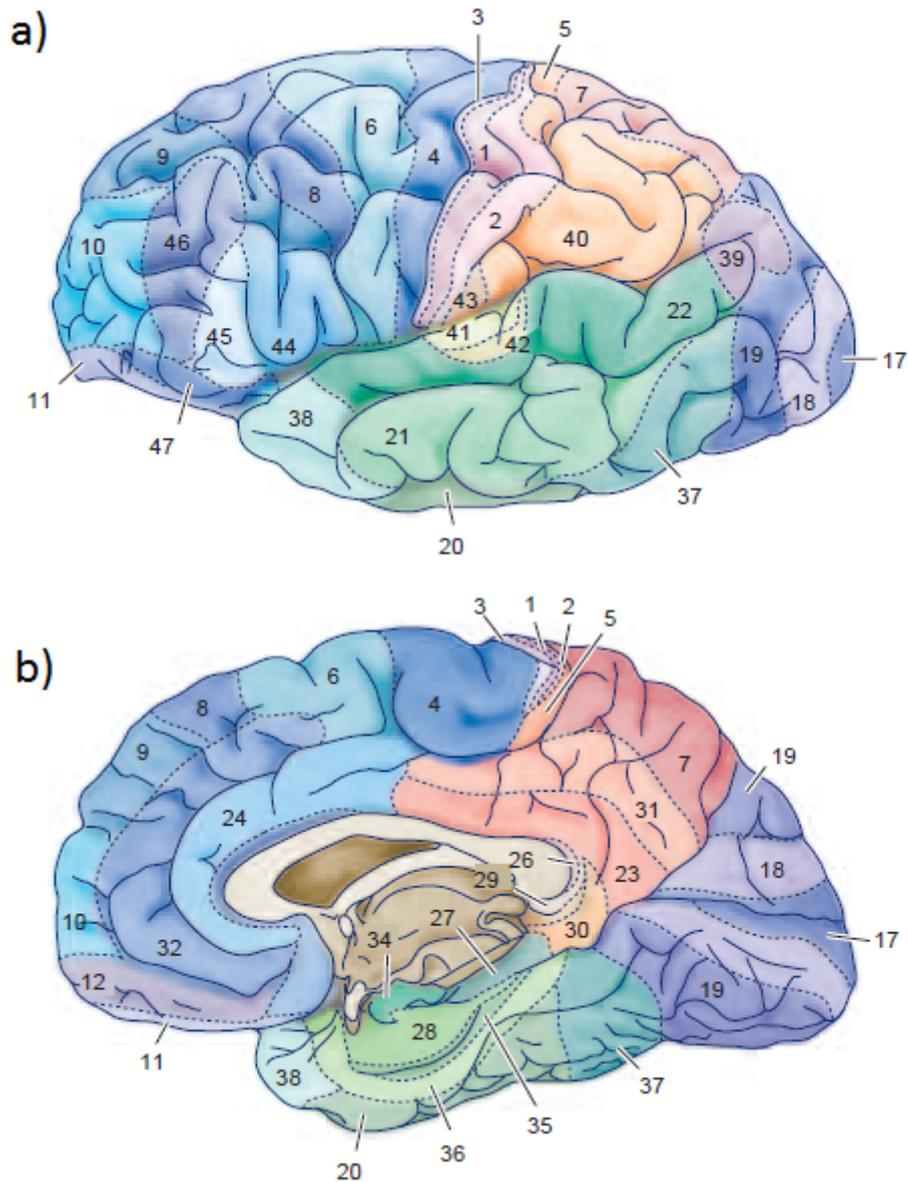


Figure 1.1 – The structure of the human neocortex. Areas in the lateral surface (a) and in the medial surface (b) as identified by the neuroanatomist Korbinian Brodmann based on the cytoarchitectural organization of neurons [Brodmann, 1909]. Adapted from [Purves et al., 2004].

processing pathways: the ventral visual pathway and the dorsal visual pathway (see Figure 1.2)[Kandel et al., 2013, Purves et al., 2004]. The ventral stream (also called "what" pathway) is responsible for object recognition. The dorsal stream (also called "where" or "how" pathway) is responsible for the spatial aspects of vision such as the object's relative locations and the analysis of motion [Ungerleider and Mishkin, 1982, Goodale and Milner, 1992]. The two pathways are hierarchically organised, each level has strong projections to the next one and the selectivity to complex features increases along the pathways. However, these connections are also widely reciprocal, with feedback pathways running in the opposite direction of the two streams.

The *ventral visual pathway* projects ventrally and anteriorly from V1 through V2, V4, and the temporal-occipital junction to the inferior temporal cortex (IT, constituted of areas TEO and TE). This stream carries information about the objects perceived, neurons in this stream are sensitive to colour, shape and texture, which are important features for object recognition. Inferior temporal cortex represents the highest level of cortical visual processing. Here, a large amount of different visual information is integrated to encode complex visual stimuli and recognize the objects, regardless of the different viewing conditions.

The *dorsal visual pathway* goes through several areas in the parietal lobe (V3, V3A, middle temporal area MT, medial superior temporal area MST and additional areas in the inferior parietal cortex) to the prefrontal cortex. Neurons within this stream show selectivity for spatial aspects of stimuli such as the direction and speed of movement, providing visual guidance towards where an object is.

The two visual pathways are not strictly segregated, but are interconnected. Along them information is shared to create a unified and coherent visual percept of what an object is and where it is. Moreover, the neuronal object representation does not rely only on visual perception, it also incorporates different

sensory modalities, object categorization, associated concepts, memories, and emotional meanings. In fact, the IT cortex is connected with several cortical areas (parahippocampal region, prefrontal cortex, amygdala and multimodal sensory areas) that, all together, participate in object recognition and representation. A large part of the projections from IT cortex are directed to the parahippocampal and perirhinal cortices of the medial temporal lobes. From there, information is then sent through the entorhinal cortex to the hippocampus, contributing in forming long-term memory of visual objects and their context. Notably, from IT cortex information also reaches the prefrontal cortex (important area for working memory) and the amygdala (emotional valence) [Kandel et al., 2013, Purves et al., 2004].

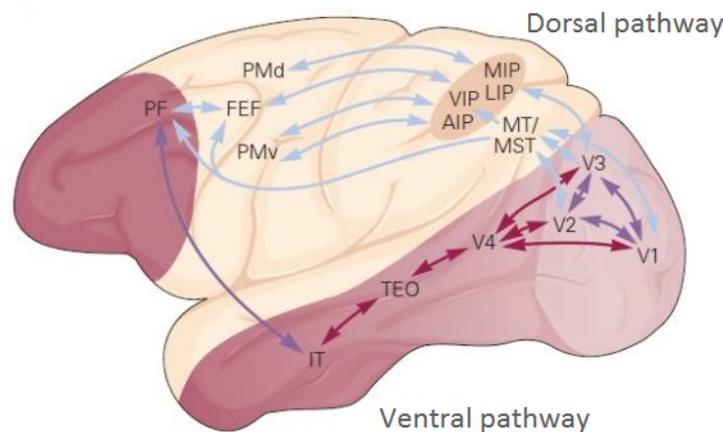


Figure 1.2 – Dorsal (cyan) and ventral (red) visual pathways in the cerebral cortex of the macaque monkey. The ventral visual pathway is involved with object recognition, while the dorsal visual pathway is involved with spatial aspects of visual perception. Abbreviations: AIP, anterior intraparietal cortex; FEF, frontal eye field; IT, inferior temporal cortex; LIP, lateral intraparietal cortex; MIP, medial intraparietal cortex; MST, medial superior temporal cortex; MT, middle temporal cortex; PF, prefrontal cortex; PMd, PMv, dorsal and ventral premotor cortices; TEO, occipitotemporal cortex; VIP, ventral intraparietal cortex; V1–V4, areas of visual cortex. Adapted from Kandel et al. [2000].

1.2 Medial temporal lobe structure and connections

The *medial temporal lobe* (MTL) consists of the hippocampus (or hippocampal region), the adjacent regions of parahippocampal cortex, perirhinal cortex and entorhinal cortex (called, all together, parahippocampal region), and the amygdala [Squire and Zola-Morgan, 1991, Eichenbaum et al., 2007]. The parahippocampal structures are located in the parahippocampal gyrus and temporal pole. Figure 1.3 shows a lateral view of the human brain and approximate position of the hippocampus and parahippocampal structures within.

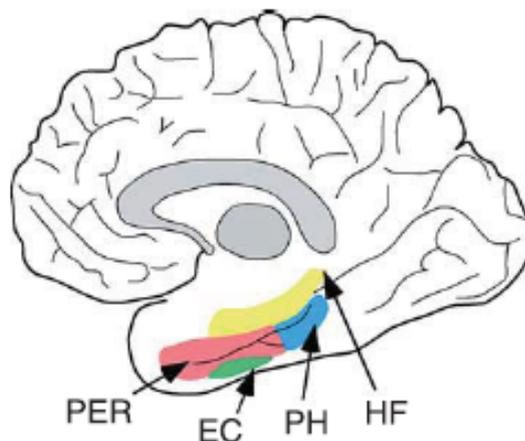


Figure 1.3 – Hippocampal system structures in lateral view of the human brain. Abbreviations: HF, hippocampal formation; PH, parahippocampal cortex; EC, entorhinal cortex; PER, perirhinal cortex. Adapted from Burwell and Agster [2008].

The *hippocampus* or *hippocampal formation* comprises the *dentate gyrus*, *Ammon's horn fields* (CA1, CA2 and CA3, also called *hippocampus proper*), and the *subiculum*. The hippocampus, greek for seahorse, is named for its shape that resembles the small marine fish. Figure 1.4 shows a microscopy photograph of the human hippocampus formation where it is easy to recognise the characteristic C-shape of the dentate gyrus interlocked with the CA fields structures. The components considered part of the hippocampus have a typical

trilaminar structure (also known as paleocortex), as opposed to the six-layered neocortex [Purves et al., 2004] and are mainly connected by unilateral pathways starting from the dentate gyrus input to the CA3 field. CA3 pyramidal cells project unidirectionally to the CA1 area, that in turn provides input to the subiculum. The unidirectionality of this circuit is unique, as corticocortical connections in the brain are usually widely reciprocal [Burwell and Agster, 2008].

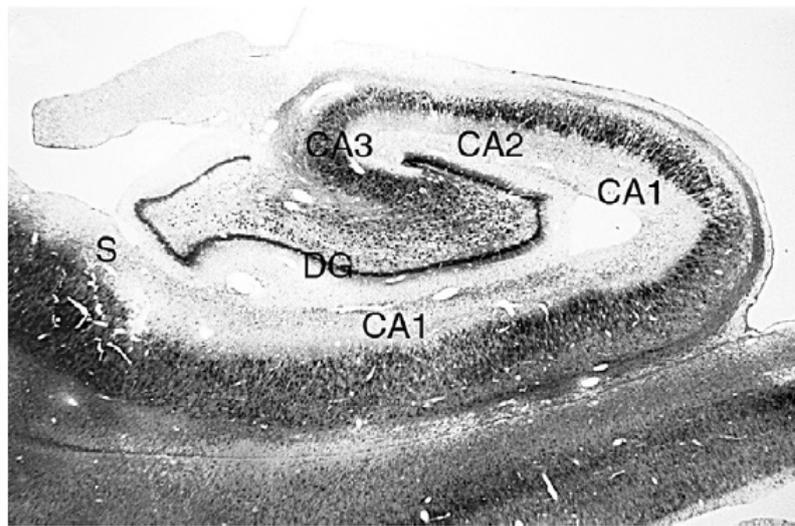


Figure 1.4 – Bright-field microscopy photograph of immunohistochemical preparations from the human hippocampal formation, stained for non-phosphorylated neurofilament, from Insausti and Amaral [2012]. Abbreviations: DG, dentate gyrus; CA, Ammon’s horn fields ; S, subiculum.

The *entorhinal cortex* (EC) is identified as Brodmann area 28 [Brodmann, 1909](see Figure 1.1), situated medial to the caudal half of the rhinal sulcus. It is the highest order area of the parahippocampal region. The EC receives most of its inputs (95%) from other polymodal association areas, in particular two-thirds of them come from the parahippocampal and perirhinal cortices, and no input from the unimodal visual areas TE, TEO and V4 [Insausti et al., 1987, Suzuki, 1996].

The *perirhinal cortex* is composed of Brodmann areas 35 and 36 [Brodmann, 1909], situated lateral to the rostrocaudal extent of the rhinal sulcus. Both

area 35 and 36 of perirhinal cortex receive similar cortical inputs, with robust projections from unimodal visual areas TE and TEO (64% of the received cortical inputs) and from other sensory modality areas [Suzuki, 1996]. The *parahippocampal cortex* is located caudal to the entorhinal and perirhinal cortices and it consists of two areas: TH and TF.

Perirhinal and parahippocampal cortices receive massive projections from the IT cortex and from the other sensory cortical areas. However, visual object information (“what” stream) from areas TE/TEO is mainly directed to the perirhinal cortex, while visuospatial information (“where” stream) from the dorsal pathway (in particular retrosplenial cortex, posterior parietal cortex and area 46) reaches, predominantly, the parahippocampal cortex. Inputs from the auditory cortex arrive mainly to area TH of the parahippocampal cortex, while the signal from somatosensory association areas reaches, equally, both the cortices [Squire et al., 2004]. Parahippocampal and perirhinal cortices project respectively to the medial and lateral areas of the entorhinal cortex, that in turn project to the hippocampus. The described connections are reciprocal: feedback projections go back from the hippocampus to the entorhinal cortex, then to the rest of parahippocampal region and finally back to neocortical areas [Lavenex and Amaral, 2000, Eichenbaum et al., 2007, Burwell, 2000].

The medial temporal lobe exhibits a unique hierarchy of associational networks that characterizes its neuroanatomy. The perirhinal and parahippocampal cortices provide a first level of associativity and integration of multimodal polysensory information. In the entorhinal cortex the information reaches a higher level of integration before it is transmitted to the hippocampus. The hippocampus represents the ultimate level of integration and abstraction of information thanks to its several associational networks. Finally, its output is distributed via feedback connections back to the neocortex [Lavenex and Amaral, 2000]. The connectivity and structure of the medial temporal lobe is shown in Figure

1.5.

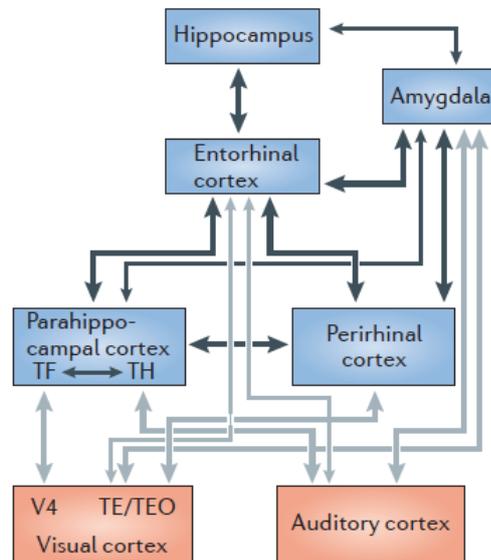


Figure 1.5 – Structure and connectivity of the human medial temporal lobe (regions in blue), connections within the medial temporal lobe (marked in black) and with the visual and auditory cortex (marked in grey). From Quian Quiroga [2012].

The *amygdala* (or amygdaloid complex) is an almond shaped complex of grey matter located rostral to the hippocampus in the anterior-medial portion of the temporal lobe. Its structure comprises different subnuclei and cortical regions. The amygdala is part of the limbic system (together with the hippocampus, cingulate gyrus, hypothalamus and anterior thalamic nuclei) and it is involved in a wide range of functions including the processing and expression of emotions, especially fear, and some forms of memory. Sensory inputs to the amygdala come mainly from the olfactory bulb, olfactory cortical areas, visual IT cortex, auditory association areas and somatic sensory areas. The amygdala also has reciprocal connections with the brain stem and hypothalamus, from which it receives visceral inputs, and with other structures of the limbic system. Hence the amygdala connects cortical regions that process sensory information with both somatic and visceral effector systems, playing an important role in the expression of emotional behaviour. The amygdala has several interconnections

with other MTL regions, in particular the hippocampus subiculum, but also the entorhinal and perirhinal cortex. However, these connections are weaker than the ones described above between the rest of the MTL [Suzuki, 1996, Purves et al., 2004].

1.3 Memory functions

Learning and memory are complex functions of the human brain, that are essential to our survival and everyday life. Learning is defined as a change of behaviour following the acquisition of new information, while memory consists of the ability to encode, store and retrieve learned information. Until the middle of the 20th century memory functions had not been localised in the brain; the hippocampus was thought to simply be a part of the limbic emotional system and some psychologists even doubted that memory was a discrete function of the brain. These beliefs changed completely during the mid-1950s following several studies on amnesic patients with medial temporal lobe resections, with the first and best-known case being patient H.M. [Scoville and Milner, 1957]. Nowadays we recognise memory as a distinct psychophysiological function and the hippocampus as central to that function. Unlike other mental functions (e.g. language or voluntary movement) mediated by discrete areas of the brain, memory functions involve different parts of the brain. These parts are not equally important and cover distinct roles related to different types of memory [Kandel et al., 2013, Eichenbaum, 2013].

1.3.1 Classification of memory

Memory can be classified according to the time over which information is stored and to the nature of the information stored. Temporally we distinguish between short-term and long-term memory. Short-term memory is the ability

to maintain a transient representation of information from new experiences that can last from seconds to minutes. Long-term memory is a more permanent form of storage that can last from days to years. Through the process of consolidation, information from short-term memory is selectively transferred to long-term memory [Purves et al., 2004].

Long-term memory has been classified in two qualitatively different types of memory: declarative (or explicit) and nondeclarative (or implicit) memory. These two memory types differ in whether conscious awareness is required to recall them. They were identified for the first time by Milner [Milner, 1962], while working with amnesic patient H.M., and labelled later (as explicit vs implicit) by Graf and Schacter [Graf and Schacter, 1985], examining both healthy and amnesic subjects.

Implicit memory is an unconscious form of memory (also known as procedural memory). Implicit memory involves skills, habits and associations acquired unconsciously (or without consciousness being necessary to the acquisition) and manifested in an automatic manner, by improved performance or altered bias [Schacter, 1987]. Examples of nondeclarative memory are procedural skills, priming, conditioning and habituation (see Figure 1.6 for a schematic of long-term memory classification).

Explicit memory is closer to our intuitive notion of memory, it regards information that can be expressed by language (hence “declarative”) and are stored and retrieved consciously. Unlike implicit memory, explicit memory is very flexible, as information can easily be associated together under different circumstances. Explicit memory can be further classified into episodic and semantic [Tulving and Donaldson, 1972]. Episodic refers to the memory of personal experiences and autobiographical memory, while semantic memory refers to factual knowledge (e.g. facts, concepts or the meaning of words).

Explicit memory processing comprises four different operations: encoding,

storage, consolidation and retrieval. Factors like attention, motivation to remember and emotional valence can influence the encoding process. Moreover, memory is a constructive process much like perception, as shown by Sir Fredrick Bartlett [Bartlett, 1932]. During retrieval we use cues and several cognitive strategies like comparison, inference or guessing to create a memory representation that is coherent to us. For this reason explicit memory, at least the episodic one, is subject to transformation over time (omissions, deletions and distortions) [Kandel et al., 2013, Purves et al., 2004].

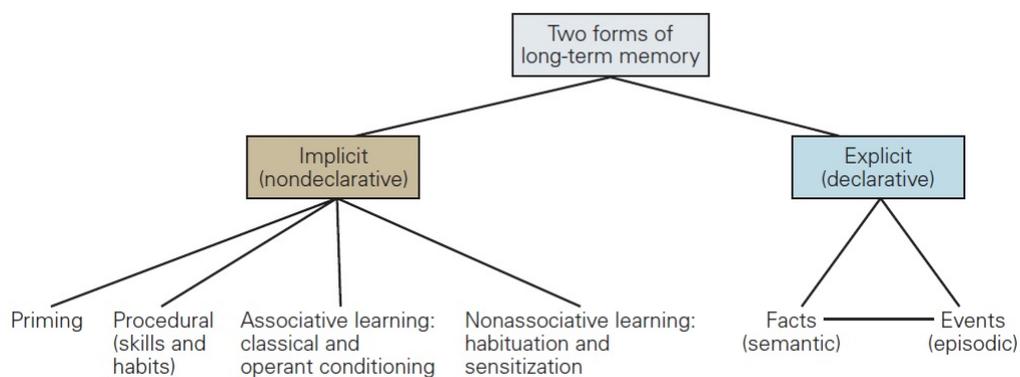


Figure 1.6 – Long-term memory classification. Adapted from Kandel et al. [2013].

1.3.2 The case of H.M.

The case of patient H.M. was studied by the psychologist Brenda Milner and the surgeon William Scoville [Scoville and Milner, 1957] and later also by many of their students and colleagues. His real name (Henry Molaison) was revealed after his death in 2008. H.M. was a 27-year-old man suffering from untreatable and debilitating temporal lobe epilepsy. On September 1953, to control his seizures, H.M. underwent brain surgery where a large portion of his medial temporal lobe was removed bilaterally. In particular, the resection involved the amygdala, hippocampal gyrus, and most of the hippocampus region. After the surgery, despite his seizures being more manageable, he manifested a

severe anterograde amnesia. He was unable to retain new long-term memories for the events following the surgery and performed very poorly in standard psychological memory tests. For instance, even after having met Milner every month for several years, every time she arrived in the room he thought he had never met her before. In contrast to this, H.M. had intact perceptual and cognitive abilities, including language, and his IQ stayed unchanged at the level of his preoperative score [Kandel et al., 2013, Purves et al., 2004]. His long-term memory was intact for information acquired long time before the surgery, even though memories from the years just before the surgery were very weak [Milner et al., 1968, Corkin et al., 1983]. Moreover, he seemed to be unaware that he had undergone surgery and of his subsequent deficit. On the other hand, his short-term memory was still working; he could retain information for short periods of time and even engage in a conversation as long as he was not distracted by interfering mental activity [Corkin, 1984]. Further tests on H.M. showed that his implicit memory was not impaired either. For instance, over three days of training, he improved significantly his score in a mirror drawing test [Milner, 1962]. He was still able to learn how to perform new tasks through procedural or reflexive learning, even though he did not consciously remember being trained in the task [Corkin, 1965, 1968].

1.3.3 MTL and declarative memory

The case of patient H.M. with his global and yet not generalised memory impairment, opened the way to a new era for memory investigation. Similar cases of patients with extensive bilateral lesions in the hippocampal area contributed to our understanding of the role of medial temporal lobe structures in the memory process. These different cases of impairment, following MTL damage, share three fundamental aspects [Squire et al., 2004]:

1. Memory is affected regardless of the sensory modality in which information

is presented (i.e. verbal, nonverbal, visual or auditory)[Corkin, 1984]. This is in accordance with the fact that in the MTL there are convergent inputs from all different sensory modalities (as seen in Section 1.2).

2. Short-term memory remains intact. Even when the size of a MTL lesion is large, patients can retain a normal amount of information over a brief time course (for instance they have a normal digit span) [Milner et al., 1968].
3. Memory impairment happens despite perceptual and intellectual abilities remaining intact. In particular, patients with more localised damage (e.g. limited to the hippocampus) have been shown to perform similarly to controls [Schmolck et al., 2002]. Patient H.M. himself, despite his lesion being more extensive, scored normally in IQ tests and exhibited only mild impairment on a few semantic knowledge tests [Milner et al., 1968].

The studies on amnesic patients, together with numerous works on animal models, contributed to very important findings about the memory process and the role of MTL. First of all, it is now clear that memory is an actual distinct psychological function and not simply a function related to each particular modality of processing. In fact impairment in amnesic patients involved all sensory modalities. Next, all those evidences prove that the hippocampus and adjacent areas are specifically important for declarative memory (both semantic and episodic), while they are not necessary to the processing of nondeclarative memory (skills, habits, conditioning, priming) [Squire, 1992]. In particular, the MTL seems to support the consolidation of new information from short-term memory to long-term memory [Alvarez and Squire, 1994]. In fact, H.M., as the other patients, had most of his older memories intact and was still able to retain information in working memory over a short period of time, but he dramatically lacked the ability to encode new information in a more permanent

way [Eichenbaum, 2013].

A still debated issue is whether the hippocampus is uniquely involved in episodic, or also in semantic memory. One view is that the hippocampal region is required for both episodic and semantic memory functions [Squire and Zola, 1998, Manns et al., 2003b], while another possible view affirms that the role of the hippocampal region is confined to episodic memory functions while adjacent structures are involved in semantic memory [Vargha-Khadem et al., 1997, Tulving and Markowitsch, 1998]. In the computational model described by Kesner and Rolls [2015] the hippocampus and neocortex are seen as complementary memory systems, with the hippocampus being used for rapid and unstructured storage of information, while the neocortex gradually builds and adjusts semantic representations. In this model, episodic information, stored in the hippocampus, can be recalled to the neocortex and be used to help form semantic memories.

Moreover, it appears clear that all the different structures in the MTL contribute to the declarative memory process [Squire et al., 2004]. Clinical evidence for this comes from several patients with damage strictly limited to the hippocampus (see for example the case patient R.B. in Zola-Morgan et al., 1986). These patients exhibit quantitatively less severe memory impairment than that affecting patients with damage extended to the parahippocampal area (e.g. H.M), even though the deficits are qualitatively very similar. Similar findings arrive from a comparison of studies in monkeys [Zola-Morgan et al., 1994], where damages limited to the hippocampus were found to cause much milder impairment than lesions extended to the adjacent area. On the other hand, even damage limited to the perirhinal cortex has been shown to cause severe recognition memory impairment in multiple versions of the object recognition memory task [Nemanic et al., 2004]. These findings support the idea that different MTL structures contribute in different ways to declarative memory

[Eichenbaum et al., 2007]. The neuroanatomical structure of MTL supports this view, since information from neocortex reaches the medial temporal lobe at different locations. As seen in Section 1.2, information about the object (“what”) from ventral cortical stream converges mainly in the perirhinal cortex and then in the lateral entorhinal cortex. Spatial visual information (“where”) from the dorsal cortical stream reaches mainly the parahippocampal cortex and then the medial entorhinal cortex. Evidence from animal models [Squire and Zola, 1996, Parkinson et al., 1988, Malkova and Mishkin, 2003] and neuroimaging studies in humans [Bellgowan et al., 2003] lead to the hypothesis that the two classes of information are mainly encoded separately along these two MTL streams. Subsequently, the hippocampus that receives information from both streams may cover the special role of integrating the “what” information (objects, people, actions) with the spatial and temporal one (“where”) to encode composite representation of an episode. The comparison between new representations and pre-existent memories would lead to reorganising the memories according to the relevant interaction between them. In this view, the combination of these processing functions is thought to constitute the basis of declarative memory [Eichenbaum et al., 2007].

Finally, the amygdala seems to be involved in the acquisition and storage of memories of emotional experiences, playing a complementary role in the memory processes. The encoding of emotional episodic memories compared to non-emotional ones is characterized by increased resources and arousal-driven enhancements that improve consolidation [Kensinger et al., 2001, Dolcos et al., 2004, Ritchey et al., 2008]. According to the modulation theory [McGaugh, 2004, LaBar and Cabeza, 2006], the emotional enhancement of memories is due to the modulatory effect of the amygdala on the MTL that boosts consolidation of arousing stimuli via noradrenergic projections.

1.3.4 Spatial memory

In 1971 place cells in the rat hippocampus were discovered by O'Keefe and Dostrovsky [O'Keefe and Dostrovsky, 1971]. Hippocampal place cells are neurons that dramatically increase their firing rate when the animal is in a specific location (place field). They are observed in extracellular recordings from CA1 and CA3 of freely behaving rats. A task-free environment provides a cleaner observation of place cells, since the introduction of a task in the experimental setup brings these neurons to respond to features of the task [Wood et al., 1999, Eichenbaum et al., 1999]. Much later, similar cells, known as grid cells, have been discovered in the rat entorhinal cortex [Hafting et al., 2005]. Grid cells also exhibit a spatial-related firing, but their firing fields cover a spatial grid instead of a single place. In the primate hippocampus 'spatial view' neurons have been found to respond to the place at which the monkey is looking during the execution of an object-place memory task [Rolls et al., 1989, 1999]. In addition, several findings from neuroimaging studies in humans have shown spatial activation of parahippocampal cortex and hippocampus during learning and memory tasks while navigating in a virtual reality environment [Maguire et al., 1997, 1998, Shelton and Gabrieli, 2002, Hartley et al., 2003]. Further evidence supporting the link between the hippocampal area and spatial navigation in humans is illustrated by the impairment of navigation skills in an amnesic patient [Maguire et al., 2006] and the activation of single neurons in epileptic patients while moving around a virtual map [Ekstrom et al., 2003].

The discovery of place cells brought to the hypothesis that the hippocampus' primary function is to support spatial memory and that it does this through a neural representation of physical space, a so called 'cognitive map' [Tolman, 1948, O'keefe and Nadel, 1978]. This spatial view of hippocampus' function contrasts with the memory view, which suggests a broader role played by the hippocampus in declarative memory. An agreement between these two views

has been difficult to find, due to the fact that most experiments on memory have investigated declarative memory in humans, while spatial cognition has been mainly studied in behaving animals [Schiller et al., 2015]. To reconcile the spatial and declarative memory views of MTL one possibility is to view spatial memory as a subset of the broader category of declarative memory [Squire, 1992]. In the model proposed by Treves and Rolls [1994] hippocampal CA3 neurons store episodic memories through an autoassociation mechanism. Such episodic memory functions are often required to perform spatial computations (in agreement with the existence of rat place cells and primate ‘spatial view’ cells), but the actual spatial computation is thought to happen in the neocortex [Kesner and Rolls, 2015]. In line with this interpretation, Eichenbaum et al. [1999] suggested that place cells might be encoding the experience of the animal in each place, rather than the location itself. According to this view, the hippocampus supports declarative memory generating a “memory space”, that is a relational representation in which the different elements of memories are bound by means of their common features [Konkel and Cohen, 2009, Eichenbaum and Cohen, 2014].

1.3.5 Consolidation models

An open question about the medial temporal lobe concerns the consolidation process within the cortical-hippocampal system. There are three main hypotheses about how this process happens that I will outline here (see McKenzie and Eichenbaum [2011] for a review).

One of the findings from patients with hippocampal damage is a selective impairment in “recent memory” over a period extending up to 10 years before the damage occurred. On the contrary, knowledge obtained remotely prior to the damage is intact [Corkin, 1984, Manns et al., 2003a, Bayley et al., 2006]. Temporally retrograde graded amnesia is often taken as evidence in support of

the standard model of consolidation [Squire and Alvarez, 1995]. According to this model, long-term memory is created as outcomes of processing between the MTL and the regions of neocortex specialised for the kind of information remembered [Mishkin, 1982, Squire, 1987]. Initially, the hippocampus and adjacent structures of the MTL bind new information from the neocortex in a memory trace. Then the hippocampus works together with neocortical areas to create a representation of information in cortex (hippocampus-dependent stage). With time and recalls, the corticocortical connections become stronger, until eventually the memory becomes independent from hippocampal representation. In this view, the hippocampus plays only a temporary role and it is no longer needed once memories are permanently stored in the neocortex. This is a dynamic process that can extend for years. The different degrees of consolidation would be responsible for the temporally graded amnesia findings. Cases of intact remote autobiographical memories in patients with MTL damage [Bayley et al., 2006] also seem to support to this model.

On the other hand, the multiple trace theory [Nadel and Moscovitch, 1997, Moscovitch et al., 2005] proposes that memories are initially stored as episodic (defined as context-specific) and then converted into semantic during consolidation. At the beginning, episodic memories are stored in the cortical-hippocampal circuitry. Each reactivation of a memory trace, because it is happening in a different neuronal and experimental context, results in the creation of a new trace. These multiple traces created share information about the initial episode. From the multiple traces factual information (free of contextual details) is extracted and integrated with pre-existent semantic memories. In this view, semantic memories are independent from the hippocampus, while, in contrast with the standard model, the involvement of the MTL remains necessary for recalling the richness of episodic details. Evidences in support of the multiple traces model come from cases of patients with temporally ungraded retrograde

amnesia (e.g. Steinvorth et al., 2005) and functional imaging studies that have shown the activation of MTL for both recent and remote episodic (and autobiographical) memories [Ryan et al., 2001, Maguire et al., 2001]. However, in contrast with this model there are reports of modulation in MTL activity related to the age of semantic memories (e.g. Smith and Squire, 2009).

A third possible model, the “schema modification” was originally proposed by Bartlett [1932] and recently extended from a consolidation point of view by McClelland et al. [1995]. According to this model, the hippocampus (with its rapid synaptic modification) acquires the new memories and gradually stores them through modification of a pre-existent cortical schema. This schema is the organisation of related knowledge, that contains both semantic and episodic information (in contrast with the multiple trace theory). Memories in the schema are interleaved according to common features and elements, but they are still hippocampus-dependent for detailed information. Moreover the modification of our schemas continues as long as new memories need to be integrated [McKenzie and Eichenbaum, 2011]. In support to this model, Tse et al. [2007] have shown how rats can develop a schema of food/location associations where new elements can be integrated within a single trial (in contrast to a slower learning of new associations in a different environment). Several other studies in humans and animals support the existence of networks of related memories, showing that hippocampal neurons encode both distinct experiences and their common overlapping elements (for a review see Eichenbaum [2004]).

1.4 Single cell recordings in human MTL

Most of the evidences about the role of the MTL in explicit memory processing in humans (reviewed in Section 1.3) comes from indirect measurements (e.g. lesion studies, functional imaging or scalp EEG). While knowledge about the

neuronal level processing is mainly based on data from animal models [Engel et al., 2005]. The development of single cell recording techniques offered the opportunity to observe the direct activation of neurons in behaving patients (see Engel et al. [2005], Mukamel and Fried [2012], Rey et al. [2014b] for reviews on invasive recordings in humans). Invasive recording in patients are sometimes necessary for the treatment of certain conditions (e.g. Parkinson’s disease or epilepsy) when the pharmacological approach fails. In particular, for certain forms of temporal lobe epilepsy intracranial recordings in the human MTL are carried out to precisely localise the seizure focus [Kandel et al., 2013]. Intracranial electrodes and microelectrodes allow us to record the local field potentials (LFPs, reflecting the coherent activity of cell assemblies) and the activity of individual neurons [Quian Quiroga and Panzeri, 2009]. Technical details of this technique will be described in Section 2.1.

Experimental studies based on human single cell recordings in the human MTL started almost 40 years ago, with the investigation of memory-related single-cell activity [Halgren et al., 1978]. In the following years several studies have explored language processing, memory and perception in the temporal lobe [Ojemann et al., 1988, Heit et al., 1988]. Fried et al. [1997] reported the presence of neurons in the human MTL responding selectively to either faces or objects with an increase (excitatory) or decrease (inhibitory) in their firing rate. More recently, MTL neurons have been found to fire selectively in response to specific categories of visual stimuli (e.g. animals, food, famous faces, landscapes) [Kreiman et al., 2000b] and even during imagery [Kreiman et al., 2000a].

1.4.1 Concept cells

The recent discovery [Quian Quiroga et al., 2005] of neurons that fire selectively and invariantly at the presentation of different photos of a particular person,

animal or place, has been a major breakthrough in the study of single cell activity in human MTL. These neurons are known as "Jennifer Aniston neurons" or, more recently, "concept cells". I will briefly outline here some of the more salient findings surrounding concept cells, for a more complete review see Quian Quiroga [2012].

Concept cells exhibit a high degree of visual invariance [Quian Quiroga et al., 2005]; they respond to the presentation of a specific subject, regardless of the particular features of the image (viewing angle, contrast etc.), as long as the subject of the image is recognised [Quian Quiroga et al., 2008b]. A following study found that these neurons could also be activated by different modalities of presentation, i.e. written or spoken name of the individual/object [Quian Quiroga et al., 2009]. Interestingly, this study also revealed different levels of invariance among the hippocampal areas, with neurons in the parahippocampal cortex being responsive only to visual stimuli, and those in the entorhinal cortex, amygdala and hippocampus being responsive to both visual and auditory ones. Moreover, neurons in the parahippocampal cortex have been shown to be more responsive (they fired in response to a higher number of stimuli in the session) but less invariant than the other MTL areas, while the hippocampus exhibits the highest selectivity of the MTL [Mormann et al., 2008, Ison et al., 2011]. The latency of responses in the MTL has been reported to be around 300-400 ms, with the exception of parahippocampal responses that happen about 50-100 ms earlier [Mormann et al., 2008, Quian Quiroga et al., 2009]. Since these responses occur much later than expected if they were direct projections from the IT cortex (latency in IT cortex is about 100-150 ms after stimulus onset), it leads to the belief that some lateral processing has to happen before the information reaches the MTL. These findings, about multimodal invariance, selectivity, and latency, are all in agreement with the hierarchical structure of the MTL seen in Section 1.2, with the hippocampus

lying at the end of a processing hierarchy that starts in the parahippocampal cortex.

The involvement of MTL neurons during memory retrieval has also been shown in single cell studies in human patients [Cameron et al., 2001, Gelbard-Sagiv et al., 2008]. In particular Gelbard-Sagiv et al. [2008] have shown that neurons activated selectively during the viewing of specific short videos also increase their firing rate right before the subjects' verbal recall of viewing the specific clip. In addition, MTL neurons have been shown to more likely fire to personally relevant concepts [Viskontas et al., 2009], in line with their involvement in memory processing.

Because of their peculiar features, concept cells are thought to encode the abstract content of the perception (its meaning) rather than the physical stimulus presented [Quian Quiroga et al., 2008a, Quian Quiroga, 2012]. Further evidence in support of an explicit representation in the MTL comes from Quian Quiroga et al. [2007] and Cerf et al. [2010]. In the first study the authors reported that it was possible to use a decoder to infer the identity of the presented stimulus from the firing patterns of MTL neurons, while it was generally not possible to distinguish between several pictures of the same subject. The latter study demonstrated that humans could voluntarily modify the firing of their MTL neurons to control an external hybrid image (superposition of two familiar stimuli), just by focusing onto a single image. In addition, other studies suggest directly that neurons in the MTL are related to the conscious aspect of perception [Kreiman et al., 2002, Quian Quiroga et al., 2008b].

Concept cells have been proposed to constitute the basis of declarative memory functions in the MTL [Quian Quiroga, 2012]. In this view, the level of abstraction is thought to vary according to the hierarchical organisation of the MTL, with the hippocampus encoding the highest level. Each concept is

represented by a network of cells whose activation brings the related concept into awareness and the association between related concepts is encoded in the partial overlap of the cell networks. In this configuration the creation of a new link between pre-existing concepts can happen very quickly, in accordance with experimental evidence.

1.4.2 Encoding of associations

One of the key mechanisms by which we create and retrieve declarative memory is association. Our brain has the ability to link novel information to pre-existing representations and to create new relationships between previously unrelated items. Therefore, MTL declarative memory functions (discussed in Section 1.3) involve, in particular, the encoding of associations between different elements [Rolls, 1996, Wallenstein et al., 1998, Eichenbaum, 2004]. Several neurophysiological and lesion studies in animals have proved the importance of the MTL for associative learning [Miyashita et al., 1989, Cahusac et al., 1993, Erickson and Desimone, 1999, Bunsey and Eichenbaum, 1996, Sakai and Miyashita, 1991, Day et al., 2003, McKenzie et al., 2014, Hattori et al., 2015]. In particular, during the execution of an object-place association task, Rolls et al. [2005] reported the presence, in the monkey hippocampus, of neurons that responded to: objects (independently of their locations); spatial views (independently of the object shown); and combinations of objects and places. Analogously to primate object-spatial view neurons, concept cells in the human MTL have been found to fire to a specific person or place [Quiñan Quiroga et al., 2005], and, recently, to combinations of a person and a spatial view [Ison et al., 2015]. The existence of separate as well as combined representations of places and objects (or people) is a fundamental requirement for an episodic memory system like the MTL [Rolls et al., 2005].

The previously mentioned work by Ison et al. [2015], based on single cell

recordings, reported that new associations could be formed very quickly in the MTL. In this study two unrelated pictures (a person and a place) were merged together to create a meaningful association. A large percentage of neurons initially firing in response to a single picture was found to start firing to the associate one as well within a few trials. Such fast creation of new associations is also found to be essential to episodic memory formation [Quian Quiroga, 2012].

Hippocampal representations of well-learned association have been described in both monkeys [Yanike et al., 2004] and rats [McKenzie et al., 2014]. Activation of hippocampal neurons have been reported also during retrieval of already consolidated associations in rabbits [Hattori et al., 2015].

Despite the evidences of MTL involvement in the encoding of associations, all of these studies were based on associative learning tasks performed by the subject. Therefore, they were not able to answer whether the MTL representations are only temporary and newly created each time or whether they persist in time (see discussion in Section 1.3.5).

Consistent with these findings, single cell recordings in human MTL have shown several cases of neurons firing to associated concepts (such as: two characters of the Star Wars saga; two co-stars in a TV show; some of the researchers performing the experiment with the patient etc.) [Quian Quiroga et al., 2005, 2009]. Starting from this observation, the first aim of this thesis is to perform a systematic analysis on MTL neurons responding to more than one stimulus, in order to evaluate whether these units respond more likely to related concepts (encode association), rather than to randomly paired stimuli.

Chapter 2

Experimental Methods

In this chapter, the experimental techniques and procedures employed for my PhD project are reported. In the first section I describe the equipment used to record intracranial activity in the human brain and the processing of the signal. Following this, I explain the experimental paradigm and the criteria to identify responsive neurons. In the last section I describe the two metrics used to measure the degree of association between items, their definition and validation.

2.1 Intracranial recordings

As seen in Section 1.4, intracranial recordings offer the opportunity to directly study neural activity in the MTL. The use of this technique in human subjects is possible in some exceptional circumstances, when patients suffering from pre-existing conditions (e.g. epilepsy, dystonia, Parkinson's disease) need invasive recordings for clinical purposes. Specifically, in the case of intractable epilepsy (in which seizures are not controllable with medication), surgical removal of the epileptic focus could be necessary. When seizures originate from deep regions of the brain and non-invasive monitoring cannot provide a concordant conclusion, the patient is implanted with depth electrodes and monitored for several days

in order to localise the epileptic focus [Engel et al., 2005].

The electrodes' locations are based on clinical criteria and preliminary information obtained by non-invasive techniques. Precise targeting is obtained through preoperative magnetic resonance imaging (MRI) and the use of a stereotactic frame. The positioning is verified after surgery by a co-registration of computed tomography (CT) scan and pre-operative MRI [Rey et al., 2014b].

Depth clinical electrodes (macroelectrodes) consist of several low impedance platinum-iridium contacts arranged along a polyurethane probe (see Figure 2.1). They allow the recording of an intracranial electroencephalogram (iEEG) and to localise the origin of epileptogenic discharges or seizures. Microwires can be inserted into the core of the clinical probe, allowing the acquisition of single cell activity and depth local field potential (LFP) [Rey et al., 2014b].

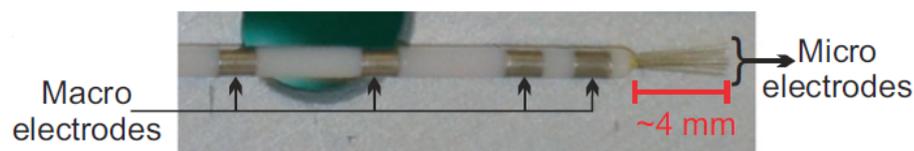


Figure 2.1 – Microwires for single cell recording are inserted along the inner part of the clinical depth macroelectrodes. Adapted from Rey et al. [2014b]

These microelectrodes are able to record the extracellular field potential, that is, the mean electrical potential in that point of the extracellular medium, generated by the contribution of all active processes within a volume of neural tissue. The different contributions from individual sources depend on: the distance between source and recording site (potential's amplitude scales with the inverse of the distance), the inclusion of interfering signals, and the attenuation characteristic of the neural tissue. Since passive neurons act as capacitive low-pass filters, this attenuation affects fast-rising events (e.g. action potentials) much more than slower ones (e.g. postsynaptic potentials) that are therefore able to propagate much farther in extracellular space [Logothetis, 2003, Buzsaki,

2006, Buzsaki et al., 2012].

The extracellular signal recorded can be processed in two different ways in order to extract the LFP or the spiking activity (see Figure 2.2):

- By low-pass filtering the recorded extracellular signal (typically up to 100-300 Hz), we obtain the LFP signal [Quian Quiroga and Panzeri, 2013]. LFPs capture the slow components of neural assembly dynamics in the tissue surrounding the electrode. They reflect several subthreshold integrative processes, including synchronised synaptic potentials [Mitzdorf, 1985, Logothetis, 2003], afterpotentials of somatodendritic spikes [Gustafsson, 1984], and voltage-gated membrane oscillations [Harada and Takahashi, 1983, Kamondi et al., 1998].
- By high-pass filtering the signal (typically over 300 Hz) we obtain the superimposed spiking activity of neurons close-by the electrode tip, plus the background activity of neurons further away [Quian Quiroga and Panzeri, 2009, Rey et al., 2015]. In fact, a neuron firing a fast action potential produces a large-amplitude deflection (spike) in the electrical field recorded near its soma [Buzsaki et al., 2012]. The extracellular action potential recorded from each neuron has a particular shape. The spike amplitude and waveform are determined by the cell characteristics (size of the soma, morphology of dendritic tree, and ionic channel distribution), the relative positioning of recording electrode and cell, and the properties of the extracellular medium [Gold et al., 2006]. These spikes can be detected from the high-pass filtered signal and then classified to separate the activity of different putative neurons (a process known as spike sorting [Quian Quiroga, 2007]) as we will see, in detail, in the following section.

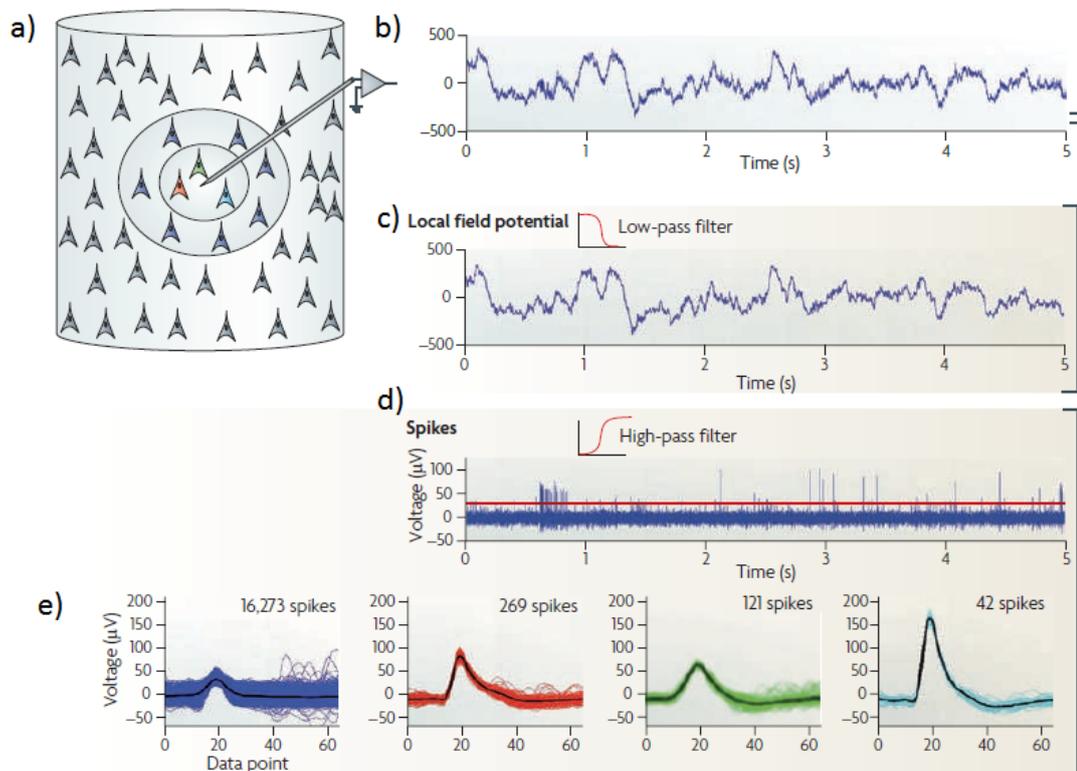


Figure 2.2 – Procedure for the processing of extracellular recordings. (a) Tip of the microelectrode recording the activity of nearby neurons. The amplified raw signal (b) can be low-pass filtered to obtain the local field potential (c) or high-pass filtered to obtain the single cell activity (d). In the high-pass filtered trace (d) the firing of nearby neurons appears as spikes on top of the background noise. The red horizontal line is the amplitude threshold used for spike detection. (e) Spike shapes are sorted in four clusters using a sorting algorithm. For neurons closer to the electrode tip (50–100 μm), the activity of each single unit can be distinguished (inner circle in panel a; spikes in red, green and cyan in panel e). For more distant neurons (up to approximately 150 μm), spikes can be detected but the signal-to-noise ratio is too small to differentiate them. These spikes are grouped together in a multi-unit cluster (outer circle of panel a, spikes in blue in panel e). Spikes from neurons further away from the tip (shown in light grey in panel a) cannot be detected and contribute to the background noise. The example data shown corresponds to five seconds of recording in the left hippocampus of an epileptic patient. Adapted from Quian Quiroga and Panzeri [2009].

2.1.1 Spike detection and sorting

In our experimental setup we employed platinum-iridium microelectrodes with a diameter of about $40\mu\text{m}$ and high impedance (300-1000 $\text{k}\Omega$). Each clinical probe had a total of nine microwires protruding from its tip, eight active recording channels and one reference channel, used to obtain a differential signal. The differential signal from the micro-wires was amplified, filtered and sampled [Rey et al., 2014b]. Different recording systems were used across the two studies to acquire the signal. Details about the recording systems and numbers of sessions/patients recorded with each of them will be given in Chapter 3 and 4, respectively, for the two separate studies presented. Once the preprocessed signal was acquired, it was then processed offline to extract the single cell activity. The spike detection and sorting were performed using the Wave_clus algorithm. The algorithm procedure will be outlined here, but further details and performance evaluation can be found in Quian Quiroga et al. [2004].

The first step of offline spike detection processing consists of high-pass filtering the signal (between 300 Hz and 3000 Hz) to separate the neural action potential from the low frequency activity. After filtering, spikes are detected using an amplitude threshold. The threshold (Thr) is based on the estimation of the standard deviation of the background noise (σ_n) and is defined as:

$$Thr = 4\sigma_n; \quad \sigma_n = \text{median} \left(\frac{|x|}{0.6745} \right) \quad (2.1)$$

where x is the filtered signal (see Rey et al. [2015] for details on the definition of σ_n). The standard deviation of background noise is preferred to the standard deviation of the signal because it reduces the interference of spikes that could lead to very high threshold [Quian Quiroga et al., 2004]. All the detected spikes are aligned to their maximum to reduce temporal jitter and for each spike 64 samples (corresponding to $\sim 2.2\text{ms}$) are saved.

The following step is the extraction of relevant features from the spike shapes. For this purpose, `Wave_clus` employs the wavelet transform, a time-frequency decomposition of the signal that gives optimal resolution in both time and frequency domains [Quian Quiroga et al., 2004]. Briefly, the spike waveforms are convolved with wavelet functions of different sizes to obtain a set of coefficients that represents the signal decomposition at the various scales [Rey et al., 2015]. For each spike a four-level multi-resolution decomposition [Mallat, 1989] using Haar wavelets is calculated, obtaining 64 wavelet coefficients. Afterwards, the ten most informative coefficients (i.e. the ones that give the best separation between spike classes) are automatically selected. The selection is based on the criterion of largest deviation from normality, since a coefficient able to separate different spike shapes is expected to have a multimodal distribution [Quian Quiroga et al., 2004].

Finally, a superparamagnetic clustering (SPC) [Blatt et al., 1996] procedure is used to separate the spikes in classes [Quian Quiroga et al., 2004]. The SPC algorithm implements simulated interactions between each data point (i.e. each spike) and its K -nearest neighbours. The m selected features of each spike are used to calculate the interaction between spikes in an m -dimensional space, while a Monte Carlo procedure simulates the iterations. The SPC is based on a generalisation of the Ising model, where instead of spins with values $\pm 1/2$, there are q different possible states per particle [Potts, 1952, Heermann and Binder, 1988]. In this model, the temperature parameter controls the probability of neighbouring points to change state together, with lower temperatures corresponding to higher probability. Changing the temperature, the system transitions from the ferromagnetic phase (low temperature), where all the spikes change state together (i.e. grouped in a single cluster), to the paramagnetic phase (high temperature), where the spikes mainly change state independently and are grouped in many clusters with a few elements each. In between the two

there is the superparamagnetic phase, where only the points grouped together change their state simultaneously; in this phase the algorithm identifies clusters. The clustering results depend mainly on the temperature and are robust to small changes in the other parameters (e.g. k and q) [Blatt et al., 1996, Quian Quiroga et al., 2004].

The algorithm automatically localises the superparamagnetic phase using a criterion based on cluster sizes. A manual refinement of the final clustering is possible using the graphic user interface. After clustering, the SPC runs a template matching to assign all the unclassified spikes to a cluster.

The whole procedure leads to group the detected spikes into different clusters. Ideally, each cluster corresponds to a single neuron activity. However, spikes coming from neurons not close enough to the electrode tip have usually a low signal-to-noise ratio, therefore they cannot be separated accurately. These spikes typically converge in a “multi-unit” cluster (e.g. Figure 2.2e, blue cluster) whose spikes have a relatively low amplitude and violate the refractory period of single neurons (i.e. spikes appear within less than 2.5 ms) [Rey et al., 2014b].

The spike sorting procedure is essential when working with extracellular recordings as it allows identification of the simultaneous activity of several neurons. Moreover, it is particularly important for the study of very sparsely-firing neurons, since, without spike sorting, their activity could not be untangled from that of more active neurons [Quian Quiroga, 2007, Rey et al., 2014b].

2.2 Experimental paradigms

As seen in Section 1.4.1, neurons in the MTL can fire very selectively to a few stimuli. In order to identify responsive neurons and the stimuli eliciting responses in these neurons, we used a simple visual paradigm (*screening session*, see Figure 2.3). During a screening session, the patient sat in bed, facing a

laptop computer on which about 100 different pictures of known people and places (for example, actors, politicians and landmarks) were shown for one second, six times each, in pseudo-random order. At the end of each trial the patient had to answer whether the image depicted a person or not. This simple task was totally irrelevant to the goal of our study and it was used only to make sure the subject paid attention during the stimulus presentation. Each screening session lasted about 30 minutes [Quian Quiroga et al., 2005, Rey et al., 2014b].

The pictures in the set were partially chosen according to each subject's interests and preferences. This was done in order to maximise the probability of finding responsive stimuli, as MTL neurons are known to respond preferentially to personal relevant concepts [Viskontas et al., 2009].

Data from screening sessions were employed for the project presented in Chapter 3. The paradigm used in the study described in Chapter 4 was a longer version of the one described above (*long screening session*). In this setup, fewer images (about 20) were presented for one second each in pseudo-random order but many more times (25 to 30 trials). The images for the long screening sessions were selected among the responsive stimuli identified from a previous screening session.

2.2.1 Responsiveness criteria

Data from each screening session were processed as detailed in Section 2.1.1. Neurons were classified in single- or multi-units based on the following: 1. the spike shape and its variance; 2. the ratio between the spike peak value and the noise level; 3. the interspike-interval (ISI) distribution of each cluster; 4. the presence of a refractory period for single units (<1% spikes within <3 ms ISI) [Quian Quiroga et al., 2009].

For the study presented in Chapter 3, the units responding significantly to

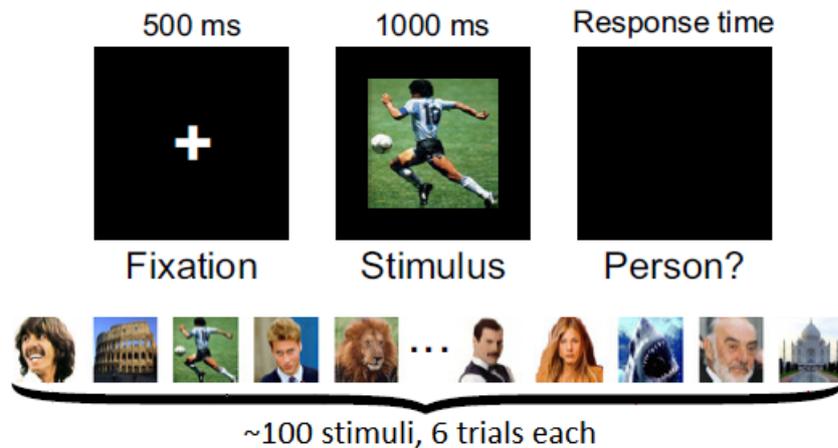


Figure 2.3 – Screening session paradigm. In each trial a fixation cross is presented on the screen for 500 ms. Next, a picture is shown for 1000 ms. Then the screen goes black, and the patient has to press a key to indicate whether he saw a person or not. The trial ends with random inter-trial interval (between 600 and 800 ms). Each screening session contains a set of ~ 100 pictures of animals, celebrities and landmarks, each of them presented six times. Adapted from Rey et al. [2014b].

a particular picture were identified using a heuristic criterion as in previous works [Quiñ Quiroga et al., 2009]. For each stimulus, spikes were aligned with the stimulus onset time, taken as time 0, for all the trials. A *response window* between 200 and 1000 ms, and a corresponding *baseline window* between -1000 and -200 ms were defined. A unit was considered to have a significant response to the given picture if it fulfilled the following criteria: (i) the median number of spikes in the response interval was higher than the average baseline (across all stimuli) plus 5 s.d.; (ii) the median number of spikes in the response window was at least 2; and (iii) the number of spikes in the response period was significantly higher than the one in baseline, according to a paired t-test ($p < 0.05$).

As in previous studies based on the long-screening sessions [Rey et al., 2014a,b], in our second study (Chapter 4) we applied a responsiveness criterion similar to the previous one. In this case the *response window* was defined between 200 ms and 700 ms while the *baseline window* was between -900 ms and -100 ms. A unit response was considered significant if: (i) the median

number of spikes in the response window was higher than the average baseline plus 5 s.d.; (ii) the median number of spikes in the response window was at least 1; (iii) the p-value of a sign-test between the spike count on each trial for the post stimulus and baseline windows was less than 0.01; and (iv) the instantaneous firing rate (calculated by convolving the spike train with a Gaussian kernel) crossed over a threshold for at least 75 ms. The threshold was computed as the average baseline value plus 4 s.d.

2.3 Association metrics

The main aim of this project was to analyse the responses from MTL neurons in relation to the degree of association between items eliciting multiple responses in the same neuron. Therefore, we needed a way to quantify the degree of association between the stimuli presented in our recording sessions. For this purpose, we employed two different metrics: one based on evaluations given by the patients themselves, and the other one based on the results of web-searches.

2.3.1 Personal association metric

After all recordings were completed, 12 of our patients (that participated in 24 screening sessions from study in Chapter 3) were asked to fill a ‘personal association matrix’. In this matrix they ranked between 0 and 10 how much a subset of between 10 and 15 pictures from the recording sessions were related to each other (between 45 and 105 comparisons in total). Each of the matrices filled by the subjects was normalised to standard z-score according to its mean μ and standard deviation σ (*normalised score* = $(x - \mu)/\sigma$). Figure 2.4 contains an example of the personal association matrix obtained by one patient.

The personal metric is, in theory, the best way to evaluate the associations between pictures. In fact, it captures the personal perception and valence of the

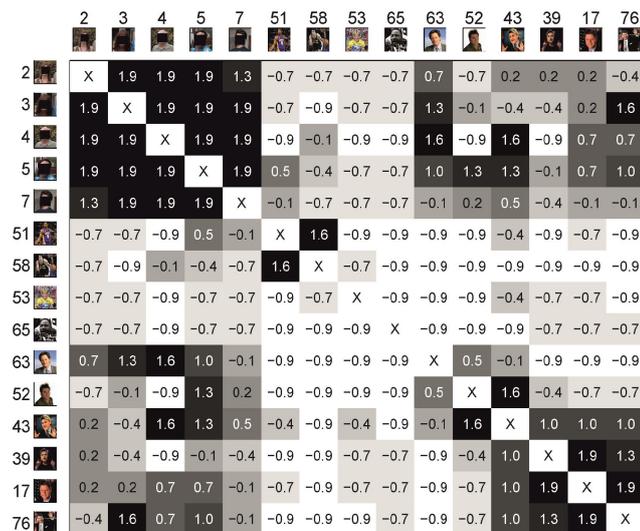


Figure 2.4 – Personal association matrix filled by one patient, with z-score normalisation. Note the cluster of associated stimuli in the top-left corner corresponding to 5 different patient’s relatives (stimuli 2, 3, 4, 5, and 7; pictures covered for confidentiality reasons).

relationship between concepts related to personal experiences and memories. However, this metric has a big limitation since we can only obtain ratings for a small subset of stimuli (for a single screening session with 100 stimuli, the complete association matrix should contain about 5000 ratings). To overcome this limitation and to include sessions where we did not have personal scores by the patients, we decided to use a second, more generic, way to estimate the degrees of association between concepts, employing web-search results.

2.3.2 Web-based association metric

Today, the World Wide Web represents the largest easy-to-access source of information about virtually everything. Web search engines, such as Google, Bing or Yahoo! have huge searchable databases of web documents updated by regularly crawling the web. The *hit count* is the number returned by a search engine in response to a submitted query. It is an estimation of the number of web pages indexed that contain that specific query. Search engine hit counts

are currently widely adopted for several research and commercial purposes and their reliability has been investigated by a few studies [Thelwall, 2008, Funahashi and Yamana, 2010].

Our web-based metric employs the hit counts to estimate the degree of association between two concepts, comparing the number of hits of a joint web-search with the numbers of hits obtained from the individual searches. This measure is based on the Pointwise mutual information (PMI) [Fano, 1961]. PMI is a well-known measure of association widely used in information theory, statistics and linguistics that has also been employed in conjunction with web-search results, for example, in the context of word synonyms [Turney, 2001]. Our metric relies on the idea that the names of associated concepts will appear more often together in a database of web pages rather than non-associated ones. The association score for each pair of pictures i and j was defined as:

$$a_{ij} = \log_2 \left(\frac{\text{hits}(\text{concept}_i \text{ AND } \text{concept}_j)}{\text{hits}(\text{concept}_i) \cdot \text{hits}(\text{concept}_j)} \right) \quad (2.2)$$

where *concept* refers to the searched name for each picture and the AND operator used in an internet search gives the number of pages containing both searched names.

We employed the Bing search engine for our metric, as it gave a more reliable estimation of associations compared to Google (see next section). We limited the web search to famous people and places; that is, to those concepts that are ‘searchable’ on the web and give a reasonable number of hits to have a reliable statistic (excluding names of family members). As with the personal matrices, the values for each recording session were normalised using a z-score.

Considering the whole set of 611 different stimuli (from the dataset of 99 screening sessions in the study presented in Chapter 3) whose degrees of association were estimated using the defined web-base metric, we obtained a web association matrix of 611×611 elements (Figure 2.5). The stimuli were

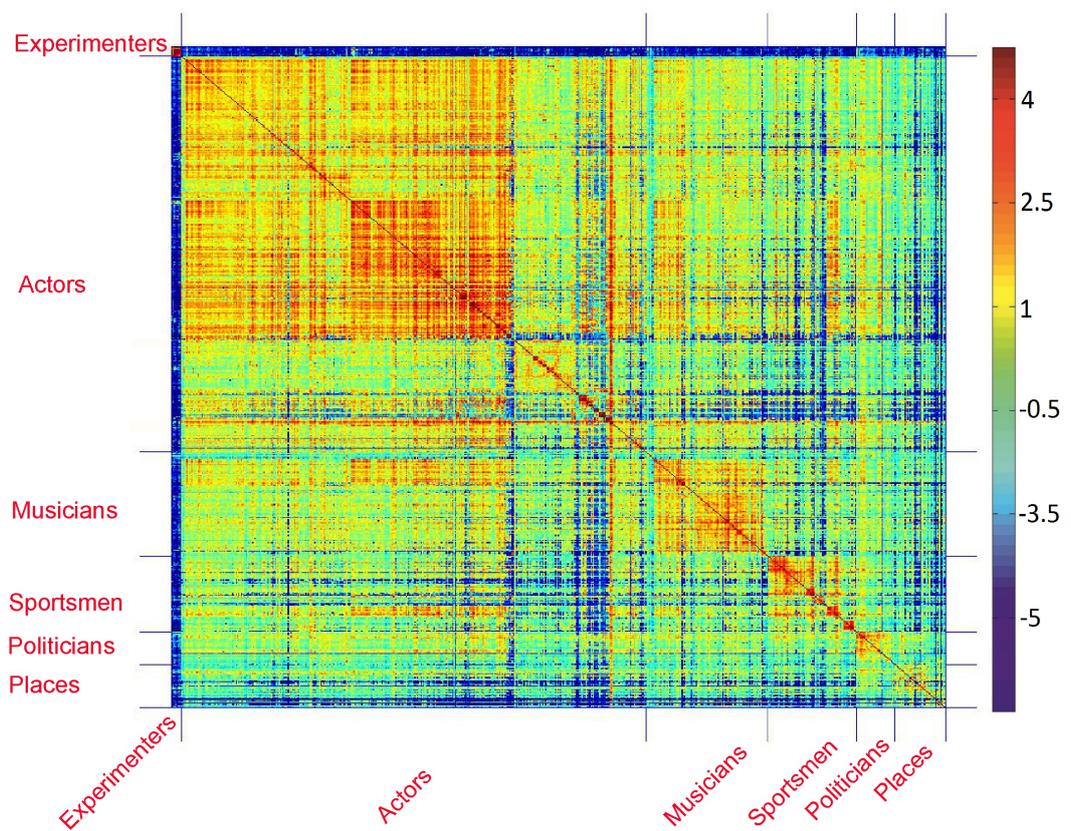


Figure 2.5 – Web-based association values between the 611 stimuli presented in the screening sessions from the study in Chapter 3 with z-score normalisation. Note the clustering of values according to the broad categories of the images (manually assigned categories are shown on the left/bottom). The colour bar on the right shows the strength of association values in arbitrary units.

manually assigned to broad semantic categories (actors, sportsmen and so on, see categories listed in Figure 2.5) and sub-categories when possible (e.g. the stimuli in the sportsmen category were further classified according to each particular sport). The association matrix was sorted according to this classification and the underlying cluster-like structure of the data, with blocks of highly associated elements along the main diagonal, emerged. A clustering analysis using a modularity optimisation algorithm (Louvain community detection, [Blondel et al., 2008, Rubinov and Sporns, 2010]) with the resolution parameter tuned to consider the same number of broad categories led to an 84% overlap with our manual classification.

2.3.3 Correlation personal and web-based metrics

The scores shown in Figure 2.5 reflect ‘universal associations’ between items, given by the shared inputs of billions of internet users, whereas the inputs by the patients reflect not just universal associations, but also relationships given by the subject’s preferences and personal experiences. We tested the reliability of the web-based association metric by measuring the cross-correlation between the normalised association scores given by the subjects and the ones obtained with the web searches for the same pairs of items (excluding, in both cases, family members). For each subject we calculated the scalar product between personal and web-based evaluations. To assess statistical significance, we compared, using a rank test, the original cross-correlation value with the ones obtained from a distribution 1,000 surrogates, created by randomly permuting the index of one of the association scores. At the single subject level, we found a positive and significant ($p < 0.05$) correlation in 11 out of 12 subjects, with a mean correlation coefficient of 0.4.

At the population level we found that the overall correlation for the whole set of evaluations (i.e. the one obtained by pulling together all the data points from

the 12 subjects) was significantly higher than chance ($p < 10^{-21}$, Spearman's rank correlation test). Moreover, when we compared the original correlation values with the median of the surrogates for each subject, using a right-sided (that is, our hypothesis is that the original values are higher than the ones of the surrogates) Wilcoxon signed-rank test, we found a significant effect ($p < 10^{-3}$).

In order to illustrate the correlation between the personal and web-based association scores, we considered all the web-based association scores for which we also had the personal score given by the patient; then we binned these web-based scores into 10 equally spaced intervals, and plotted the mean (\pm SEM) value in each bin (x axis) against the mean (\pm SEM) value of the corresponding personal score (y axis) (see Figure 2.6). In this plot we observed a strong correlation for highly associated items (for example, Bill Clinton and Hilary Clinton were clearly associated both for the patients and for the web) and more variability for those that are less associated. On the basis of these results, we could consider the universal web-association scores as a proxy for the personal scores by the patients, which, in spite of individual differences, gave a reasonable approximation on average.

The results reported so far refer to the web-based metric obtained employing Bing search engine. We also tested the results obtained with Google search engine. However, the Google-based metric did not correlate well with the personal one (only 4 matrices out of 12 were significantly correlated, with a mean correlation coefficient of 0.14). This was probably due to the instability of its hits count that does not seem to be accurate enough Funahashi and Yamana [2010], therefore we chose to use Bing for our study.

2.3.4 Definition of mean association scores

For each neuron with a significant response to two or more images (pluri-responsive neuron) we defined a mean association score between the pair of

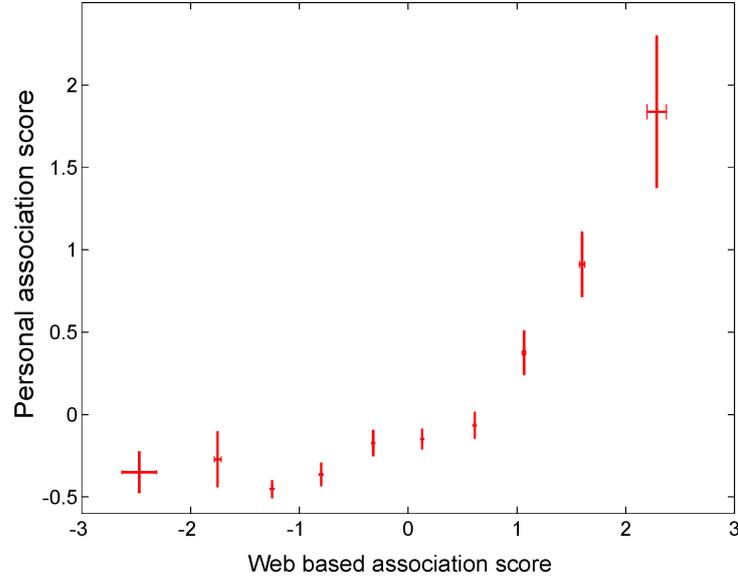


Figure 2.6 – Correlation between the personal and the web-based association scores (mean \pm SEM). The positive correlation between the two metrics was found significant both at the single subject level and at the population level.

images eliciting responses (AS_{R-R}), and a mean association score between the pair of images, where one elicited a response and the other did not (AS_{R-NR}). In other words, for each responsive picture we consider its association scores with the other pictures eliciting responses and with the ones not eliciting responses. Specifically, if N is the total number of pictures presented in the session, and k is the number of responsive stimuli for the given neuron then $N-k$ is the number of non-responsive ones. We identify with indices $[1, \dots, k]$ the responsive pictures and with indices $[k+1, \dots, N]$ the non-responsive ones. Then, for each pluri-responsive neuron the mean association scores were defined as follows:

$$AS_{R-R} = \frac{1}{\binom{K}{2}} \sum_{i=1}^{k-1} \sum_{j=i+1}^k a_{ij} \quad (2.3)$$

where a_{ij} are the z-score-normalised association scores from the personal matrices (or from equation 2.2), and the binomial coefficient $\binom{K}{2}$ is the total number of possible pair combinations with k responsive stimuli. Analogously,

we defined the mean association score for the other pairs as:

$$AS_{R-NR} = \frac{1}{k \cdot (N - k)} \sum_{i=1}^k \sum_{j=k+1}^N a_{ij} \quad (2.4)$$

where the notation is the same as above.

For each neuron, the first score (AS_{R-R}) indicated the average strength of association between pairs of responsive pictures that we compared with the average strength of association between all pairs of responsive and non-responsive pictures (AS_{R-NR}). The results obtained will be presented in Chapter 3 and 4 for the two separate studies.

Chapter 3

Long-term coding of personal and universal associations in the human MTL

In this chapter I describe the results of my main PhD project about the encoding of long-term associations in the human MTL. These results have been recently reported in a peer-reviewed journal [De Falco et al., 2016].

As seen in Section 1.4.2, the role of hippocampus and its neighbouring MTL structures in declarative memory functions involves, in particular, the encoding of associations between items. However, this involvement has only been shown through studies based on associative learning tasks, leaving unclear whether the MTL engagement is confined to the task execution or goes beyond it. In fact, MTL neurons could be showing their involvement (i.e. responding to learned associations) during the execution of associative tasks, but then carry out a different function afterwards. Moreover, there is, so far, no direct evidence of MTL neurons coding naturally-formed associations (i.e. not acquired through an associative learning task). The available evidences leave open the question whether associations are encoded temporarily in the MTL or persist in a more

stable representation that outlasts the task execution (see discussion in Section 1.3.5). To address this issue, following previous observations of MTL neurons firing to well-known associated concepts (e.g. two co-stars in a TV show, Quian Quiroga et al. [2005]), we performed a systematic study to determine if these co-activations were just random coincidences or if there is a consistent tendency for MTL neurons to encode meaningful associations, independent of the execution of an associative learning task. To do this, we analysed MTL neurons' responses to presented images in relation to two different association metrics (Sections 2.3.1 and 2.3.2) measuring the degree of association between stimuli.

3.1 Subjects and recordings

The data came from 99 experimental sessions in 49 patients with pharmacologically intractable epilepsy recorded at University of California, Los Angeles. Patients were implanted with chronic depth electrodes for 7–10 days to determine the seizure focus (see Section 2.1). All studies conformed to the guidelines of the Medical Institutional Review Board at University of California, Los Angeles, and all patients provided signed consent forms. I performed the data processing and analysis described in this chapter, but I was not personally involved in the data recordings.

The electrode location was based exclusively on clinical criteria; here we report findings from sites in the hippocampus, amygdala, entorhinal cortex and parahippocampal cortex. For the first 41 patients (84 sessions), the differential signal from the micro-wires was amplified and recorded using a 64-channel Neuralynx system (filtering between 1 and 9,000 Hz and sampling at 28 kHz), while for the remaining patients we used a 128-channel Blackrock system (filtering between 0.3 and 7,500 Hz and sampling at 30kHz). The experimental

paradigm employed (screening session) is described in Section 2.2. For all the 99 screening sessions the signal was processed as described in Section 2.1.1 to detect and separate the neural spikes. Through the spike sorting process, we identified a total of 4078 units; 1648 single units (SU) and 2430 multi-units (MU). Single- and multi-units were classified according to the spike shape and cluster properties as detailed in Section 2.2.1. For each unit, we identified the significant responses according to the criteria in Section 2.2.1. In particular, we found 260 single- and 290 multi-units that were responding significantly to at least one picture, and, among them, 129 single- and 132 multi-units responsive to more than one picture. We will refer to these units as pluri-responsive neurons in the following text.

3.2 Personal association metric

From the 24 screening sessions (from 12 patients), for which we had the personal evaluations of associations (see Section 2.3.1), we found a total of 32 units firing significantly to more than one stimulus (mean: 3.1 pictures per neuron; s.d.: 4.8); 19 of them were classified as single-units and 13 as multi-units. For each of these pluri-responsive neurons, from the related personal association matrix, we calculated a mean association score across pairs of responsive stimuli (AS_{R-R} , equation 2.3) and a mean association score across pairs of responsive and non-responsive stimuli in the session (AS_{R-NR} , equation 2.4).

Figure 3.1a shows an example of a single unit firing to the pictures of three relatives of the patient and two celebrities. Note that although these responses do not appear very strong in terms of absolute and sustained firing level, their firing rates were significantly higher compared to the neuron's spontaneous activity (as assessed with the responsiveness criteria in Section 2.2.1). From the patient's association matrix shown underneath (Figure 3.1b), the relatives

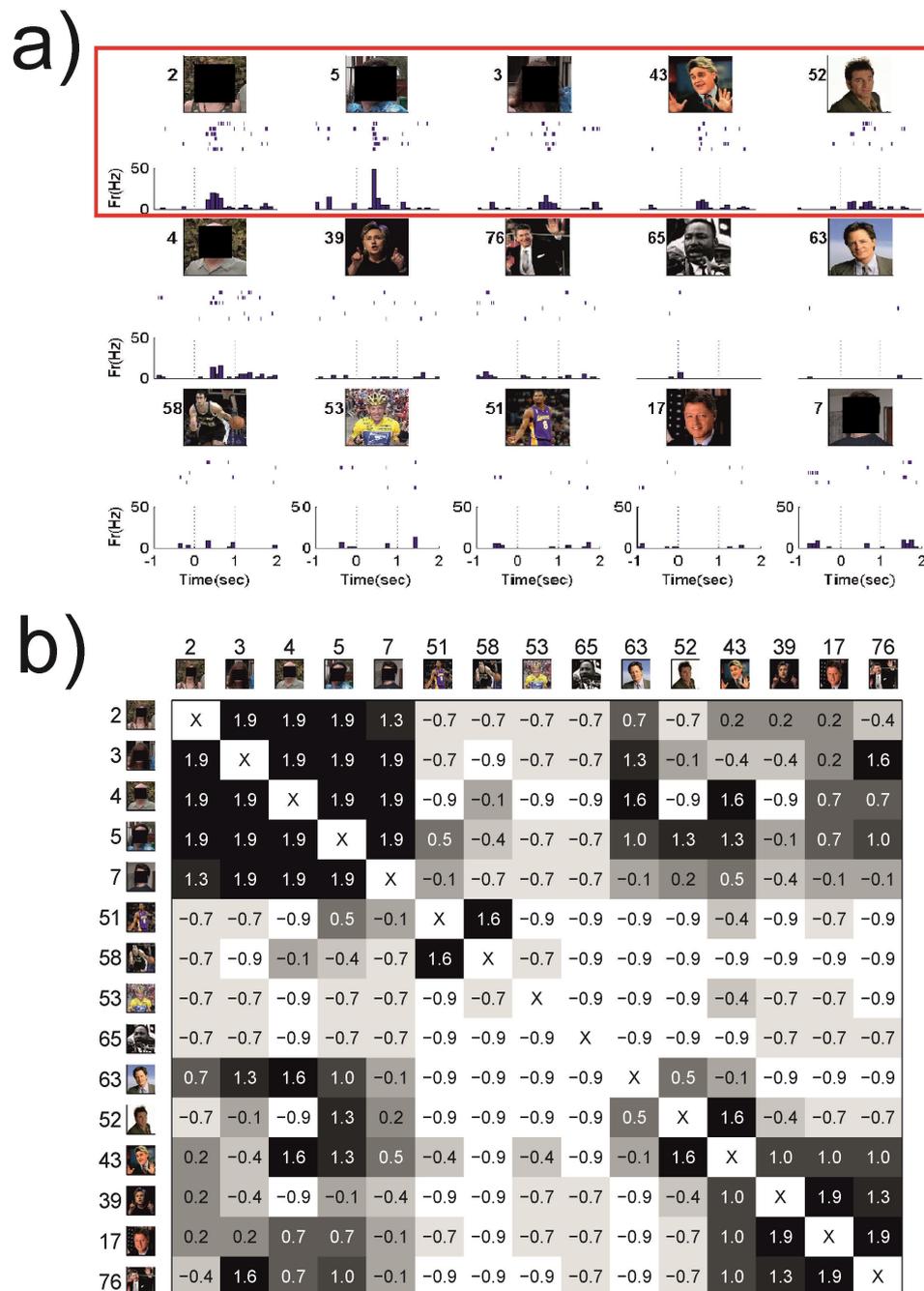


Figure 3.1 – (a) A single neuron in the hippocampus that fired to pictures of the patient’s relatives (pictures covered for confidentiality reasons) and of two celebrities that were related, according to the patient’s report, to one of the relatives (stimulus 5). For each stimulus, the raster plot containing the single spikes detected at each time (each row is associated to a trial, starting from the top) and the firing rate histogram are shown. Time zero marks the onset of picture presentations which lasted for 1 sec. (b) Personal association matrix filled by the patient with a z-score normalisation. The association scores for the pairs of pictures to which the neuron fired (mean: 0.89) were higher than the scores obtained for the other pairs of pictures (mean: 0.12).

appear highly related to each other and the two celebrities appear related to one of the relatives (stimulus 5). In this example, the Association Score for pairs of responses ($AS_{R-R} = 0.89$) was higher than the score obtained for the other pairs ($AS_{R-NR} = 0.12$), meaning that the stimuli eliciting response in this neuron were more associated between each other than with the rest of the stimuli in the session. Moreover, it is interesting to notice that this neuron did not fire to all family members (for example, it did not fire to stimulus 7) or celebrities, but only to a subset of the people that the patient considered to be related.

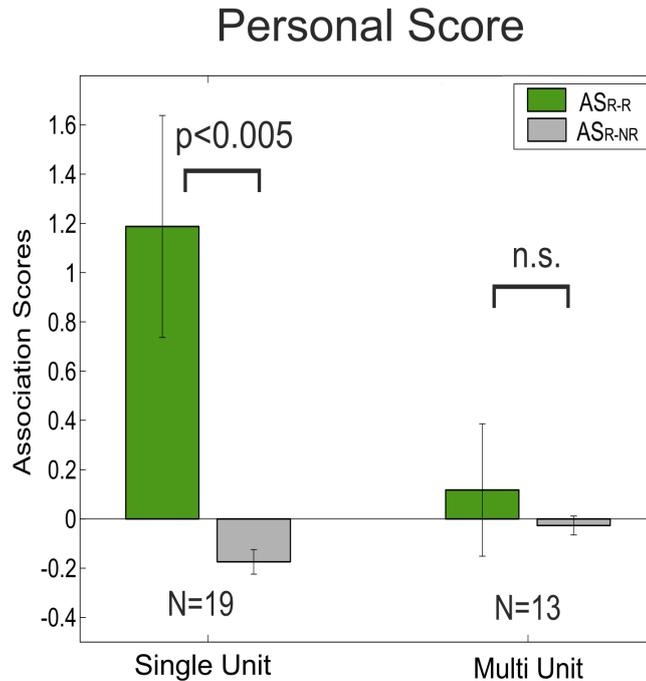


Figure 3.2 – Mean across units ($\pm SEM$) of association scores for pairs of responses (AS_{R-R} , green) and for the other pictures pairs (AS_{R-NR} , grey) based on the patients’ association scores. For single units, the population of AS_{R-R} values was significantly higher than the population of AS_{R-NR} values according to a right-sided Wilcoxon signed-rank test.

At the population level, for the 19 single-units (SU) and 13 multi-units (MU) with more than one significant response, we compared the association scores on responsive pairs (AS_{R-R} values) with the association scores for the

other pairs (AS_{R-NR} values). Figure 3.2 shows this comparison for each of the two groups of units: here the green bar and grey bars represent the mean values of AS_{R-R} and of AS_{R-NR} , respectively, across the units considered. For the single-unit population the AS_{R-R} values were significantly higher than the AS_{R-NR} , according to a right-sided Wilcoxon signed-rank test ($p < 0.005$). The same tendency was observed for the population of 13 multi-units, although in this case the difference was not significant (which can be attributed to the fact that some of the response pairs may correspond to the firing of different neurons).

Table 3.1 contains a summary of the numbers of neurons tested against the personal association metric. Frequency distribution of number of images to which these neurons responded to is shown in Figure 3.3.

	Units	Single units	Multi units
Recorded units	777	396	381
Units responding to at least 1 stimulus	86	54	32
Units responding to at least 2 stimuli	32	19	13
Units responding to related stimuli according to the personal association metric ($AS_{R-R} > AS_{R-NR}$)	19	14	5
Units responding to unrelated stimuli according to the personal association metric ($AS_{R-R} \leq AS_{R-NR}$)	13	5	8
Units responding to a single stimulus	54	35	19
Units not significantly responding to any stimulus	691	342	349

Table 3.1 – Number of units tested against the personal association metric (24 recording sessions).

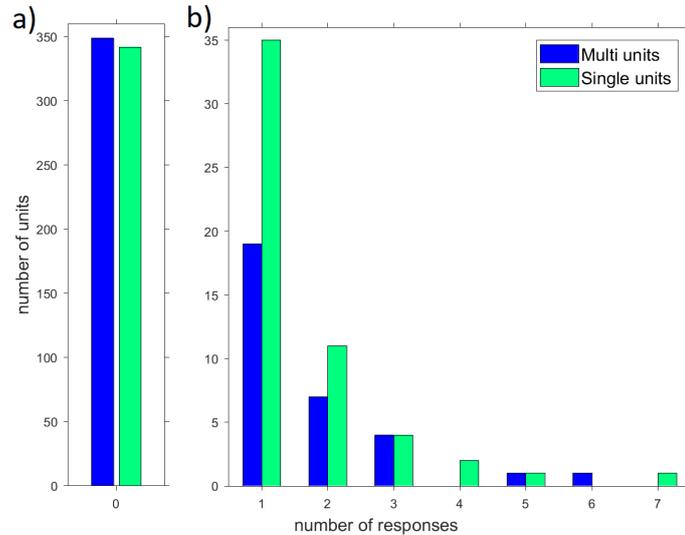


Figure 3.3 – Distribution of the numbers of responses across the neurons tested against the personal association metric (396 single-units and 381 multi-units). (a) Non-responsive units. (b) Units responding to one or more stimuli in the personal association metric (mean numbers of responses ($\pm SD$): $1.8(\pm 1)$ for multi-units and $1.6(\pm 1)$ for single-units). Note that the two panels have different y-axis scales for better visualisation.

3.3 Web-based association metric

To overcome the limitations of the patients’ association scores we defined a web-based association metric. This metric, discussed in detail in Section 2.3.2, estimates the degree of association between two concepts from the numbers of hits obtained as results of web searches. It allowed us to consider the whole set of “searchable” stimuli (famous people and famous places) across all the recorded screening sessions. The web-based metric was applied to the whole dataset of 99 experimental sessions (recorded in 49 patients), in which we had identified 261 units (129 single-units and 132 multi-units) with more than one picture eliciting a response (mean: 3.0 pictures per neuron; s.d.: 3.9). As we did before for the personal metric, we calculated the association scores AS_{R-R} and AS_{R-NR} (Section 2.3.4) for each pluri-responsive neuron using the web-based association metric. We then compared the two populations of values, for both

single and multi-units, with a right-sided Wilcoxon signed-rank test and found that pairs of images for which the neuron fired (AS_{R-R}) were significantly more associated than other pairs of images (AS_{R-NR}), for the single-units ($p < 10^{-7}$) and, in this case, also for the multi-units ($p < 10^{-5}$).

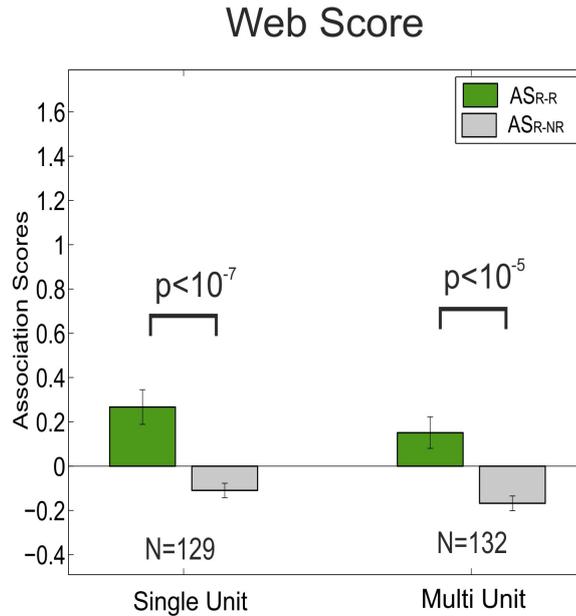


Figure 3.4 – Mean across units ($\pm SEM$) of association scores for pairs of responses (AS_{R-R} , green) and for the other pictures pairs (AS_{R-NR} , grey) based on the web-association scores. For both single- and multi-units the population of AS_{R-R} values was significantly higher than the population of AS_{R-NR} values according to a right-sided Wilcoxon signed-rank test.

Figure 3.4 shows this comparison which, in agreement with the one obtained with the personal association metric, showed that MTL neurons fire preferentially to associated items. Furthermore, for single units, the difference between the web-association values for the pairs of images to which the neurons responded and the other pairs ($AS_{R-R} - AS_{R-NR}$) (first two columns in Figure 3.4) were significantly lower than the ones obtained with the personal scores (first two columns of Figure 3.2), according to a Wilcoxon rank-sum test ($p < 0.05$); while the difference was not significant for the multi-units. This is in line with the fact that the web-based scores reflect universal associations, while the personal metric captures unique and more meaningful relationships for each

subject.

Several control tests, described in the following sections, ruled out the possibility that this result could be attributed to familiarity effects, visual similarity between the stimuli, broad semantic categorisations or recall of associated items.

	Units	Single units	Multi units
Recorded units	4078	1648	2430
Units responding to at least 1 stimulus	550	260	290
Units responding to at least 2 stimuli	261	129	132
Units responding to related stimuli according to the web-based association metric ($AS_{R-R} > AS_{R-NR}$)	187	92	95
Units responding to unrelated stimuli according to the web-based association metric ($AS_{R-R} \leq AS_{R-NR}$)	74	37	37
Units responding to a single stimulus	289	131	158
Units not significantly responding to any stimulus	3528	1388	2140

Table 3.2 – Number of units tested against the web-based association metric (99 recording sessions).

Table 3.2 contains a summary of the numbers of neurons tested against the web-based association metric. The complete frequency distribution of numbers of images to which these neurons responded to is shown in Figure 3.5. From the distribution of number of responses it was possible to calculate the mean fraction of stimuli in the set eliciting response across responsive units, which was: 2.5% ($\pm 2\%$) for single-units and 2.7% ($\pm 2\%$) for multi-units, in line with literature results on MTL neurons selectivity [Quiari Quiroga, 2012].

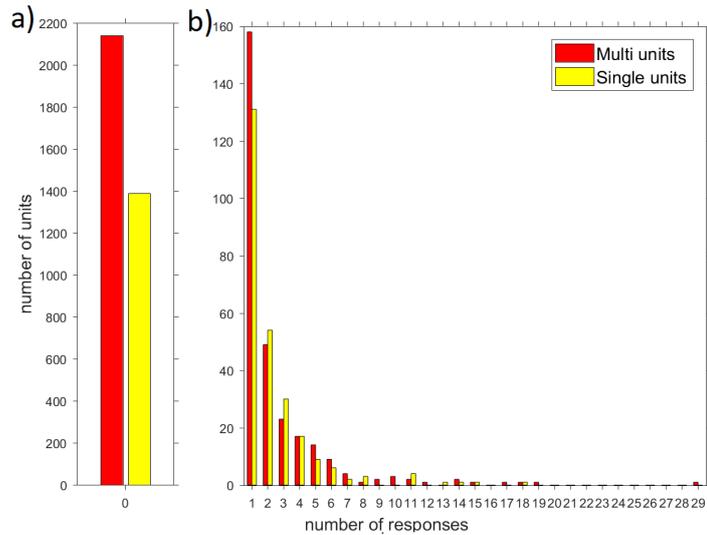


Figure 3.5 – Distribution of the numbers of responses across the neurons tested against the web-based association metric (1648 single-units and 2430 multi-units). (a) Non-responsive units. (b) Units responding to one or more stimuli (mean numbers of responses ($\pm SD$): $2.6(\pm 3)$ for multi-units and $2.4(\pm 2)$ for single-units). Note that the two panels have different y-axis scales for better visualisation.

3.3.1 Effect of relative number of hits

From previous findings [Viskontas et al., 2009], it is known that MTL neurons tend to respond more likely to familiar items (for example, a very famous actor, landscape and so on). Therefore, since familiar pictures are represented by more neurons, there is, in principle, a higher chance of finding neurons firing to pairs of familiar pictures compared with non-familiar ones. In this case it could be argued that our findings are simply due to a random overlap of cell assemblies firing to very familiar concepts, irrespective of the degree of association between them. To rule out this possibility we performed two control analyses based on the concept’s relative number of hits, where the number of hits for a single concept was taken as an estimation of its familiarity. First, for each pair of concepts we defined a joint-familiarity score as:

$$f_{ij} = \log_2(\text{hits}(\text{concept}_i) \cdot \text{hits}(\text{concept}_j)) \quad (3.1)$$

Then, we replaced a_{ij} with f_{ij} in equations 2.3 and 2.4 and calculated the familiarity scores for each pluri-responsive neuron. We found that the scores for the pairs to which the neurons responded (R-R) were not significantly different to the ones for the other pairs (R-NR) ($p = 0.8$ for single units and $p = 0.4$ for multi-units; Wilcoxon signed-rank test), thus showing that results cannot be explained by an effect introduced by the larger number of hits of familiar pictures.

Second, for each experimental session we divided the pictures of each category (actors, places, sportsmen and so on; categories as shown in Figure 2.5) into two groups: high familiarity (H; pictures with number of hits above the median) and low familiarity (L; pictures with number of hits below the median). This grouping in high and low familiarity was done by category to balance the fact that some categories were naturally more “familiar” (e.g. the actor category have on average a higher number of hits than the others). We then calculated the association scores (equations 2.3 and 2.4) for pair responses where pictures were both highly familiar (HH), both not very familiar (LL), and one highly familiar and the other one not (HL). Replicating the results shown in Figure 3.4 for each of the three subgroups (HH, LL and HL), AS_{R-R} was significantly higher than AS_{R-NR} in all groups, both for single- and multi-units ($p < 0.05$, Wilcoxon signed-rank test). Moreover, the difference in both AS_{R-R} and AS_{R-NR} across the three subgroups was not significant according to an ANOVA test ($p = 0.3$ for single- and $p = 0.08$ for multi-units), thus showing that our result could not be attributed to the familiarity of the pictures used.

3.3.2 Visual similarity analysis

To rule out the possibility that the tendency of MTL neurons to respond to associated concepts could just reflect perceptual similarities between pictures (e.g. a picture of an actor is more similar to a picture of another actor than to

one of a landscape), we performed a visual similarity analysis. For each pair of stimuli i and j , we estimated the visual similarity v_{ij} between them as the cross-correlation between the images (that is the normalised scalar product between images' pixels, each with 160x160 pixels, greyscale and z-score normalised). We then computed the visual similarity scores VS_{R-R} and VS_{R-NR} using equations 2.3 and 2.4 where a_{ij} was replaced with v_{ij} , for the whole population of pluri-responsive neurons. The difference between the scores obtained for response pairs and other pairs was significantly higher for the web-association metric ($AS_{R-R} - AS_{R-NR}$) compared to the visual similarity one ($VS_{R-R} - VS_{R-NR}$) for single-units ($p < 0.005$; Wilcoxon signed-rank test) while the difference was not significant for multi-units. Since the simple cross-correlation can be sensitive to alignment issues, we repeated this analysis employing both the 2D cross-correlation between the images and the Earth Mover's Distance [Rubner et al., 2000] as estimations of visual similarity. Both metrics gave significantly lower differences between the VS_{R-R} and the VS_{R-NR} values compared to the ones obtained with the web-based association scores ($p < 10^{-3}$ in all cases, both for the single- and the multi-units; Wilcoxon signed-rank test). This test showed that our results cannot be trivially attributed to perceptual similarity between stimuli. These results are in agreement with the fact that MTL neurons do not represent visual features, as they show visual invariance responding to completely different pictures of the same person [Quian Quiroga et al., 2005] and can also fire upon different sensory modalities of presentation of the same concept (to the person's written and spoken name) [Quian Quiroga et al., 2009].

3.3.3 Effect of semantic categories

Previous studies (e.g. Kreiman et al. [2000b]), reported neurons in the MTL responding to stimulus categories (e.g. all the animals or all the places presented). To test that the effect found in the association scores was not due to category

responses, we ran the following analyses. First, we verified that the result still stood when restricting our association scores calculation only to the stimuli presenting faces (the largest group in our set of stimuli). This excluded the possibility that the presence of place-selective cells could have driven our result. In the 99 sessions analysed we found 103 single-units and 99 multi-units (202 neurons in total) firing significantly to more than one face. For this population the degree of association for paired responses (AS_{R-R}) was still statistically higher than the degree of association for the other pairs (AS_{R-NR}) of stimuli both for single-units ($p < 10^{-5}$, Wilcoxon signed-rank test) and multi-units ($p < 0.05$).

Next, to test that our results were not simply due to broad semantic category responses, we verified that the degree of association for pairs of items eliciting responses was still higher than for the other pairs when constraining the comparisons to be within the same category (when the items eliciting responses belonged to the same category) or across the same categories (when they belonged to different categories). Specifically, considering the list of broad categories shown in Figure 2.5, for each category and each pluri-responsive neuron, we calculated AS_{R-R} and AS_{R-NR} from equations 2.3 and 2.4, considering only association values within each category (i.e. only the pairs of stimuli belonging to the specific category). Pulling together the association values for all the categories considered, the difference between AS_{R-R} and AS_{R-NR} remained significant for the single units ($p < 0.005$; Wilcoxon signed-rank test; number of values = 103) but did not reach significance for the multi-units ($p = 0.06$; Wilcoxon signed-rank test; number of values = 116). Similarly, for pairs of items eliciting responses that belonged to different categories (for example, an actor and a place), we constrained the comparisons to be across the same categories (in our example, comparing the original association value with the ones between other actors and places). Also in this case, the difference

between AS_{R-R} and AS_{R-NR} remained significant for the single units ($p < 0.01$; Wilcoxon signed-rank test; number of values = 102) and showed a tendency for the multi-units ($p = 0.06$; Wilcoxon signed-rank test; number of values = 112). These results proved that the association score difference described in Section 3.3 could not be just attributed to semantic categorisations.

3.3.4 Response latency

Another possible objection to our finding is that it is, in principle, possible that the responses to associated pictures were just caused by a cue-recall effect. In other words, it is possible that the neurons we recorded from encoded only one of the pictures, and then the presentation of a second associated stimulus acted as a cue, evoking the one encoded - thus making the neuron fire indirectly. This case is actually very unlikely given the large number of stimuli presented in a session (about 100 images) and since no associative learning task is performed. In fact, with such a large number of images it does not seem realistic that the subjects will spontaneously recall the same specific relationships at each trial. To further support our view, we ran a latency analysis on our responses. A study based on single neuron responses in human MTL in a cued-recall paradigm [Kreiman et al., 2000a], reported that the mean latency of human MTL responses obtained by evoking an image from a cue was 409 ms (s.d.: 291 ms), about 130 ms longer than the one they estimated for the visual responses (i.e. triggered by the picture presentations). Using the same method reported by Kreiman et al. [2000a], we estimated the latencies of our responses. Specifically, for each significant response, we obtained the spike density function (sdf) convolving the spike train with a Gaussian of 100 ms width and then averaging across trials. The latency was then computed as the time where the sdf crossed the baseline plus 2 s.d. value, and stayed up for at least 50 ms. The baseline value was computed as the average value of the sdf on all stimuli

estimated in the time window preceding the stimulus presentation (between -900 ms and -100 ms, where time zero is the time of stimulus onset). We obtained that the mean latency of our responses was 253 ms (s.d.: 129 ms), which is well below the one reported by Kreiman et al. [2000a] for cue-recall responses. In addition, the mean latency difference of the responses to different stimuli (in the same neurons) was of 78 ms (s.d.: 89 ms), while the latency difference reported for recall responses was more than 60% larger.

3.4 Location analysis

As seen in sections 1.2, hippocampus and parahippocampal regions are structured in a hierarchical way. Moreover, MTL neurons exhibit different levels of selectivity [Ison et al., 2011], visual invariance, and multimodal convergence [Quian Quiroga et al., 2009] across the distinct MTL areas, suggesting an increase of abstraction along the MTL hierarchy [Quian Quiroga, 2012]. In order to find out whether there were differences in the encoding of association across MTL areas, we separated our results according to the area from where they were recorded. First, for each of the four areas we recorded from (hippocampus, amygdala, entorhinal cortex and parahippocampal cortex), we compared the association scores AS_{R-R} and AS_{R-NR} (equations 2.3 and 2.4). For all the areas the association scores for pairs of responsive stimuli were significantly larger than the ones for other pairs (Figure 3.6, $p < 0.05$ in all cases; Wilcoxon signed-rank test).

Next, we compared the difference between association scores for response pairs and other pairs ($AS_{R-R} - AS_{R-NR}$) across the different MTL areas. For this analysis we pulled together single- and multi-units to increase statistical power (i.e. to have a good number of units in each area), given that results obtained with the web-based scores in the two groups were similar. We found

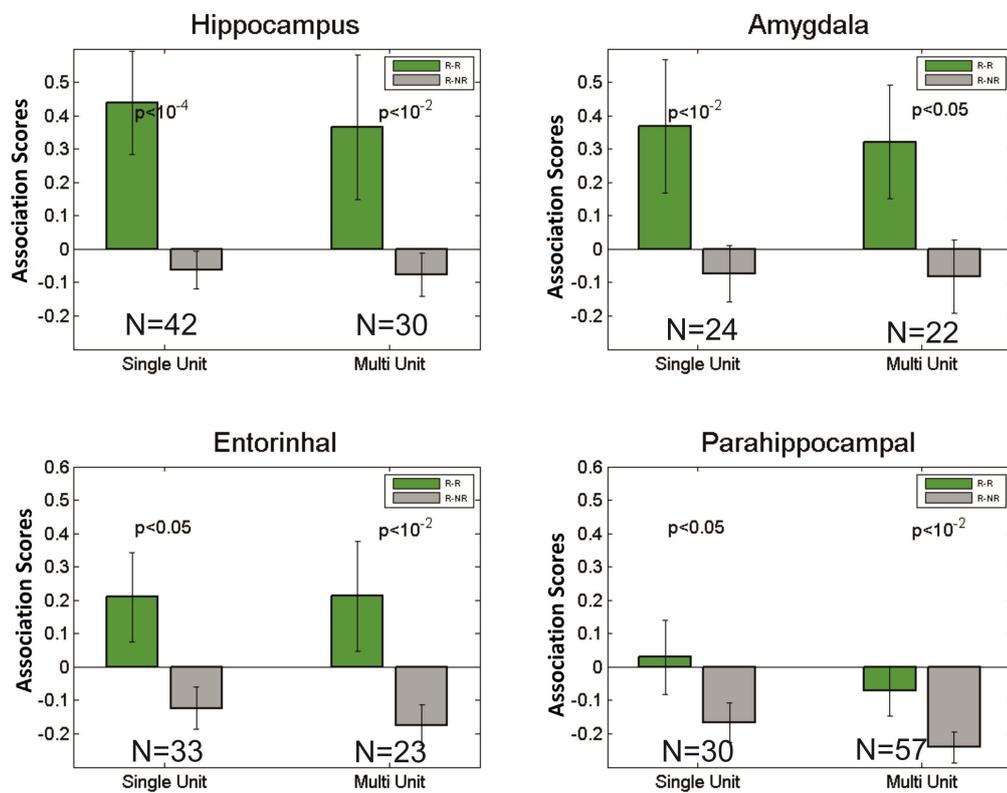


Figure 3.6 – Mean association scores for pairs of responses AS_{R-R} (green) and other pairs AS_{R-NR} (grey) in the different MTL areas. Difference between the two scores was assessed with a Wilcoxon signed-rank test.

a tendency for the score difference to vary across regions, being largest in the hippocampus and lowest in the parahippocampal cortex (Figure 3.7), which, however, did not reach statistical significance ($p = 0.07$; ANOVA test).

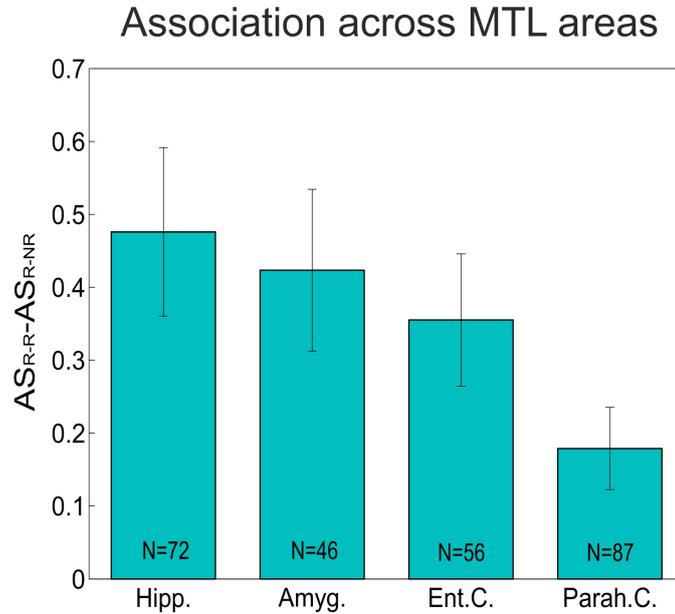


Figure 3.7 – Difference between mean association scores for the pairs of images eliciting responses across the different MTL areas for single- and multi-units together.

Furthermore, when looking at the responses separated by categories, we observed a larger proportion of responses to scenes in the parahippocampal cortex compared to other areas. This result was supported by a significant modulation of the proportion of responses to scenes across the different MTL areas (analysis of variance, $p < 10^{-6}$), while no effect was found for the other categories (categories defined as in Figure 2.5).

Neurons responding to view of landscapes or places have been reported before in the human MTL (e.g. Ekstrom et al. [2003], Quian Quiroga et al. [2005], Rey et al. [2014b], Kreiman et al. [2000b]). These scene-responding neurons are reminiscent of ‘spatial view’ neurons found in the monkey hippocampus [Rolls et al., 1989, 1999]. Spatial view neurons respond selectively when the monkey is looking at a particular part of a spatial environment [Rolls et al.,

1999] and, in line with findings from concept cells [Quian Quiroga et al., 2005], they exhibit a low spontaneous activity [Rolls et al., 1999]. Moreover, spatial view cells represent allocentric information (they respond to the where the monkey is looking independently from the animal location, relative position or head direction). In humans, a similar feature has been reported for view neurons during a virtual navigation task [Ekstrom et al., 2003], where 70% of the view cells were reported to respond independently from the subject's location. Such representation of space in the MTL could be fundamental in respect to its memory functions by providing the spatial context component to episodic memories (i.e. remember where a particular object/person was seen) [Rolls et al., 1999].

3.5 Non-topographic organisation

We investigated whether there is a topographic organisation of responses in the MTL, namely, that nearby neurons tend to respond to associated concepts. For this, we focused on the responses from 72 electrodes that had more than one unit (single- or multi-unit) separated after spike sorting, with at least one significant response each. For each electrode we defined a mean association score between the images eliciting responses in the different units (recorded from the same electrode) using equation 2.3, and compared it with the mean association score between the images eliciting responses in one of the neurons but not in the other one (equation 2.4). In other words, we assessed whether nearby neurons (recorded from the same electrode) tended to fire to associated items. In line with previous evidence from studies in the rodent hippocampus [Redish et al., 2001], as well as from illustrative cases showing that neighbouring human MTL neurons tend to fire to completely unrelated things [Quian Quiroga, 2012], the mean association score between the stimuli eliciting responses in these close-by

units was not significantly different from the one for the other stimulus pairs (Wilcoxon signed-rank test; Figure 3.8), both when considering only the single units ($N=49$) or all units together ($N=159$), thus arguing against a topographic organisation of responses.

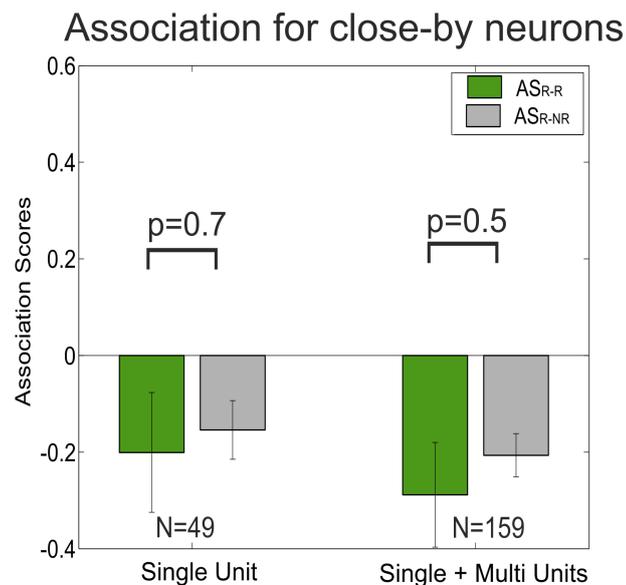


Figure 3.8 – Average value ($\pm SEM$) across electrodes of the association scores AS_{R-R} (green) and AS_{R-NR} (grey) between pictures eliciting responses in nearby neurons, which were recorded from the same electrode and separated after spike sorting. In this case N indicates the number of electrodes with at least two responsive units separated after spike sorting. The difference between AS_{R-R} and AS_{R-NR} was not significant according to a Wilcoxon signed-rank test.

3.6 Probability of pair responses

Furthermore, using the web-based scores, we assessed how the probability of neurons responding to a pair of items depends on the degree of association between the items. That is, given a neuron firing to one stimulus, we estimated the probability of firing to a second stimulus as a function of the degree of association with the first one. Differently from the previous analyses presented (focused on neurons with more than one response), in this case we also considered neurons with a single significant response. In fact, when calculating the

probability of neurons firing to two pictures with a given association, we had to consider not only the cases where neurons fired to both stimuli, but also the cases where neurons fired to one stimulus (even if this was the neuron's unique response) but not the other. Therefore, for this analysis we considered all 550 neurons (260 single-units and 290 multi-units) with at least one response found in our dataset. Briefly, the associations scores for all the pairs of stimuli presented together in the 99 sessions were pulled together, z-score normalised, and binned into 10 equally-spaced intervals (first and last bins were forced to contain first and last 2.5% of points to handle outliers). For each responsive neuron and bin, we calculated the ratio between the number of "pair responses" and the number of responses to one but not the other image of the pair. We then averaged across neurons to obtain the percentage (and standard error) of paired responses for each bin.

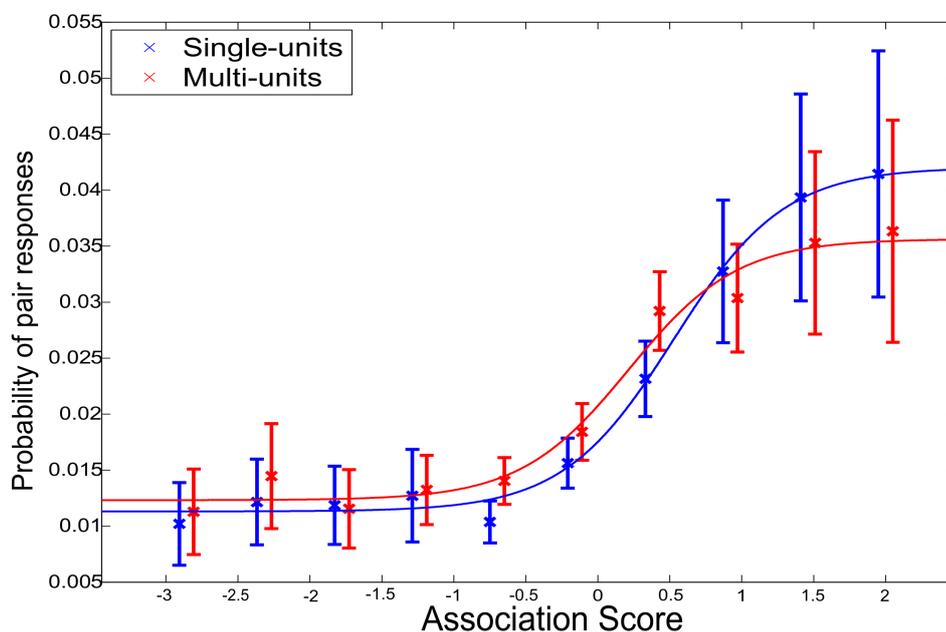


Figure 3.9 – Probability of responses to a pair of pictures as a function of their degree of association. Mean value ($\pm SEM$) on 260 single-units and 290 multi-units with psychometric fit of the data-points.

The results obtained, for both single- and multi-units (Figure 3.9) showed a psychometric-like non-linear increase of the probability of pair responses

with the degree of association between items. In particular, the probability of neurons responding to highly associated items saturated at about 4%, going down to about 1% for weakly-associated ones. This difference might, however, be larger considering that items that were not associated according to the web metric may have had a particular relationship for the patient, whereas there is more consistency between the subjective and web-metric scores for the highly associated items (Figure 2.6).

3.7 Cell-by-cell analysis

For each responsive neuron we performed a cell-by-cell analysis comparing the neural activity in response to the different pairs of stimuli with their degrees of association. For this analysis we did not use any criterion to separate responses and non-responses (as opposite to what we did in the previous tests), but employed directly a filtered version of the neural activity. In particular, for each of the 550 responsive units detected we calculated a matrix of joint neural responses F , with elements $F_{ij} = F_i F_j$. Here F_i and F_j are the neural responses (i.e. the median number of spikes in the ‘response window’; defined as in Section 2.2.1 between 200 ms and 1000 ms post stimulus onset) to pictures i and j , respectively, normalised by the maximum response. Response values below the neuron’s mean baseline activity plus 2 spikes were capped to zero, in order to decrease the influence of background noise. Basically, for each unit, the matrix F contained the combined neural activity in response to each pair of stimuli presented in the corresponding session. Its calculation resulted in 151 units with a matrix of joint neural responses filled with zeros (i.e. there were not two or more responses different from zero) and we therefore focused on the remaining 399 units (174 single- and 225 multi-units) with non-zero matrices. For each neuron we also computed the web-association scores between all the

pictures shown in the corresponding experimental session, and created the matrix of association scores A . Each row of this matrix was normalised using a z-score, given that, in spite of the overall z-score normalisation, some items (e.g. a very famous actor) tended to have higher association scores than others.

3.7.1 Correlation between neural responses and association

As first step, we performed a cell-by-cell analysis to measure the correlation between joint neural responses and association matrices. For each of the 399 units (174 single- and 225 multi-units) with at least one non-zero value in F , we computed the scalar product between the matrix of joint neural responses (F) and the matrix of web-association scores (A):

$$r = \sum_{i=1}^N \sum_{j \neq i} A_{ij} F_{ij} \quad (3.2)$$

To assess significance of the correlations between the neural and web-association matrices, for each neuron we statistically compared the original correlation value from equation 3.2 with a distribution of 1000 surrogates values (rank test). Each surrogate in the distribution was obtained by randomly permuting the indexes of the responses and generating a surrogate version of the joint neural responses matrix (F_{si}). F_{si} was then employed in equation 3.2 to calculate each surrogate correlation value r_{si} for the distribution.

Figures 3.10 and 3.11 show two examples of pluri-responsive neurons where the described analysis was performed. Note that in these figures significant responses are not marked, but they are reported in caption. In particular, from panels b) and c) that contain the heat-maps for the web-based association matrix and the joint-neural responses, it is already possible to visualise in both cases that the pattern in the joint-neural matrix partially matches the

clustering in the association matrix. However, this matching is just partial, due to the fact that the neurons are not firing to all the associations presented in the session, but only to few of them. Despite of this limitation, in both cases this partial matching was enough to give a significant correlation (above chance) between the two matrices (the real value for the correlation was higher than 95% of surrogate values, as shown in panel d of the figures). On the other hand, the neuron shown in Figure 3.12 responded to two clearly associated concepts (Superman and Mr. Incredible), but the correlation between the joint-neural and the web-association matrices was not significant due to the sparseness of the joint-neural matrix. Considering all the 399 units, we obtained a significant correlation between the neural and web-association matrices only in 19% of the cases (38 single- and 38 multi-units; rank test with $p < 0.05$). Nevertheless, at the population level we compared the original correlation values with the median of the surrogates for each neuron using a right-sided (i.e. our hypothesis is that the original values are higher than the ones of the surrogates) Wilcoxon signed-rank test and found that the correlation values were significantly larger than chance ($p < 10^{-6}$, $n=399$). Furthermore, to control for any semantic category effect (as for the control in Section 3.3.3), we performed the same statistical comparisons but constrained the permutations to be within the same category (broad categories defined as in Figure 2.5). In this case, 35 units (17 single- and 18 multi-units) showed a significant correlation between the neural and web-association matrices. Considering the whole population, the original correlations were still significantly larger than the ones obtained with the category-constrained surrogates ($p < 0.05$; Wilcoxon signed-rank test).

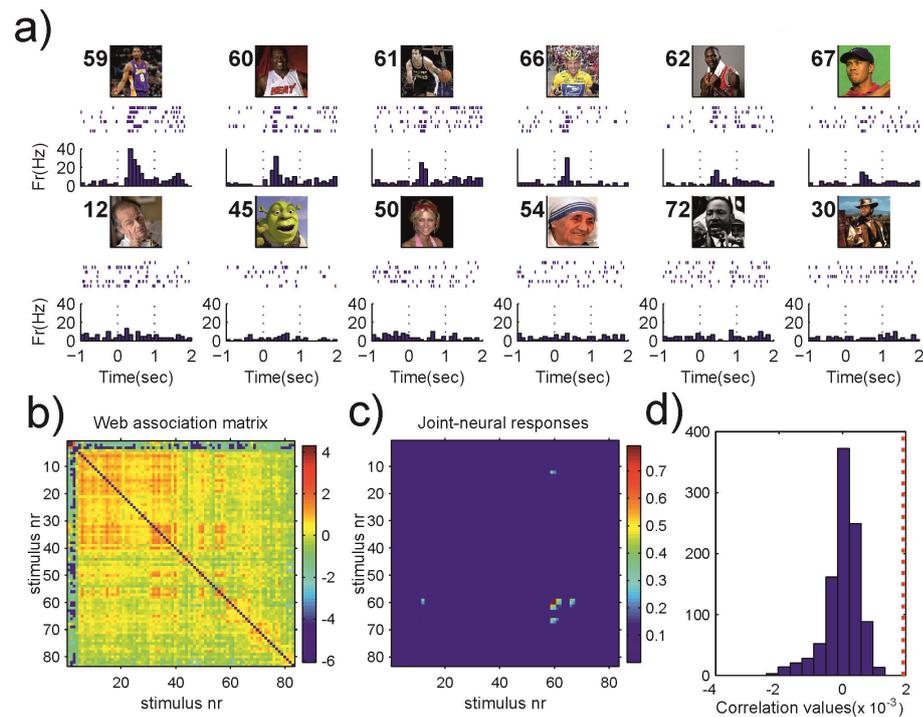


Figure 3.10 – a) A single neuron in the hippocampus that fired to pictures of basketball players and to other sportsman (significant responsive stimuli: 59, 60, 61, 66 and 62). The panel contains the neural responses to the first 12 most responsive stimuli; significant responses are not marked as no responsive criterion was employed for this analysis. (b,c) Web-association and joint neural response matrices. Colour bars on the right denote association and join-response strength, respectively, in arbitrary units. (d) Correlation between the web-association and joint-response matrices (red dotted line) and for a distribution of 1,000 surrogates, which were obtained by randomly shuffling the responses. The original correlation value was significantly larger than chance (rank test compared with the population of the 1,000 surrogates) with $p < 0.005$.

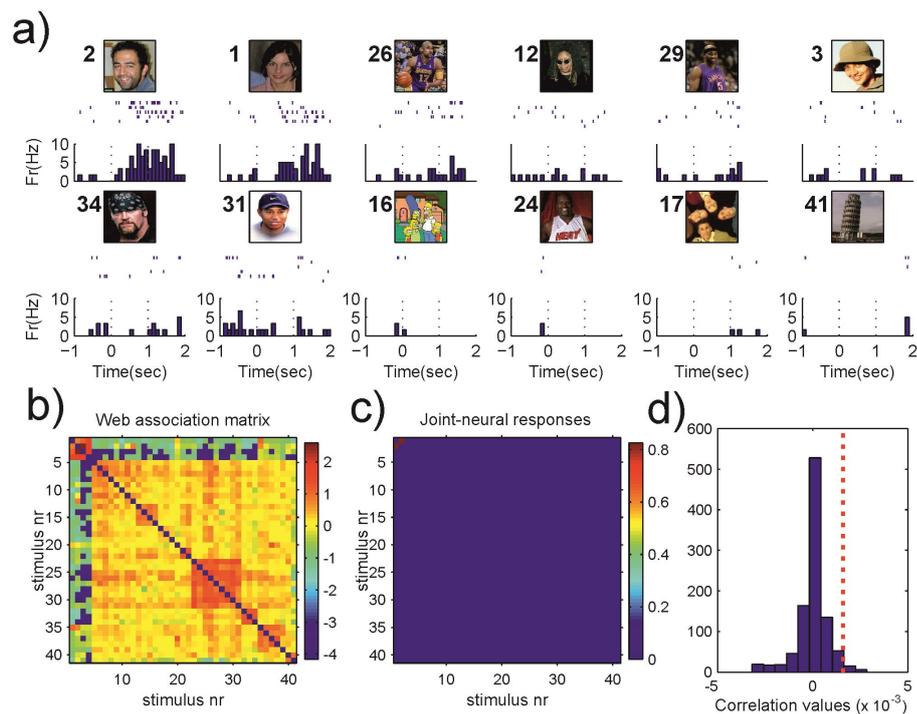


Figure 3.11 – A single neuron in the parahippocampal cortex that fired to pictures of two researchers performing the experiments with the patient (significant responsive stimuli: 2 and 1), but not to the picture of a third researcher also working with the patient (stimulus 3). Conventions are the same as for Figure 3.10. In this case the correlation between the web-association and joint-response matrices was significantly larger than chance ($p < 0.05$, rank test, 1,000 surrogates).

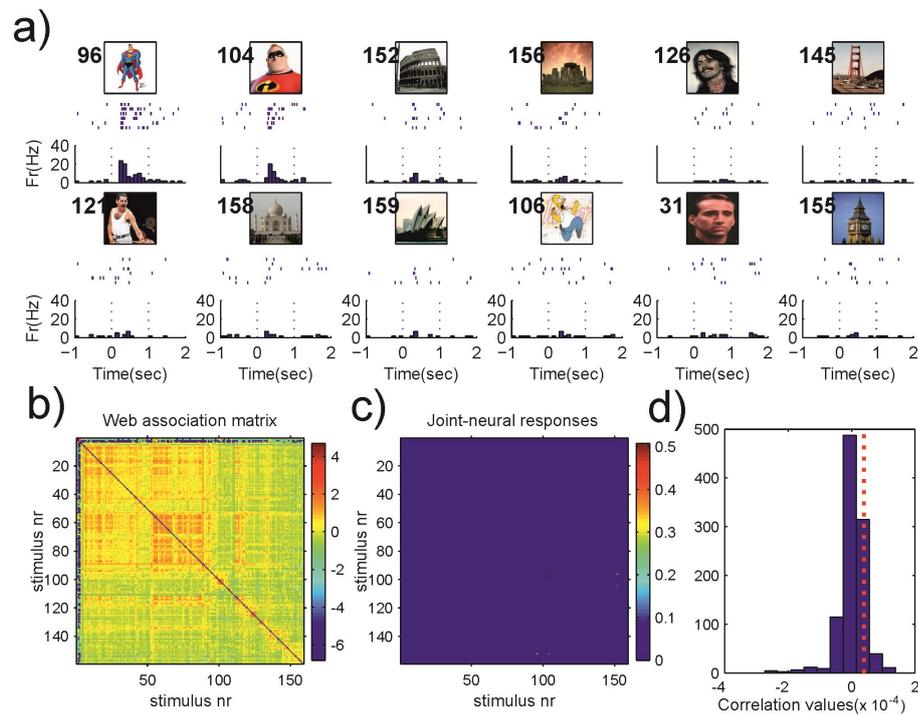


Figure 3.12 – A single neuron in the entorhinal cortex that fired to pictures of two cartoons, Superman and Mr. Incredible (significant responsive stimuli: 96 and 104), but not to other cartoons, like Homer Simpson (stimulus 106). Conventions are the same as for Figure 3.10. In this case, the correlation between the web-association and the joint-response matrices was not significant ($p = 0.13$, rank test, 1000 surrogates), due to the fact that the neuron encoded relatively few of the associations in the web-association matrix.

Finally, an average joint-neural activation matrix (Figure 3.13) was obtained by averaging all the matrices of joint neural responses obtained from the 550 responsive neurons. This matrix contained some holes (empty values) corresponding to untested pairs, since not all the possible pairs of stimuli were shown together in at least one session. In line with the other results in this section, we found that the correlation between this average joint-neural matrix and the web-association matrix shown in Figure 2.5 (obtained when considering all responses and pictures) was also significantly larger than chance ($p < 0.01$, rank test, 1000 surrogates).

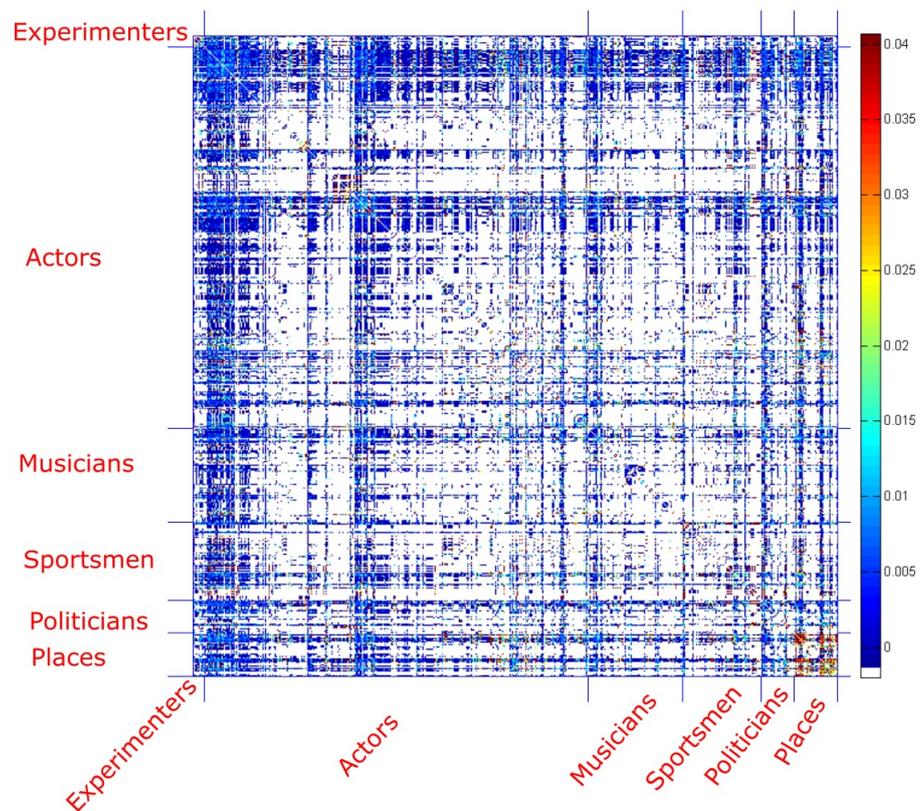


Figure 3.13 – Average joint-neural activation map containing the neural co-activation for each pair of stimuli averaged on 550 units. The white points in the map are empty entries in the matrix corresponding to pairs of stimuli that were never tested together.

3.7.2 Decoding analysis

We applied a decoder approach to prove that we could predict the association score between responsive stimuli, from the corresponding neural activity, better than chance. For every unit, we considered each non-zero entry F_{ij} in the joint-neural response matrix and predicted the association score between the corresponding items (i, j) using a nearest-neighbour approach with a leave-one out validation (one at a time, each value is predicted based on the rest). That means, if we denote by F'_{ij} the closest value to F_{ij} in the joint-neural response matrix, and by A'_{ij} its corresponding association score, we took A'_{ij} as an estimation of A_{ij} . Since we only employed the upper triangular portion of both matrices (otherwise the closest value would trivially tend to be the symmetric counterpart), the web-association matrices were normalised row-by-row and then symmetrised: $\widehat{A}_{ij} = (A_{ij} + A_{ji})/2$ (for $i > j$). Note that this analysis could only be applied to joint-neural response matrices with at least two entries different than zero in the upper triangular portion, which was the case for 345 units (136 single- and 209 multi-units). We quantified the accuracy of the predictions with the mean square error,

$$\varepsilon = \frac{1}{K} \sum \left(\widehat{A}'_{ij} - \widehat{A}_{ij} \right)^2 \quad (3.3)$$

where \widehat{A}'_{ij} and \widehat{A}_{ij} are the predicted and real association scores, respectively, and K is the number of non-zero values in the joint-neural response matrix for $i > j$. As before, for each neuron we used a rank test to compare the mean square error of the original estimations with the ones obtained from a distribution of 1000 surrogates, created by randomly permuting the indexes in the responses F . For each neuron tested the level of chance was equal to the median value of the surrogate distribution of mean square errors. Considering the whole population of 345 units, we found that predictions were significantly larger than

chance ($p < 0.05$, surrogate rank test) for 62 (18%) of them (27 single- and 35 multi-units). At the population level, we compared the original errors with the median of the surrogates for each neuron using a left-sided (i.e. our hypothesis is that the original errors are lower than the ones of the surrogates) Wilcoxon signed-rank test and found that predictions were larger than chance ($p < 0.05$).

3.8 Firing rate distributions and sparseness

In order to analyse the firing characteristic of the single-units identified we looked at their spontaneous and peak firing rates. For each unit, the mean baseline firing rate was calculated as mean number of spikes across trials fired by the unit in a time window between -1000 ms and -300 ms preceding the stimulus onset (time zero). The spontaneous firing rate was obtained as the mean of the baseline firing rates across all presented stimuli. Likewise, the response firing rate was calculated as the mean number of spikes across trials in a time window between 100 ms and 800 ms. The peak response firing rate corresponds to the maximum response firing rate (i.e. the firing rate in response to the most effective stimulus). Figure 3.14 shows the spontaneous and peak firing rate distributions for the 260 responsive single-units identified. From the distribution of peak response firing rates it appears that several responses have a low response activity (15% of values < 3 Hz). However, it is important to note that these values are calculated on a time window of 700 ms seconds; therefore, transient responses lasting no more than 100 ms (see for example the response to stimulus 5 in Figure 3.1) will have a peak response firing rate much smaller than the maximum value of their instantaneous firing rate. At the population level, the mean baseline firing rate was 1.9 Hz (SD: 3 Hz) with only 9.3% of values over 6 Hz, while the mean value for the peak response firing rate was 9.3 Hz (SD: 7.6 Hz); in line with the values reported by Ison et al. [2011] for a

similar population of responsive units in the human MTL (mean baseline firing rate 1.3 Hz, mean peak firing rate 7.8 Hz).

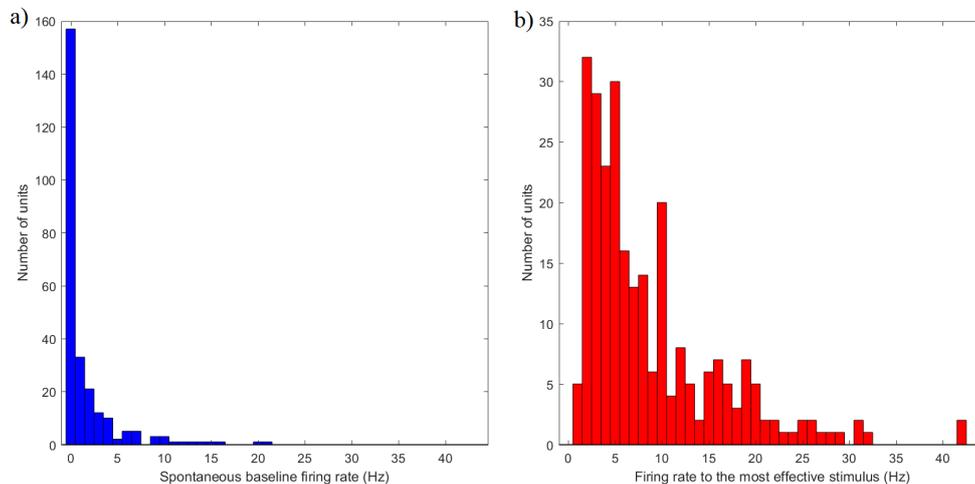


Figure 3.14 – Distribution of firing rates for the 260 responsive single-units. a) Spontaneous firing rates. b) Peak response firing rates (in response to the most effective stimulus).

3.8.1 Sparseness of the representation

As seen in section 1.4.1, MTL neurons are highly selective, in the sense that each unit fires to very few of the pictures presented. Typically these neurons are found to respond to about 2–3% of the stimuli presented [Quian Quiroga, 2012], as reported in several studies (e.g. Quian Quiroga et al. 2005, 2007), and in agreement with what we obtained from our dataset (Figure 3.5) where only 2.5% of the stimuli presented elicited a response in the neural population considered (single-units). These findings suggest that MTL neurons provide a rather sparse representation, in which each unit responds to a small proportion of stimuli.

For a set on S stimuli, the single neuron sparseness a^s can be quantified as proposed by Rolls and colleagues [Rolls and Tovee, 1995, Rolls and Treves,

2011] as:

$$a^s = \frac{\left(\sum_{i=1}^S y_i/S\right)^2}{\left(\sum_{i=1}^S y_i^2\right)/S} \quad (3.4)$$

where y_i is the mean firing rate of the neuron in response to the stimulus i . Single cell sparseness has a maximum value of 1 and approaches 0 for very selective units. Using this definition, we calculated the single cell sparseness for the whole population of 260 responsive single-units, the complete distribution is shown in figure 3.15(panel a). We found a mean value for a^s of $0.35(\pm 0.31)$, in agreement with the selectivity values reported by Quiñ Quiroga et al. [2007] for a similar population of MTL neurons (median value for $a^s = 0.39$, median selectivity $S = 0.71$, with S approaching 1 for very selective units).

Furthermore, when looking at the responses of a population of N neurons to a given stimulus, we can estimate the population sparseness a^p [Rolls and Treves, 2011], defined as:

$$a^p = \frac{\left(\sum_{i=1}^N y_i/N\right)^2}{\left(\sum_{i=1}^N y_i^2\right)/N} \quad (3.5)$$

where, in this case, y_i is the mean firing rate of neuron i to the specific stimulus. Using this definition, the population sparseness a^p was calculated for the set of 300 stimuli that were shown to a population of at least 10 single-units (across the total 260 units). The mean value for a^p was $0.36 (\pm 0.11)$, which is very close to the mean value obtained for the single cell sparseness. The complete distribution of a^p values is shown in Figure 3.15(panel b).

The values found for a^s and a^p suggest a rather sparse representation in this area, and are in line with the values reported for spatial view cells in the monkey hippocampus ($\langle a^s \rangle = 0.34$; $\langle a^p \rangle = 0.33$, [Rolls et al., 1998, Rolls and Treves, 2011]). In a sparse representation, information about a specific stimulus is provided by the activity of a relatively small population of neurons.

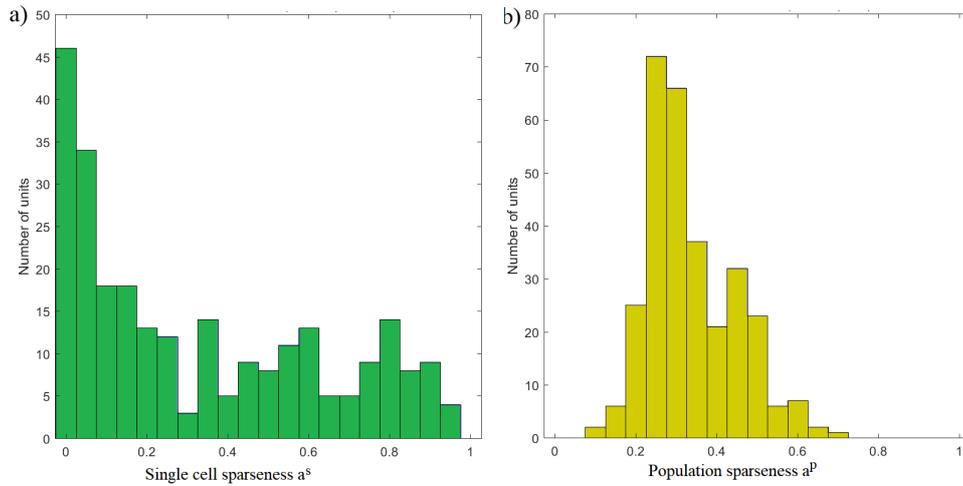


Figure 3.15 – a) Distribution of single cell sparseness (a^s) for the 260 responsive single-units identified. Mean: 0.35 ± 0.31 . b) Distribution of population sparseness (a^s) for the 300 stimuli that were shown to at least 10 single-units across the recording sessions. Mean value: 0.36 ± 0.11

A sparse representation is consistent with the MTL’s memory functions, as low values of sparseness increase the number of memories that the system is able to store and retrieve [Rolls and Treves, 1990, 2011]. In contrast, a more distributed representation means that information about a specific stimulus is provided by the firing activity of a large population of neurons [Abbott et al., 1996]; a distributed representation is more likely to be expected in sensory systems [Rolls, 2008].

On the other extreme of sparse representation there is local representation, in which a single cell activity provides all the information about a particular stimulus. These hypothetical cells are also referred to as grandmother cells (a grandmother cell activates when a person sees a specific subject, such as his grandmother) [Barlow, 1995]. Although the coding found in the MTL is very sparse and explicit (see Section 1.4.1), it is far from the one-to-one mapping proposed by a grandmother cell representation [Quiñ Quiroga et al., 2008a]. In fact, the idea that there is only one cell responding to a given concept would make it highly improbable to find such neurons in a limited experimental setup

like ours, and contrasts with the population sparseness values found in our data. Moreover, several MTL neurons are found to respond to more than one concept (see distribution of responses in Figure 3.5), while a grandmother cell is expected to be activated by a single identity. In line with the model described in Section 1.4.1, our findings here support the idea that concepts are encoded in the MTL by very sparse, but not local, cell assemblies.

3.9 Summary on the distribution of responses

This section contains a summary about the numbers of units and responses used in this study and their distribution across patients, locations, and semantic categories. Because of the different requirements for each kind of analysis performed, the number of neurons effectively available to each test varied. Table 3.3 contains a recap of the numbers of neurons employed in the different cases.

	Units	Single units	Multi units
Total number of recorded units	4078	1648	2430
Number of units with more than 1 response for which we had the patient's association scores (Figure 3.2)	32	19	13
Number of units with more than 1 response for which we used the web-based metric scores (Figures 3.4, 3.6 and 3.7).	261	129	132
Number of responsive units (with at least 1 response) used to study the probability of pair responses (Figure 3.9). The same units were used for the study of topographic representation (Figure 3.8).	550	260	290
Number of units with a non-zero joint-neural response matrix used for the cell-by-cell correlation analysis (Section 3.7.1).	399	174	225
Number of units with at least two non-zero values in the upper half of the joint-neural response matrix used for the decoding analysis (Section 3.7.2).	345	136	209

Table 3.3 – Number of units recorded and used for the different analyses.

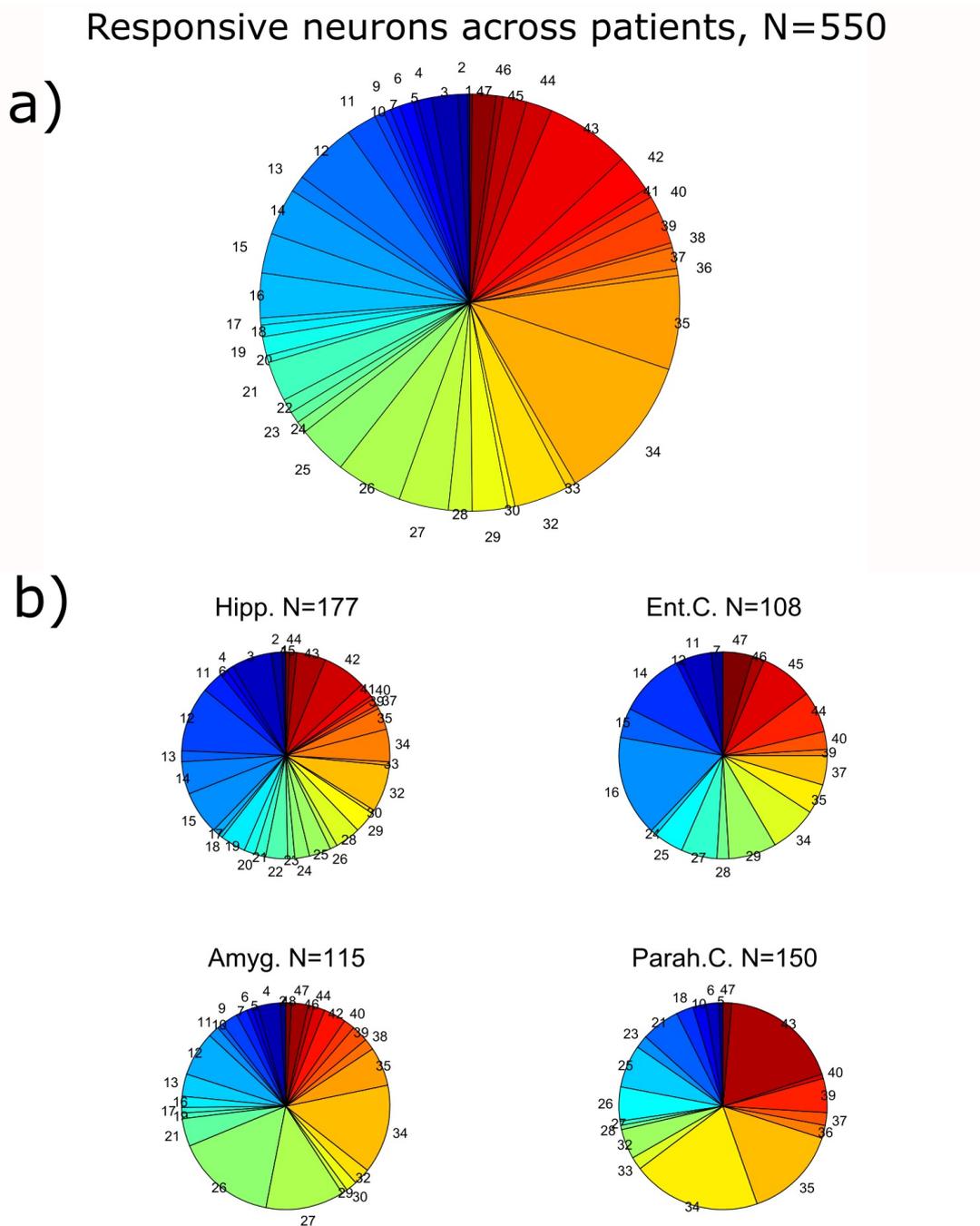


Figure 3.16 – a) Distribution of responsive neurons (Table Summary on the numbers of units per analysis, row 4) across patients. b) Distribution of responsive neurons per MTL area across patients. The numbers around the pie-charts identify the patients.

Figure 3.16 contains the proportions of responsive neurons across patients (all together in panel a and separated by MTL area in panel b). These plots

show that the responsive units were distributed across the whole cohort of patients. In line with this observation, the percentages of responses per category across patients (Figure 3.17c) also appeared spread across the different patients.

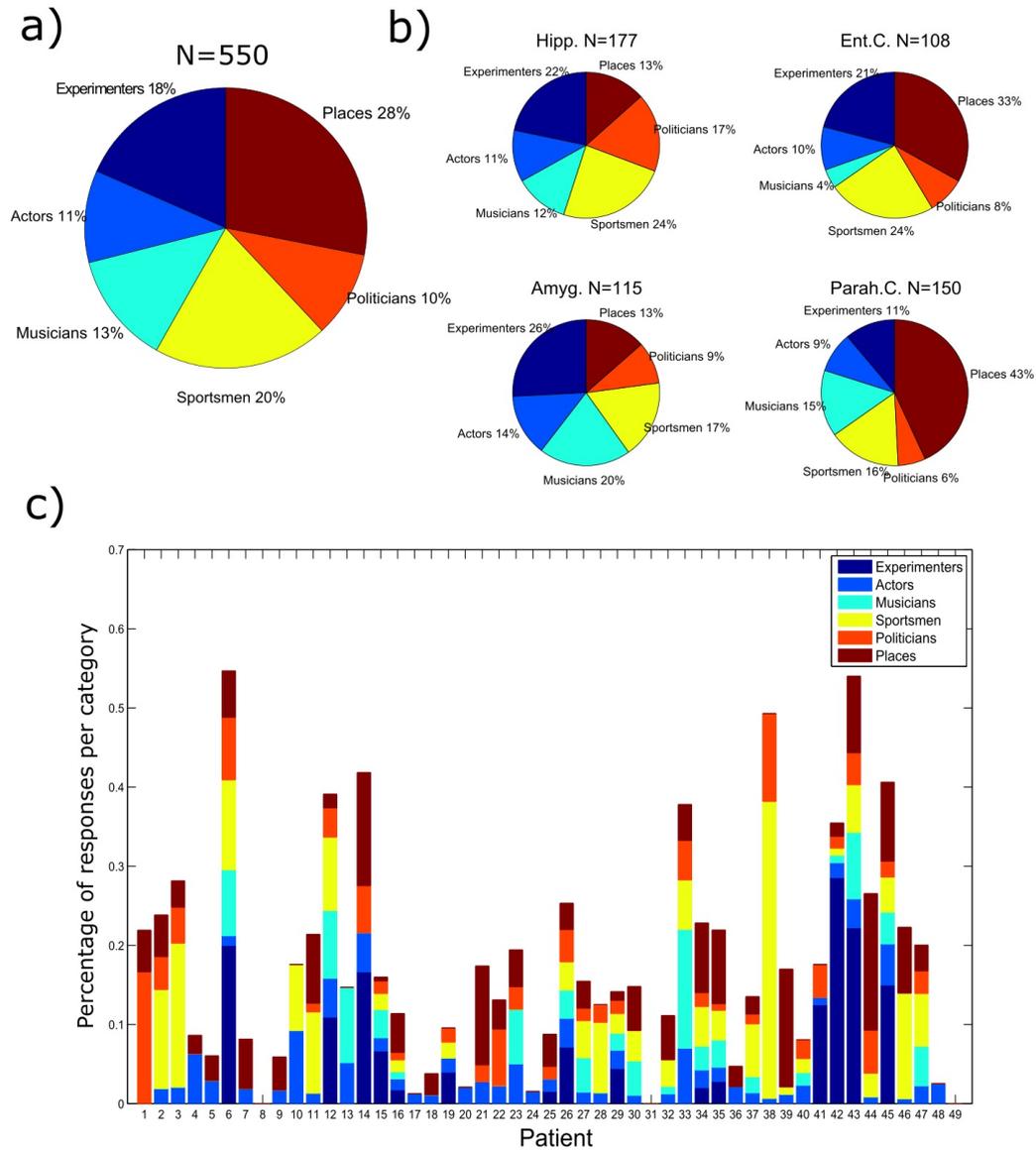


Figure 3.17 – Proportion of responses to each defined category for all responsive neurons (a) and for each MTL area (b). c) Percentage of responses per category for each patient. The percentage of responses for each category was calculated separately for each category (number of responsive stimuli divided by the number of stimuli presented for that category) and it is plotted with stacked bars only for space reasons.

Figure 3.17 shows the proportions of responses to each category considering

all the MTL areas together (panel a) or separately (panel b). We found a larger proportion of responses to scenes in the parahippocampal cortex compared to other areas, in line with the well-known scene-selectivity of the parahippocampal neurons [Epstein and Kanwisher, 1998, Mormann et al., 2017]. This observation was supported by a significant modulation of the proportion of scenes across the different MTL areas (ANOVA test, $p < 10^{-6}$), while no effect was found for the other categories.

3.10 Conclusions to Chapter 3

Summarising, we demonstrated that whenever MTL neurons fire to more than one concept, these concepts tend to be associated. To begin with, this tendency was proved for concepts considered associated by the subjects themselves: showing, in particular, that the patients' association scores for pairs of pictures to which neurons fired were significantly larger than the ones for other pairs of pictures. Next, three different analyses based on a web-based association metric further supported our claim: first, in line with the personal metric results, we found that the web-based association scores for responsive pairs of stimuli were higher than for other pairs. Second, a cell-by-cell analysis showed a correlation above chance between the neural responses and the association matrices (without using a criterion to define a response). Third, based on the neuron's responses we could predict the degree of association between pictures significantly better than chance. Moreover, we showed that these results could not be trivially attributed to an effect given by the visual similarity of the images, their relative familiarity, or the recall of a single picture triggered by associated ones. Following on from these results, we also showed that the responses to associated items were more correlated to the subjects' scores, determined by their own personal experiences, than to 'universal' web-association values.

In addition, the neurons' responses reflected specific relationships between individual items and could not be merely attributed to broad semantic categories. In fact, MTL neurons tended to fire to some, but not all of the pictures in a category (see for example the neuron shown in Figure 3.1a that fired only to some of the family members presented). This observation was quantified at the population level, showing that the association scores for pairs of responsive stimuli was still higher than for the other pairs when restricting the comparison to be in the same category. In other words, the items eliciting responses in a specific category (for example two actors) were on average more associated to each other than to the rest of the items in the category (e.g. the other actors). Moreover, we showed that neurons also responded to pairs of pictures belonging to different semantic categories (for example, an actor and a place) and that, also in this case, the association scores for the specific items the neuron fired to were, on average, significantly larger than the association scores for other items across the same categories (other actors and places).

The trend to fire more likely to associated concepts showed a tendency to be different among neurons in the different MTL areas, in particular, being highest in the hippocampus and lowest in the parahippocampal cortex. Moreover, as reported in the rodent hippocampus [Redish et al., 2001], and in contrast to findings in monkey cortical areas [Tanaka, 1996], close-by neurons did not tend to fire for associated concepts, thus supporting a non-topographical organisation of the MTL. Such non-topographically organised representation is in line with the random synaptic connections reported by Li et al. [1994] between neurons in CA3 (the hippocampal area proposed to be involved in the creation of associations [Treves and Rolls, 1994]), and it is indeed ideal for the fast creation of associations between arbitrary concepts [Quiñones Quiroga, 2012].

Our results showed that MTL neurons tend to encode well-established relationships between concepts. Furthermore, we showed that the probability

of getting responses to pairs of items increases with the strength of association between stimuli, going from 1% for non-associated items to 4% for highly associated ones. Findings from a recent study [Ison et al., 2015], where human MTL neurons rapidly encoded newly learned associations between (at first) unrelated items, showed that the proportion of neurons that changed their selectivity to encode the new associations was relatively large (about 40%). Therefore, we can argue that from a relatively high proportion of neurons initially encoding new associations, only a small fraction of these will consolidate this information into long-lasting representations in the MTL.

In line with previous reports from human MTL [Quian Quiroga et al., 2007], we found that the single cell and population sparsenesses in our set of neurons/stimuli ($\langle a^s \rangle = 0.35$; $\langle a^p \rangle = 0.36$, Section 3.15) suggested a very sparse representation in this area, which is consistent with what is expected for a memory system [Rolls and Treves, 1990, 2011], but it is still far from a grandmother cell representation [Quian Quiroga et al., 2008a].

Our results are in line with a recent interpretation about the role of MTL neurons in declarative memory [Quian Quiroga, 2012]. According to this view, each concept is represented in the MTL by a network of cells that, when activated, brings the concept into awareness. In this framework the associations between concepts are encoded via partially overlapping assemblies; meaning that a proportion of neurons firing to a certain concept may also fire to related ones (as shown in our study).

The long ongoing debate about the specific role of the MTL in the consolidation of episodic memory (see Section 1.3.5) opposes two alternative views about whether MTL representations are just temporary or more permanent. According to the standard consolidation model [Squire et al., 2004], the MTL has only a temporary role during learning and MTL representations are no longer held once memories consolidate in the cortex. Alternatively, the multiple

trace theory [Nadel and Moscovitch, 1997] argues that the MTL provides a long-lasting representation that continues to play a critical role for declarative (and particularly episodic) memory after learning. Evidence in support of one or the other model has been largely based on lesion studies and the investigation of human amnesic patients [Squire and Zola-Morgan, 1991, Moscovitch et al., 2005, Rosenbaum et al., 2005, Bayley et al., 2006], and has provided mixed results. There is little direct evidence from neuron’s recordings. In particular, about 15% of place cells in the rodent hippocampus have been shown to maintain their tuning for up to 30 days in a familiar environment [Ziv et al., 2013], even though it is still an open issue whether, and to what extent, rodent place cells can be taken as a model of declarative memory. Further results in support of a long-term involvement of the MTL in encoding associations come from Miyashita and colleagues that showed how memory-retrieval activation appeared earlier in perirhinal cortex than in visual cortical areas [Naya et al., 2001], in line with their previous finding of a disruption of pair coding neurons in visual cortical areas on lesions in the entorhinal and perirhinal cortices [Higuchi and Miyashita, 1996]. Closer to our study, selective responses to well-learned associations were described in the monkey [Yanike et al., 2004] and the rat hippocampus [McKenzie et al., 2014]. Moreover, using an eye-blink conditioning paradigm, neurons in the rabbit CA1 were shown to be active at the recall of remotely acquired associations [Hattori et al., 2015]. In line with these findings, we have shown, in humans, the presence of a long-term coding of associations between known concepts. As seen in Section 1.3.3, the role of MTL in the formation of semantic memories is still debated. Therefore, note that our results do not rule out the possibility that the long-term associations found in the MTL (that are beyond semantic category relationships) might be also encoded or formed in neocortex, and that they might just reflect the semantic associative structure of the inputs coming into the MTL from neocortical areas.

In conclusion, the fact that MTL neurons code previously existing associations that are independent of the specific task performed by the subjects, shows a long-term representation that goes beyond a temporary and malleable coding and offers new insights to our understanding of the role of the MTL in memory coding, its stability and capacity, which further support the role of these neurons in long-term memory.

Chapter 4

Firing properties of pluri-responsive neurons in the human MTL

In this chapter I describe the results of my follow-up study about the characterisation of spiking responses in single neurons firing to different concepts. In Chapter 3, we proved that MTL neurons encode associations. In fact, we found that whenever MTL neurons fired significantly to more than one concept, these concepts were usually related. However, it remained unclear how the individual neurons encoded these different concepts and associations. In this study we aimed to test whether the single units exhibited different firing properties in response to different stimuli and if those differences were eventually related to the relationship between concepts. In particular, we characterised the neural responses in terms of their electrophysiological characteristics: the response strength (the firing rate in response to a given stimulus) and the latency of spiking response (the time when the spiking response begins). We employed a modification of the paradigm used in Chapter 3 with a larger number of trials (25-35), which allowed us to measure the firing properties in a better way than

with only six trials.

The Chapter is organised as follows: First, I describe the data employed in the study and their processing. Next, I illustrate the analyses and results concerning the variability of response strength and latency across responses. The following section contains the results obtained in relation to our web-based association metric. Finally, I present the results obtained by comparing the firing properties in units firing to one or more stimuli.

4.1 Subjects and recordings

The data for this project were recorded at King's College Hospital in London during 21 recording session from six different patients with pharmacologically intractable epilepsy. Patients were implanted with intracranial electrodes for clinical reasons for 7–10 days. The study was approved by King's College Hospital Research Ethics Committee and all patients gave their written informed consent to participate. I was personally involved in setting up and performing these investigations; recording and pre-processing the data; and I performed all the analyses presented here. The electrodes were implanted bilaterally in the areas of hippocampus (24 probes) and amygdala (12 probes), their locations were based exclusively on clinical criteria. For the first 2 patients (7 sessions), the micro-wires signal was recorded with a 64-channel Neuralynx system (filtered between 0.1 and 9,000 Hz, and sampled at 32,556 Hz); for the remaining patients and sessions we used a 64-channel Blackrock system (filtered between 0.3 and 7,500 Hz, and sampled at 30,000 Hz). The experimental paradigm used was a long screening session in which 10–25 pictures were shown 25–35 times each in pseudorandom order. These pictures were selected following the results of a screening session recorded a few hours before (see Section 2.2 for the paradigm's details).

As for the study described in Chapter 3, the data recorded were processed using Wave_clus (as described in Section 2.1.1) to detect and separate the neural spikes. For each of the 393 single units detected, we identified the significant responses based on the spiking activity in the baseline and response windows (the responsiveness criteria are detailed in the second part of Section 2.2.1). Overall, we identified a total 165 significant responses. These responses came from 81 units in 69 different microelectrodes (11 from amygdala, 55 from anterior hippocampus, and 15 from posterior hippocampus), firing significantly to at least one stimulus. Among them, 37 units were responsive to more than one concept (7 from amygdala, 23 from anterior hippocampus, and 7 from posterior hippocampus). We will refer to them as pluri-responsive units. Considering these 37 pluri-responsive units we identified 208 pairs of stimuli eliciting responses in the same unit (responsive pairs). The complete frequency distribution of number of images to which the neurons responded to is shown in Figure 4.1.

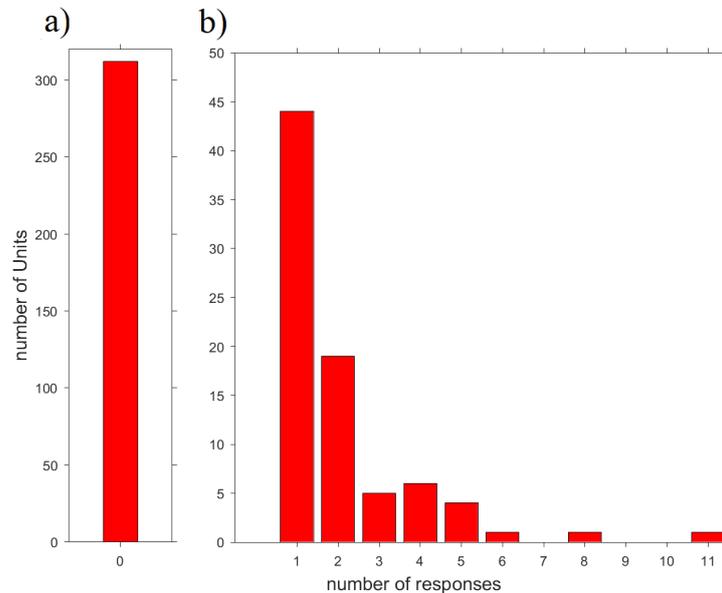


Figure 4.1 – Distribution of the numbers of responses across the 393 recorded units. a) Non-responsive units. b) Units responding to one or more stimuli. Note that the two panels have different y-axis scales for better visualisation.

4.1.1 Spike-shape surrogate test

As a first control, we assessed whether the spike sorting correctly separated the spikes from different units, or if there was any contamination in the clusters (i.e. an incorrect spike sorting procedure could have merged in one cluster spikes fired from different units despite significant differences in their spike-shapes).

Given a pair of stimuli st_1 and st_2 eliciting response in the same unit, we considered all the spikes S_1 and S_2 fired in the response window [100 ms - 800 ms] following each stimulus, across all trials. Each spike-shape consists of a vector of 64 elements (as sampled during the spike detection processing, Section 2.1.1). If n_1 and n_2 are the numbers of spikes fired in response to stimuli st_1 and st_2 , then S_1 and S_2 are two matrices of sizes $n_1 \times 64$ and $n_2 \times 64$, respectively. The spike-shape difference ΔS , associated to the two responses, was evaluated by means of the Mahalanobis pairwise distance between the two populations of spikes S_1 and S_2 as follows:

$$\Delta S = \sqrt{\sum_{j=1}^{n_1} D_M(S_{1j}, S_2) + \sum_{k=1}^{n_2} D_M(S_{2k}, S_1)} \quad (4.1)$$

where the notation S_{xj} indicates the j -esim row of S_x (i.e. the vector of 64 elements sampling the shape of the j -esim spike fired in response to st_x); while D_M indicates the Mahalanobis distance between a single vector and a distribution of vectors, defined as:

$$D_M(X_j, Y) = \sqrt{(X_j - \mu) \cdot \sigma^{-1} \cdot (X_j - \mu)^T} \quad (4.2)$$

with μ and σ being the rowwise mean and covariance of the matrix Y .

Statistical significance of the spike-shape difference was evaluated with a permutation test where the original value ΔS was compared with a distribution of 1000 surrogate values calculated on random permutations of the spike labels between the two stimuli in the pair. Specifically, for each re-arrangement of

the labels, we obtained two surrogate populations of spikes and a surrogate spike-shape distance ΔS_i . The ranking of the original value (ΔS) among the population of surrogate values, gave the p-value for the null hypothesis that the shape for the two spike populations was the same. We found that only 5 out of 208 responsive pairs (2.4%) exhibited a statistically significant difference in spike-shape ($p < 0.05$). Therefore, we concluded that the sorting procedure was done correctly and other potential differences between pairs of responses could not be explained by poor spike sorting.

4.2 Quantification of spiking activity

In order to analyse how the neural spiking activity differed before and after the stimulus presentation, we compared the spike counts in the time window preceding and following the stimulus onset. Considering the stimulus onset as time zero, we set the baseline window between -1000 ms and -300 ms and the response window between 100 ms and 800 ms. The mean strength of spiking activity (in Hz) was defined as mean number of spikes, across trials, fired by one unit in the specific time window, normalised by the window length.

Figure 4.2 shows the strength of spiking activity distribution for the 81 responsive units. For each unit, the spontaneous firing rate was calculated as mean strength of activity in the baseline window across all presented stimuli, while the maximum response firing rate corresponds to the maximum strength of activity in the response window across all presented stimuli. The mean baseline firing rate was 3.1 Hz (SD: 4.6 Hz) and the mean value for the maximum response firing rate was 11.6 Hz (SD: 9 Hz). These values are slightly higher than the ones found for the dataset in Chapter 3 (Figure 3.14). This effect is probably due to the more restrictive response criteria employed in this case to identify the responsive units (see Section 2.2.1) and to the different experimental

paradigm (in which we targeted the more strongly responding cells, see Section 2.2).

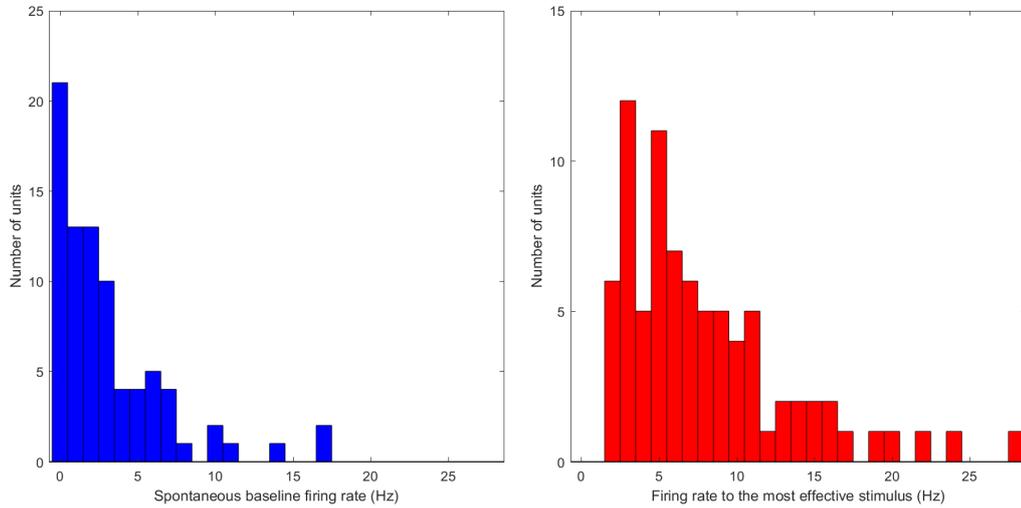


Figure 4.2 – Distribution of firing rates for the 81 responsive units. a) Spontaneous firing rates. b) Maximum response firing rates (in response to the most effective stimulus).

In order to investigate whether these neurons encode a graded type of representation or a more binary one, we looked at the distributions of firing rates in response to different stimuli. For each responsive unit, we calculated the baseline corrected strengths of response to each of the stimuli presented in the session; these strengths were then normalised between 0 and 1 (with 1 corresponding to the strength of response to the most effective stimulus). The distribution of normalised firing rates for all the responsive units is reported in Figure 4.3, separated for responsive and non-responsive stimuli. From the two histograms we can see that neurons show a zero strength in response (corresponding to their spontaneous firing rate) to most of the non-responsive stimuli, whereas they respond maximally and similarly to most of the responsive stimuli. This result suggests a nearly binary coding for these neurons, which will be further supported by the following analyses described in this section. A binary coding would have strong implications for the capacity of the MTL to

encode and retrieve memories (as discussed in Section 4.6).

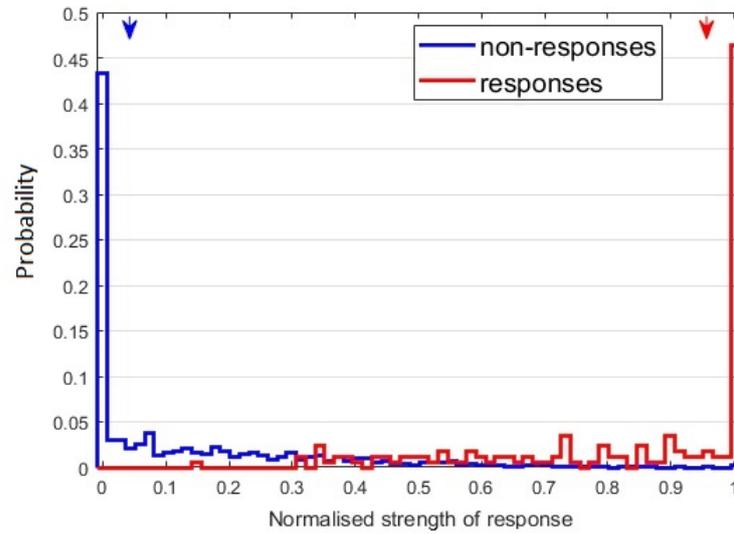


Figure 4.3 – Histograms for the normalised strength of activity in all the responsive units. Strength was computed for all the stimuli presented in each session, and separated according to whether or not they were responsive. The total areas of the two histograms were normalised to 1.

Next, for each of the 165 significant responses, we calculated the mean strength of baseline activity as the mean number of spikes in the baseline window. The mean strength of response was calculated both for individual responses (mean number of spikes across trials in the response period for each of the 165 individual responses) and multiple responses (mean number of spikes across all trials of all responsive stimuli for each of the 37 pluri-responsive neurons). Specifically, given a pluri-responsive neuron, all the trials from the different responsive stimuli were put together. The mean strength for each multiple response was then calculated as mean number of spikes in the response window across all the collapsed trials.

Similarly, from the spike counts in the three cases (baseline, individual responses, and multiple responses) we calculated the strength variability (as the mean absolute deviation across trials) and the coefficient of variation (standard deviation divided by mean across trials) of spiking activity.

The comparison results are reported in Figure 4.4, where the empirical

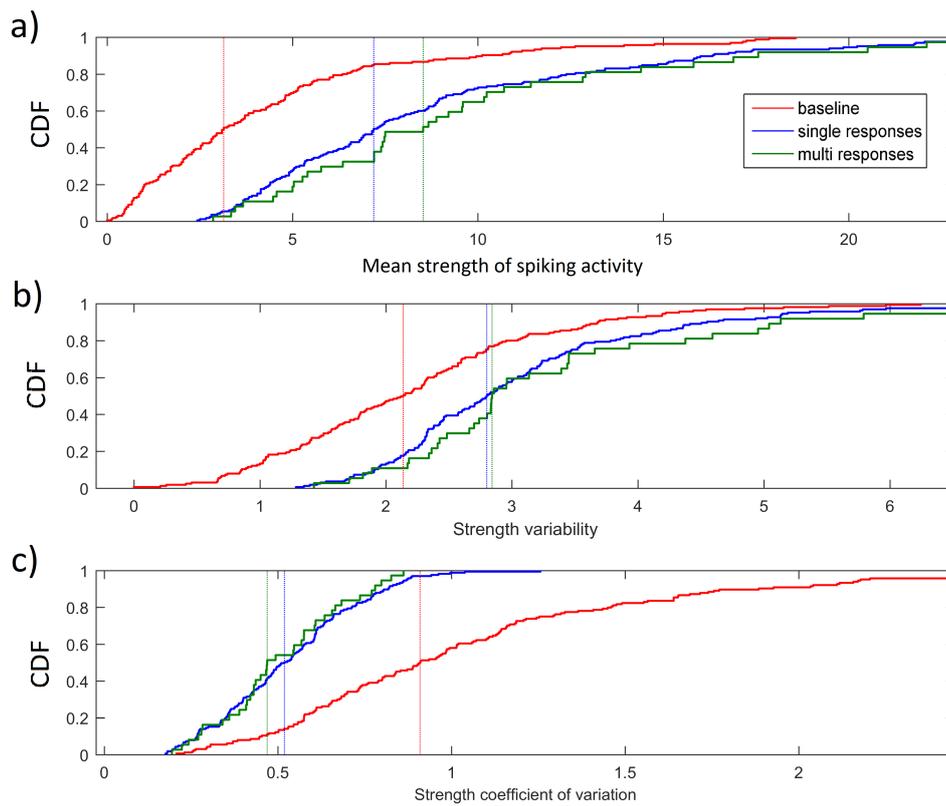


Figure 4.4 – Empirical cumulative distribution function (CDF) for strength (a), variability (b), and coefficient of variation (c) of the spiking activity. Vertical dotted lines denote the median of each distribution. In each case, the quantities were computed during baseline (red, 165 points) and response period (in blue for the 165 individual responses, and in green for the 37 multiple responses). In all three cases we observed significant differences between baseline and individual responses, but no difference between individual and multiple responses.

cumulative distribution functions (CDF) for the three cases are shown. We observed a significant difference between the strength of spiking activity for baseline and for individual responses (paired sign test, $p < 10^{-36}$). A significant difference was also found in the variability and coefficient of variation of spiking activity, when comparing baseline and individual responses (paired sign test, $p < 10^{-12}$ for the variability and $p < 10^{-34}$ for the coefficient of variation). The difference between baseline and individual responses was indeed expected, considering the criteria we used to identify the responses (based on the comparison of baseline and response activities). When comparing the distributions for individual and multiple responses we did not find significant differences in any of the quantities considered (Kolmogorov Smirnov test, $p = 0.3$ for the strength of spiking activity, $p = 0.4$ for the variability, and $p = 0.9$ for the coefficient of variation), suggesting that the spiking activity across different responses is not dissimilar from the one for individual responses. In addition, the median value of the coefficient of variation for baseline activity was very close to 1 (0.9 ± 0.3), which is consistent with a Poisson process [Goh and Barabasi, 2008]. On the contrary, we found that the coefficient of variation was smaller than 1 for both individual and multiple responses; indicating that, despite the increase in the strength of response, the spike count variability of the responses was smaller than their spike count average.

4.2.1 Pairwise response strength comparison

Subsequently, we investigated differences in strength of spiking activity of pluri-responsive neurons responding to different stimuli. The spike counts in the response period, for each of the 208 responsive pairs identified, were compared using a surrogate test (similar to the one employed on the spike-shapes).

Specifically, given a responsive pair, we compared their difference in response strength (mean value of the spike count in the response period, as defined in

Section 4.2) to a distribution of 1000 surrogate values, created by randomly permuting the trial labels for the two responses. From the ranking of the original response strength difference among the population of surrogate values we obtained a p-value for the null hypothesis that the two responses had the same strength.

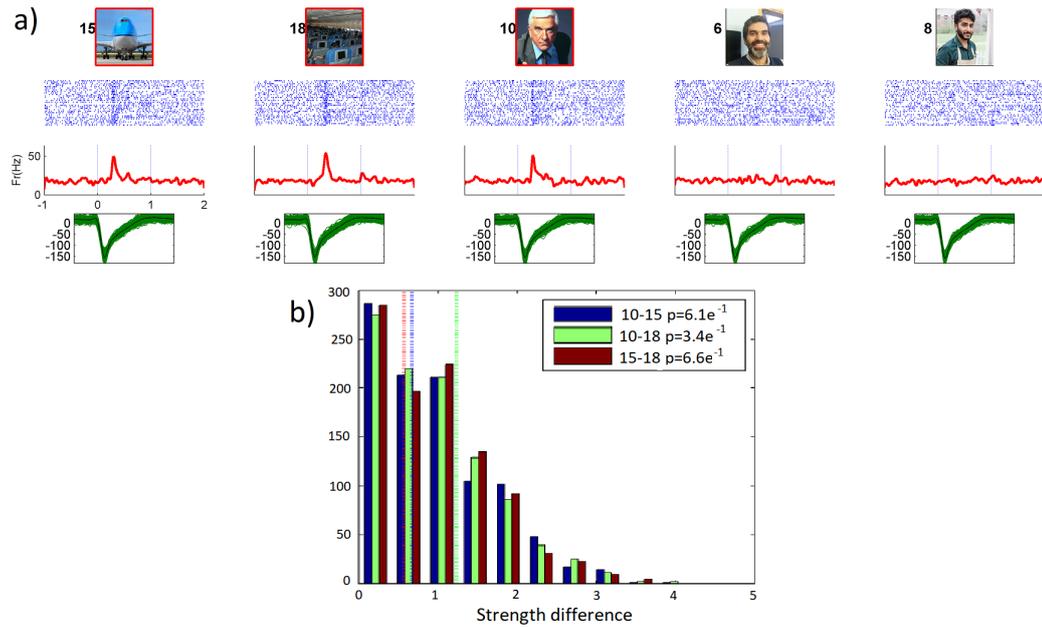


Figure 4.5 – a) A pluri-responsive unit in the posterior hippocampus, responding to the pictures of a Boeing, the interior of a plane, and Leslie Nielsen (leading actor in the movie “Airplane!”). Time zero marks the onset of picture presentations which lasted for 1 sec. For each stimulus, the raster plot (blue lines) containing the single spikes detected at each time (each row is associated to a trial, starting from the top), the instantaneous firing rate (red curve) and the spike shapes (green) in the response period (100 ms to 800 ms) are shown. The red frame around the picture marks a significant response. The responsive stimuli were strongly related to each other according to the unit association scores ($AS_{R-R} = 0.3$, $AS_{R-NR} = -0.06$, see Section 4.4). b) Distribution of response strength differences of the surrogate populations (histograms) for the three responsive pairs. Vertical dotted lines mark the original difference values.

Figure 4.5 shows an example of a pluri-responsive neuron significantly firing to the presentation of three different pictures. In this example, the surrogate test for each of the three responsive pairs gave no significant difference in the response strength (panel b; stimuli 10-15, $p = 0.61$; stimuli 10-18, $p = 0.34$;

stimuli 15-18, $p = 0.66$).

At the population level, only 24% of the pairs (50 out of 208) exhibited a significant difference ($p < 0.05$) in the response strength. Figure 4.6 shows the complete distribution of p-values for this test.

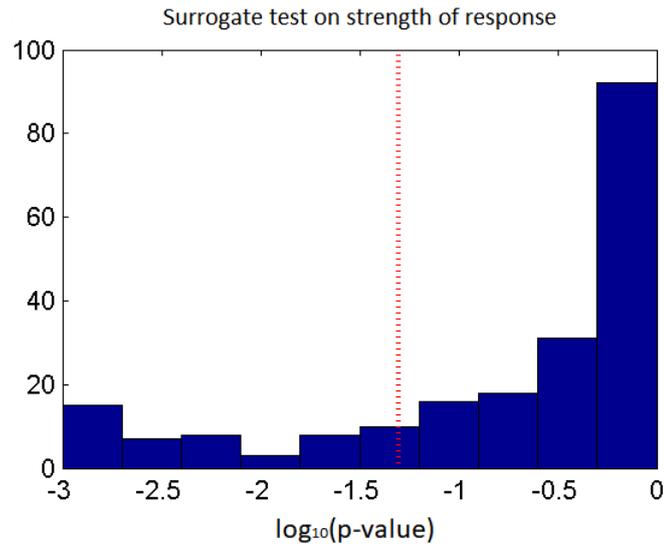


Figure 4.6 – Distribution of p-values for the surrogate test on response strength for the 208 responsive pairs; only 24% of them exhibited a significantly different response strength. Red dotted line marks significance threshold ($p = 0.05$).

4.2.2 Strength difference between responsive pairs

As an alternative way to look at the population results, we compared the response strength differences between responsive pairs and between random re-arrangements of the pairs. In particular, for each responsive pair we calculated the normalised response strength difference as the strength difference of the pair divided by the maximum strength value in the pair (this normalisation corrected for the different levels of activity across neurons). The shuffled pairs were obtained by randomly re-pairing the 122 significant responses coming from pluri-responsive neurons. We created 1000 surrogate pair distributions (each one consisting of 208 pairs), and found that the original distribution of

strength differences was significantly smaller than the random pair distributions (surrogate rank test on the medians, $p < 10^{-3}$). Figure 4.7 shows that the strength differences for the responsive pairs were significantly smaller than the differences for a representative random pair distribution, selected as the closest to the median value of surrogate distributions (rank-sum test, $p < 10^{-33}$).

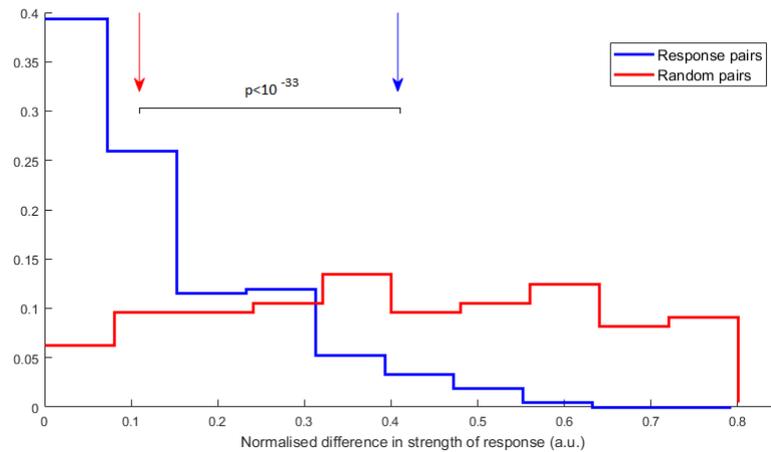


Figure 4.7 – Distribution of normalised response strength difference for the 208 responsive pairs (blue line) and for a representative population of random pairs of significant responses (red line). The two distribution are statistically different according to a rank-sum test ($p < 10^{-33}$). Arrows mark the median value of each distribution.

4.2.3 Decoding analysis of response strength

Next, using a decoder approach, we compared the response strength at the single trial level. A naive Bayesian decoder with leave-one-out cross-validation was run on each pluri-responsive unit to test whether it could discriminate single trial differences in the response strength across stimuli. For a single pluri-responsive unit, the trials from all different responsive stimuli were pulled together. The spike count in the response period (C_i) for the single trials and the corresponding label (l_i , telling which stimulus the trial belonged to) were considered. For each trial i the Bayesian decoder was trained with the spike counts and labels for all trials except i , and the label for the left-out trial was

then predicted by the classifier. The decoding performance was estimated as percentage of trials predicted correctly, and its statistical significance (for the significance level $p = 0.05$) was assessed in comparison to the performances obtained on a population of 1000 surrogates created by randomly shuffling the trial labels. For each neuron tested, the level of chance for the decoding performance was equal to $1/N$, with N being the number of its responsive stimuli.

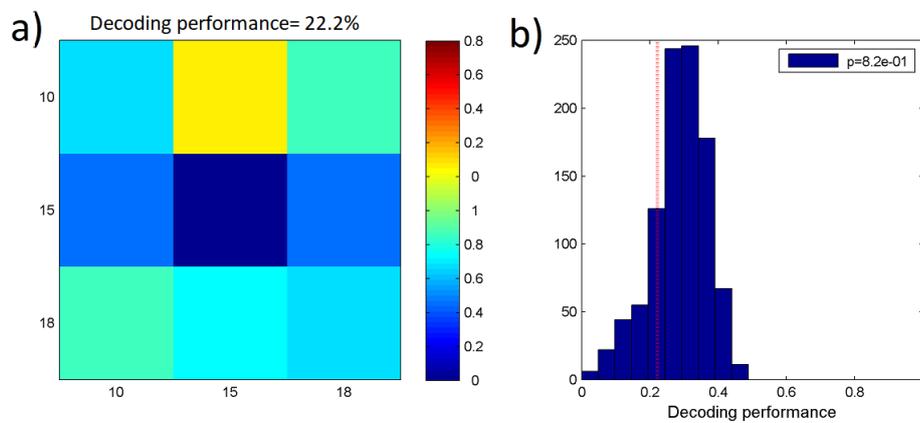


Figure 4.8 – Decoding analysis results for the exemplary unit in Figure 4.5. a) Confusion matrix for the decoder based on the single trial spike counts of the three responsive stimuli. b) Distribution of the decoding performance for the population of 1000 surrogates (histogram) and real decoding performance (red dotted line). The decoding performance was 22.2%, not significantly different from the chance level of performance (33.3%), $p = 0.82$.

For the exemplary unit considered (Figure 4.5), the decoding analysis results are shown in Figure 4.8: a confusion matrix (panel a) illustrates the decoding performance (22%), which was in this case not significantly different from chance (panel b, $p = 0.82$, chance level of performance = 33.3%).

Considering all 37 pluri-responsive units, only 7 (19%) showed statistical significance for this test ($p < 0.05$). The complete distribution of p-values is reported in Figure 4.9.

These results, together with the ones reported in the previous sub-sections, show that, in most cases, pluri-responsive neurons do not exhibit differences in

response strength across different responsive stimuli; suggesting that, in general, it is not possible to discriminate a certain responsive stimulus from the single unit strength of activity.

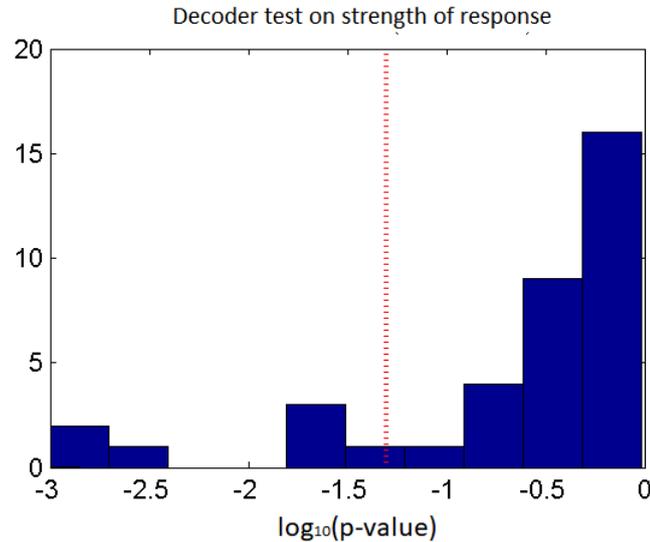


Figure 4.9 – Distribution of p-values for the significance of decoding performance based on single trial response strength for the 37 pluri-responsive units. In only 19% of units the decoding performance was above chance. Red dotted line marks significance threshold ($p = 0.05$).

4.2.4 Response strength difference vs number of trials

The difference in response strength could, in principle, depend on the number of trials considered (e.g. we may not find a significant difference when using, for example, only 6 trials, but the difference could appear when considering a larger number of trials). In order to rule out the possibility of having larger differences if more or less trials were used, we assessed how the estimation of strength difference between responses changed with the number of trials. For a given number of trials (nt), the strength of the single response was defined as the mean number of spikes fired in the response window (100 ms to 800 ms) across trials 1 to nt . For each pair of responses (208 pairs in total), we estimated the normalised absolute strength difference as a function of number

of trials, spanning from 1 to the maximum number of trials in the session (for each pair the strength difference was normalised by the maximum strength in the pair, as in Figure 4.7). We then computed the grand average across all pairs and found a systematic decrease in the response strength difference with stimulus repetition. Finally, the data points were fitted with a negative exponential function to verify the existence of a plateau (i.e. if the estimation remained stable after a certain number of trials nt).

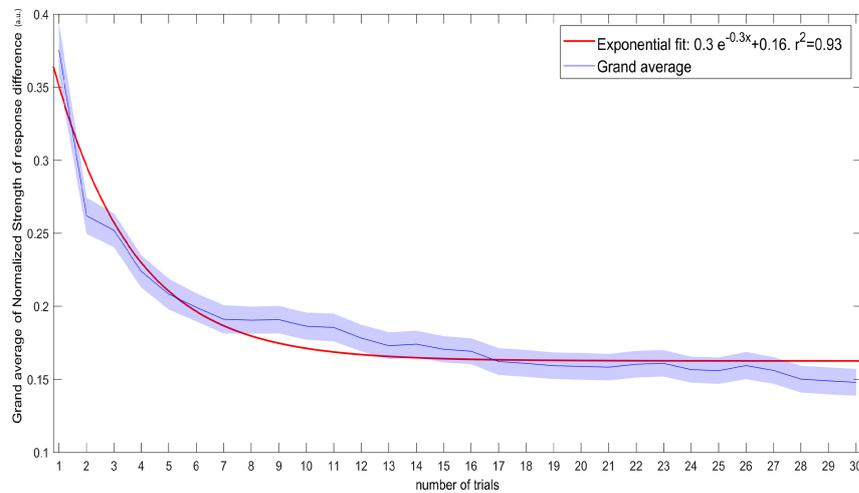


Figure 4.10 – Strength difference estimation with the number of trials used. Grand average of the normalised response strength difference across responsive pairs $\pm SEM$ (blue). The exponential fit of the data (red line) shows that the estimation reached an asymptotic value (0.16) when using at least 17 trials.

Figure 4.10 shows how the estimation reached an asymptotic value and became stable when using more than about 16 trials (less than 1% change for more trials), far fewer than the minimum number of trials we used (25). Therefore, the lack of differences found in the response strength cannot be attributed to a low number of trials.

4.3 Spiking response latency

Next, we moved to examine the latency of spiking responses. For all individual responses we computed a response onset - that is, the time following the stimulus presentation when the spiking response started. The response onset was defined as the time when the rate (IFR) crossed over a threshold for at least 75 ms (short periods of less than 20 ms going below threshold were ignored). The IFR was obtained from the convolution of the spike train with a Gaussian kernel with $\sigma = 10ms$ (truncated at 1% amplitude). The threshold was set to the mean plus 4 standard deviations of the IFR in the baseline window (from -900 ms and -100 ms) computed across all stimuli (for neurons with low baseline activity, a minimum of 5 Hz was imposed for the threshold). Similarly, the response offset was defined as the first time, following the response onset, when the IFR went below the set threshold for at least 100 ms.

4.3.1 Latency estimation with number of trials

In order to verify the stability of our estimation, we examined how the response latency changed with the number of trials used for its computation. For each of the 165 responses, we estimated the response onset (using the calculation described above) considering a number of trials (nt) spanning from 6 to the maximum number of trials in the session. The trials were considered in the order they were presented (i.e. considering nt trials here meant to consider trials from 1 to nt). The average onsets across all the responses as a function of nt were then fitted with a negative exponential function. Note that the number of estimated onsets for each nt varied because for certain responses a spiking onset could not be effectively estimated, or the stimulus was presented less than nt times in the session. Figure 4.11 shows that the average onset estimation became stable when using more than about 12 trials (less than 1% change

with more trials). In our case, since we considered a minimum of 25 trials, any eventual lack of differences in latency cannot be due to a low number of trials.

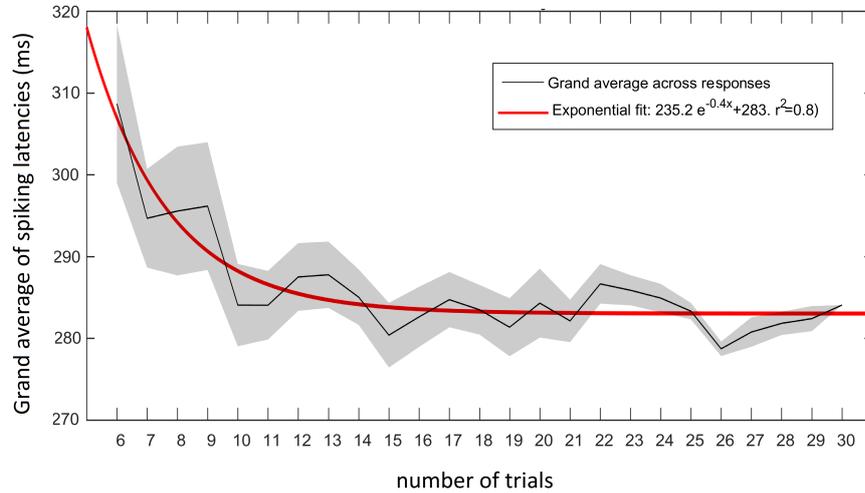


Figure 4.11 – Grand average of the estimated latency across responses $\pm SEM$ (black line) with the number of trials used for the calculation. The exponential fit shows the existence of an asymptotic value in the estimation. The estimate became stable when using at least 12 trials (less than 1% change).

Moreover, we looked at the variation of latency estimation when moving from the first to the last trial. We calculated the response onsets on the whole population (165 responses) considering a sliding window with fixed number of trials ($nt = 6$) moving from the top to the bottom of the trial list (i.e. if N is the maximum number of trials for the given response, the window starts in trials 1 to 6 and slides to trials $N-5$ to N). The average latency as a function of the last trial number in the window is reported in Figure 4.12. The estimate did not change with the position of the window. Statistically, latency estimations based on trials 1 to 6 were not different to the ones based on trials 20 to 25 (paired sign test, $p = 0.93$). Therefore, the estimation bias seen with 6 trials (panel a) was due to the number of trials and not to the presentation order.

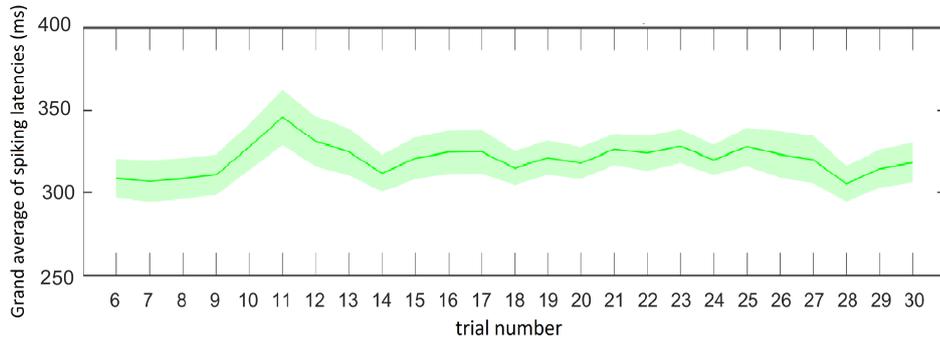


Figure 4.12 – Latency estimation with a sliding window of 6 trials (for each trial number, the estimation was done with the preceding 6 trials). No significant difference was found comparing the estimations based on trials 1 to 6 with the ones based on trials 20 to 25 (paired sign test, $p = 0.93$).

4.3.2 Pairwise latency comparison

Next, we looked at differences in spiking response latency for pluri-responsive units responding to different stimuli. Similar to the analysis performed with the strength difference, for each pair of significant responses from the same unit, the difference in response latency (calculated as described above) was compared to the population of 1000 surrogate values. Each surrogate value was the latency difference estimation on a surrogate pair of responses (obtained by randomly permuting the trials from the two original responses).

Figure 4.13 shows an exemplary unit recorded from the hippocampus responding to two stimuli. The spiking onsets and offsets for the significant responses are marked on top of the instantaneous firing rate (IFR) plot (magenta dashed lines). There was no difference in the response latency between the two responses according to the surrogate test performed ($p = 0.3$).

When looking at the population of 208 responsive pairs, we found that only 19% of them (40 out of 208) presented a significant difference in spiking latency. Figure 4.14 shows the complete distribution of p-values.

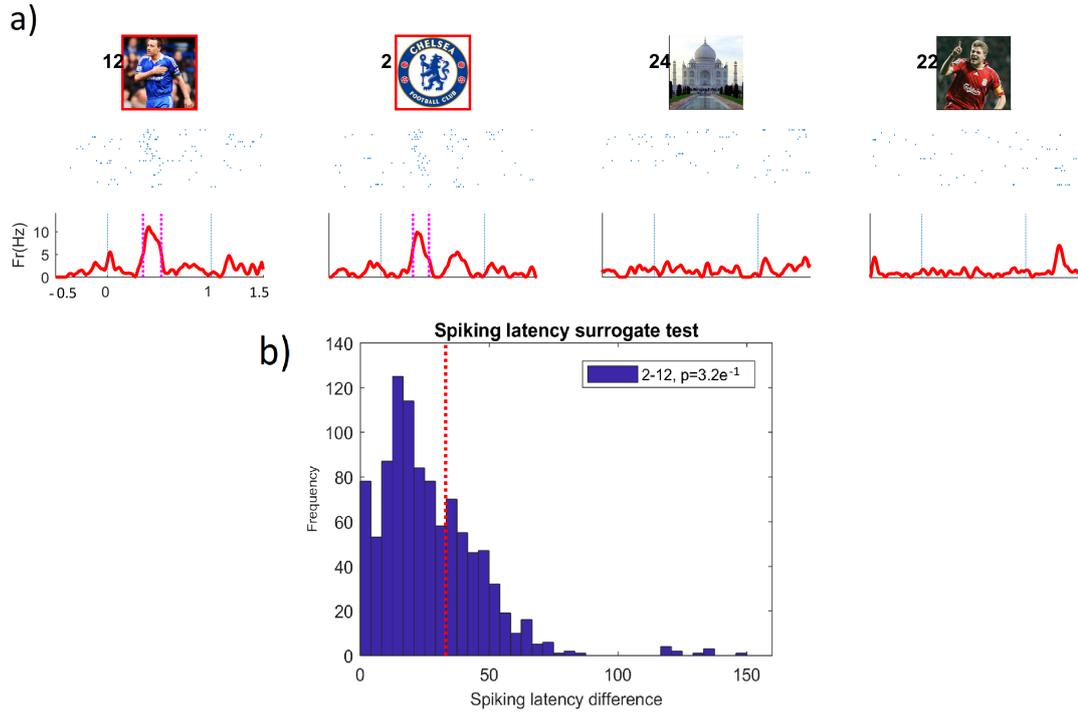


Figure 4.13 – A single unit recorded from left anterior hippocampus responding to the pictures of John Terry (player and captain of Premier League club Chelsea) and of the Chelsea club badge. Conventions are as in Figure 4.5. Magenta dotted line on top of the IFR mark the response onsets and offsets. The responsive stimuli were strongly related to each other according to the unit association scores ($AS_{R-R} = 1.8$, $AS_{R-NR} = 0.19$, see Section 4.4). b) Distribution of latency differences for the surrogate populations (histogram) and original difference value (red dotted line). The original value was not significantly different from the surrogate population ($p = 0.3$).

4.3.3 Latency difference on responsive pairs

Similar to what was observed for the response strength differences (Section 4.2.2), we found that the population of latency differences for the responsive pairs was smaller than for random pairs of responses. Considering the 122 significant responses coming from pluri-responsive units, we randomly permuted their pair labels to obtain surrogate pair populations (containing 208 pairs each). The original distribution of latency differences was significantly smaller than the random pair distributions (rank test on median with 1000 surrogates, $p < 10^{-3}$). Figure 4.15 shows the latency differences for the responsive pairs compared to a representative random pair distribution, selected as the closest to

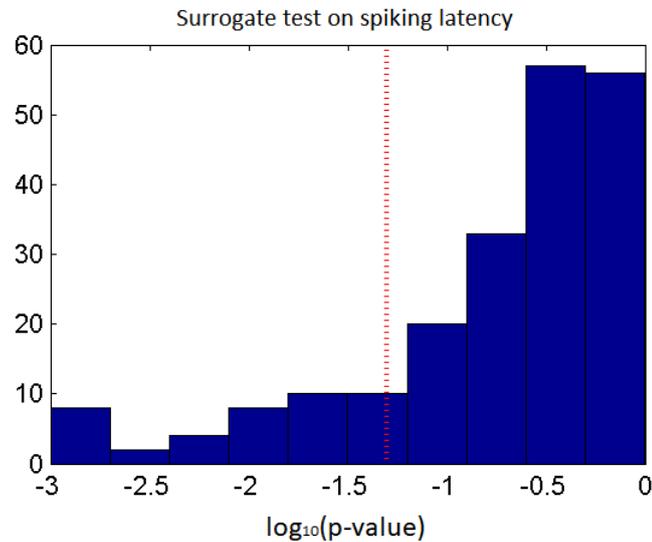


Figure 4.14 – a) Distribution of p-values for the surrogate test on latencies for the 208 responsive pairs; only 19% of them exhibited a significantly different latency. Red dotted line marks significance threshold ($p = 0.05$)

the median of surrogate distributions. The original distribution was significantly smaller than the shuffled one (rank-sum test, $p < 10^{-9}$).

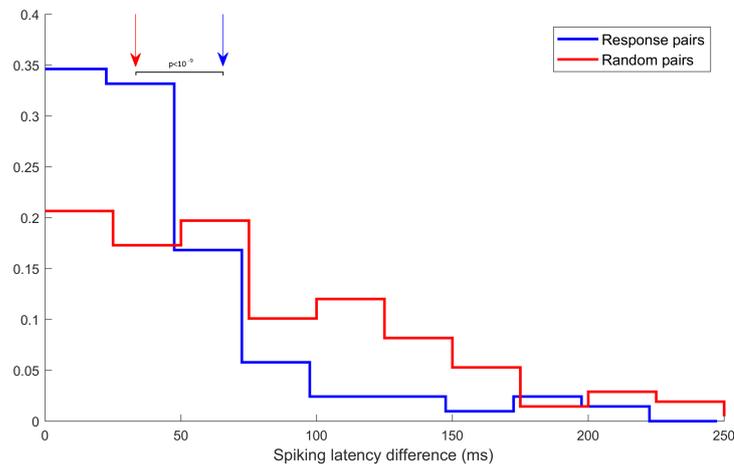


Figure 4.15 – Distribution of latency difference for the 208 responsive pairs (blue line) and for a population of 208 random pairs of significant responses (red line). The two distributions were statistically different according to a rank-sum test ($p < 10^{-9}$). Arrows mark the median value of each distribution.

4.4 Association metric results

The relationship between responsive and non-responsive stimuli in pluri-responsive units was analysed with the same approach used for the study in Chapter 3. Essentially, for each pair of stimuli we defined an association score based on the numbers of hits found with a web search engine (see Section 2.3.2 for details), which we took as a measure of the association strength between them. The values obtained were z-score normalised on a session-by-session basis.

4.4.1 Association scores

For each pluri-responsive unit, the normalised association scores between pairs of stimuli were averaged within two groups, depending on whether both stimuli led to significant responses (R–R), or one was a response and the other not (R–NR) (see Section 2.3.4). Therefore, for each pluri-responsive unit, we calculated a mean association score for pairs of responses (AS_{R-R} , eq. 2.3) and for other pairs (AS_{R-NR} , eq. 2.4) .

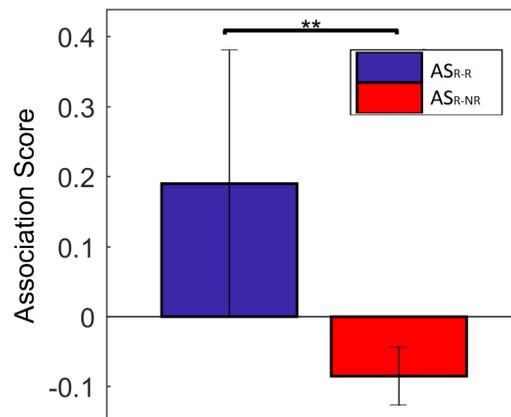


Figure 4.16 – Mean ($\pm SEM$) of association scores across 35 pluri-responsive single units for pairs of responses (AS_{R-R} , blue) and for the other pictures pairs (AS_{R-NR} , red) based on the web-association metric. The population of AS_{R-R} values was significantly higher than the population of AS_{R-NR} values according to a right-sided Wilcoxon signed-rank test ($p < 0.05$).

Considering all the 35 pluri-responsive units with at least two searchable responsive stimuli, we found that the association scores for responses were significantly higher than the scores for other pairs (right sided Wilcoxon signed-rank test, $p < 0.05$, Figure 4.16). This result reproduces the finding from our previous study (Section 3.3), that MTL pluri-responsive neurons fire more likely to associated concepts.

For the exemplary units shown in Figures 4.5 and 4.13 the association scores (reported in the figures' captions) reflected the population trend, with AS_{R-R} being larger than AS_{R-NR} in both examples. Hence, for these particular units, the responsive stimuli were strongly associated to each other. On the other hand, some pluri-responsive units did not follow the population trend and exhibited an association score for responses smaller than for other pairs ($AS_{R-R} < AS_{R-NR}$). In particular, when looking only at the units showing a significant decoding performance (Section 4.2.1), we found that $AS_{R-R} < AS_{R-NR}$ in 5 out of 7 cases. Moreover, in some cases, this opposite trend effect could be explained by the fact that the responsive stimuli were related by personal associations (only meaningful for the patient) that the web-score could not capture.

Table 4.1 contains a summary of the number of neurons tested against the personal association metric.

	Units
Recorded units	393
Units responding to at least 1 stimulus	76
Units responding to at least 2 stimuli	35
Units responding to related stimuli according to the web-based association metric ($AS_{R-R} > AS_{R-NR}$)	23
Units responding to unrelated stimuli according to the web-based association metric ($AS_{R-R} \leq AS_{R-NR}$)	12
Units responding to a single stimulus	41
Units not significantly responding to any stimulus	317

Table 4.1 – Number of units tested against the web-based association metric (21 recording sessions).

4.4.2 Differences in neural responses and association

We further analysed the relationship between association strength and similarity in the electrophysiological responses. First, we studied the correlation between association scores and strength/latency differences for the responsive pairs. Considering all the pairs of significant responses with a valid association score (i.e. the pairs of stimuli that were searchable on the web, 139 pairs from 35 units), we calculated the Pearson correlation coefficient (r) between the normalised differences in response strength and the corresponding association scores. As in Figure 4.7, the strength difference was normalised by the highest strength value in the pair. The same test was used to assess the correlation between absolute differences in spiking latencies and association scores on pairs of responses (139 pairs, as before).

Figure 4.17 shows these correlations. Panel a illustrates that the association score had an inverse correlation with the normalised difference in the response strength (Pearson correlation, $r = -0.21$, $p < 0.05$). A similar correlation value was found between the association score and spike latency difference (panel b, Pearson correlation, $r = -0.28$, $p < 10^{-3}$). The negative value of the correlation coefficients indicates that more associated stimuli exhibited smaller differences in both electrophysiological quantities examined (strength and latency), meaning that they triggered a more similar neural response compared to less associated stimuli.

Furthermore, the set of responsive pairs was split in two halves (Low and High association groups) based on the median value of the normalised association scores between all responsive pairs. Consistently with the correlation results, the High association group showed significantly smaller differences than the Low association one, in both normalised strength and spike latency (rank-sum test, $p < 10^{-2}$ and $p < 10^{-3}$ respectively).

For each group of responsive pairs, separated by association score, we

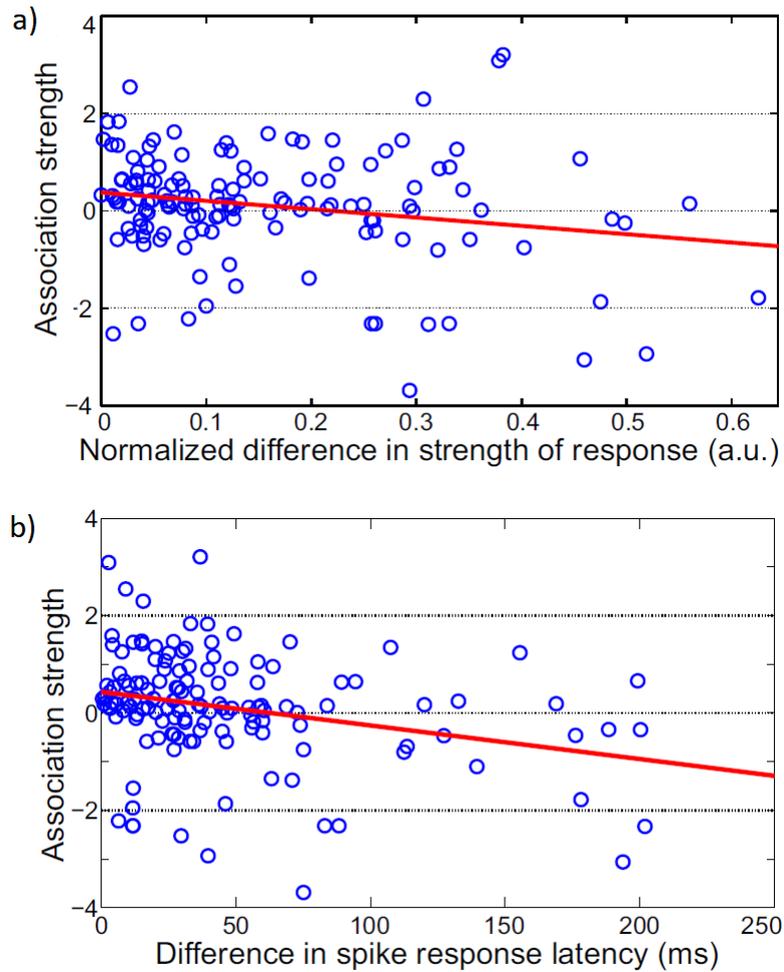


Figure 4.17 – a) Normalised difference in response strength vs association strength (association score for each pair of responsive stimuli) for 139 responsive pairs plus linear regression line (red). Pearson correlation $r = -0.21$ ($p < 0.05$). b) Difference in spike response latency vs association strength for 139 responsive pairs. Pearson correlation $r = -0.28$ ($p < 10^{-3}$).

calculated the proportion of significant comparisons (pL and pH) obtained from the surrogate tests (Figure 4.18). The two proportions pL and pH were compared using a right-sided Z test to test the null hypothesis that the high association group contained fewer differences. We found that the pairs of responses highly associated exhibited a smaller percentage of differences in response strength (pH=17%) compared to the ones with low scores (pL=32%, Z test, $p < 0.05$). The percentage of difference in spike latency followed the same trend (pH=14% and pL=20%), but the comparison did not reach the significance threshold (Z test, $p = 0.17$). Moreover, when considering differences in both strength or spike latency, the percentage of pairs with at least one significant difference went down from 33% overall to 25% in the high association group (and reached 42% in the low association group).

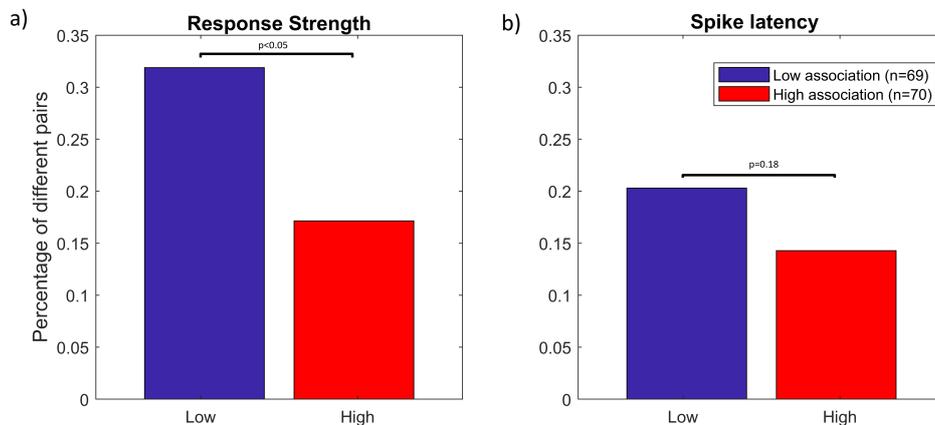


Figure 4.18 – a) Percentage of responsive pairs exhibiting significant difference in response strength (according to the surrogate test) separated by Low and High association scores (groups split according to the median association values). The number of different pairs is higher in the Low association group (Z test, $p < 0.05$). b) Percentage of responsive pairs exhibiting significant difference in spike latency separated by Low and High association scores. The difference in percentage between the two groups did not reach statistical significance (Z test, $p = 0.17$).

4.5 Response characteristics with number of responses

Finally, we analysed whether responsive units exhibited variations in the response characteristics according to the number of significant responses.

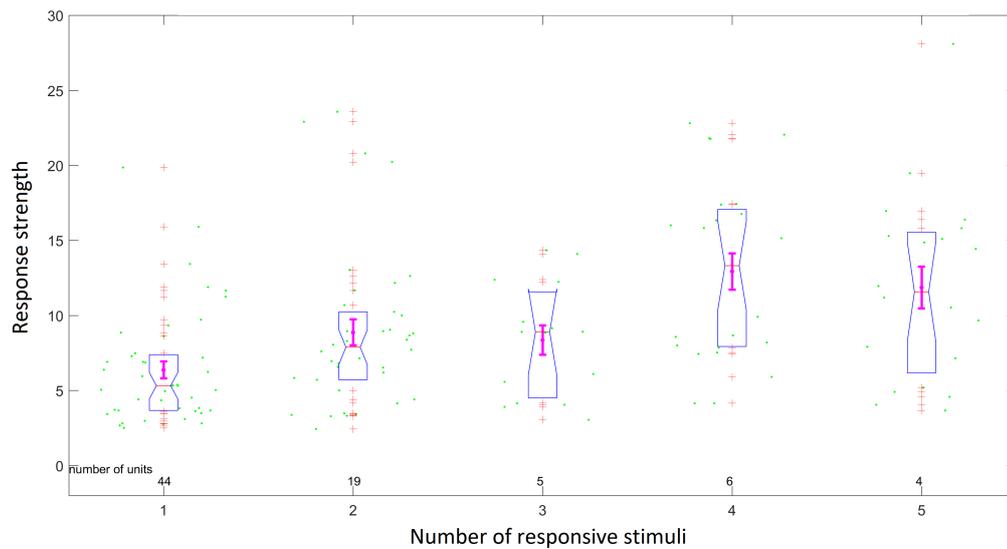


Figure 4.19 – Response strength of individual responses (green dots), separated by number of responsive stimuli in the unit. Pluri-responsive units exhibited a larger response strength (Spearman correlation, $\rho = 0.43$, $p < 10^{-6}$). Mean values \pm SEM across units in each group are shown in magenta, while boxplots are shown in blue.

The whole set of 81 responsive units was grouped according to the number of stimuli eliciting a response. The three units with more than five responses were excluded from this computation as they led to groups with a single element, leaving a total of 78 units classified in five groups (from 1 to 5 significant responses). For each unit, we considered the response strength of individual responses. The correlation between the response strength values and the number of significant responses for the 78 units was assessed with a non-parametric correlation analysis. We found a direct correlation between response strength and number of responsive stimuli (Figure 4.19, Spearman correlation, $\rho = 0.43$, $p < 10^{-6}$). That is, units firing to more stimuli exhibited higher firing rates

than units firing to fewer ones. The fact that pluri-responsive units responded with higher firing-activity could suggest that they were either more strongly wired into the network [Yassin et al., 2010, Barth and Poulet, 2012], or that they had a higher intrinsic excitability (since more excitable neurons might have a higher chance of being recruited to encode associations, e.g. [Epsztein et al., 2011]).

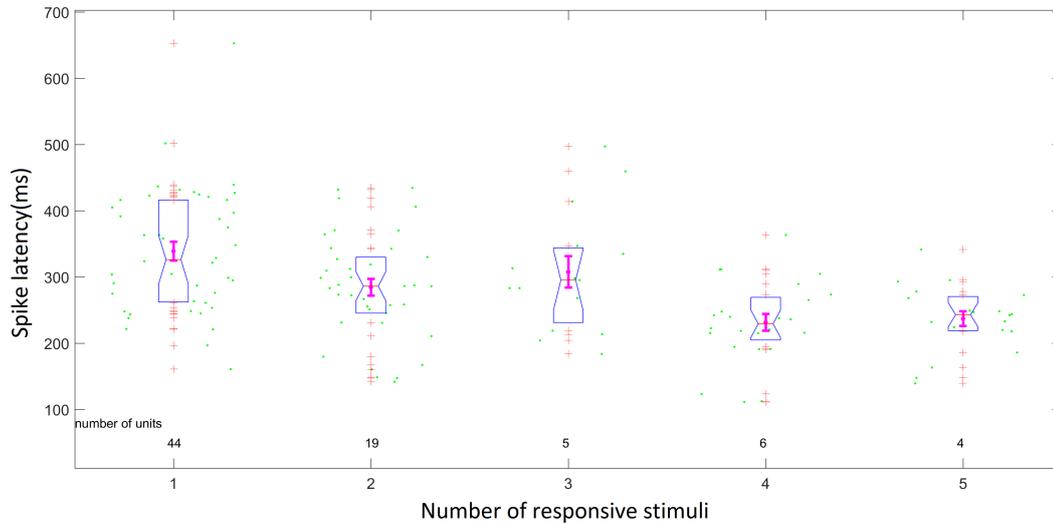


Figure 4.20 – Spike response latencies for individual responses (green dots) separated by number of responsive stimuli in the unit. Latencies were significantly correlated with the number of responses in the unit (Spearman correlation, $\rho = -0.45$, $p < 10^{-7}$). Mean values \pm SEM across units in each group are shown in magenta, while boxplots are shown in blue.

In the same way, we tested the correlation between the number of significant responses and spike latency for the 78 units. In this case, we found that the two quantities had an inverse correlation (Figure 4.20, Spearman correlation, $\rho = -0.45$, $p < 10^{-7}$), meaning that pluri-responsive units tended to fire earlier than units firing to just one stimulus. This result could be simply related to the higher excitability of pluri-responsive neurons found before (i.e. more excitable units might fire more quickly) or could be the effect of a process of neural unitisation (i.e. when two stimuli become encoded in a unitised manner the unit adopts the earlier latency between them, as discussed below). To rule

out the possibility that the effect was due to the higher firing rate found for pluri-responsive units, we run an extra control test. We chose 20 units from the single and pluri-responsive groups with similar baseline strengths (estimated in the window from -1,000 ms to -300 ms, centred around the average of the median strength of each distribution) and still found that the latencies for the pluri-responsive group were significantly earlier than the others (rank-sum test, $p < 10^{-2}$). Therefore, the earlier latencies found for pluri-responsive units are not related to the neuron excitability, but could actually be the result of a process of neural unitisation. In this context we refer to neural unitisation in the sense that a neuron responds equally (in terms of firing rate and latency) to different associated stimuli; as reported by Fujimichi et al. [2010], where neurons in area 35 of the macaque perirhinal cortex have been observed to respond similarly to pairs of stimuli during a pair-association task. When more stimuli encoded by a single neuron become unitised, the neural responses to each of the associated stimuli will have a similar electrophysiological signature (i.e. showing similarity in terms of response strength and latency). Therefore, a neuron encoding unitised responses will adopt a single latency, which we hypothesise should be the earliest one among the latencies of the individual stimuli.

4.6 Conclusions to Chapter 4

The first finding of this study was that when MTL single neurons responded to more than one stimulus, the spiking activity across the different responses was generally the same. In particular, we showed that: 1) the distribution of normalised firing rates in response to different stimuli was nearly binary, and the distribution of spiking activity in response to individual responsive stimuli and across responsive stimuli did not differ in mean, variability, or coefficient

of variation; 2) the firing rates of single units responding to different stimuli was the same in most cases (76% of responsive pairs exhibited no difference in the response strength); 3) this effect was not due to already small differences between the individual response (as proved by the random pairing analysis); and 4) the response spiking activity at the single trial level was not distinguishable across different responses (the stimuli could not be significantly decoded in 81% of pluri-responsive units). These results are in line with the all-or-none character of spiking response for a unit firing to the same consciously perceived stimulus [Quiñones Quiroga et al., 2008b], and extend this feature to the case of different responsive stimuli. Therefore, the type of representation encoded by these neurons is nearly binary, as opposed to the graded coding reported from cortical areas (e.g. Rolls and Tovee [1995], Logothetis and Sheinberg [1996], Tanaka [1996], Freiwald and Tsao [2010], Pasupathy and Connor [2002]). It has been shown that a graded coding can potentially store more information per pattern, but it reduces the number of patterns that can be stored in the network [Rolls et al., 1997]. The existence of a relatively binary encoding in the hippocampus represents an advantage in terms of robustness to noise, efficacy for retrieval and effective capacity of this memory system [Rolls et al., 1997, Rolls and Treves, 2011].

An analogous result was obtained analysing the response latencies of single units firing to different stimuli. In fact, 81% of responsive pairs exhibited no significant difference in latency and, again, the effect was not due to small difference between the population of responses (random pairing analysis on latencies). In addition, when considering differences in both response strength or latency, we found that more than half of cases did not exhibit any difference (67% of responsive pairs were not different in either strength or latency).

These findings suggest that, in general, the activity of a single unit is not enough to distinguish between different concepts. Therefore, they enforce the

interpretation [Quian Quiroga, 2012] that the information about the stimulus identity is encoded in the neural assembly.

According to the MTL associative model proposed by Quian Quiroga [2012] (discussed in Sections 1.4.1 and 3.10), the association between two items is created by overlapping of the neural assemblies encoding the individual concepts. Therefore, when a new association is formed some neurons from the individual assemblies will be recruited to be part of the overlap encoding it. In this regard, we have here replicated the main result from the study in Chapter 3 that MTL neurons in the overlap of two assemblies (firing to more than one stimulus) encode associations. In addition, we have now shown that the differences in firing characteristic were smaller for strongly associated stimuli (Section 4.4.2). Moreover, when looking only at the highly associated pairs of responses we found that the proportion of total differences dropped from 33% to 25% (i.e. more than three quarters of pairs showed no difference in either strength or latency).

Previous evidence [Ison et al., 2015] demonstrated that it is possible for MTL neurons to learn new associations very quickly (in just one trial). However, these newly-encoded associations showed more differences in strength and response latency compared to our dataset: 71% differences in strength of spiking activity ; and 38% of new responses with significantly later response onset (8 out of 21 units), assessed with rank-sum test on Poisson latencies.

A possible explanation to the discrepancies between newly-formed and long-term associations is in the process of association unitisation. The idea of neural unitisation presented here should not be confused with the concept of unitisation described in psychology literature, which defines unitisation as the encoding of previously separate items as a single unit [Graf and Schacter, 1989]. In our work, we refer to unitisation as whether different stimuli can be discriminated at the single neuron level, by the strength and latency of the neural response,

but without implying whether such stimuli can be discriminated or not at the behavioural level. At the single neuron level, when two items are encoded in a unitised manner, the neuron responds equally to both (as notably reported by Fujimichi et al. [2010] in area 35 of monkey perirhinal cortex). Besides, we know that only a small fraction of newly-encoded associations consolidates in a long-term representation in the MTL. In fact, as already noted in Section 3.10, the proportion of neurons that changed their selectivity to encode the new associations (about 40%, as reported by Ison et al. [2015]) was substantially larger than the maximum proportion of long-term association found in our previous study (4%, Section 3.6). Therefore, we expect that once associations are well consolidated in time, the different elements will become unitised and their neural responses will become more similar, which is in agreement with our results (fewer response differences in long-term associations). Unitisation of the representations at the single neuron level increases the overall capacity of the MTL and the robustness of the information encoded. Moreover, the existence of a unitisation process does not reduce the ability of MTL to distinguish the individual items, since, as we noted before, single concepts seem to be encoded in neural assemblies rather than at the single neuron level.

Finally, we showed a significant correlation between spiking activity and number of responsive stimuli, and an inverse correlation between response latency and number of responsive stimuli. The observed electrophysiological characteristics of pluri-responsive neurons (higher responsive firing rates and earlier latencies) can be explained in the framework of partially overlapping assemblies encoding associations in the MTL [Quiari Quiroga, 2012]. In particular, the first result could reflect the fact that the more excitable units in an assembly are more easily recruited to be part of the overlap and encode new associations (in agreement with findings from rodent's place cells [Epsztein et al., 2011]), and/or that the stronger connectivity of the network (following the recruitment

in the overlap) gives them more excitatory drive [Barth and Poulet, 2012]. In addition, the earlier latencies of pluri-responsive units compared to single responsive ones, are consistent with the possibility of a unitisation process of MTL representations. In fact, once more responsive concepts become unitised in a single neuron, the unit will start responding equally to them and will adopt a single latency for the unitised representation. This latency has to be the earliest across the latencies of individual encoded concepts in order to allow the representations of individual concepts to be integrated in the respective neural assemblies (i.e. if the representation of an individual concept was delayed in a particular neuron, that neuron could not participate in the activation of the assembly encoding the concept).

Chapter 5

Conclusions and future directions

The main objective of this thesis was to deepen our understanding of the role of human MTL neurons in the encoding of associations. Specifically, we conducted a systematic study to determine whether the activation of single neurons in response to different stimuli was just coincidental, or whether there was a significant tendency for MTL neurons to respond to associated items. Subsequently, we analysed the differences between spiking responses elicited by different stimuli in a single neuron, and how these differences related to the degree of association between stimuli. We proposed and employed a method to measure the degree of association between concepts based on the results of web searches. This approach (described in Chapter 2) was proved to be a good approximation of the patients' evaluations of associations and allowed us to extend our analysis to recording sessions for which we did not have a patient's report.

5.1 MTL neurons encode well-established associations

The main finding from the study described in Chapter 3 was that pluri-responsive neurons in the MTL tend to respond preferentially to associated items. This tendency was shown, at first, using the patient’s evaluation of associations (comparing the ‘personal’ association scores for stimuli eliciting and not eliciting response in a single neuron, Section 3.2). This finding was further supported by several analyses based on our web-based association metric (described in Sections 3.3, 3.6, and 3.7) and replicated in our second study (Section 4.4.1). Moreover, we ruled out alternative explanations for the results observed. In particular, we controlled for the effect of relative familiarity of the stimuli, visual similarity of the images, semantic category coding, and cue-recall responses (detailed in Sections 3.3.1, 3.3.2, 3.3.3, and 3.3.4, respectively). In addition, in agreement with the MTL hierarchical structure, we showed a tendency for the level of association between responsive stimuli to vary across MTL regions (being the highest in the hippocampus and the lowest in the parahippocampal cortex). Therefore, we concluded that MTL neurons encode well-established naturally-formed associations (the associations were pre-existent and not acquired through an associative learning task, since no such task was involved) and that the coding is not homogeneous among the different MTL areas.

In the ongoing debate about the role of the MTL in the consolidation of episodic memory (detailed in Section 1.3.5), our result supports the idea that the involvement of MTL in encoding episodic memories goes beyond a temporary coding, in line with the prediction of the multiple trace model [Moscovitch et al., 2005] (further details about the models and evidences in favour of the different views have been discussed in section 3.10). It is important to notice

that our findings do not rule out the possibility that the same associations might be still encoded in cortex, but rather provide new evidence about the MTL coding capacity, the stability of its representations, and its long-term involvement in memory processing.

5.2 Overlapping assemblies encode associations

As discussed in Section 1.4.1, it is well established that some neurons in the MTL respond in a very selective and invariant way to specific concepts [Quian Quiroga et al., 2005]. These so-called concept cells are thought to encode the ‘meaning’ of the presented stimuli rather than the perceptual details [Quian Quiroga et al., 2008a] and have been proposed to constitute the basis of declarative memory functions in the MTL [Quian Quiroga, 2012]. According to this view [Quian Quiroga, 2012], each concept is encoded in a cell assembly that, when activated, brings the specific concept into awareness. Concept cells represent single nodes of these neural assemblies. In addition, the association between concepts is thought to be encoded in the partial overlap of these cell networks. Specifically, two assemblies encoding related concepts will have some nodes (neurons) in common that encode both items. In line with this view, new associations can be created very quickly in the MTL, by changing the tuning of few neurons in the assemblies to encode the new information (in line with recent findings by Ison et al. [2015]).

Our results are in agreement with this framework of interpretation of the MTL processing. In fact, we have shown that when MTL neurons fire to more than one stimulus, these stimuli tend to be associated, suggesting that these pluri-responsive neurons belong to the overlap of two or more assemblies. Moreover, in agreement with findings from the rodent hippocampus [Redish et al., 2001], we found a lack of topographical organisation in the

MTL (close-by neurons do not tend to fire to associated concepts, Section 3.5). A non-topographical organisation of representations is indeed ideal for the rapid integration of new memories, since it allows the quick creation of new associations between arbitrary concepts [Quian Quiroga, 2012].

From the study presented in Chapter 4 we found that in most cases (67%), when a neuron responds significantly to more than one stimulus, its neural activity does not differ either in strength or latency in response to the different stimuli. In other words, this means that MTL neurons exhibit an all-or-none response (in line with evidence for responses to the same stimulus [Quian Quiroga et al., 2008b]) and that, in general, the spiking activity of a single unit is not enough to distinguish between different responsive concepts. This finding supports the idea [Quian Quiroga, 2012] that information about the stimulus identity is encoded in the neural assembly rather than at single neuron level.

In addition, when analysing the activity of neurons firing to one or multiple stimuli, we found a significant correlation between strength spiking activity and number of responsive stimuli (Section 4.5), that is: neurons in the overlap of assemblies (pluri-responsive) exhibit a higher responsive firing rate. This effect could mean that the more excitable units in an assembly are more likely to be recruited to encode new associations, and/or that the neurons that are part of the overlap, being more strongly connected, receive more excitatory drive [Barth and Poulet, 2012].

5.3 Unitisation of the neural responses

In Section 3.6, we have shown that the probability of getting responses to pairs of items increases with the strength of association between stimuli. Specifically, going from 1% for non-associated items to 4% for highly associated ones. The highest percentage value still seems to be quite small when compared to the pro-

portion of neurons that changed their selectivity to encode the new associations reported by Ison et al. [2015] (about 40%). This discrepancy suggests that only a small fraction of the newly-encoded associations consolidates into long-lasting representations in the MTL. Moreover, the newly-encoded associations (in Ison et al. [2015]) exhibited larger proportions of differences in both strength and response latency compared to the ones we found for the long-term ones (Sections 4.2.1, 4.2.3 and 4.3.2).

The existence of a process of neural unitisation of the responses could explain this difference between newly-formed and long-term associations. As detailed in Section 4.6, the concept of neural unitisation postulated here should not be confused with the one described in psychology literature [Graf and Schacter, 1989]. In this context, neural unitisation indicates that the responses to each of the associated stimulus have similar electrophysiological signatures. In other words, that a neuron responds equally (in terms of firing rate and latency) to different associated stimuli. A neuron responding to two items could initially exhibit distinct neural responses to each of them (i.e. different strength of response or latency), and then start to respond equally to the two items once they become unitised. Responses become unitised with consolidation in the MTL, in line with our finding that different elements of well consolidated associations trigger a more similar neural response than those of newly-encoded associations. Consistent with this process, we found that pluri-responsive neurons tend to fire earlier than neurons responding to just one stimulus (Section 4.5). This suggests that once representations become unitised in a single neuron, the unit will adopt a single latency which is the earliest among the latencies of the single responsive stimuli.

The existence of a unitisation process is in agreement with the small percentage of different neural responses found in Chapter 4. Unitisation of the representations at the single neuron level makes the encoded information more

robust to noise. In this regard, it increases the overall capacity of the MTL and does not lessen its ability to separate individual items (since concepts are encoded at the assembly level).

5.4 Future directions

The work presented in this thesis has addressed the questions asked at the beginning of the manuscript, helping to uncover unknown features of association coding in the MTL. However, many other open issues about the MTL functionality are left to investigate. Some of the possible future directions to explore are suggested here.

Firstly, it would be interesting to investigate the activity of several neurons firing to the same concepts (in other words, neurons belonging to the same neural assembly encoding that specific concept). Comparing their firing characteristics could help understand how information is transmitted through the assembly. For instance, we could ask if there is a temporary alignment of the firing activity in these neurons or whether single trial differences seen for one unit are replicated in the rest of the assembly.

Another possible future development of this line of research has to do with the plasticity of hippocampal representations. As we have seen before, MTL neurons can form new associations very rapidly and some associations are stably encoded in time (long-term associations). We might then ask what happens in between the two time-scales, how newly-encoded associations become strengthened, weakened or altered with the passing of time, and if we can interfere in the process and manipulate this outcome. Modern recording systems already allow the continuous recording of patients' intracranial activity up to several days. However, the current challenge is to be able to track the activity of single neurons over many hours of recording. In order to do so, we

will need to develop adaptive sorting algorithms able to take into account the slow happening changes in the spike-shape of single units (e.g. due to electrode drifting or to changes in the intrinsic cell characteristics), the sudden appearance or disappearance of active units, variation of noise level, and the contamination artefacts. Moreover, these algorithms will have to be computationally and time efficient in dealing with the large size of the datasets recorded (in the order of hundreds of gigabytes).

Once we are able to track the activity of single units for a longer time span we will also be able to explore whether, and in which manner, concept cells are involved in the different phases of memory processing. For example, we could investigate what happens in the single neurons during the conscious recall of a memory or whether there are endogenous reactivations of those neurons during consolidation (e.g. single neurons activated during the task might fire spontaneously during off-task periods), and how these events impact on memory performance. More generally, longer recordings of single neuron activity will unlock many new possibilities in the design of experimental paradigms that go beyond the execution of a single task.

Bibliography

- L. Abbott, E. T. Rolls, and M. J. Tovee. Representational capacity of face coding in monkeys. *Cerebral Cortex*, 6(3):498–505, 1996.
- P. Alvarez and L. R. Squire. Memory consolidation and the medial temporal lobe: a simple network model. *Proceedings of the National Academy of Sciences*, 91(15):7041–7045, 1994.
- H. Barlow. The neuron doctrine in perception. 1995.
- A. L. Barth and J. F. Poulet. Experimental evidence for sparse firing in the neocortex. *Trends in neurosciences*, 35(6):345–355, 2012.
- F. C. Bartlett. *Remembering*. Cambridge Univ. Press.[LRS], 1932.
- P. J. Bayley, R. O. Hopkins, and L. R. Squire. The fate of old memories after medial temporal lobe damage. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 26(51):13311–13317, Dec 20 2006.
- P. Bellgowan, E. Buffalo, J. Bodurka, and A. Martin. High resolution imaging of the anterior medial temporal lobe during object and spatial memory. In *Soc. Neurosci*, volume 556, 2003.
- M. Blatt, S. Wiseman, and E. Domany. Superparamagnetic clustering of data. *Physical review letters*, 76(18):3251, 1996.

- V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.
- K. Brodmann. *Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues*. Barth, 1909.
- M. Bunsey and H. Eichenbaum. Conservation of hippocampal memory function in rats and humans. *Nature*, 379(6562):255–257, Jan 18 1996.
- R. D. Burwell. The parahippocampal region: corticocortical connectivity. *Annals of the New York Academy of Sciences*, 911(1):25–42, 2000.
- R. D. Burwell and K. L. Agster. *Anatomy of the Hippocampus and the Declarative Memory System*, chapter 3.03. Elsevier Ltd, 2008.
- G. Buzsaki. *Rhythms of the Brain*. Oxford University Press, 2006.
- G. Buzsaki, C. A. Anastassiou, and C. Koch. The origin of extracellular fields and currents eeg, ecog, lfp and spikes. *Nature reviews neuroscience*, 13(6):407–420, 2012.
- P. Cahusac, E. T. Rolls, Y. Miyashita, and H. Niki. Modification of the responses of hippocampal neurons in the monkey during the learning of a conditional spatial response task. *Hippocampus*, 3(1):29–42, 1993.
- K. A. Cameron, S. Yashar, C. L. Wilson, and I. Fried. Human hippocampal neurons predict how well word pairs will be remembered. *Neuron*, 30(1):289–298, 2001.
- M. Cerf, N. Thiruvengadam, F. Mormann, A. Kraskov, R. Quiñan Quiroga, C. Koch, and I. Fried. On-line, voluntary control of human temporal lobe neurons. *Nature*, 467(7319):1104–1108, 2010.

- S. Corkin. Tactually-guided maze learning in man: Effects of unilateral cortical excisions and bilateral hippocampal lesions. *Neuropsychologia*, 3(4):339–351, 1965.
- S. Corkin. Acquisition of motor skill after bilateral medial temporal-lobe excision. *Neuropsychologia*, 6(3):255–265, 1968.
- S. Corkin. Lasting consequences of bilateral medial temporal lobectomy: Clinical course and experimental findings in hm. In *Seminars in Neurology*, volume 4, pages 249–259. Thieme Medical Publishers, Inc., 1984.
- S. Corkin, N. Cohen, and H. Sagar. Memory for remote personal and public events after bilateral medial temporal lobectomy. In *Society for Neuroscience Abstracts*, volume 9, page 28, 1983.
- M. Day, R. Langston, and R. G. Morris. Glutamate-receptor-mediated encoding and retrieval of paired-associate learning. *Nature*, 424(6945):205–209, 2003.
- E. De Falco, M. J. Ison, I. Fried, and R. Quiñan Quiroga. Long-term coding of personal and universal associations underlying the memory web in the human brain. *Nature Communications*, 7:13408, 2016.
- F. Dolcos, K. S. LaBar, and R. Cabeza. Dissociable effects of arousal and valence on prefrontal activity indexing emotional evaluation and subsequent memory: an event-related fmri study. *Neuroimage*, 23(1):64–74, 2004.
- H. Eichenbaum. Hippocampus: cognitive processes and neural representations that underlie declarative memory. *Neuron*, 44(1):109–120, 2004.
- H. Eichenbaum. What h.m. taught us. *Journal of cognitive neuroscience*, 25(1):14–21, 2013.
- H. Eichenbaum and N. J. Cohen. Can we reconcile the declarative memory and

- spatial navigation views on hippocampal function? *Neuron*, 83(4):764–770, 2014.
- H. Eichenbaum, P. Dudchenko, E. Wood, M. Shapiro, and H. Tanila. The hippocampus, memory, and place cells: is it spatial memory or a memory space? *Neuron*, 23(2):209–226, 1999.
- H. Eichenbaum, A. R. Yonelinas, and C. Ranganath. The medial temporal lobe and recognition memory. *Annual Review of Neuroscience*, 30:123–152, 2007.
- A. D. Ekstrom, M. J. Kahana, J. B. Caplan, T. A. Fields, E. A. Isham, E. L. Newman, and I. Fried. Cellular networks underlying human spatial navigation. *Nature*, 425(6954):184–188, 2003.
- A. K. Engel, C. K. Moll, I. Fried, and G. A. Ojemann. Invasive recordings from the human brain: clinical insights and beyond. *Nature Reviews Neuroscience*, 6(1):35–47, 2005.
- R. Epstein and N. Kanwisher. A cortical representation of the local visual environment. *Nature*, 392(6676):598–601, 1998.
- J. Epsztein, M. Brecht, and A. K. Lee. Intracellular determinants of hippocampal ca1 place and silent cell activity in a novel environment. *Neuron*, 70(1):109–120, 2011.
- C. A. Erickson and R. Desimone. Responses of macaque perirhinal neurons during and after visual stimulus association learning. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 19(23):10404–10416, Dec 1 1999. LR: 20061115; JID: 8102140; ppublish.
- R. M. Fano. *Transmission of Information: A Statistical Theory of Communications*. M.I.T. Press, Cambridge, MA., 1961.

- W. A. Freiwald and D. Y. Tsao. Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science*, 330(6005):845–851, 2010.
- I. Fried, K. A. MacDonald, and C. L. Wilson. Single neuron activity in human hippocampus and amygdala during recognition of faces and objects. *Neuron*, 18(5):753–765, 5 1997.
- R. Fujimichi, Y. Naya, K. W. Koyano, M. Takeda, D. Takeuchi, and Y. Miyashita. Unitized representation of paired objects in area 35 of the macaque perirhinal cortex. *European Journal of Neuroscience*, 32(4):659–667, 2010.
- T. Funahashi and H. Yamana. *Reliability Verification of Search Engines' Hit Counts: How to Select a Reliable Hit Count for a Query*, pages 114–125. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.
- H. Gelbard-Sagiv, R. Mukamel, M. Harel, R. Malach, and I. Fried. Internally generated reactivation of single neurons in human hippocampus during free recall. *Science*, 322(5898):96–101, 2008.
- K.-I. Goh and A.-L. Barabasi. Burstiness and memory in complex systems. *EPL (Europhysics Letters)*, 81(4):48002, 2008.
- C. Gold, D. A. Henze, C. Koch, and G. Buzsáki. On the origin of the extracellular action potential waveform: a modeling study. *Journal of neurophysiology*, 95(5):3113–3128, 2006.
- M. A. Goodale and A. D. Milner. Separate visual pathways for perception and action. *Trends in neurosciences*, 15(1):20–25, 1992.
- P. Graf and D. L. Schacter. Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11(3):501, 1985.

- P. Graf and D. L. Schacter. Unitization and grouping mediate dissociations in memory for new associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(5):930, 1989.
- B. Gustafsson. Afterpotentials and transduction properties in different types of central neurones. *Archives italiennes de biologie*, 122(1):17–30, 1984.
- T. Hafting, M. Fyhn, S. Molden, M.-B. Moser, and E. I. Moser. Microstructure of a spatial map in the entorhinal cortex. *Nature*, 436(7052):801–806, 2005.
- E. Halgren, T. L. Babb, and P. H. Crandall. Activity of human hippocampal formation and amygdala neurons during memory testing. *Electroencephalography and Clinical Neurophysiology*, 45(5):585–601, 1978.
- Y. Harada and T. Takahashi. The calcium component of the action potential in spinal motoneurons of the rat. *The Journal of physiology*, 335:89, 1983.
- T. Hartley, E. A. Maguire, H. J. Spiers, and N. Burgess. The well-worn route and the path less traveled: distinct neural bases of route following and wayfinding in humans. *Neuron*, 37(5):877–888, 2003.
- S. Hattori, L. Chen, C. Weiss, and J. F. Disterhoft. Robust hippocampal responsivity during retrieval of consolidated associative memory. *Hippocampus*, 25(5):655–669, 2015.
- D. W. Heermann and K. Binder. *Monte Carlo Simulation in Statistical Physics*. Springer-Verlag Berlin Heidelberg, 1988.
- G. Heit, M. E. Smith, and E. Halgren. Neural encoding of individual words and faces by the human hippocampus and amygdala. *Nature*, 333(6175):773–775, 1988.
- S. Higuchi and Y. Miyashita. Formation of mnemonic neuronal responses to visual paired associates in inferotemporal cortex is impaired by perirhinal

- and entorhinal lesions. *Proceedings of the National Academy of Sciences of the United States of America*, 93(2):739–743, Jan 23 1996.
- R. Insausti and D. G. Amaral. *Hippocampal Formation*, chapter 24. Elsevier Inc., 2012.
- R. Insausti, D. G. Amaral, and W. Cowan. The entorhinal cortex of the monkey: li. cortical afferents. *Journal of Comparative Neurology*, 264(3):356–395, 1987.
- M. J. Ison, F. Mormann, M. Cerf, C. Koch, I. Fried, and R. Quiñan Quiroga. Selectivity of pyramidal cells and interneurons in the human medial temporal lobe. *Journal of neurophysiology*, 106(4):1713–1721, 2011.
- M. J. Ison, R. Quiñan Quiroga, and I. Fried. Rapid encoding of new memories by individual neurons in the human brain. *Neuron*, 87(1):220–230, 7/1 2015.
- A. Kamondi, L. Acsády, X.-J. Wang, and G. Buzsáki. Theta oscillations in somata and dendrites of hippocampal pyramidal cells in vivo: Activity-dependent phase-precession of action potentials. *Hippocampus*, 8(3):244–261, 1998.
- E. R. Kandel, J. H. Schwartz, and T. M. Jessell. *Principles of neural science*. McGraw-Hill, Health Professions Division, New York, 2000. ISBN 0838577016 9780838577011 0071120009 9780071120005. ID: 42073108.
- E. R. Kandel, J. H. Schwartz, and T. M. Jessell. *Principles of Neural Science, Fifth Edition*. McGraw-Hill Education, 2013.
- E. A. Kensinger, M. T. Ullman, and S. Corkin. Bilateral medial temporal lobe damage does not affect lexical or grammatical processing: Evidence from amnesic patient hm. *Hippocampus*, 11(4):347–360, 2001.
- R. P. Kesner and E. T. Rolls. A computational theory of hippocampal function,

- and tests of the theory: new developments. *Neuroscience & Biobehavioral Reviews*, 48:92–147, 2015.
- A. Konkel and N. J. Cohen. Relational memory and the hippocampus: representations and methods. *Frontiers in neuroscience*, 3:23, 2009.
- G. Kreiman, C. Koch, and I. Fried. Imagery neurons in the human brain. *Nature*, 408(6810):357–361, 2000a.
- G. Kreiman, C. Koch, and I. Fried. Category-specific visual responses of single neurons in the human medial temporal lobe. *Nature neuroscience*, 3(9):946–953, print 2000b. 10.1038/78868.
- G. Kreiman, I. Fried, and C. Koch. Single-neuron correlates of subjective vision in the human medial temporal lobe. *Proceedings of the national academy of sciences*, 99(12):8378–8383, 2002.
- K. S. LaBar and R. Cabeza. Cognitive neuroscience of emotional memory. *Nature Reviews Neuroscience*, 7(1):54–64, 2006.
- P. Lavenex and D. G. Amaral. Hippocampal-neocortical interaction: A hierarchy of associativity. *Hippocampus*, 10(4):420–430, 2000.
- X.-G. Li, P. Somogyi, A. Ylinen, and G. Buzsáki. The hippocampal ca3 network: an in vivo intracellular labeling study. *Journal of comparative neurology*, 339(2):181–208, 1994.
- N. K. Logothetis. The underpinnings of the bold functional magnetic resonance imaging signal. *Journal of Neuroscience*, 23(10):3963–3971, 2003.
- N. K. Logothetis and D. L. Sheinberg. Visual object recognition. *Annual review of neuroscience*, 19(1):577–621, 1996.

- E. A. Maguire, R. S. Frackowiak, and C. D. Frith. Recalling routes around london: activation of the right hippocampus in taxi drivers. *Journal of neuroscience*, 17(18):7103–7110, 1997.
- E. A. Maguire, N. Burgess, J. G. Donnett, R. S. Frackowiak, C. D. Frith, and J. O’keefe. Knowing where and getting there: a human navigation network. *Science*, 280(5365):921–924, 1998.
- E. A. Maguire, R. N. Henson, C. J. Mummery, and C. D. Frith. Activity in prefrontal cortex, not hippocampus, varies parametrically with the increasing remoteness of memories. *Neuroreport*, 12(3):441–444, 2001.
- E. A. Maguire, R. Nannery, and H. J. Spiers. Navigation around london by a taxi driver with bilateral hippocampal lesions. *Brain*, 129(11):2894–2907, 2006.
- L. Malkova and M. Mishkin. One-trial memory for object-place associations after separate lesions of hippocampus and posterior parahippocampal region in the monkey. *Journal of Neuroscience*, 23(5):1956–1965, 2003.
- S. G. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE transactions on pattern analysis and machine intelligence*, 11(7):674–693, 1989.
- J. R. Manns, R. O. Hopkins, J. M. Reed, E. G. Kitchener, and L. R. Squire. Recognition memory and the human hippocampus. *Neuron*, 37(1):171–180, 2003a.
- J. R. Manns, R. O. Hopkins, and L. R. Squire. Semantic memory and the human hippocampus. *Neuron*, 38(1):127–133, 2003b.
- J. L. McClelland, B. L. McNaughton, and R. C. O’reilly. 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(3):419, 1995.
- J. L. McGaugh. The amygdala modulates the consolidation of memories of emotionally arousing experiences. *Annu. Rev. Neurosci.*, 27:1–28, 2004.
- S. McKenzie and H. Eichenbaum. Consolidation and reconsolidation: two lives of memories? *Neuron*, 71(2):224–233, 2011.
- S. McKenzie, A. J. Frank, N. R. Kinsky, B. Porter, P. D. Rivière, and H. Eichenbaum. Hippocampal representation of related and opposing memories develop within distinct, hierarchically organized neural schemas. *Neuron*, 83(1):202–215, 2014.
- B. Milner. Les troubles de la mémoire accompagnant des lésions hippocampiques bilatérales. *Physiologie de l’Hippocampe*, 1962.
- B. Milner, S. Corkin, and H.-L. Teuber. Further analysis of the hippocampal amnesic syndrome: 14-year follow-up study of hm. *Neuropsychologia*, 6(3):215–234, 1968.
- M. Mishkin. A memory system in the monkey. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 298(1089):85–95, 1982.
- U. Mitzdorf. Current source-density method and application in cat cerebral cortex: investigation of evoked potentials and eeg phenomena. 1985.
- Y. Miyashita, E. T. Rolls, P. M. Cahusac, H. Niki, and J. D. Feigenbaum. Activity of hippocampal formation neurons in the monkey related to a conditional spatial response task. *Journal of Neurophysiology*, 61(3):669–678, 1989.
- F. Mormann, S. Kornblith, R. Quiñero, A. Kraskov, M. Cerf, I. Fried, and C. Koch. Latency and selectivity of single neurons indicate hierarchical

- processing in the human medial temporal lobe. *Journal of Neuroscience*, 28 (36):8865–8872, 2008.
- F. Mormann, S. Kornblith, M. Cerf, M. J. Ison, A. Kraskov, M. Tran, S. Knieling, R. Quiñan Quiroga, C. Koch, and I. Fried. Scene-selective coding by single neurons in the human parahippocampal cortex. *Proceedings of the National Academy of Sciences*, page 201608159, 2017.
- M. Moscovitch, R. S. Rosenbaum, A. Gilboa, D. R. Addis, R. Westmacott, C. Grady, M. P. McAndrews, B. Levine, S. Black, and G. Winocur. Functional neuroanatomy of remote episodic, semantic and spatial memory: a unified account based on multiple trace theory. *Journal of anatomy*, 207(1):35–66, 2005.
- R. Mukamel and I. Fried. Human intracranial recordings and cognitive neuroscience. *Annual review of psychology*, 63:511–537, 2012.
- L. Nadel and M. Moscovitch. Memory consolidation, retrograde amnesia and the hippocampal complex. *Current opinion in neurobiology*, 7(2):217–227, 1997.
- Y. Naya, M. Yoshida, and Y. Miyashita. Backward spreading of memory-retrieval signal in the primate temporal cortex. *Science (New York, N.Y.)*, 291(5504):661–664, Jan 26 2001.
- S. Nemanic, M. C. Alvarado, and J. Bachevalier. The hippocampal/parahippocampal regions and recognition memory: insights from visual paired comparison versus object-delayed nonmatching in monkeys. *Journal of Neuroscience*, 24(8):2013–2026, 2004.
- G. A. Ojemann, O. Creutzfeldt, E. Lettich, and M. M. Haglund. Neuronal activity in human lateral temporal cortex related to short-term verbal memory, naming and reading. *Brain*, 111(6):1383–1403, 1988.

- J. O'Keefe and J. Dostrovsky. The hippocampus as a spatial map. preliminary evidence from unit activity in the freely-moving rat. *Brain research*, 34(1): 171–175, 1971.
- J. O'keefe and L. Nadel. *The hippocampus as a cognitive map*. Oxford: Clarendon Press, 1978.
- J. Parkinson, E. Murray, and M. Mishkin. A selective mnemonic role for the hippocampus in monkeys: memory for the location of objects. *Journal of Neuroscience*, 8(11):4159–4167, 1988.
- A. Pasupathy and C. E. Connor. Population coding of shape in area v4. *Nature neuroscience*, 5(12), 2002.
- R. B. Potts. Some generalized order-disorder transformations. In *Mathematical proceedings of the cambridge philosophical society*, volume 48, pages 106–109. Cambridge Univ Press, 1952.
- D. Purves, G. J. Augustine, D. Fitzpatrick, W. C. Hall, A.-S. Lamantia, J. O. McNamara, and S. M. Williams. *Neuroscience*. Sinauer Associates Inc, 2004.
- R. Quian Quiroga. Spike sorting. 2(12):3583, 2007. revision 137442.
- R. Quian Quiroga. Concept cells: the building blocks of declarative memory functions. *Nature Reviews Neuroscience*, 13(8):587–597, 2012.
- R. Quian Quiroga and S. Panzeri. Extracting information from neuronal populations: information theory and decoding approaches. *Nature Reviews Neuroscience*, 10(3):173–185, 2009.
- R. Quian Quiroga and S. Panzeri. *Principles of neural coding*. CRC Press, 2013.

- R. Quian Quiroga, Z. Nadasdy, and Y. Ben-Shaul. Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural computation*, 16(8):1661–1687, 2004.
- R. Quian Quiroga, L. Reddy, G. Kreiman, C. Koch, and I. Fried. Invariant visual representation by single neurons in the human brain. *Nature*, 435(7045):1102–1107, Jun 23 2005.
- R. Quian Quiroga, L. Reddy, C. Koch, and I. Fried. Decoding visual inputs from multiple neurons in the human temporal lobe. *Journal of neurophysiology*, 98(4):1997–2007, 2007.
- R. Quian Quiroga, G. Kreiman, C. Koch, and I. Fried. Sparse but not ‘grandmother-cell’ coding in the medial temporal lobe. *Trends in cognitive sciences*, 12(3):87–91, Mar 2008a.
- R. Quian Quiroga, R. Mukamel, E. A. Isham, R. Malach, and I. Fried. Human single-neuron responses at the threshold of conscious recognition. *Proceedings of the National Academy of Sciences*, 105(9):3599–3604, 2008b.
- R. Quian Quiroga, A. Kraskov, C. Koch, and I. Fried. Explicit encoding of multimodal percepts by single neurons in the human brain. *Current biology : CB*, 19(15):1308–1313, Aug 11 2009.
- S. Ramón y Cajal. *Textura del sistema nervioso del hombre y de los vertebrados*. Moya, Madrid, 1904.
- S. Ramón y Cajal. 1911. *Histologie du Système Nerveux de l’Homme et des Vertèbrès*, 1909.
- A. D. Redish, F. P. Battaglia, M. K. Chawla, A. D. Ekstrom, J. L. Gerard, P. Lipa, E. S. Rosenzweig, P. F. Worley, J. F. Guzowski, and B. L.

- McNaughton. Independence of firing correlates of anatomically proximate hippocampal pyramidal cells. *J Neurosci*, 21(5):RC134, 2001.
- H. G. Rey, I. Fried, and R. Quian Quiroga. Timing of single-neuron and local field potential responses in the human medial temporal lobe. *Current Biology*, 24(3):299–304, 2014a.
- H. G. Rey, M. J. Ison, C. Pedreira, A. Valentin, G. Alarcon, R. Selway, M. P. Richardson, and R. Quian Quiroga. Single-cell recordings in the human medial temporal lobe. *Journal of anatomy*, Aug 28 2014b.
- H. G. Rey, C. Pedreira, and R. Quian Quiroga. Past, present and future of spike sorting techniques. *Brain research bulletin*, 119:106–117, 2015.
- M. Ritchey, F. Dolcos, and R. Cabeza. Role of amygdala connectivity in the persistence of emotional memories over time: An event-related fmri investigation. *Cerebral Cortex*, 18(11):2494–2504, 2008.
- E. Rolls, Y. Miyashita, P. Cahusac, R. Kesner, H. Niki, J. Feigenbaum, and L. Bach. Hippocampal neurons in the monkey with activity related to the place in which a stimulus is shown. *Journal of Neuroscience*, 9(6):1835–1845, 1989.
- E. T. Rolls. A theory of hippocampal function in memory. *Hippocampus*, 6(6):601–620, 1996.
- E. T. Rolls. *Memory, attention, and decision-making: a unifying computational neuroscience approach*. 2008.
- E. T. Rolls and M. J. Tovee. Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *Journal of neurophysiology*, 73(2):713–726, 1995.

- E. T. Rolls and A. Treves. The relative advantages of sparse versus distributed encoding for associative neuronal networks in the brain. *Network: computation in neural systems*, 1(4):407–421, 1990.
- E. T. Rolls and A. Treves. The neuronal encoding of information in the brain. *Progress in neurobiology*, 95(3):448–490, 2011.
- E. T. Rolls, A. Treves, D. Foster, and C. Perez-Vicente. Simulation studies of the ca3 hippocampal subfield modelled as an attractor neural network. *Neural Networks*, 10(9):1559–1569, 1997.
- E. T. Rolls, A. Treves, R. G. Robertson, P. Georges-François, and S. Panzeri. Information about spatial view in an ensemble of primate hippocampal cells. *Journal of Neurophysiology*, 79(4):1797–1813, 1998.
- E. T. Rolls, J. Xiang, and L. Franco. Object, space, and object-space representations in the primate hippocampus. *Journal of neurophysiology*, 94(1):833–844, 2005.
- E. T. Rolls et al. Spatial view cells and the representation of place in the primate hippocampus. *Hippocampus*, 9(4):467–480, 1999.
- R. S. Rosenbaum, S. Köhler, D. L. Schacter, M. Moscovitch, R. Westmacott, S. E. Black, F. Gao, and E. Tulving. The case of kc: contributions of a memory-impaired person to memory theory. *Neuropsychologia*, 43(7):989–1021, 2005.
- M. Rubinov and O. Sporns. Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage*, 52(3):1059–1069, 9 2010.
- Y. Rubner, C. Tomasi, and L. J. Guibas. The earth mover’s distance as a metric for image retrieval. *International journal of computer vision*, 40(2):99–121, 2000.

- L. Ryan, L. Nadel, K. Keil, K. Putnam, D. Schnyer, T. Trouard, and M. Moscovitch. Hippocampal complex and retrieval of recent and very remote autobiographical memories: evidence from functional magnetic resonance imaging in neurologically intact people. *Hippocampus*, 11(6):707–714, 2001.
- K. Sakai and Y. Miyashita. Neural organization for the long-term memory of paired associates. *Nature*, 354(6349):152–155, 11/14 1991. 10.1038/354152a0.
- D. Schacter. Implicit memory: History and current status. *J. Experimental Psychol. Learning, Memory, Cognition*, 1987.
- D. Schiller, H. Eichenbaum, E. A. Buffalo, L. Davachi, D. J. Foster, S. Leutgeb, and C. Ranganath. Memory and space: towards an understanding of the cognitive map. *Journal of Neuroscience*, 35(41):13904–13911, 2015.
- H. Schmolck, E. A. Kensinger, S. Corkin, and L. R. Squire. Semantic knowledge in patient hm and other patients with bilateral medial and lateral temporal lobe lesions. *Hippocampus*, 12(4):520–533, 2002.
- W. B. Scoville and B. Milner. Loss of recent memory after bilateral hippocampal lesions. *Journal of neurology, neurosurgery, and psychiatry*, 20(1):11–21, 02 1957. J1: J Neurol Neurosurg Psychiatry.
- A. L. Shelton and J. D. Gabrieli. Neural correlates of encoding space from route and survey perspectives. *Journal of Neuroscience*, 22(7):2711–2717, 2002.
- C. N. Smith and L. R. Squire. Medial temporal lobe activity during retrieval of semantic memory is related to the age of the memory. *Journal of Neuroscience*, 29(4):930–938, 2009.
- L. Squire and S. Zola-Morgan. The medial temporal lobe memory system. *Science*, 253(5026):1380–1386, September 20, 1991 1991.

- L. R. Squire. *Memory and brain*. New York: Oxford, 1987.
- L. R. Squire. Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. *Psychological review*, 99(2):195, 1992.
- L. R. Squire and P. Alvarez. Retrograde amnesia and memory consolidation: a neurobiological perspective. *Current opinion in neurobiology*, 5(2):169–177, 1995.
- L. R. Squire and S. M. Zola. Structure and function of declarative and non-declarative memory systems. *Proceedings of the National Academy of Sciences*, 93(24):13515–13522, 1996.
- L. R. Squire and S. M. Zola. Episodic memory, semantic memory, and amnesia. *Hippocampus*, 8(3):205–211, 1998.
- L. R. Squire, C. E. Stark, and R. E. Clark. The medial temporal lobe. *Annu.Rev.Neurosci.*, 27:279–306, 2004.
- S. Steinvorth, B. Levine, and S. Corkin. Medial temporal lobe structures are needed to re-experience remote autobiographical memories: evidence from hm and wr. *Neuropsychologia*, 43(4):479–496, 2005.
- W. A. Suzuki. Neuroanatomy of the monkey entorhinal, perirhinal and parahippocampal cortices: organization of cortical inputs and interconnections with amygdala and striatum. In *Seminars in Neuroscience*, volume 8, pages 3–12. Elsevier, 1996.
- K. Tanaka. Inferotemporal cortex and object vision. *Annual Review of Neuroscience*, 19(1):109–139, 1996.
- M. Thelwall. Extracting accurate and complete results from search engines: Case study windows live. *J.Am.Soc.Inf.Sci.Technol.*, 59(1):38–50, 2008.

- E. C. Tolman. Cognitive maps in rats and men. *Psychological Review*, 55(4). (4):189–208, July 1948.
- A. Treves and E. T. Rolls. Computational analysis of the role of the hippocampus in memory. *Hippocampus*, 4(3):374–391, 1994.
- D. Tse, R. F. Langston, M. Kakeyama, I. Bethus, P. A. Spooner, E. R. Wood, M. P. Witter, and R. G. Morris. Schemas and memory consolidation. *Science*, 316(5821):76–82, 2007.
- E. Tulving and W. Donaldson. *Organization of memory*. 1972.
- E. Tulving and H. J. Markowitsch. Episodic and declarative memory: role of the hippocampus. *Hippocampus*, 8(3):198–204, 1998.
- P. D. Turney. *Mining the Web for Synonyms: PMI-IR versus LSA on TOEFL*, pages 491–502. Machine Learning: ECML 2001. Springer-Verlag, Berlin Heidelberg, 2001.
- L. G. Ungerleider and M. Mishkin. *Two Cortical Visual Systems*. MIT Press, 1982.
- F. Vargha-Khadem, D. G. Gadian, K. E. Watkins, A. Connelly, W. V. Paesschen, and M. Mishkin. Differential effects of early hippocampal pathology on episodic and semantic memory. *Science (New York, N.Y.)*, 1997.
- I. V. Viskontas, R. Quiñones Quiroga, and I. Fried. Human medial temporal lobe neurons respond preferentially to personally relevant images. *Proceedings of the National Academy of Sciences of the United States of America*, 106(50): 21329–21334, Dec 15 2009.
- G. V. Wallenstein, M. E. Hasselmo, and H. Eichenbaum. The hippocampus as an associator of discontiguous events. *Trends in neurosciences*, 21(8):317–323, 1998.

- S. Wirth, M. Yanike, L. M. Frank, A. C. Smith, E. N. Brown, and W. A. Suzuki. Single neurons in the monkey hippocampus and learning of new associations. *Science (New York, N.Y.)*, 300(5625):1578–1581, Jun 6 2003.
- E. R. Wood, P. A. Dudchenko, and H. Eichenbaum. The global record of memory in hippocampal neuronal activity. *Nature*, 397(6720):613–616, 1999.
- M. Yanike, S. Wirth, and W. A. Suzuki. Representation of well-learned information in the monkey hippocampus. *Neuron*, 42(3):477–487, 2004.
- L. Yassin, B. L. Benedetti, J.-S. Jouhanneau, J. A. Wen, J. F. Poulet, and A. L. Barth. An embedded subnetwork of highly active neurons in the neocortex. *Neuron*, 68(6):1043–1050, 2010.
- Y. Ziv, L. D. Burns, E. D. Cocker, E. O. Hamel, K. K. Ghosh, L. J. Kitch, A. E. Gamal, and M. J. Schnitzer. Long-term dynamics of ca1 hippocampal place codes. *Nature neuroscience*, 16(3):264–266, 2013.
- S. Zola-Morgan, L. R. Squire, and D. Amaral. Human amnesia and the medial temporal region: enduring memory impairment following a bilateral lesion limited to field ca1 of the hippocampus. *Journal of Neuroscience*, 6(10):2950–2967, 1986.
- S. Zola-Morgan, L. R. Squire, and S. J. Ramus. Severity of memory impairment in monkeys as a function of locus and extent of damage within the medial temporal lobe memory system. *Hippocampus*, 4(4):483–495, 1994.