

On the neuronal activity in the human brain during
visual recognition, imagery and binocular rivalry

Thesis by
Gabriel Kreiman

In partial fulfillment of the requirements
for the degree of
Doctor of Philosophy



California Institute of Technology

Pasadena, California

2002

(Defended August 29, 2001)

© 2002

Gabriel Kreiman

All Rights Reserved

To all those teachers
who taught me to enjoy learning

Acknowledgments

Haec ego non multis scribo, sed tibi: satis enim magnum alter alteri theatrum sumus¹.

All of the work described in the following chapters would not have been possible without the enthusiastic help of a large and nice group of colleagues and friends.

First of all, I would like to thank all the patients who participated in these experiments. We have not paid them and they have cooperated for the advancement of science.

Candi Maechtlen and Irene Wainwright have provided valuable and prompt help throughout these years in several of the fundamental details that keep things going. Irene has also provided excellent editorial assistance in several of our publications.

I have been fortunate to be able to interact and learn from a large number of teachers, colleagues and friends throughout my entire life. While a number of them have not contributed directly to this thesis work, it seems evident that they have enormously influenced my education and my preparation. Among them, I would like to especially mention my chess teacher Cesar Corte who was perhaps among the first who showed me the way to indulge in the wonderful task of solving new problems. The long hours of scientific exploration clearly have a root in me in the uncountable hours that we spent in front of a chessboard trying to figure out the solution to a problem or the best move. I would also like to express my gratitude to Rebeca Muller, Daniel Goldstein, Alberto Kornblihtt, Ernersto Timmermann, Rodolfo Ugalde and the math Olympiad team in Buenos Aires. I would like to also mention the continuous support from my family Raul, Lidia, Lauri, Elias, Haydee, Dora, Mariela and friends Damian Scherlis, Dan Rozenfarb, Diego Tisman and Dario Winter.

When I arrived to Caltech, I learnt the first steps in Neuroscience during my rotation in the lab of Mark Konishi to whom I am greatly indebted. Fabrizio Gabbiani immersed me in the wonderful world of quantitative spike train analysis. I would like to

¹ I write this not to the many, but to you only, for you and I are surely enough of an audience for each other.

also thank Erin Schuman who introduced me to the hippocampus and taught me the first steps in electrophysiology.

A number of people from UCLA and Caltech have helped me with the recordings including Anatol Bragin, Charles Wilson, Rick Staba, Erick Behnke, Jack Morrow and Eve Isham. Laurent Itti built the push buttons used in all our experiments.

The data obtained during sleep (Chapter 5) were recorded by the skilful and laborious endeavor of Charles Wilson and Rick Staba. Jonathan Lin, a Surf student from Caltech, provided fundamental help for the analysis and study of the correlations during sleep.

Most of the experiments to try to understand the psychophysics of flash suppression (briefly mentioned in Appendix 2) were carried out by SURF student Nick Knouf and by the enthusiastic collaboration of Pasadena High School student Shaun Lee. Michael Herzog was extremely kind in his discussions and suggestions about flash suppression, particularly in facilitating his programs to measure orientation discrimination. He was also key in the genesis of the idea to objectively measure suppression (see Appendix 2).

The comparison of different methods of spike sorting (Appendix 1) was carried out by Angela Yu and Jason Davis. Maneesh Sahani and John Pezaris provided their algorithm (spikesort) for comparison purposes. Part of the spiker software developed by Angela Yu and myself was written by Christophe Pouzat.

The experiment using repetitive transcranial magnetic stimulation during binocular rivalry (Appendix 3) was carried out in Harvard Medical School in collaboration with Julian Keenan and Alvaro Pascual Leone.

Geraint Rees and Nikos Logothetis have been very kind in providing advice for the binocular rivalry and flash suppression experiments (Chapter 6).

I would also like to thank a number of people for scientific discussions and for insightful comments to our writings: Mariela Zirlinger, Dan Rozenfarb, Terry Sejnowski and Francis Crick. My thesis committee, conformed by Mark Konishi, Richard Andersen, Shim Shimojo and John Allman (in addition to Itzhak Fried and Christof Koch) has provided valuable feedback.

The support of my wife Mariela Zirlinger was also key to the completion of all this work.

Itzhak Fried is the neurosurgeon at UCLA who has made all of these experiments recording the activity of single neurons possible. He has been a pioneer in the technique of chronic single unit recordings in epileptic patients. His scientific enthusiasm has been highly motivating and encouraging at the same time of helping the patients with their clinical problems. This is remarkable in itself but, in addition to that, Itzhak has proved to be extremely helpful in the design, analysis and discussion of our experiments.

Finally, the inspiring driving force behind all the work here described is my advisor Christof Koch. His enthusiasm for research is contagious, imparting the basement in Beckman Institute a strong scientific momentum. Together with Francis Crick, he has had the insight that we can and we should rigorously study the neuronal correlates of consciousness. His curiosity and motivation has pushed the field of consciousness research (among other fields) into the mainstream of neuroscience exploration. He has created an excellent environment to learn and explore new questions. He has made the already interesting adventures of scientific exploration during my doctorate work even more fascinating, enlightening them with new views, new questions, new solutions and new challenges every time. I am immensely grateful to him for allowing me the opportunity to work throughout these years in the lab and for his continuous motivation and encouragement.

Abstract

How does the neuronal activity in our brains give rise to our perceptions? We recorded the electrophysiological activity of over one thousand individual neurons in the human brain during object recognition, binocular rivalry, visual imagery and sleep. Subjects were patients with intractable epilepsy implanted with depth electrodes in targets including the amygdala, entorhinal cortex and hippocampus to localize the seizure focus for potential surgical resection. This has allowed us to explore the neuronal responses during visual processing in humans at an unprecedented level of spatial and temporal resolution. We observed a high degree of selectivity in the responses to complex visual stimuli. Some units were selective to categories of pictures including faces, houses, objects, famous people and animals while others responded only to one or a few stimuli, suggesting a sparse representation of visual information in the medial temporal lobe. Most of the selective neurons modulated their responses depending on the subject's percept during flash suppression. To further explore the correlation between perception and neuronal activity we investigated the vivid images that can be voluntarily generated in our minds in the absence of concomitant visual input. Our study revealed neuronal correlates of visual imagery and supports a common substrate for the processing of visual input and recall. Since visual memory is also prominent during dreams, we investigated the neuronal responses during different stages of the sleep-wake cycle. We observed an increase in synchrony during slow wave sleep compared to the wake and rapid-eye-movement sleep states. Our results suggest that neuronal activity in the human medial temporal lobe correlates with perception, shows a strong degree of invariance to changes in the input and could be involved in processing, storing and recalling visual information.

Table of Contents

Acknowledgments	v
Abstract	ix
Table of Contents	xi
List of Figures	xix
List of Tables	xxiv
List of abbreviations	xxv
1 Introduction	1
1.1. Informal definitions: statement of the questions	1
1.2. Brief history of theoretical ideas regarding perception	3
1.3. Why an experimental approach is necessary	6
1.4. Brief comments about epilepsy	7
1.5. Brief history of single neuron recordings	9
1.6. Summary and organization of this thesis	10
1.7. Figure legends	14
2 Methods and neuronal properties	16
2.1 Introduction	16
2.2 Recordings	18
2.2.1 <i>Electrode implantation</i>	18
2.2.2 <i>Location of implanted electrodes</i>	19
2.2.3 <i>Location of seizure onset focus</i>	20
2.2.4 <i>Electrophysiological recordings</i>	21
2.3 Comparison with other techniques	22
2.4 On the probability of finding a neuron	26

2.5 Spike sorting.....	31
2.5.1 <i>Criteria used to define a unit</i>	31
2.5.2 <i>Separating clusters based on the waveform shape</i>	32
2.6 Neuronal types.....	32
2.7 Spiking properties.....	34
2.7.1 <i>Firing rates</i>	35
2.7.2 <i>Interspike interval distributions and CVs</i>	36
2.7.3 <i>Bursting</i>	36
2.8 Anatomical connections to and from the MTL.....	38
2.9 Summary and conclusions	41
2.10 Figure legends	44
2.11 Tables	49
3 Visual Properties.....	64
3.1 Introduction	64
3.2 Introduction	67
3.2.1 <i>Types of stimuli</i>	69
3.2.2 <i>Behavior: visual recognition</i>	71
3.2.3 <i>Neuronal recordings</i>	72
3.2.4 <i>Distribution of the number of presentations</i>	73
3.3 Visual selectivity, definitions and examples.....	75
3.3.1 <i>Category specificity, criteria</i>	75
3.3.2 <i>Some relevant technical issues about the analysis</i>	77
3.3.3 <i>Category specificity, examples</i>	80
3.3.4 <i>Responsive but non-selective units</i>	81
3.3.5 <i>Visual responses to individual stimuli, criteria</i>	83
3.3.6 <i>Visual responses to individual stimuli, examples</i>	83
3.4 Summary of visual responses	86
3.4.1 <i>Distribution of category-specific responses</i>	86

3.4.2 <i>Distribution of responses to individual stimuli</i>	91
3.5 Is there any clustering of selective units?	93
3.6 Latencies and durations	96
3.7 Guessing which stimulus was presented.....	99
3.8 Change in neuronal response with time	102
3.9 Bursts and visual responses	104
3.10 Neuronal representation: rate codes, timing codes and other codes	106
3.11 Other questions and technical issues	107
3.11.1 <i>Seizure focus</i>	107
3.11.2 <i>Eye movements and blinks</i>	107
3.11.3 <i>Behavioral task</i>	110
3.11.4 <i>Color versus black and white and other inhomogeneities</i>	111
3.11.5 <i>Error trials</i>	113
3.11.6 <i>Right versus left</i>	114
3.11.7 <i>Stability of recordings: Monday/Tuesday neurons</i>	116
3.11.8 <i>Variability across subjects</i>	122
3.12 Summary and discussion	123
3.12.1 <i>Lesions and neurological studies</i>	124
3.12.2 <i>Comparison with fMRI and ERP studies in humans</i>	127
3.12.3 <i>Comparison with monkey electrophysiological experiments</i>	129
3.12.4 <i>Speed of processing in the visual system</i>	133
3.12.5 <i>Recognition, memory and familiarity</i>	135
3.12.6 <i>Neural codes</i>	136
3.12.7 <i>What do these neurons really “like”?</i>	138
3.12.8 <i>Categories, exemplars, both or none?</i>	141
3.12.9 <i>Final remarks</i>	143
3.13 Figure legends	144
3.14 Table legends.....	157

4 Visual Imagery	181
4.1 Introduction	181
4.2 Brief historical remarks	182
4.3 Methods	185
4.3.1 <i>Tasks</i>	185
4.3.2 <i>Electrode implantation, recordings and stimulus presentation</i>	189
4.3.3 <i>Data analysis</i>	189
4.4 Neuronal responses during visual imagery	190
4.4.1 <i>Comparison of baselines</i>	191
4.4.2 <i>Selectivity during vision and imagery: examples</i>	191
4.4.3 <i>Selectivity during vision and imagery: summary</i>	192
4.5 Comparison of neuronal activity during vision and imagery	194
4.5.1 <i>Response latencies and durations during vision and imagery</i>	194
4.5.2 <i>Comparison of firing rates</i>	195
4.5.3 <i>Guessing the stimulus from a single neuron's activity</i>	196
4.6 Other questions and technical issues	197
4.6.1 <i>Tones and auditory responses</i>	197
4.6.2 <i>Motor responses during vision</i>	198
4.6.3 <i>Right versus left</i>	199
4.6.4 <i>Statistical significance of the results</i>	200
4.7 Summary and conclusions	202
4.7.1 <i>Is visual imagery indeed visual?</i>	202
4.7.2 <i>The lack of pungency and tang in imagination</i>	208
4.7.3 <i>Alternative interpretations</i>	209
4.7.4 <i>Regional differences: how much longer will it take?</i>	213
4.7.5 <i>An image is worth a thousand words</i>	214
4.7.6 <i>Cause or correlation?</i>	215
4.8 Outlook and open questions	216

4.9 Figure legends	219
4.10 Tables	224
5 Neuronal activity during sleep	236
5.1 Introduction	236
5.2 Methods	238
5.2.1 Polysomnographic studies	238
5.2.2 Single neuron recordings in the human brain during sleep	239
5.2.3 Study of significance in correlations	240
5.3 Single neuron activity during sleep	246
5.4 Synchronous activity during sleep	248
5.4.1 Increased synchrony during slow wave sleep	248
5.4.2 Some technical issues about the analysis	253
5.5 Summary and discussion	256
5.5.1 Sleep and awareness	256
5.5.2 How are the specific contents of dreams represented?	257
5.5.3 Synchrony, EEG and oscillatory rhythms	259
5.5.4 Sleep and memory	260
5.6 Figure legends	263
5.7 Tables	269
6 Binocular Rivalry and Flash Suppression	281
6.1 Introduction	281
6.2 Brief historical remarks	284
6.2.1 Binocular rivalry	284
6.2.2 Flash suppression	285
6.3 Experimental procedures	286
6.3.1 Neuronal recordings and stimulus presentation	286
6.3.2 Tasks	288
6.3.3 Data analysis	292

6.4 Neuronal activity during binocular rivalry	293
6.4.1 Behavior	293
6.4.2 Neuronal selectivity during binocular rivalry	294
6.4.3 Amount of data obtained during binocular rivalry	295
6.5 Neuronal activity during flash suppression	296
6.5.1 Behavioral responses during flash suppression	297
6.5.2 Examples of neuronal responses during flash suppression	298
6.5.3 There is life also outside the temporal lobe	299
6.5.4 Comparison of monocular responses and flash responses	301
6.5.5 Predicting the percept during flash suppression	302
6.5.6 Summary of responses during flash suppression	303
6.5.7 Seizure activity	306
6.5.8 Inter-hemispheric switching	306
6.6 Chasing the percept inside the brain	307
6.6.1 Comparison with data from other methods and models	307
6.6.2 Perceiving IT?	313
6.6.3 Are monkeys conscious during binocular rivalry?	313
6.7 Figures legends	316
6.8 Tables	325
7 Summary and Reflections	345
7.1 Brief summary of the results	345
7.2 Invariance in perception and neuronal response	346
7.3 Perceptual processing, memory or language?	349
7.3.1 Memory	349
7.3.2 Language and concepts	351
7.4 Neural codes	353
7.5 Response duration	354
7.6 H.M., cause and correlations	356

7.7 Neuronal causal correlates of consciousness	359
1 Appendix 1: Spike sorting	360
1.1 Introduction	360
1.2 Brief review of available methods	362
1.2.1 <i>Dual window discrimination</i>	362
1.2.2 <i>Template-matching</i>	363
1.2.3 <i>Cluster cutting (Datawave)</i>	363
1.2.4 <i>Automatic algorithms</i>	364
1.2.5 <i>Spiker</i>	365
1.3 Comparison of spike sorting methods	367
1.3.1 <i>Which clusters should we compare?</i>	368
1.3.2 <i>Number of clusters</i>	369
1.3.3 <i>Within user repeatability</i>	370
1.3.4 <i>Comparison with models</i>	371
1.3.5 <i>Comparison of visual selectivity between Spiker and Datawave spikes</i>	371
1.4 Summary and further questions	373
1.5 Figure legends	376
1.6 Tables	379
2 Appendix 2. Flash suppression: competition between eyes or patterns?	388
2.1 Introduction	388
2.2 Experimental methods	390
2.2.1 <i>Flash suppression with patchy images</i>	391
2.2.2 <i>Swapped flash suppression</i>	392
2.2.3 <i>Behavioral responses</i>	394
2.2.4 <i>Subjects and stimulus presentation</i>	395
2.3 Results	396
2.3.1 <i>Flash suppression with patchy images</i>	396
2.3.2 <i>Swapped flash suppression</i>	397

2.4 Discussion	398
2.5 Figure legends	401
3 Appendix 3. The frontal lobe and binocular rivalry – Repetitive trans-cranial magnetic stimulation studies	410
3.1 Introduction	410
3.2 Experimental procedures	412
3.2.1 <i>Safety of rTMS</i>	412
3.2.2 <i>Stimulus presentation and subject responses</i>	413
3.2.3 <i>Motor threshold discrimination</i>	414
3.2.4 <i>Magnetic stimulation</i>	415
3.2.5 <i>Experimental design</i>	415
3.2.6 <i>Anatomical localization</i>	415
3.2.7 <i>Data Analysis</i>	417
3.3 Results	418
3.3.1 <i>Normalized distribution of dominance periods</i>	418
3.3.2 <i>Piecemeal rivalry duration</i>	418
3.3.3 <i>Comparison with a simple binocular rivalry model</i>	419
3.3.4 <i>Experiment 2</i>	420
3.4 Discussion and very preliminary conclusions	421
3.5 Figure legends and tables	423
3.6 Tables	425
References	431

List of Figures

Figure 2-1:	Perceptual invariance	14
Figure 2-1:	Schematic of the electrodes	44
Figure 2-2:	Patient with electrodes implanted	44
Figure 2-3:	MRI localization of electrode position	44
Figure 2-4:	Sample of extracellular recording	45
Figure 2-5:	Probability of finding a neuron	45
Figure 2-6:	Lack of influence of age, gender, handedness and IQ on the overall probability of finding a neuron	46
Figure 2-7:	Attempt to characterize neuronal types.....	46
Figure 2-8:	Distribution of firing rates	46
Figure 2-9:	Distribution of CV and CV_2	47
Figure 2-10:	Sample interspike interval distributions.....	47
Figure 2-11:	Distribution of burst duration and proportion of spikes in bursts.....	47
Figure 2-12:	Block diagram of the input to and main connections in the MTL	48
Figure 3-1:	Sample of stimuli	144
Figure 3-2:	Distribution of the number of presentations per experiment	144
Figure 3-3:	Selectivity to categories, example 1.....	145
Figure 3-4:	Selectivity to categories, example 2.....	145
Figure 3-5:	Visually responsive but non-selective unit, example in the MTL	146
Figure 3-6:	Neuronal responses outside the medial temporal lobe.....	147
Figure 3-7:	Selectivity to individual stimuli, example 1 (Clinton neuron).....	147

Figure 3-8:	Selectivity to individual stimuli, example 2 (Curly neuron).....	148
Figure 3-9:	Distribution of neuronal response latencies	148
Figure 3-10:	Distribution of neuronal response durations	149
Figure 3-11:	ROC analysis and p_e distribution	149
Figure 3-12:	Change in firing rate upon repeated presentation of stimuli.....	150
Figure 3-13:	Non-parametric analysis of variance	151
Figure 3-14:	Stability of recordings (Monday and Tuesday neurons).....	152
Figure 3-15:	Distribution of category-specific responses	152
Figure 3-16:	Number of categories that neurons were selective to	153
Figure 3-17:	Distribution of selective responses to individual stimuli per stimulus category and location	153
Figure 3-18:	Number of individual stimuli that neurons were selective to	154
Figure 3-19:	Eye movements.....	154
Figure 3-20:	Stability of recordings: visual selectivity across days	154
Figure 3-21:	Variability in spike counts and first spike times.....	155
Figure 3-22:	Spike timing. A glimpse of what we are missing and what's ahead.....	155
Figure 4-1:	Visual imagery. Single trial example and experiment design.....	219
Figure 4-2:	Selective neuron, example 1	219
Figure 4-3:	Selective neuron example 2	220
Figure 4-4:	Correlation of firing rates during the baseline	220
Figure 4-5:	Comparison of activity during vision and imagery.....	220
Figure 4-6:	Comparison of firing rates for individual stimuli	221
Figure 4-7:	Non-selective unit, example.....	221

Figure 4-8:	Distribution of latencies and response durations for vision and imagery	222
Figure 4-9:	Summary comparison of vision and imagery	222
Figure 4-10:	ROC analysis	222
Figure 4-11:	Regional differences, extrapolation	223
Figure 5-1:	Schematic EEG characterization of the different sleep stages.....	263
Figure 5-2:	Firing rates of single units during different states	263
Figure 5-3:	Coefficient of variation during different states	264
Figure 5-4:	Spiking and bursting	264
Figure 5-5:	Assessment of statistical significance in cross-correlograms	265
Figure 5-6:	Sample of spike trains	265
Figure 5-7:	Examples of two-point and three-point cross-correlograms	266
Figure 5-8:	Comparison of pairwise correlations in one subject	266
Figure 5-9:	Summary of results	267
Figure 5-10:	Anatomical location of synchronized pairs.....	268
Figure 5-11:	Robustness of the results to parameter changes.....	268
Figure 6-1:	Necker cube: example of a bistable percept	316
Figure 6-2:	Binocular (and monocular) rivalry.....	316
Figure 6-3:	Behavior during rivalry and the gamma distribution	316
Figure 6-4:	Binocular rivalry: single trial example	317
Figure 6-5:	Binocular rivalry: selective neuron example	317
Figure 6-6:	Flash suppression paradigm	318
Figure 6-7:	Distribution of the number of presented stimuli	318
Figure 6-8:	Behavior results during flash suppression	318

Figure 6-9:	Flash suppression: single trial example	319
Figure 6-10:	Example 1, responses to categories	319
Figure 6-11:	Example 2, responses to individual stimuli	320
Figure 6-12:	Example 3, responses to individual stimuli	321
Figure 6-13:	Example in the SMA.....	321
Figure 6-14:	Comparison of response properties during monocular presentation and flash period.....	321
Figure 6-15:	ROC analysis	322
Figure 6-16:	Normalized SDF summary for flash suppression	323
Figure 6-17:	Summary of number and proportion of neurons that follow the percept	323
Figure 6-18:	Comparison with data from monkeys	324
Figure A1-1:	Manual spike waveform separation (Datawave).....	376
Figure A1-2:	Two-dimensional projection plots	376
Figure A1-3:	Two-dimensional plots after separating the clusters.....	376
Figure A1-4:	Spiker menus.....	377
Figure A1-5:	Example of waveforms obtained from different sorting procedures ..	377
Figure A1-6:	Same-user repeatability.....	377
Figure A1-7:	Example of PSTHs obtained with two different spike sorting methods 378	
Figure A2-1:	Flash suppression with patchy images	401
Figure A2-2:	Swapped flash suppression	401
Figure A2-3:	Predictions for swapped flash suppression	402
Figure A2-4:	Swapped flash suppression with flickering.....	402

Figure A2-5:	Patchy images, results	402
Figure A2-6:	Swapped flash suppression, results	403
Figure A3-1:	Helmholtz/Wheatstone mirror haploscope and TMS round coil	423
Figure A3-2:	Scheme of the experiments	423
Figure A3-3:	Average normalized distribution of dominance periods	423
Figure A3-4:	Average changes in dominance duration with TMS	424
Figure A3-5:	Schematic of second order rivalry model	424

List of Tables

Table 2-1:	Comparison of single unit recordings in humans to other techniques	49
Table 2-2:	Percentage of units that showed bursting.....	51
Table 3-1:	Description of stimuli	157
Table 3-2:	Total number of units and number of selective units.....	158
Table 4-1:	Number of selective neurons in the medial temporal lobe	224
Table 5-1:	Number of units recorded in each location	269
Table 5-2:	Number of pairs and triplets analyzed during each epoch	269
Table 6-1:	Number of units analyzed during binocular rivalry	325
Table 6-2:	Number of units and selectivity during flash suppression test	325
Table A1-1:	Proportion of spikes in each cluster	379
Table A1-2:	Comparison of Datawave to Spiker	380
Table A3-1:	Localization procedures	425

List of abbreviations

General abbreviations

2-AFC	two-alternative forced choice. This term is used in several psychophysical experiments where the subject has to respond by making a forced decision between two possible alternatives.
cpd	cycles per degree. These units are used to describe the spatial frequency of gratings and other stimuli.
EEG	electroencephalography. Throughout the text we use the term EEG to refer to the electrical potentials recorded typically in the scalp but also intracranially (see for example Allison <i>et al.</i> , 1994.) The technique was introduced by Berger in 1930 (Berger, 1930).
EPSP	excitatory post-synaptic potential (Kandel <i>et al.</i> , 2000)
ERP	evoked response potential
ISI	interspike interval. Time difference between successive action potentials.
IQ	Intelligence quotient
LTP	long-term potentiation (Bliss and Lomo, 1973, Kandel <i>et al.</i> , 2000)
MEG	magnetoencephalography
MRI	magnetic resonance imaging. Also used in the text are the terms structural MR imaging, and fMRI, standing for functional magnetic resonance imaging
PET	positron emission tomography
r^2	correlation coefficient
REM	rapid-eye movement (see Chapter 5)
SWS	slow-wave sleep (see Chapter 5)
TMS	transcranial magnetic stimulation (Pascual-Leone <i>et al.</i> , 1998, Ruohonen, 1998, Kamitani and Shimojo, 1999). Also used in the text is the term rTMS that stands for repetitive TMS (see Appendix 3).

Anatomical abbreviations

Amy	amygdala. Almond-shaped mass of gray matter in the anterior portion of the temporal lobe. Also called amygdaloid nucleus. (Latin, almond, from Greek amugdale)
BA	Brodmann's area. (Brodmann, 1909)
EC	entorhinal cortex (from Greek entos, within.) Structure within the rhinal cortex. Brodmann's area 28.
Hip	hippocampus. A ridge in the floor of each lateral ventricle of the brain that consists mainly of gray matter (Late Latin, a sea horse with a horse's forelegs and a dolphin's tail (from its shape in cross section), from Greek hippokamos : hippos, horse. + kampos, sea monster).
PHG	parahippocampal gyrus
Fr	frontal cortex
Sub	subiculum
Tem	anterior temporal cortex
Occ	occipital cortex
Par	Parietal cortex
IT	inferior temporal cortex
STS	superior temporal sulcus
Prh	perirhinal cortex. Brodmann areas 35 and 36.

1 Introduction

1.1. Informal definitions: statement of the questions

What we see is not necessarily what we get. That our visual perceptions can be dissociated from the light intensity patterns impinging on our retinae has been known for a long time. Physicists and psychologists have discovered several visual illusions that are commonly used by magicians to trick the eye. What we end up “seeing” is a heavily modified and processed version of what we “get” at the eyes.

If we recorded the activity at the level of the rods and cones in the retinae we would obtain a rather precise transduction of the photons reaching the photoreceptors into an electrical signal. However, at some point higher up in the visual hierarchy, we suppose that the neuronal activity will more accurately represent our perceptions and impressions. Much effort has been devoted throughout history to try to understand the very nature of light itself. More recently, investigators have largely advanced our understanding of the molecular changes and phenomena that go into the conversion of the message from photons to electrical signals in the rods and cones. The next big challenge is to try to understand what happens next in the brain. The pattern of light intensity reaching the retinae depends on the properties of the outside world. Physicists have been very successful in describing in a quantitative manner how the wavelength and intensity of the light, the reflectance, refractance and other physical and chemical properties of the objects and the medium, the properties of the lenses (and eyes) determine the number and wavelength of the photons that reach the retinal.

Our perceptions are by nature subjective and thus inherently more difficult to study. Perhaps due to this, several physicists have tried to skew this problem for a long time. Helmholtz was one of the pioneers in trying to understand vision both at the level of

what happens outside our brains and also venture into the more mysterious terrain of perception. It is this complicated but fascinating side of vision that I have delved into for my graduate work. I have studied three different phenomena where there is a clear and strong dissociation between vision and perception: visual imagery, dreams and a phenomenon called binocular rivalry. Visual imagery and dreams constitute perhaps some of the most obvious situations where this strong dissociation takes place. In both cases, there is no activity at the level of the eyes (assuming visual imagery and dreaming with closed eyes)². Yet, we “feel” the presence of a visual image in our “mind’s eye.” If we imagine or dream about a specific visual stimulus, we conjecture there must be some neurons somewhere in our brains that are representing that percept. Both phenomena are extremely simple and constitute a major component of our everyday lives. Yet, our understanding of these phenomena at the neuronal level is extremely poor.

As in many other scientific fields, progress in Neuroscience has been and continues to be limited by the experimental techniques and approaches. For example, one could argue that a major advance could be made in the search for extraterrestrial life (assuming that we agree on what to look for) if we could look into a large fraction of the universe at a very high temporal and spatial resolution. Yet, this is quite complicated and the business of those interested in the search for extraterrestrial life is to make educated guesses based on the available information in addition to try to push the limits of the available technology. Likewise, in trying to understand the function of the brain, a large understanding could be gained if we had access to the detailed molecular composition and electrical potential throughout the cerebrum at a high spatial and temporal resolution. The work in the following approximately 400 pages attempts to make educated guesses and at the same time push the limits of the available technology to study the function of the human brain during these situations where what we perceive or get is dissociated from the retinal input.

² More precisely, there is no externally induced activity in the retinae. In the primate brain there is no feedback from the thalamus to the retina; therefore any activity that arises in the retina when the eyes are closed can be considered to be spontaneous changes.

1.2. Brief history of theoretical ideas regarding perception

Humanity has wondered in some way or another about the problem of how perceptions are represented for as long as we have written records and possibly before as well. Largely, at least until the end of the nineteenth century, we did not have the right tools to develop an empirical program for the rigorous study of these questions. Therefore, except for honorable and heroic exceptions, most of the discussions about the functioning of the brain were theoretical and philosophical in nature with very few empirical data or experimental attempts. Looking back as a Neuroscientist at the beginning of the twenty first century it seems quite unclear how much these philosophical discussions have brought to our current understanding³. The number of ideas about how perception is related to the physical world is too vast to enumerate, but I would like to bring just a few examples illustrative of the whole range. I am not going to be fair to these ideas here since I will just describe in a few words the work that illustrious thinkers performed in the course of years or decades. I therefore apologize to them beforehand.

To begin with, it took several millennia to recognize that the brain was the major center controlling the body. The heart was the "acropolis of the body" according to the Egyptians, the Mesopotamians, the Hebrews, the Hindus and the early Greeks (Finger, 2000). An ancient Egyptian study of neurological cases, the Edward Smith papyrus, has been found that dates back to the third millennium BCE. There, it is possible to find several descriptions of brain injuries correlated with effects that are far from the site of injury. In spite of this, the dominant view was that the heart and not the brain constituted the seat of the soul (Finger, 2000). Hippocrates and Plato were among the first to believe that the brain was responsible for "... joys, delights, laughter and sports, sorrows, grief, ...". That not all the Greeks embraced this notion is evident, for example, from the writings of Aristotle, one of Plato's students, who still adhered to the cardiocentric position based on his own anatomical studies and the study of other cultures.

³ To be completely fair, since our understanding of Neuroscience is at the dawn of its history, it is likewise legitimate to ask what value our current primitive notions, biases and experimental attempts, will contribute ultimately to the development of a serious understanding of the brain comparable to the understanding of the molecular world that Chemists have or the laws of nature as described by Physicists. But the enormous advance in the last fifty years certainly looks promising.

Several centuries later, Descartes argued that there exist two distinguishable and separate worlds, that of the soul and the physical realm⁴. He further proposed that the pineal gland in the brain is key to the communication between the perceptions in the soul and the visual world obtained from the physical senses. The pineal gland was chosen because, he argued, it is unique as conscious thought is, it is located in the middle of the brain and he thought that its damage leads to death. While the latter observation turns out to be wrong, the legacy and influence of Descartes' thoughts are enviable (Bogen, 1994). To this day, a large number of philosophers, lay people and also scientists hold the dualist view of a "soul-world" and a "physical-world" (although few may like to use those same words).

Not few neuroscientists argue that perception is vaguely distributed in the whole brain or at least in a vast area of the brain. This seems to be some form of revival of the pervasive attempt to explain multiple complex questions by postulating the existence of something called "ether" a few centuries ago. These ideas take several different forms. One suggestion is that a single neuron cannot be conscious; therefore it may not be possible to construct consciousness out of a small group of these unconscious neurons. This is somewhat akin to the idea that a single line does not have surface so how can you create a surface from several lines? Well, this is indeed quite simple and not mysterious at all. Others argue along these lines because they observe that a large part of the brain is activated (in scalp electroencephalographic, scalp magnetoencephalographic or functional imaging studies using magnetic resonance or positron emission tomography) in some task such as showing a picture. But this does not imply that all parts are required to actually perform the task of recognizing the picture. I consider this relationship between cause and correlation to be a crucial issue for the experimental study of perception and therefore I will discuss it more extensively in the last Chapter.

⁴ One should of course be very careful in evaluating Descartes' proposition. I am here quoting him for his writings. Given that the Inquisition had assassinated one of his dearest colleagues and the arduous problems Galileo was facing, it would not be surprising that several of his writings do not actually reveal his true thoughts. Indeed, after Galileo's sentence he cancelled the publication of "Du Monde" in 1634 (Descartes, 1994, Finger, 2000). We also would like to believe that we are more than what our papers say. And referees are not as tough as the Inquisition was, even those in Science and Nature. Furthermore, his famous "Discourse de la méthode pour bien conduire la raison et chercher la verité dans les sciences" was published in anonymous form (can you imagine many scientists that publish their work anonymously nowadays?).

Several other proposals have been advanced including the notion that language is an essential component of conscious perception or that motor output is key to consciousness. Others insist that there is no such a problem as consciousness, that consciousness is merely an illusion, a human construction or a mere by-product of brain function. Roger Penrose has put forward the idea (that I do not fully understand) that quantum uncertainties in microtubuli are to blame for conscious processing. Crick and Koch have proposed that a small subset of neurons with very specific properties, likely to be situated in a higher part of the visual brain that communicates with frontal cortex, could represent the neuronal correlates of visual consciousness (Crick and Koch, 1990, Crick and Koch, 1998, Crick, 1994). Since this astonishing hypothesis constitutes a central motivation for the work in these pages, it will be discussed more thoroughly later in the Thesis. The reader will also distinguish that these ideas ubiquitously permeate the discussions and ideas described in the next several hundred pages. There have been many other very interesting ideas, hypotheses and frameworks that I will not discuss in here for lack of space but not of interest (see for example Bogen, 1995, Logothetis, 1998, Libet, 1982, Cotterill, 1998, Tononi and Edelman, 1998, Penfield, 1937, Chamberlain, 2001.)⁵

Several philosophers have even argued that it will not be possible to understand the characteristics of perception along the lines that a system cannot understand itself. They suggest that our minds are simply at or near the cognitive limits in trying to solve this problem. Some philosophers distinguish between two aspects of the problem of perception, one being comparatively "easy" and the other one "hard." It is the latter that may be beyond our capabilities according to this subgroup. It does not matter how much knowledge is accumulated about the anatomy, physiology and even molecular biology of a given organism, say a bat; we would never be able, they argue, to feel what it feels like to be a bat. The subjective experience of knowing what it is to be a bat is reserved for bats and not for human scientists, no matter how clever and hard working. In spite of this negative outlook from some philosophers, the next six hundred pages will attempt to explore these questions in a detailed and rigorous scientific manner and I will let History

⁵ There are many others. I hope the many thinkers that have contributed to the discussion will forgive me for the omissions.

and Time decide whether it is or not possible to understand perception and consciousness at the neuronal level. Let us commence.

1.3. Why an experimental approach is necessary

Our perception remains largely unaltered even after large changes in the input. Seeing the world through colored filters, for example, can dramatically change the responses of cones in the retinae and earlier visual areas; yet we can still recognize a flower as such. Large position changes, size changes, even rotations do not alter in the least our capacity to recognize the flower. Other objects can occlude a large part of the flower and we can still recognize it. The flower can be defined by the synchronous movement of dots in a random dot pattern or by a simple line contour, or by changes in texture or contrast and we would still recognize it. We can see the flower in an afterimage after it has disappeared from the retina. We can see it in a binocular rivalry experiment even in the presence of a completely different stimulus in the other retina. Furthermore, we can close our eyes, effectively reducing any possible retinal activity, and still perceive the flower in our minds. We speculate that there should be a neuronal representation that shows the same type of invariance as our perception.

Philosophers and theoreticians can argue back and forth for several centuries as they have done in the past whether there should only be one representation, multiple ones, or none. But a final answer to such questions will come from an empirical exploration of the neuronal activity in the brain under different stimulus manipulations. This is not to say that philosophy and theoretical models cannot help design and analyze the results from these experiments. There are currently widely different levels of analysis ranging from the study of the molecular structure of ion channels to the study of patients with large lesions or the assessment of electrical activity averaged over billions of neurons. Being trained as a physical chemist, I cannot help but have a strong bias to try to understand the details. Yet, this approach will be time consuming and some of us are curious to try to start understanding some of these questions within our lifetimes. The other extreme is the view that by looking at the average output of huge ensembles of

neurons (on the order of several millions), one can gain insight into the inner workings of the cerebrum. Although this may very well turn out to be true, I find this quite unclear at this point. An intermediate and somewhat accessible compromise is the electrical spiking activity of individual neurons or small groups of neurons. Before I review the history of recording the electrical activity of individual neurons in animals as well as in humans, I would like to say a few words about epilepsy because the electrophysiological data from all the patients described in the current thesis come from patients with epileptic seizures.

1.4. Brief comments about epilepsy

Epilepsy was long thought to be a curse rather than a disease. Indeed, the English word used nowadays to describe epileptic attacks, “seizure,” originates from the idea that something/somebody from outside the patient is taking control his body. Epilepsy was one of the most feared diseases for the early Greeks. Hercules, for all his strength, supposedly suffered from epilepsy and in fact epilepsy was even called *morbus Hercules*. Various forms of seizures were recognized by the Greeks who coined the word epilepsy conveying the idea that something or somebody seizes the man (Penfield and Jasper, 1954). Hippocratic physicians were the first to propose that the seizures have a natural cause and the divine connotation that was given to them arose only from ignorance (Finger, 2000). Some centuries later, the famous Roman physician Galen proposed the usage of discharges from electric fish in the treatment of epilepsy which was still one of the best-documented and most feared brain disorders. Thomas Willis in the XVII century argued extensively against the still pervasive association of epilepsy with witchcraft. It was not until the investigations of John Hughlings Jackson starting in 1861 that a serious study and characterization of epilepsy began (Finger, 2000).

A 1999 report showed that epilepsy and seizures affect approximately 2.3 million US people of all ages (Epilepsy Foundation, 1999). Throughout history, there have been a multitude of attempts of treatment. There have been significant advances in the medical treatment that allows many epileptics to live normal lives free of seizures (McNamara, 1999, Epilepsy Foundation, 1999). Unfortunately, to this day, there is an important

fraction of patients (estimated at approximately 25%) who are refractory to pharmacological treatment. Patients who have temporal lobe epilepsy refractory to antiepileptic drugs can be evaluated for resective surgery (Ojemann, 1997, Penfield and Jasper, 1954, Engel, 1996). The first documented report of surgical treatment for epilepsy dates back to 1886 when Victor Horsley resected part of cortex and thus cured a patient with focal motor seizures (Kandel *et al.*, 2000). The modern surgical treatment has its roots in the extensive efforts of Wilder Penfield and Herbert Jasper in the 1950s (Penfield and Jasper, 1954). Resective surgery requires a careful localization of the seizure focus or foci and assessment of the potential adverse effects of the surgery. The benefit from the surgery is directly related to the adequacy of the resection and therefore precise localization is essential. In several cases, non-invasive techniques including structural magnetic resonance imaging and scalp electroencephalographic (EEG) recordings can be used in order to map the foci and guide a potential surgery. Non-invasive scalp EEG measurements are biased towards registering the activity of particular sets of neurons adjacent to the skull and deep structures potentially responsible for seizures can therefore be missed. If the structural information from MRI fails to show evidence of neuroanatomical abnormalities, doctors can still use the information from non-invasive Positron Emission Tomography (PET) and Single Photon Emission Computed Tomography (SPECT) to attempt to find large metabolic changes that occur during seizures. In other cases, however, all this information does not yield a precise localization. A large fraction of these patients have temporal lobe epilepsy; it is estimated that there are 200,000 to 250,000 patients with medically refractory temporal lobe epilepsy in the United States (Ojemann, 1997). It is in this subset of epileptic patients that the implantation of depth electrodes can be an alternative to attempt to map the seizure focus for potential resection. The factors leading to epilepsy are still unclear. A normal individual can have a seizure under special circumstances but this does not lead to epilepsy. It is now believed there is a genetic predisposition at least to some forms of seizures. Twenty-five single gene mutations that are linked to an epileptic phenotype have been identified in recent studies (McNamara, 1999, Kandel *et al.*, 2000). However, epilepsy often develops after a specific cortical injury or trauma such as a deep head

wound. There also seems to be a correlation between seizures and early insults such as prolonged febrile convulsion or encephalitis during the early years.

1.5. Brief history of single neuron recordings

Edgar Adrian in England and Wade Marshall and Philip Bard in the United States pioneered the work on recording the electrical activity of individual neurons in the 1920s and 1930s (Kandel *et al.*, 2000, Rieke *et al.*, 1997). Adrian noted the highly stereotyped nature of action potential waveforms and observed that this did not depend on the sensory modality and that similar action potentials were used to convey information from the brain to the muscles to direct movements. He further suggested based on his studies of sensory nerves innervating the muscle spindle receptor that it is the frequency of action potentials that conveys information; while this notion has been contended in the last decade and it is apparent that there could be other ways of conveying signal, this notion has pervaded Neuroscience for the remainder of the twentieth century. In the 1950s and 1960s techniques for studying the activity of single cells in the brains of intact and behaving animals including primates were developed by Ed Evarts and Vernon Mountcastle. Microelectrodes had just become available in the 1950s and Mountcastle used in behaving monkeys by recording the neuronal activity in the motor cortex. In contrast with intracellular recordings, these extracellular recordings proved to be much more stable to the pulsating movements of the brain in awake animals. Electrophysiological exploration of neural tissue was extended in the second part of the twentieth century to different species including rats, cats and monkeys. As in humans, the visual system is the main sensory modality in the latter and monkeys can be trained in relatively complex tasks; it is therefore from monkeys that a large fraction of our single cell understanding about the visual system comes from.

It turns out that it is possible to record the activity of individual neurons in the human brain. This is done in patients implanted with electrodes for clinical reasons, mostly in epileptic patients. Investigators have recorded the activity of individual neurons in other patients as well. For example, recordings have been performed during bilateral

cingulotomy in patients with obsessive-compulsive disorder or depression (Davis *et al.*, 2000, Hutchison *et al.*, 1999) and during surgery in Parkinson patients (Levy *et al.*, 2000, Magarinos-Ascone *et al.*, 2000). But most of the data from individual neurons in the human brain comes from epileptic patients. If the evidence from non-invasive techniques (see previous Section and Chapter 2) fails to yield convergent evidence of a seizure focus, patients can be considered for placement of intracranial electrodes. These are depth electrodes implanted that are typically implanted bilaterally in several targets of the medial temporal lobe including the amygdala, hippocampus, entorhinal cortex and parahippocampal gyrus. Occasionally, probes are targeted also to some extra-temporal targets including the supplementary motor area, the anterior cingulate, orbitofrontal cortex, or occipital and parietal targets (see Chapter 2). Several investigators have recorded the activity of isolated neurons from these depth electrodes. Single neuron recordings in humans have been reported in several papers starting early in the 1950s (Rayport and Waller, 1967, Ward and Schmidt, 1961, Calvin *et al.*, 1973, Rossi *et al.*, 1968, Ward and Thomas, 1955).

1.6. Summary and organization of this thesis

Throughout the course of my thesis work, I have studied the activity of more than one thousand individual neurons (see Chapter 2). Given that the human brain contains on the order of 10^{11} neurons, it seems evident that we have only explored a rather minuscule proportion of the brain. We have also investigated only a very small fraction of the total number of neurons in the medial temporal lobe or any of the structures therein. Yet, it should be noted that there are several difficulties in trying to study the activity of individual units in the human brain (many of these are discussed in more detail in Chapter 2). Furthermore, in all electrophysiological investigations, only a very small (and biased) sample of any given structure can be explored at the level of single neurons.

I will start by describing the methodological procedures and how it is possible to record the activity of single neurons in humans in Chapter 2. I will present the data

quantitatively assessing the probability of recording from a neuron with our techniques. This can potentially be relevant for those interested in potential clinical applications of electrode implantation. The remainder of Chapter 2 is concerned with the general properties of the neurons including the distribution of firing rates in each of the areas that we have studied, within-trial variability of spike times and the occurrence of spikes within short bursts of activity. The rest of the thesis is based on the methods described in this Chapter.

Chapter 3 is devoted to the properties of single neurons in response to presentation of different types of visual stimuli. Particularly within the medial temporal lobe, we found that some units respond selectively to complex visual stimuli. We describe the properties of these responses including a preliminary characterization of the type of stimuli that neurons responded to and the observed changes in neuronal activity. We briefly comment about the neuronal responses during visual presentation outside the medial temporal lobe. The rest of the thesis is concerned with studying these visual responses in cases where the percept is maintained in spite of large changes in the retinal input.

In Chapter 4, we describe the responses of these neurons while subjects formed mental images of the stimuli in their minds with their eyes closed and compare these to the neuronal activity upon observing the same stimuli presented in a monitor. We observed that some neurons modulated their responses depending on the type of stimulus the subjects were imagining. The existence of a selective response in the absence of concomitant retinal activity constitutes one of the strongest possible invariance properties of a representation that is correlated with perception. We argue that our data suggest a common mechanism for the processing of visual input and the recall of visual information.

In Chapter 5, we describe the neuronal activity in the human brain during sleep. While we have not been able yet to characterize whether selective units are specifically reactivated during dreams or other portions of sleep, we will argue that there are strong differences in the underlying neuronal activity during the different periods of the sleep-wake cycle. In contrast to the other Chapters, I was not involved in recording the data discussed in Chapter 5. The recordings were performed by Rick Staba and Charles

Wilson from UCLA while Jonny Lin and myself worked on the analysis and interpretation of the results.

During binocular rivalry, two different stimuli are presented to corresponding areas of the two retinæ. While the retinal input is largely constant, perception alternates in a seemingly random fashion between the two. These perceptual changes are accompanied by changes in neuronal activity during a related phenomenon called flash suppression as described in Chapter 6. Most of the visually selective neurons followed the perceptual alternations in spite of the fact that the retinal input was constant.

A summary of our results and a discussion of plausible interpretations are given in Chapter 7. We also reflect in Chapter 7 on other general aspects of the neuronal responses that we have observed such as their invariance properties, their duration and whether they are correlated to perception or causally linked to it.

The appendices at the end discuss some technicalities and several other projects and ideas that I have not had the time to put into a whole Chapter. Appendix 1 is concerned with the first stage after acquiring the data, that of sorting out the different neuronal sources that give rise to the extracellular changes in electrical potential. In Appendix 2, I discuss psychophysical experiments to try to better understand the flash suppression phenomenon described in Chapter 6. In Appendix 3, I show preliminary data of a separate experiment applying repetitive transcranial magnetic stimulation to the frontal cortex of normal subjects in order to attempt to elucidate the possible role of frontal areas in perceptual transitions⁶.

Preliminary and brief reports of some of our findings have been published previously (Kreiman *et al.*, 2000a, Kreiman *et al.*, 2000b) or have been submitted for publication (Kreiman *et al.*, *submitted*). Because of space constraints in the journals, I hope that the reader will find more information and detailed explanations of several of our findings and possible implications in the current thesis work. The converse is also true; for reasons of space, there are a number of problems that I have worked on during my thesis and that have been reported previously that I will not be discussing in the

⁶ If time permits, I will discuss in a separate Appendix, my preliminary explorations of the coding capabilities beyond the simple spike counts and firing rates computed in the other Chapters.

current thesis (Ouyang *et al.*, 1999, Kreiman *et al.*, 2000c, Zirlinger *et al.*, 2001, Kreiman, 2001, Krahe *et al.*, *submitted*).

1.7. *Figure legends*

Figure 1-1: Perceptual invariance

Our perception is very robust to large changes in the visual world. Here we show 15 different transformations of the famous painting of sunflowers by Van Gogh. These include blurring, thresholding, interposing a glass window before the picture, large changes in the colors, display of edge information, large decrease in contrast, display of edges plus highlighting, edge detection and thresholding, addition of gaussian noise, displaying through water, displaying the negative of the image, change of the colors to gray-scale, high contrast, 180 degree rotation and large occlusions. Our recognition of the painting remains intact in spite of the multitude of horrors imparted to Van Gogh's masterpiece.

2 Methods and neuronal properties

2.1 Introduction

An important fraction of our understanding about the functioning of brains comes from the laborious but fascinating *in vivo* (albeit some times *in anesthesia*) studies of the electrical activity of individual neurons in different parts of animal brains by inserting electrodes. *In vivo* recordings allow the exploration of neuronal activity in behaving animals, thus permitting the correlation of the responses in the brain with the performance in the perceptual or motor tasks. While there exist several other techniques to monitor directly or indirectly the function of structures within the cerebrum, the electrophysiological approach offers a high spatial and temporal resolution that cannot currently be achieved by other methods such as functional imaging, electrical or magnetic scalp recordings or the study of lesions. A more detailed comparison of the advantages of these different techniques is discussed in Section 2.3 and Table 2-1. The extracellular electrical potential is clearly not the only relevant variable to monitor the activity of brain tissue but it constitutes one of the most readily accessible ones to the experimenter. This is not to say that it is in general an easy task to perform neurophysiological recordings but it is evidently easier than trying to get a quantitative estimate of the concentration of several molecules and ions as a function of space and time at this point of time.

The exploration of the brain by the invasive procedure of surgery and placement of electrodes has largely been confined to explorations in animal models (see historical overview in Chapter 1). Yet, it could be argued, that the brain is one of the organs that distinguishes us most from (other) animals. It turns out that it is feasible to study the

activity of individual neurons in the human brain⁷. In this chapter, I will start by discussing why and how it is possible to record the activity of individual neurons in the human brain and the advantages and limitations of this technique. Then I will discuss the methodological aspects for the acquisition of the data that we have gathered throughout the course of my thesis work and address several important questions about these recordings in a quantitative manner. Specifically, I will estimate the probability of recording from a neuron with our techniques and argue that this provides an indirect measure of neuronal density. I will also analyze whether there is a correlation between this probability and the subject's gender, intelligence, age, etc. I will describe briefly our procedure for sorting the different neuronal sources that give rise to the extracellular potential; more details about this and a comparison with other methods is given in Appendix 1. I will present a brief description of some of the overall properties of the activity of neurons in the human brain. I will here discuss the distribution of their firing rates, the variation within a spike train and whether spikes occur within bursts or isolated spikes. The properties reported in the current chapter correspond to average values and we will argue that they are indicative of the properties of these neurons during “spontaneous” activity. The responses during different behavioral states and tasks upon presentation of visual stimuli, during imagination, during sleep and during binocular rivalry are described in the subsequent chapters. Finally, I will give a brief overview of the main anatomical connections that convey visual information to the MTL and the main output connections of the MTL structures.

⁷ I learnt about the possibility of recording the activity of single neurons in the human brains entirely by chance. I was attending a lab meeting in an animal electrophysiology lab at Caltech and one of the post-docs in the lab briefly commented on the paper published by Fried *et al.* in *Neuron* in 1997 (Fried *et al.*, 1997). My understanding of English was even worse than now and I clearly do not know the names of most animal species in English. I therefore initially thought they were talking about recordings in some strange animal. I ran back to the lab to check the manuscript. It was true; it was indeed *Homo sapiens*.

2.2 Recordings

2.2.1 Electrode implantation

Subjects are patients with pharmacologically resistant epilepsy⁸. Unfortunately, in approximately 25% of epileptic patients, the drugs that we know today do not seem to be efficient in combating the seizures (Epilepsy Foundation, 1999, Ojemann, 1997, Engel, 1996). This actually amounts to a rather large number of cases given that it is estimated that in the US alone epilepsy and seizures affect approximately 2.3 million people. A resection of the brain tissue responsible for generating the seizures is considered as an important way of treatment in these cases. In order to map the seizure focus or foci, patients are submitted to a battery of non-invasive techniques that include the study of the semiology of the seizures, neuropsychological testing, structural magnetic resonance imaging (MRI), positron emission topography (PET), and scalp EEG and MEG recordings. In all the patients that we have worked with, extensive non-invasive evaluation failed to yield concordant data corresponding to a single resectable epileptogenic focus and therefore the patients were stereotactically implanted with up to 12 chronic intracranial depth electrodes to determine the focus of their seizures for possible surgical resection (Fried *et al.*, 1997, Fried *et al.*, 1999, Kreiman *et al.*, 2000a). Following electrode placement, patients are monitored in a special unit on the neurosurgical ward. The amount of time that the patients remain in the hospital depends on clinical criteria; in particular, on whether the doctors have acquired enough information for the localization of the seizures. This typically involves approximately seven to ten days. The first one or two days after surgery, patients are under different types of medications and are not apt to doing any tests; all the tests are typically performed on days two through seven after the implantation of the electrodes.

The surgeries were performed by neurosurgeon Itzhak Fried at the University of California at Los Angeles. All the experiments reported here were carried out at the

⁸ It is not legal (at least in the US) to implant electrodes in the brain of a healthy human subject for the sake of scientific exploration. Even if the subject is a scientist and signs a written consent in the absence of exposure to alcohol or any drug, neurosurgeons would not even dream of being involved in this kind of procedure.

UCLA hospital ward. Through the lumen of the electrodes of approximately 1.25 mm diameter, up to 8 microwires (40 μm diameter) were inserted. The polyurethane electrode probes contained 6 - 7 Pt contacts approximately 1.5 mm wide with separations of 1.5-4 mm (see Figure 2-1). From these contacts, continuous EEG data are acquired 24 hours a day. The microwires were composed of foamvar insulated platinum / 20% iridium and typically had 1 kHz impedances in the range of 0.2 to 1 $\text{M}\Omega$. Two different types of electrodes were used. In some cases, the microwires were disposed in a linear array separated by 500 μm ; the separation between microwires ranged in these cases from 500 μm to 4 mm. The linear array of microwires was used in all the extra-temporal probes and in 11% of the MTL probes. In other cases, the electrodes were lying on a cone with an opening angle that we estimate to be smaller than 45 degrees. A schematic depiction of these electrodes is shown in Figure 2-1. Microwires typically extended 1 to 3 mm beyond the tip of the electrodes. We estimate that microwires end up within a radius of approximately 1.5 mm. Electrodes are placed stereotactically with MR imaging and angiographic guidance. After placement, they are sealed into the shell. Therefore, no movement is possible except for possible uncontrolled movement of the electrode relative to the brain.

All studies described here conformed to the guidelines of the Medical Institutional Review Board at UCLA. Subjects signed a consent and an assent form before participating in the experiments. They can voluntarily withdraw from the study and terminate any test if they wish to. The present thesis work describes data from 36 subjects (18 males; 28 right-handed, 2 ambidextrous; 36 ± 9 years old, ranging from 19 to 51). A photograph showing one of the patients⁹ with the electrodes implanted while he was in the ward is shown in Figure 2-2.

2.2.2 Location of implanted electrodes

The number as well as the site of implantation of the electrodes was based exclusively on clinical criteria. After carefully studying the results acquired by scalp EEGs, MRI and other clinical information, the doctors decide on those areas that they

⁹ Permission from the patient was requested before publishing this photograph in here.

suspect could be potential seizure foci and decide on the targets to implant the electrodes. The location of the electrodes was verified by structural magnetic resonance images obtained before removing the electrodes and post-operatively (Figure 2-3). MR images are acquired at 1.5 Tesla. In most of the patients with pharmacologically intractable epilepsy the MTL seems to be one of the main suspects. Therefore, electrodes were implanted in some target areas within the MTL in all the patients that we have studied in this thesis. Other potential targets include the supplementary motor area, the orbitofrontal cortex and the anterior cingulate. Very rarely do doctors implant electrodes in the parietal or occipital lobes. The proportion of electrodes implanted in each area is shown in Figure 2-5.

The spatial resolution of the MR images did not allow us to accurately determine in which CA fields the hippocampal probes were placed or what layer of the entorhinal cortex we recorded from. Neurons from anterior, middle and posterior parts of the hippocampus were pooled together as hippocampus neurons throughout the thesis. Most neurons in the amygdala were in the basolateral nuclear complex. It should be noted that in most cases it is not possible to observe the traces of the microwires. Therefore, the electrode locations that we report are based on the location targets, the observations during implantation and the MR data obtained with the electrodes implanted and post-operatively based on the position of the tip of the electrode. On top of the spatial resolution limits of MR imaging, this gives an additional error in our estimate of the location of the microwires.

2.2.3 Location of seizure onset focus

The information recorded during seizures from the depth electrodes was used to localize the seizure focus (Engel, 1996, Ojemann, 1997, Fried *et al.*, 1999). In a subsample of 11 patients¹⁰, 17% of the implanted electrodes were within the clinically determined seizure focus area (that is, either in the other hemisphere or in a different brain area on the same side). Thirty-four percent of the electrodes were in the same brain lobe as the seizure focus and 58% of the electrodes were in the same hemisphere. Note

¹⁰ I have not received from UCLA the seizure focus for the last 10 patients.

that this does not reflect the actual proportions in the number of neurons that we recorded from because there was a bias towards choosing recording sites outside the putative seizure foci. On a sub-sample of 40% of the recorded units¹⁰, 11% of the units were located within the seizure focus area. Twenty-two percent of the units were in the same brain lobe as the seizure focus and 42% were in the same hemisphere. As we will show in Chapters 3, 4 and 6, only a small proportion of the visually selective units were within the seizure focus. Overall, we did not observe clear differences in the waveforms or firing rates of those units inside compared to those outside the seizure focus (see also Section 2.7).

2.2.4 Electrophysiological recordings

Our data acquisition equipment was limited to the possibility of recording from only 15 channels from four electrodes. Thus, in each experiment we recorded from at most 15 microwires in 4 different brain regions. Therefore, before starting the experiments we checked the quality of the data in every single microwire by monitoring the activity in a single channel oscilloscope. The observations from this analysis are further discussed in Section 2.4. In 38% of the experiments we observed more than 15 microwires with units and we could have recorded an average of $51 \pm 29\%$ more data. After deciding which electrodes and microwires to record from, the data from those channels were amplified (the amplification factor ranged from 1000 to 5000 and was changed throughout the course of this thesis work), notch filtered at 60 Hz, high-pass filtered using a 4-pole Butterworth filter with a corner frequency of 300 Hz (custom hardware and Grass Instruments, West Warwick) and digitally stored using commercial data acquisition software (Datawave, Denver, Colorado). The sampling rate ranged from 10 kHz to 20 kHz; again this changed throughout the course of the thesis. We did not record the continuous signal. This was a limitation of the data acquisition system that we used. Snippets of 2 ms duration¹¹ were selected based on the crossing of a manually set threshold and stored for off-line analysis. The strategy was to set this threshold very low so as not to miss any possible unit. This necessarily involves acquiring a large amount of

¹¹ In a few experiments, we used snippets of 3 ms.

noise and artifacts in the waveforms. These were carefully discriminated based on the shape of the waveforms during the procedure of spike sorting (see Section 2.5 and Appendix 1). Figure 2-4a shows a sample of the continuous data that we obtained. The dashed line shows the threshold and two samples of the 2 ms waveforms that cross this threshold are shown. The larger amplitude waveform corresponds to a unit and was discriminated from the smaller noise waveform during the spike sorting process.

2.3 Comparison with other techniques

It should be noted that there are several advantages and limitations in this technique (see Table 2-1). I would like to briefly present this and compare the pros and cons of recording the activity of individual neurons in the human brain with other techniques. I will refer to functional MR imaging, scalp EEG recordings and monkey electrophysiology. Of course, fascinating data have been and will continue to be obtained by recording the neuronal activity in other animals and using other techniques such as PET in humans but I think these four models constitute the closest ones to our approach.

To begin with, as indicated above, the electrodes positioning is based exclusively on clinical criteria. We therefore cannot choose where to record from. It turns out, however, that in most of these patients with intractable epilepsy the presumed origin of the seizures is the medial temporal lobe, a fascinating area to study high-level vision and memory. Some medical doctors have actually implemented a different type of probe system that allows them to record from multiple contacts along the electrode (Kawasaki *et al.*, 2000, Kawasaki *et al.*, 2001, Howard III *et al.*, 1996, Adolphs, personal communication). This allows them to actually record the activity in sites along the trajectory to the target area, which is extremely exciting because it extends the number of sites that can be studied at the single neuron level in the human brain. This could potentially enable the study of lower visual areas as well. In addition to the fact that we are unable to choose the recording areas, we cannot move the electrodes once they are implanted. In contrast, basically the whole brain is amenable to study for the monkey electrophysiologist (although some areas may be harder to reach and record from than

others). Functional MR imaging in humans can yield information about basically any area in the brain where there is an appreciable hemodynamic change. Furthermore, the number of different areas that can be studied at a time is also limited in our case by clinical constraints. This is somewhat less of a problem in monkey electrophysiology but time constraints and the difficulty of electrode implantation also limit the number of areas that are studied simultaneously (for the most part, monkey electrophysiologists study one or at most two different areas at a time). In contrast, fMRI and scalp EEG studies in humans can simultaneously sample large areas of the brain.

Moving an electrode necessarily causes some damage to the underlying tissue and this cannot be done in the patients when there is no clinical justification for this. Monkey electrophysiologists typically spend hours (days, weeks, months) of their lives moving an electrode trying to find cells. This is done in order to achieve a large signal-to-noise (*SNR*) ratio in single or dual unit recordings and because it seems unlikely that blindly inserting an electrode into the target area will yield a neuron close to the electrode. This is indeed exactly what we observe: in more than 80% of the implanted microwires we do not observe a neuron (Figure 2-5). Our approach is therefore based on the existence of several tens of wires where a neuron can potentially be recorded. The number of microwires per patient as well as the exact frequency at which we observed units is described in detail in Section 2.4.

In contrast with animal experiments, the number and frequency of patients to study is limited. At least in principle, fly electrophysiologists can do as many experiments in as many subjects and cells as they wish. This is not exactly true in monkeys in practice; because of a combination of monetary, time, technological and other constraints, there is also some limitation in the number of subjects and neurons that can be recorded from. There are typically less than eight patients a year. Each patient stayed in the hospital for approximately one week. The recording time is further limited by several clinical and practical requirements. Doctors and nurses need to visit and take care of the patients. Family and friends wish to spend time with the patients and so on. Therefore, not only the number of individuals but also the time to do the experiments is severely constrained in our case.

One question of interest in electrophysiological experiments is where the recorded units are located and what type of neurons they are. There are neurons with very different morphologies, different pattern of connections and different sets of expressed genes. All this information is potentially very important since it could be that the physiological responses from interneurons that make local connections are very different from those of pyramidal cells that show feed-forward or feedback projections to other cortical areas. It is therefore important to distinguish whether one is recording from an interneuron or a pyramidal cell, what layer in cortex one is recording from, which CA field in the hippocampus, etc. One would ideally like to know for each neuron that was recorded its detailed pattern of afferent and efferent connections. In animals, this is typically done by injecting a tracing or causing some electrolytic lesion and studying the animal post-mortem. We cannot of course do this in our patients¹². It is not easy to distinguish between all the different possible neurons based on the signatures from extracellular electrical activity (see also Section 2.6).

The amplitude of the signal ranged from 55 to 1680 μV (see Figure 2-4d for the distribution). Electrical isolation is a critical issue to improve the quality of the recordings in any electrophysiology lab. Because of this, several precautions are taken in animal electrophysiological laboratories such as enclosing the subject (animal, slice, etc.) in a Faraday cage, limiting as much as possible the number of electrical equipment near the recording setup, preventing the subject from moving, etc. The clinical environment that we work in is far from ideal from this standpoint. In spite of this, we have managed to acquire data at an acceptably good signal-to-noise ratio (*SNR*) in our recordings. For each recording, I computed the *SNR* for each neuron cluster as the ratio of the average height of the waveforms to the noise estimate¹³. The *SNR* ratios ranged from 2.38 to 14.76 (see Figure 2-4c for the distribution).

¹² It should be noted, however, that this is becoming more and more rare in recordings from awake monkeys that have been extensively trained for specific tasks. Therefore, the localization capacity may not be as poor compared to those cases, but it is clearly very far from those other scenarios where there is a well identified neuron (or group of neurons) that one can record from (Bialek *et al.*, 1991, Hatsopoulos *et al.*, 1995, MacLeod and Laurent, 1996, Kreiman *et al.*, 2000c.)

¹³ A word is relevant here about how we estimated the noise. As we mentioned above, we have not stored the continuous trace but a series of 2-ms snippets of waveforms beyond a low threshold. After spike sorting (see Section 2.5 and Appendix 1) we have an assignment of every waveform to a specific cluster. When using Datawave as spike sorting procedure we used as noise estimate the standard deviation of the voltage around the mean source for each cluster (the noise estimation using the Bayesian approach in Spiker was

Given all these problems, it may be asked whether it is still worthwhile to pursue these experiments. I believe the answer is clearly yes given what we stand to gain. Macaque monkeys require several months of arduous training in order to perform even simple tasks that can be done by our patients after a very brief explanation. Furthermore, we can study questions that are very difficult to address in animals such as volitional visual imagery, self-awareness, language or complex emotions¹⁴. It could be argued then that other techniques are available to study the brain in humans. But the temporal and spatial resolution that we can achieve is several orders of magnitude better than that from any other technique (see Table 2-1). I hope to convince the reader of the fascinating new insights that can be gained by studying the activity of individual neurons in human patients in spite of all the drawbacks mentioned in this section.

One further point needs to be discussed. All the data that I will be describing shortly originated from recordings in epileptic patients and it could be argued that their brains are not completely normal. Therefore, it could be questionable to what extent the results obtained from these patients can be extrapolated to understanding the function of the normal human brain. First of all, most of the patients are completely “normal” from any possible behavioral measure. Some are not, but that is of course also true within the “normal” population. If the concern is the extrapolation to the brains of “normal humans,” it is quite evident that, in general, the brains of epileptic patients are much more similar to a human “normal” brain than those of monkeys (*let alone* other animals). Second, most of the data that we have studied comes from areas well separated from the

different; see Appendix 1 and http://www.cns.caltech.edu/~gabriel/academia/spike_sorting.html for details). It should be noted that we are assuming here that the same sources of noise are applicable when the waveform is occurring as in the no-spike trace (Fee *et al.*, 1996, Sahani, 1999). We have also compared this estimate to the following two other procedures: (i) concatenating the waveforms of all the clusters which did not contain neurons (see Section 2.5 for the criteria used to determine whether a cluster represents a neuron or not) and computing the root-mean-square (*rms*) of this signal and (ii) in those cases where we had the raw continuous trace, we computed its *rms* as is traditional.

¹⁴ This is not to say that it is not possible to design interesting experiments to attempt to get at these problems in monkeys. Indeed, several investigators have achieved fascinating progress in monkeys, but it seems evident that studying humans opens the doors to very rich and complex behavioral questions. In particular, for those of us interested in consciousness, it could be argued that it is still very unclear to what extent other animals express this phenomenon. We hypothesize and we want to believe that they do and I personally consider it very likely that this will turn out to be the case, but we must logically reckon that this is merely an assumption for the moment.

seizure focus (see Section 2.2)¹⁵. This is so because a priori doctors do not know where the seizure focus is (that is why they implant the electrodes in the first place) and some of the electrodes end up being implanted in the brain structures that are contralateral to the seizure focus or in the same hemisphere but a separate brain location (such as electrodes in the MTL for a patient that has a seizure focus in the SMA). Finally, I would like to remind the reader that some of the most fascinating discoveries in Neuroscience have been made by studying epileptic patients (Broca, 1861, Sperry, 1982, Penfield and Jasper, 1954, for an overview, see Kandel *et al.* 2000).

2.4 On the probability of finding a neuron

Given that our data acquisition equipment only allowed us to record from 15 channels at a time and that the number of implanted microwires was much larger than 15, we had to decide before each experiment which units to record from. Before starting the experiments, I therefore checked out every single microwire to evaluate the quality of the recordings and the presence of units (or lack thereof) in each channel. Note that we usually did not record these data, so the only values we have are our notes on the observations on the presence or absence of units. This is a qualitative observation subject to some degree of subjectivity. However, in most cases, it is quite clear to distinguish the presence or absence of a unit. Furthermore, we can actually estimate our accuracy in these judgments by comparing our notes with the data that we record afterwards. If we do this it turns out that in 86% of the channels that we decided to record from, we actually did observe a unit. We should also note that in some cases, we were unsure about a specific channel but we still recorded its activity in order not to miss a possible unit. Also, in several cases, the recordings were done several hours (or even a day or two) after the checkout and the differences could be due to real changes or movement of the electrodes relative to the brain. The above value is an underestimation of our accuracy in

¹⁵ Approximately more than 80% of the selective units were recorded from areas outside the seizure focus; this is quantified precisely in subsequent chapters. Furthermore, we did not observe any clear difference in the specificity of those visually selective units within and outside the seizure focus.

predicting a unit from the checkout of every channel. Therefore, our accuracy in evaluating neuronal activity from the oscilloscope trace is very high.

In general, we spend approximately 20 seconds for each microwire. According to this, it can be shown assuming a Poisson distribution that it would be unlikely ($p < 0.01$) to completely miss a neuron with a spontaneous firing rate beyond 0.25 spikes/second. It can be argued that there is a neuron with a null spontaneous firing rate that only responds upon presenting a picture of, say, Clinton. In that case, we would miss the neuron in the checkout. However, for us to completely miss it, the spontaneous activity has to be extremely low¹⁶. Furthermore, we typically repeat the checkout for each channel the next day, making it even less likely that we completely miss a unit.

Here I show a summary of the quantitative evaluation of the results obtained from all these checkouts in 27 patients. I was involved in the checkouts of 21 of these patients; the rest were performed by Wes, Catherine or Eve. The number of checkouts per patient ranged from 1 to 5 with an average of 2.4 ± 1.3 . The data to be presented here corresponds to a total of 263 electrodes (127 in the left hemisphere, 136 in the right hemisphere). Overall, we observed an average of 382 microwires containing units. Therefore, this corresponds to an approximate probability of finding a neuron of 18%. This should be multiplied by the probability that a recorded microwire will actually lead to a unit as stated above. This yields an overall probability of observing a unit of approximately 14%. This value may seem low and somewhat discouraging for those unfamiliar with electrophysiological techniques. To them, we should note that this means that, since we have an average of 70 channels to record from, we should expect to be able to record neuronal from roughly 10 channels on average. Also, in each channel one can attend to more than one single unit (see below for the section on spike sorting). On the other hand, for those neurophysiologists that spend vast amounts of time moving an electrode to try to chase a neuron, this value may seem somewhat high. By randomly inserting an electrode, we observe a neuron with a probability of approximately 1 in 7. First, we should note that we are not selecting here a unit based on its firing rate. Second, the signal-to-noise ratio that we obtain is lower than what is achieved in other preparations

¹⁶ Some examples do exist in the literature of neurons with almost null spontaneous activity that show a very sparse response with only a few spikes to specific stimuli (Rieke *et al.*, 1997).

where the researcher spends time trying to chase units. If we ‘searched’ within our data set for higher firing rates and higher *SNR* then our probability would also be lower. Finally, it is possible as we discuss below that this value depends on the area that one is recording from. In particular, it may be easier to use this technique in an area like the hippocampus but perhaps not in other areas. The hippocampus may be an area where it is relatively (compared to some other areas) easier to obtain electrophysiological recordings¹⁷.

These numbers constitute a grand average over several different parameters. Ideally, one would like to study this problem in further detail. One could pose the question of what is the probability of obtaining a neuron in a female patient between 20 and 25 years old, in the right hippocampal CA3 area, in the morning, while the patient is sleeping, in a specific hospital room, and analyze each of these factors separately. The more we constrain the parameters, the smaller the pool to analyze. The number of electrodes per patient ranged from 6 to 13. There was no clear overall difference between the left and right hemispheres in the number of implanted electrodes with averages of 4.7 ± 1.5 and 5.0 ± 1.4 respectively (t test, $p > 0.4$). That is, in most cases, probes were implanted bilaterally. However, when we look at the number of microelectrodes with units, the mean values were 5.4 ± 3.0 and 8.7 ± 4.9 for the left and right hemispheres respectively (2 tailed t test, $p < 0.005$). That is, we observed more microwires with units in the right hemisphere. The origin of this difference is unclear¹⁸. One detail that could be relevant is that we always start checking out the units in the right hemisphere (this may seem to be completely trivial but, historically, the piece of paper where we mark down our observations has the right hemisphere on top). It is therefore possible that overall we have inadvertently spent less attention and time to scrutinize the data in the left hemisphere¹⁹. Our observation does not seem to depend on the seizure focus. That is, 43% of patients had a focus on the right side, 38% on the left side, 14% showed a

¹⁷ Even an experimentally challenged person like myself could obtain electrophysiological recordings in the hippocampus during my rotation in the Schuman lab (Ouyang *et al.*, 1997).

¹⁸ Note that the values reported here correspond to the check-out analysis before any of the experiments. Therefore, they cannot be influenced by the proportion of stimuli in the experiments or the selectivity of the neurons to these stimuli.

¹⁹ It should be noted that this is a relatively long process (70 channels x (20 seconds + 5 seconds) is almost 30 minutes. In practice, due to troubleshooting, and scrutinizing more time in some channels, it can take 60-120 minutes if there are no other problems with the equipment.

bilateral focus and in the remaining 5% the focus could not be univocally determined²⁰. While the statistical tests suggest that the probability is low, it is still conceivable that these observations are due to chance. Alternatively, there may indeed be a difference in the probability of observing a neuron in the two hemispheres.

One of the most important factors (from the point of view of the brain) that influence the probability of obtaining a neuron with our approach is the neuronal density in each area. It is therefore of interest to examine separately the probability of obtaining a neuron in each area (obtained after normalizing for the number of probes in that area). This value ranges from 1% in left frontal cortex to 38% in right entorhinal cortex (Figure 2-5)²¹.

Other general parameters could be relevant in determining the probability of finding a neuron. We correlated the average probability per patient with the gender, age, and IQ of the subjects. It should be noted that, given the low numbers, it is important to be careful in the assessment of the significance of the correlations. Several methods to assess the significance of the correlations between two variables assume a normal distribution. This may be hard to justify given the small numbers and our lack of understanding of the distribution of some of these parameters (for example, why assume that the age of the subjects is normally distributed around the mean?²²). We therefore used a non-parametric bootstrap procedure to assess the probability of obtaining a given r^2 value by chance. I take a small parenthesis here to describe this method. Let $\mathbf{x}, \mathbf{y} \in \mathfrak{R}^n$ (that is, n is the number of entries in each vector). Let the correlations coefficient between \mathbf{x} and \mathbf{y} be r^i ($-1 \leq r^i \leq 1$). The idea is very simple, if there is something special about the correlation in the values of \mathbf{x} and \mathbf{y} , then the specific order of the values in \mathbf{y} should be very important. However, if the correlations are just random, one should be able to obtain a similar correlation after randomly shuffling the order of the entries in \mathbf{y} . Let Π_y represent such a random

²⁰ Need to update these values. Need seizure focus data from patients 341 thru 350 from I.F.

²¹ Check out if there are other measure of neuronal density and how they correlate with these values.

²² Note that we are not looking at the mean of several measurements of the mean age; this could be claimed to be normally distributed according to the law of large numbers (Keeping, 1995). Indeed, the distribution of the ages of the subjects looks quite uniform and very far from normal. I am not aware of data showing that there is a differential prevalence of prevalence of intractable temporal lobe epilepsy at different ages (this would justify a non-uniform distribution).

permutation. For each permutation, we can compute our statistic variable of interest, in this case the coefficient of correlation $r(\Pi_y)$ between the two variables. The probability p that the observed correlation coefficient r^l can be obtained randomly is given by

$$p = \#\{\Pi_y / r(\Pi_y) > r^l\} / \#\{\Pi_y\}$$

where $\#$ indicates the cardinal of elements in the set (Efron and Tibshirani, 1993).

There does not seem to be any clear correlation between the overall probability of finding a neuron (P) and the age of the subject (Figure 2-6a). This conclusion holds also for the probability restricted to the left or right hemispheres. There is no clear difference either in P for males or female patients. Again, the conclusion also holds for the probabilities restricted to the left or right hemispheres. We also analyzed whether there was any change in P for left- or right-handed patients²³. Overall, we did not see any clear change. Upon restricting the analysis to the hemispheres, it seemed that there was a small (albeit statistically significant, $p < 0.05$) increase in the probability of finding a neuron in the left hemisphere for left-handed patients. This was not the case for the complementary comparison: there was no significant difference in the probability of finding a neuron in the right hemisphere for right- or left-handed patients. Another variable that could conceivably influence the probability of finding a neuron is the subject's intelligence. It is not the purpose of this section to discuss in detail how to measure intelligence or compare the different quantitative methods available to assess it. We used the traditional IQ ratio (Sternberg and Kaufman, 1998, Duncan *et al.*, 2000) measured as a Full Scale IQ, Verbal IQ and Performance IQ. Overall there does not seem to be any clear correlation between P and these IQ measures. Intelligence may not be in the total number of cells in the brain but in a more complex connection pattern.

It would be interesting to ask these same questions about specific brain areas. For example, it is possible that intelligence correlates with neuronal density in the frontal cortex but not in the entorhinal cortex. Unfortunately, it is hard to give a precise value for each subject and for a specific location. If there are 8 microwires, then the probability of finding a neuron can only take the discrete values $0/8, 1/8, \dots, 8/8$. This, together with the small numbers, makes this analysis difficult to interpret.

²³ Note that here we are referring to the handedness as reported by the patients and not to the results of hemisphere dominance as assessed by the Wada test.

2.5 Spike sorting

Extracellular recordings do not give us access to the intracellular voltage changes of the neurons we record from. However, most people who perform intracellular recordings still end up analyzing only the spiking activity of their neurons. In addition, intracellular recordings require moving the electrode to find the neuron and are typically quite unstable compared to extracellular ones (see, however, Fee, 2000 and Leonardo, personal communication). Moreover, extracellular recordings allow us the exciting possibility of recording more than one neuron at the same time. Every neuron around the electrode will contribute to some degree to the recorded signal. The weight in the contribution will depend on several factors including the distance and position of the electrode with respect to the neuron (Johnston and Wu, 1995, Holt and Koch, 1999). This requires an extra step to separate the recordings into the different neuronal sources. This might not seem to be a difficult problem given all the mathematical algorithms available for studying the clustering of data. However, there are several practical issues that make the problem quite complicated. A number of different investigators have developed algorithms and heuristics to tackle this problem (Lewicki, 1998, Fee *et al.*, 1996, Sahani, 1999, Henze *et al.*, 2000, Harris *et al.*, 2000, Pouzat, personal communication).

I am going to give a brief description here of the approach we have used to spike sorting; a somewhat more detailed description and a comparison with the “Spiker” algorithm that we have written is given in Appendix 1.

2.5.1 Criteria used to define a unit

Noise (electrical noise, movement artifacts, etc.) could also give rise to specific clusters; it is therefore important to set differentiate these from neuronal sources (see example in Figure 2-4b and Appendix 1). We used the following criteria to define a unit:

- (i) Lack of significant peaks at 60 Hz and harmonics in the power spectrum.

(ii) Lack of interspike intervals (ISIs) shorter than an absolute refractory period. These short ISIs are characteristic of multi-unit activity or noise. Specifically, a threshold of <2% of interspike intervals shorter than 2 ms was imposed.

(iii) SNR>2 (see above for the definition of SNR).

Note that no restriction (no bias!) was imposed on the overall firing rate (several electrophysiology labs only study units with more than 1-2 spikes/sec).

2.5.2 Separating clusters based on the waveform shape

All the data shown in the current thesis was submitted to a simple manual clustering procedure. Eight parameters were selected to compare the signal shape and evaluated for each single waveform. These parameters implemented were the first three principal components of the data (accounting for 52 to 99% of the variance), the signal height (maximum – minimum), width (time of trough – time of peak), valley before peak voltage, time of peak and time of trough. These eight values were displayed in two-dimensional plots ($8 \times 7 / 2 = 28$ possible plots) and typically separated into clusters in some of these projections. This method was implemented in the Datawave commercial software and examples are given in Appendix 1. Given the subjective nature of this procedure, we also compared the results for one third of all the data using our custom made semi-automatic “Spiker” algorithm; the results of this comparison are described in Appendix 1.

2.6 Neuronal types

It is quite clear that not all neurons in the brain are equal. Neurons show different electrical properties (Mountcastle, 1959, Mountcastle, 1995) and different shapes (Albright *et al.*, 2000), they emit different neurotransmitters (Kandel *et al.*, 2000) (Kandel and Squire, 2000) and express different genes (see for example Zirlinger *et al.*, 2001.) It would therefore be very interesting to know what type of neurons we record from. In contrast to studies *in vitro* particularly in non-primate animals, we cannot use

staining and histological techniques for this purpose. There have been some reports that suggest that it would be possible, at least, to make some basic classification based on the firing rate and waveform width of the neuronal activity. A clear separation between pyramidal cells and interneurons was obtained in recordings from the rat hippocampus by Buzsaki's group (Csicsvari *et al.*, 1999) and in recordings from the macaque prefrontal cortex (Rao *et al.*, 1999). Recently, Sheinberg and Logothetis reported that there is an anticorrelation of the attenuation of the signal height within a burst and the waveform width (Sheinberg and Logothetis, 2000). These reports suggest that it should be possible to make a rough characterization of interneurons and pyramidal cells based on these parameters. Interneurons, they claim, show a constant waveform height, lack of bursting, short widths and typically display higher maintained firing rates. Pyramidal cells, in contrast, show bursting and consequently signal attenuation, comparatively longer widths and lower maintained firing rates.

In our sample, we have been unable to observe a clear distinction of different clusters of neuronal types based on these parameters (Figure 2-7**a,b**). For each interspike interval shorter than 10 ms, we computed the ratio of the amplitude of the first spike to that of the second spike. The width was computed as the time from peak to trough and the firing rate was computed over the entire experimental session. As can be seen in the 3D plot, it is hard to distinguish two groups as indicated in the previous paragraph. The 2D projections show non-significant r^2 values very close to 0. It is conceivable that the above statements only hold for specific brain regions. We therefore plotted the same parameters within each location (Figure 2-7**c,d**). Overall, however, we still did not see any clear differentiation of two neuronal types. One possible reason for the difference with the strong separation shown by Buzsaki's group may be the *SNR* of the data. The larger amount of noise in our samples may make our measurements of width and height somewhat variable and imprecise, thereby blurring a distinction between these neuronal types. Another possible reason is the possible clustering of bursts into two separate units (see also discussion in Section 2.7.3).

2.7 Spiking properties

Before presenting the data obtained in response to specific stimuli, imagination, dreams or flash suppression, I would like to briefly describe the general properties of the neurons that we have recorded from. Specifically, here I will discuss the distribution of their firing rates, the shape of the interspike interval distributions and the fitting of simple models to them, their power spectra and the bursting properties of these neurons. I would like to present this in a manner that is independent of the behavior and responses that we studied. To do this, ideally, one would study recordings of “spontaneous activity.” This is a fairly commonly used concept in the animal electrophysiological literature. However, when we think about humans, the concept becomes quite more complicated. Functional imaging researchers typically like to invoke control situations where subjects “put their minds in a blank state” or “empty their heads from any thoughts” or “think of nothing.” No matter how hard I tried, I have not been able to reach this blank or empty state (with the possible exception of sleeping states but that is a whole different issue to be discussed in Chapter 5).

The data that I will present in this section corresponds to the average activity of neurons over the whole experimental sessions (31.8 ± 13.5 minutes). I have estimated that during all of the tasks, the recorded activity that is reported in subsequent chapters amounts to approximately 20% of the total recorded activity (due to intervals between pictures, rest periods, etc). However, if a neuron responds to a particular type of stimulus, this stimulus-induced activity amounts to only approximately 2% of the total recording. Therefore, for the purpose of analyzing the “spontaneous activity,” whatever that means, the data that I will present here can be considered to be somewhat contaminated with these stimulus-induced responses. One alternative would be to report only the firing rate after chopping off the stimulus presentation time, a short interval before the stimulus (due to possible expectancy issues), a short period after stimulus presentation (due to the decision and preparation to make a response), the behavioral response time, etc. The remaining spike train would contain several somewhat short stretches of data and somewhat arbitrary decisions would have to be made to compute interspike intervals

distributions, CVs, burst proportions, etc. I have opted to simply compile the activity over the entire spike train. Please bear this in mind in the remaining paragraphs.

2.7.1 Firing rates

The overall distribution of firing rates is shown in Figure 2-8. The mean rate was 3.3 ± 4.9 spikes/sec (range 0.03 to 64.5 spikes/sec). There were 21% of units with firing rates below 0.5 spikes/sec. Many electrophysiologists typically discard these neurons and go on searching for more vigorous responses. This clearly introduces some bias in the reported literature²⁴. In those cases where our approach is taken (Wilson and McNaughton, 1993, McNaughton *et al.*, 1983, Nicolelis *et al.*, 1999, Nicolelis *et al.*, 1997), it seems evident that several neurons fire at extremely low firing rates. Firing is energetically expensive. Unless reductions in firing rate are intended to encode information (see Chapter 3), it would seem more convenient to display low spontaneous activity. This seems to be the rule in most neurons. Notable exceptions include some neurons that display very high maintained rates of activity such as the P-receptor afferents in the *Eigenmannia* weakly electric fish (Kreiman *et al.*, 2000c) and cerebellar Purkinje cells (Kandel *et al.*, 2000).

Figure 2-8b shows a detail of the distribution of firing rates separated for each location. In order to compare these, we performed a one-way analysis of variance comparing the variance in the firing rates across locations to that within locations. This yielded a p value < 0.01 , suggesting that there are indeed differences between the different locations. A post-hoc multiple comparison procedure using Scheffe's method, (Scheffe, 1953), comparing each location to the rest, suggested that the neurons in the amygdala seem to show an overall lower maintained firing rate while the frontal²⁵ and supplementary motor area neurons show an increased spontaneous activity.

²⁴ A bias is also introduced by the type of electrodes used. This is also true in our case, *bien sure*.

²⁵ Note, however, the very small number of clusters analyzed in the frontal group.

2.7.2 Interspike interval distributions and CVs

A common measure of the variability of the timing of spikes is the coefficient of variation (CV). CV corresponds to the ratio of the standard deviation to the mean of the interspike interval distribution. For a Poisson process, $CV=1$ while for perfectly regular interspike intervals $CV=0$. The values of CV ranged from 0.75 to 8.71 (1.8 ± 1.5). The CV is largely affected by bursting and by long intervals without firing. The CV_2 is a measure of within spike train variability that is more robust to these variations (Holt *et al.*, 1996). CV_2 ranged from 0.41 to 2.05 (1.08 ± 0.13).

An analysis of variance indicated that the CV showed location dependent changes. The amygdala showed a decreased value (1.51 ± 0.78) and the hippocampus showed an increased value (2.33 ± 1.98) compared to the rest of the locations. The values of CV_2 were much more homogeneous.

2.7.3 Bursting

A common property of many neurons is the occurrence of groups of spikes with interspike intervals typically shorter than 20-30 ms and sometimes even as short as 3 ms. The exact values are very important here. One of the proposed mechanisms for bursts is an increase in the intracellular concentration of Ca^{2+} that depolarizes the cell and therefore lowers the effective threshold for the occurrence of a second spike. This increase in Ca^{2+} is typically very short (on the order of several ms). One of the reasons for the interest in bursts is that they may constitute a more reliable way of transmitting information than isolated spikes. This is due to failures in action potential propagation, failures in release of neurotransmitter, and the increase of the EPSP in the postsynaptic neuron (Kandel *et al.*, 2000, Lisman, 1997, Koch, 1999).

Bursts typically lead to a bimodal distribution in the interspike interval distribution. If the proportion of spikes in bursts is very small, however, this may be hard to detect. We studied the occurrence of bursts by evaluating the deviations of the

autocorrelogram from that expected for a Poisson process (Abeles, 1982)²⁶. An obvious deviation occurs at short time lags near zero due to the refractory period. In other cases, a bump at short time lags can be noticed in the interspike interval distribution and the autocorrelogram. Neurons that fire short intervals of bursts typically show this behavior (see for example Bastian and Nguyenkim, 2001).

An example of the ISI distribution of a non-bursting unit is shown in Figure 2-10a. The red line corresponds to the fit from a Poisson distribution; as the figure shows, the Poisson distribution fits the data quite accurately. The autocorrelogram is shown in

²⁶ Briefly, let $P(t)$ be the probability that a spike is generated in the interval $(t, t+\Delta t)$. In general, this probability depends on the bin size Δt . Assuming the spike train to be a point process having at time t the average firing rate $\lambda(t)$ we can then write this probability as $P(t)=\lambda(t) \Delta t$ with $[\lambda(t)]=\text{spikes/sec}$. If we study the responses from N spike trains (given by N neurons or by N repetitions of an identical stimulus for a given neuron), we can expect to observe a total of n spikes where $n=NP(t)$

$$\lambda(t) = \frac{P(t)}{\Delta t} = \frac{n}{N\Delta t}$$

If we are only looking at one single spike train, we can think of N as the total number of bins when we compare against a homogeneous Poisson process where $\lambda(t)=\lambda=\text{constant}$.

Let $X(t)$ denote the spike train after binning where $X(t)=1$ iff there is a spike in the interval $(t, t+\Delta t)$. Then we can define the autocorrelation by

$$\varphi(\tau) = \sum_t X(t)X(t+\tau)$$

where the sum extends to all values of t that were recorded.

This gives the total count in units of spikes. To compute the number of spikes per bin, one has to normalize by dividing by the total number of bins that were included in the above sum. Note that, in general, this number will depend on the variable τ . If the recording length is much larger than τ , however, this effect will be negligible. In order to convert to spikes/sec, one has to further divide this value by the size of the bin. Assuming a homogeneous Poisson process, we can compute the expected value and the confidence limits for the autocorrelogram. Following the discussion above, we can compute the probability that there will be a spike in any given bin, $P(t)$. For two bins (t and $t+\tau$), the probability that there will be a spike in both is simply P^2 . Let L correspond to the total number of bins. Then we can easily compute the expected value for the autocorrelogram:

$$E[\varphi(\tau)] = LP^2.$$

Note that under the above assumptions, this value does not depend on τ . Let us denote this expected value by x . For a Poisson process, the probability of finding m spikes given the expected value x is given by

$$P_{\text{poisson}}(m | x) = \frac{e^{-x} x^m}{m!}$$

The cumulative distribution C is given by

$$C(m | x) = \sum_{i=0}^m P_{\text{poisson}}(i | x)$$

Then the confidence limits m_{inf} and m_{sup} for a significance level of α are given by

$$C(m_{\text{inf}} | x) < \alpha \quad \text{and} \quad C(m_{\text{inf}} + 1 | x) > \alpha$$

$$C(m_{\text{sup}} - 1 | x) < 1 - \alpha \quad \text{and} \quad C(m_{\text{sup}} | x) > 1 - \alpha$$

Figure 2-10c. Except for the refractory period, it is quite flat and can be fit by a homogeneous Poisson process with the parameter given by the firing rate of the unit. The situation is very different for a bursting unit (Figure 2-10b). The ISIs show a bimodal distribution with a sharp peak concentrated at low ISIs corresponding to bursts and broader peak at longer intervals for the remaining spikes. The Poisson fit (red trace) is quite poor. A peak beyond the 99.9% confidence intervals for a Poisson process is evident in the auto-correlogram.

Using these criteria we observed that an average of 26% of the units showed bursting. The mean duration of bursts was 26 ± 12 ms and the proportion of spikes in bursts was 0.25 ± 0.16 . The proportion of bursting units was not homogeneous among the different studied locations. There were more bursting units in the amygdala, entorhinal cortex and hippocampus than in other locations (see Table 2). Note that while the percentages are very high in the parietal and occipital probes, the number of units was very small. The burst duration and proportion of spikes in bursts were slightly but significantly higher in the hippocampus. I already imagine Christof, impatient, asking: “But do these bursts convey any special information?” This question will have to await the next chapter.

2.8 Anatomical connections to and from the MTL

The classic diagram of Felleman and Van Essen culminated with the projections of the last exclusively visual areas of the macaque monkey brain to, la crème de la crème, the entorhinal cortex and from there to the hippocampus (Felleman and Van Essen, 1991)²⁷. The MTL including the EC, hippocampus, parahippocampal gyrus and amygdala

²⁷ There is a very large degree of interconnectivity in the brain. Most of the anatomical staining techniques typically report the main connections and sometimes also observe some minor projections. Thus, when we say, for example, that the visual information is conveyed to the hippocampal system through the entorhinal cortex, this does not mean that there is absolutely no direct connection from inferior temporal cortex to the hippocampus. Instead, what we are saying is that the pathway through the EC is more massive. A tacit assumption that is worth remarking here is that the functional relevance of the different pathways in conveying information is to some extent correlated with the number of nerve projections. This is not necessarily true because it is conceivable that a connection with very few but highly efficient and reliable axons and synapses could be as strong (in terms of eliciting an EPSP in the post-synaptic neuron) as a

are located in the ventromedial aspect of the anterior temporal cortex²⁸. The entorhinal cortex is already a multimodal area and receives input from somatosensory, olfactory and auditory cortices (Brown and Aggleton, 2001, Suzuki, 1996). This does not necessarily mean that the same individual neurons receive multimodal input; I am not sure whether this has been investigated but there are reports that indeed different sub-structures of the entorhinal cortex are targets of input from distinct systems and modalities (Insausti *et al.*, 1995). The visual input comes both from the dorsal as well as ventral streams. Input from the frontal eye fields is conveyed to the EC through Brodmann area 46 (Felleman and Van Essen, 1991). Areas 7a and 7b project to the perirhinal cortex (Brodmann areas 35 and 36; Felleman and Van Essen, 1991) and parahippocampal area (Brown and Aggleton, 2001). There are strong projections from both the perirhinal cortex and the parahippocampal cortices to and from the EC. A schematic illustration of the main connectivity pattern responsible for conveying visual information to the MTL is illustrated in Figure 2-12. It should be emphasized that this diagram is highly oversimplified and does not accurately indicate the highly intricate interconnectivity of the brain.

In the ventral visual stream, the anterior inferior temporal cortex area TE shows divergent projections to both the perirhinal and entorhinal cortices (Suzuki, 1996, Saleem and Tanaka, 1996). There seems to be an anatomical differentiation in area TE; the ventral part of anterior TE shows strong projections to the perirhinal cortex (both areas 35 and 36) as well as the EC whereas the dorsal part shows strong projections to area 36 and only sparse projections to areas 35 and EC (Saleem and Tanaka, 1996). Labeled terminals were found in all of the cortical layers in areas 35 and EC but there was a denser concentration in middle layers in area 36. Both the perirhinal as well as the entorhinal

massive set of weak connections. This point is technically more complicated, it has been much less explored and requires further investigation. However, *a priori*, we assume that if there are more axons from area A to area C than from B to C, then the input from A to C is *functionally* more important. Thus, we have not seen any reports claiming that here is a direct input from lower visual areas such as V1 to EC. If this exists at all, the current data seem to support that this would be comprised of very few projections; following our assumption then we can presume that it will be less relevant. In the case of the amygdala, there may be other pathways that can take information from the retina through the thalamus or even through the hypothalamus or superior colliculus. We assume that these pathways are probably not the most important ones in determining the visual selectivity to complex stimuli that we discuss in our work.

²⁸ Several textbooks also use the term “limbic system” to refer to several of these structures in the MTL including the amygdala and hippocampus.

cortices project back to area TE. These “backward” projections seem to be more widely distributed than the forward ones, an observation that has been made in anatomical studies of projections within the visual cortex (Saleem and Tanaka, 1996). In addition to area TE, the entorhinal cortex projects directly to other neocortical structures (Insausti *et al.*, 1997), to the amygdala and possibly to the striatum (Suzuki, 1996, Brown and Aggleton, 2001).

The EC constitutes the main input to and output from the hippocampus (Ramon y Cajal, 1968, Brown and Aggleton, 2001). The EC shows a robust projection from layer II to the molecular layer of the dentate gyrus. Most of the adjacent areas including the perirhinal cortex, the parahippocampal area as well as input from other neocortical sites project through the entorhinal cortex and not directly into the hippocampus. For this reason, the EC has traditionally been described as a relay station in the transmission of information to and from the hippocampus but it is possible that there can be very distinct functional differences as evidenced from new data from lesions and electrophysiological recordings (Brown and Aggleton, 2001).

The amygdala lies at the interface between the cortex and subcortical structures such as the striatum and hypothalamus. It is structurally heterogeneous with at least a dozen different sub-nuclei (Ziringer *et al.*, 2001, LeDoux, 2000). The major subdivision is between the basolateral and cortical parts of the amygdala. There are massive connections to and from the amygdala to both cortical and sub-cortical structures (LeDoux, 2000, Kandel *et al.*, 2000) in addition to a intra-amygdala connections (LeDoux, 2000). Of particular interest to us, the amygdala is intricately connected to the entorhinal, perirhinal and parahippocampal cortices (Suzuki, 1996). There is also evidence, at least in rats, that the amygdala projects to the hippocampus (Pikkarainen *et al.*, 1999). There are also direct forward and feedback projections between the amygdala and area TE (Cheng *et al.*, 1997).

Structures thought to be homologous to the amygdala and hippocampal system are present in all jawed vertebrates (Butler, 1996). This should not be taken to imply that there has been little change during evolution of these structures. On the contrary, there have been important changes in the size and distribution of connections in these areas. Several studies have remarked that there are striking overall general similarities in the

anatomy of the MTL between different mammals. In particular, the entorhinal-perirhinal-hippocampal system of monkeys seems to be very similar to that in the rat brain (Insausti *et al.*, 1997, Brown and Aggleton, 2001, Butler, 1996). The amygdala also shows very similar subdivisions and patterns of afferent and efferent connections between rats, cats and primates (LeDoux, 2000, Butler, 1996).

The anatomy of the human brain is known in much less detail. Several textbooks exist that describe many of the landmarks and structures in the human brain, but it is quite difficult to study in detail the connectivity as in the studies described above for monkeys and rats. In some studies, it has been possible to carry out some of the work done in nonhuman primates in order to investigate the anatomy of the human brain in further detail, and investigators have noted remarkable similarities between the two (see for example Insausti *et al.*, 1995).

2.9 Summary and conclusions

I hope the current chapter has convinced the reader that recording the activity of single neurons in the human brain is a difficult but feasible task. I hope the following chapters will convince the readers that this is a fascinating endeavor.

It is difficult because we do not have all the conditions typically present in an animal electrophysiology research lab. It is possible because we have been able to carefully record the signal with an appropriate SNR and separate the underlying units. The data that we have gathered is comparable and indeed quite similar to recordings obtained in animals such as rats and monkeys in similar areas. This is indeed a strong but common assumption in Neuroscience research. Most of the neuroscientists use animal models in the hope that this will allow us to better understand our own brains. This is by no means trivial and it does require an extrapolation. It is therefore reassuring to know that several properties are indeed very similar in our brains. The waveform shapes, the range of firing rates, the CVs, the bursting properties in the rat or monkey hippocampus, for example, are very close to those in the human brain.

The data shown in the current chapter can also have important implications in one potential direct application of recording from neuronal activity in the human brain: neural prosthetics (see for example Chapin *et al.*, 1999.) Fascinating data is being accumulated at a fast pace that indicates that it may be possible some day to implant electrodes in the human brain to direct the movement of artificial limbs (see for example the work of Nicolelis and the Caltech prosthetic arm project). In the transition to the human brain, it will be important to know what kind of electrodes to use and the basic properties of neuronal data in the human brain as we have outlined here. Furthermore, it may be very important to know the probability of actually recording from a neuron.

What are the remaining questions and technological challenges? Well, to begin with, the overall technology that we have used is at least one decade old (probably more). We should be able to record much more data with modern equipment. We would be able to record local field potentials at the same time as single neuron activity. We would have the continuous trace for each channel. We would not have to select which channels to record from. An anecdote is pertinent here. It took several years after the advent of electroencephalographic recordings (EEG) for sleep researchers to discover paradoxical or REM sleep (during which most dreams occur). One of the main reasons for this is that there was not much money for sleep research and they could only afford the paper to record a few minutes at a time. Therefore, in most of the recordings, they simply could not observe paradoxical EEG traces. When Kaserinsky and Kleitman started to record EEG traces during the whole night, they made the paradigm-shifting observation of paradoxical sleep. To a large extent, more paper was the key. Well, I think this is going to be the case in our research as well. Recording more data will lead to open several new territories²⁹. One trivial and immediate application would be, for example, the possibility of recording the activity of single neurons preceding every single seizure. To this date, the data that we have gathered in this respect was recorded when, by chance, the subject had a seizure during one of our tests. One potentially interesting venue to explore is tetrode recordings. Tetrodes are rapidly invading the animal electrophysiology world, enabling researchers to isolate better and more units in each recording sites. Several

²⁹ As I am writing these lines, I learn that we have finally placed the order for new recording equipment that will allow us to monitor the activity from 64 channels simultaneously!

fundamental questions (clustering of neurons with similar properties, local synchronized networks) will be open with this technology. In principle, there would be no reason not to test this in the clinical setup as well.

The spontaneous firing rates of most of the neurons in the medial temporal lobe are quite low. As stated above, most of the neurons showed an overall rate of <1 spike/sec (see Figure 2-8). This suggests that most of the data that we read from animal recordings is missing a large fraction of the neuronal population. It is unclear at this point what exactly this implies since it is possible that the group of neurons that are visually selective do have higher or lower spontaneous activity rates than non-visually responsive neurons. It should also be noted that the type of electrode used for electrophysiological exploration already introduces a bias in the types of neurons that can be recorded from. The firing rates in lower visual areas such as the retina and LGN and also cortical structures like V1 are much higher (Wandell, 1995, Livingstone and Hubel, 1988). This has led Christof to suggest that there may be a radical change in the representation of information and how neurons actually use spikes to encode the visual world to a sparse representation. This will be discussed in more detail in the next chapter when considering the visual responses.

2.10 Figure legends

Figure 2-1: Schematic of the electrodes

Two types of electrodes were used in the recordings (see text). Here we show a schematic description of the type of electrodes most often used in temporal lobe targets. Platinum/iridium contacts of approximately 1.5 mm length along the electrode were used to acquire clinical wide band EEG data. Through the lumen of the 1.25 mm diameter electrodes, 8 platinum/iridium microwires were inserted. Electrodes were fabricated at UCLA. Microwires extended 1 to 3 mm from the tip of the electrode, lying inside a cone with an opening angle of less than 45°.

Figure 2-2: Patient with electrodes implanted

This is a picture taken from one of our patients while he was in the ward. He had the electrodes implanted in his brain in order to monitor the focus or foci of his seizures as described in the text. Permission from the patient was requested prior to posting the picture in here.

Figure 2-3: MRI localization of electrode position

The position of the electrodes was determined by obtaining a structural magnetic resonance image. Here we show an example of an MRI obtained in one of the patients indicating the position of one of the probes in the right hippocampus. The top image (a) shows an axial image and the bottom image (b) shows a detail of the coronal section. MRI scans were obtained during the week of chronic monitoring with the electrodes implanted; post-operative CT and MRI were also obtained. The CT was co-registered with the MRI structural information for anatomic verification. The distal end of the electrode can be seen here in the images and included the platinum/iridium microwires schematized in the previous figure. The MRI was obtained at 1.5 Tesla. This allowed us

to indicate the anatomical region of the electrode but not to specify which layer or specific sub-region within each area we recorded from.

Figure 2-4: Sample of extracellular recording

(a) Sample of extracellular recording indicating the magnitude of the signal and signal-to-noise ratios. The data saved to disk consisted of a collection of 2 ms wide waveforms that crossed a manually set threshold indicated here by the horizontal dashed line. Two individual waveforms within this trace are zoomed in; the one on the left corresponds to a neuron while the one on the right did not according to the criteria defined in the text. The scale is shown in the bottom right part of the figure. The zoomed waveforms were 2 ms wide. (b) Sample of different extracellular waveforms recorded from the same microwire. All the waveforms that cross a manually set threshold are recorded. In this example, the threshold was set at 100 A/D mV as indicated by the arrow in the y axis. The waveforms were separated by spike sorting (see Section 2.5 and Appendix 1). Four different clusters were obtained in this example, indicated here in different colors. According to the criteria specified in the text, clusters 1, 2 and 3 corresponded to neurons whereas cluster 0 did not. A random sample of 50 waveforms from each cluster is shown; the actual total number of waveforms in this experiment was > 30,000. (c) Distribution of signal-to-noise ratios. The dashed vertical line shows the $SNR=2$ lower boundary. The red arrow shows the mean value. Bin size = 0.25 (d) Distribution of the amplitude (peak-to-trough) of the signal. The red arrow shows the mean value. Bin size = 50 mV.

Figure 2-5: Probability of finding a neuron

(a) Distribution of the number of electrodes implanted in each region. These are separated into those in the left and right hemispheres in (b) and (c) respectively. (d) Proportion of microwires with units from the total number of implanted microwires in each location. These values are divided into the right and left hemispheres in (e). The data shown in these figures come from 27 patients where all these values were available (see text for details).

Figure 2-6: Lack of influence of age, gender, handedness and IQ on the overall probability of finding a neuron

(a) Probability of finding a neuron as a function of the age of the subject. The red line corresponds to a linear fit of the data; the r^2 is indicated on top together with the statistical probability of obtaining such a correlation coefficient by chance (see text for details). (b) Probability of finding a neuron for male and female subjects. (c) Probability of finding a neuron for left- and right-handed subjects. (d-f) Probability of finding a neuron as a function of different quantitative measures of intelligence: (d) Full scale IQ, (e) Verbal IQ, (f) Performance IQ. The red lines correspond to linear fits of the data; the r^2 and p values are shown on top for each figure.

Figure 2-7: Attempt to characterize neuronal types

(a) Waveform width, firing rate, and attenuation (heights ratio) for each unit. (b) 2-D projections of the 3-D plot in a. Top, signal attenuation versus width, middle, firing rate versus width, bottom, signal attenuation versus firing rate. All data. (c) 3-D plot as in (a) for each location. (d) 2-D projections as in (b) for each location.

Figure 2-8: Distribution of firing rates

(a) Distribution of spontaneous firing rates for all data. The x axis was cut at 20 spikes/sec for clarity. There were 2% of units with firing rates beyond 20 spikes/sec. The arrow indicates the mean value. Bin size = 0.5 spikes/sec. (b) Detail of the distribution of units with firing rates between 0 and 10 spikes/sec separated according to the anatomical location of the unit. Bin size = 0.5 spikes/sec. The arrow indicates the mean value. The total number of units over which the distribution was computed is indicated in the top right corner of each graph.

Figure 2-9: Distribution of CV and CV_2

(a) Distribution of coefficients of variation (CV) for spontaneous activity (see text for details) for all data. The x-axis was cut at a value of 5 for clarity. There were 3% of units with $CV > 5$. The red arrow indicates the mean value. Bin size = 0.1. (b) Distribution of CV_2 for spontaneous activity. The red arrow indicates the mean value. (c) Distribution of CV for each location. Bin size = 0.1. (d) Distribution of CV_2 for each location. Bin size = 0.1.

Figure 2-10: Sample interspike interval distributions

Examples of interspike interval distributions for a non-bursting unit (a) and a bursting unit (b). Bin size = 2 ms. The red line corresponds to the best fit of an exponential distribution to the data. The inset shows in more detail the [0,100] ms time intervals. For each unit, we computed the auto-correlogram (c and d). Bin size = 2 ms. The black dashed line corresponds to the expected autocorrelogram for a homogeneous Poisson process. The red dashed line indicates the 99.9% upper confidence limit for the Poisson assumption. The maximum interspike interval for spikes within a burst was evaluated from the time at which the peak in the autocorrelogram (d) returned within the confidence interval as indicated by the arrow.

Figure 2-11: Distribution of burst duration and proportion of spikes in bursts

(a) Distribution of the durations of bursts including all bursting units. Neurons were considered to be bursting or not based on the comparison of the interspike interval distribution and auto-correlogram with those of a Poisson process; the criteria and details are given in the text (see Section 2.7.3). Bin size = 2 ms. The arrow indicates the mean value. (b) Proportion of the total number of spikes that occurred within bursts. The maximum interspike interval to consider two spikes to belong to one burst event was computed separately for each unit (see part a). Bin size = 0.05. The arrow indicates the mean value. (c) Distribution of burst durations for each location within the medial

temporal lobe. **(d)** Distribution of the proportion of spikes occurring within bursts for each location.

Figure 2-12: Block diagram of the input to and main connections in the MTL

(a) Schematic diagram of the human brain showing the temporal lobe in red. **(b)** Very simple diagram of the main pathways conveying visual input to the MTL. This block diagram is highly oversimplified and does not even begin to illustrate the anatomical complexity of the human brain; several important pathways are missing and we are not indicating the specific layers where projections start or end up (see text for details and references). All the data comes from detailed studies in monkeys and rats but not in humans.

2.11 Tables

Table 2-1: Comparison of single unit recordings in humans to other techniques

	Single unit recordings in humans	Single unit recordings in monkeys	fMRI in humans	Scalp EEG in humans
Spatial resolution	1 neuron ($\sim 10^{-5}$ mm ³)	1 neuron ($\sim 10^{-5}$ mm ³)	$> 1\text{mm}^3$ ($\sim 10^5$ neurons)	$> 1\text{cm}^3$ ($> 10^8$ neurons)
Temporal resolution	< 1 ms	< 1 ms	> 2 secs	< 1 ms
Extrapolation to our brain	Epileptic (and other) patients	Monkeys	Can study normal humans (as well as patients)	Can study normal humans (as well as patients)
Total experiment time per subject	Approximately 1 week	Several years	Unlimited	Unlimited
Areas of study	Limited by clinical constraints	Basically unconstrained	Unconstrained	Unconstrained
Number of areas to study simultaneously	Limited by clinical constraints	Constrained by difficulty of implanting electrodes	Can study whole brain	Can study whole brain
Type of experiments	Unconstrained (except for time constraints)	Only simple experiments and even these require extensive overtraining	Unconstrained	Unconstrained
Invasiveness	Invasive	Invasive	Non-invasive	Non-invasive

Comparison of four different techniques to study the human and non-human primate brain. Some important advantages and disadvantages of each technique are outlined here (see text for detail and references). The conversion from spatial resolution to neurons is based on the simplified assumption of a spherical neuron of 20 μm diameter. The time resolution indicated for fMRI involves the limitations involved by the hemodynamic response rather than the sampling rate of the equipment. Estimating the spatial resolution of EEG is complicated and controversial (Kandel *et al.*, 2000). A reverse source analysis, which requires several assumptions about the physical structure of the brain, can lead to a

resolution of a few centimeters. Without these assumptions, activity from an electrode probe in the frontal lobe could be caused by a source in the occipital lobe thus making the localization capability unclear. The reported value in this table probably corresponds to an optimistic lower limit after a source analysis.

Table 2-2: Percentage of units that showed bursting

	Amy	AC	Hip	EC	Par	Occ	Fr	PHG	SMA	Total
right	13	9	13	26	0	0	NA	2	10	15
left	39	0	14	46	38	46	0	7	13	32
all	33	6	14	38	27	35	0	4	12	26

Percentage of the total number of units that showed bursts of spikes (see Section 2.7.3 for the methods and criteria). Units are divided within each area into those in the right and left hemispheres. The locations are Amy: amygdala; AC: anterior cingulate; Hip: hippocampus; EC: entorhinal cortex; Par: parietal; Occ: occipital; Fr: frontal; PHG: parahippocampal gyrus; SMA: supplementary motor area.

3 Visual Properties

3.1 Introduction

Before embarking on an investigation of what occurs in the brain during visual imagination, dreaming or binocular rivalry, it is essential to understand whether individual neurons in the human medial temporal lobe (MTL) show visual responses upon presentation of different kinds of stimuli. We therefore started by exploring the responses of single neurons to a large array of complex visually presented stimuli.

The study of the single neuron visual responses in different areas of the monkey visual cortex has been an intensive and productive area of research in the past decades. This line of investigation has established a large parallel and hierarchical network with almost 30 visual areas (Felleman and Van Essen, 1991). In the first cortical target of thalamic projections, primary visual cortex or V1, neurons show selectivity for comparatively “simple” properties of “simple” stimuli. Investigators have described that there are neurons that increase their firing rate upon visual presentation of stimuli such as bars within a relatively narrow range of orientations as long as the stimulus is presented within the small receptive field of the unit (Wandell, 1995, Hubel and Wiesel, 1959, Livingstone and Hubel, 1988). Note that there is nothing simple about the mechanism by which this selectivity arises and neuroscientists still debate about the characteristics of models that are capable of giving rise to these properties. But at least there are some models and computational ideas of how the transformation of visual information from the retina to V1 could take place. The visual information is conveyed from V1 to subsequent stages for further processing. An important distinction has been drawn between the “where” or “motion” dorsal pathway and the “what” or “form” ventral pathway (Livingstone and Hubel, 1988, Zeki, 1992). Neurons selective to some combination of

color and shape have been described in visual areas V2 (Hedge and Van Essen, 2000), V3 (Gegenfurtner *et al.*, 1997) and V4 (Ghose and Tso, 1997, Pasupathy and Connor, 1999, Tanaka, 1993, Tanaka, 1996) in addition to inferior temporal (IT) cortex. Although there do seem to be some neurons before IT that can show specific activity to complex shapes, particularly in V4 but also in V2 (Hedge and Van Essen, 2000), neurons in these earlier visual areas do not show all the invariance properties of IT neurons. However, neurons in area V2 already show some very interesting properties that are quite relevant to our endeavor. For example, some neurons in area V2 respond to subjective illusory contours of the type that can be seen in the Kanizsa triangle (Peterhans and von der Heydt, 1991, von der Heydt *et al.*, 1984). Neurons in visual motion area V5 (also called area MT) integrate information from beyond the classical receptive field as described by John Allman (Allman *et al.*, 1985). Neurons in area V4 are selective to the color of stimuli and are influenced by long-range interactions from both local and global chromatic surround information. Zeki has described “perceptive” neurons that respond in a context-dependent manner to the perceived color of a stimulus rather than its wavelength (Zeki, 1983). At what is normally considered the other extreme of the ventral visual pathway, inferotemporal cortex, single neurons respond in a selective manner to “complex” stimuli such as faces and hands. Neuronal responses are not restricted to these “natural” stimuli; for example neurons can respond to common objects such as a calculator, different views of a paperclip (Logothetis and Pauls, 1995), or even fractal abstract shapes created in a computer (Miyashita and Chang, 1988, Yakovlev *et al.*, 1998)³⁰. In contrast to V1, we have a much more rudimentary understanding of how this selectivity arises. In fact, several investigators would argue, quite persuasively in my opinion, that we do not fully understand V1 either. But at least in that case, we know that neurons are tuned to specific variables such as the orientation of a bar or grating. Unfortunately, very little work has been done to attempt to address how the information about edges or the orientation of bars in primary visual cortex can be transformed into the selectivity for complex shapes such as faces observed in IT. One of the very interesting aspects of the neuronal responses in IT is that the neurons show a large degree of invariance to several physical properties of the stimulus including their position within

³⁰ Selective responses to paperclips or the complex patterns are typically observed after extensive training.

the visual field, size, color and texture (Gross, 1994, Desimone *et al.*, 1984, Tanaka, 1993, Tanaka, 1996, Logothetis and Sheinberg, 1996). Furthermore, Sary *et al.* have described that the selective activation is invariant to the type of cue used to define the complex shape of the preferred stimulus (Sary *et al.*, 1993).

It should be noted that the famous van Essen diagram does not end in inferotemporal cortex (Felleman and Van Essen, 1991). IT neurons project to higher areas in the medial temporal lobe (Suzuki, 1996, Saleem and Tanaka, 1996, Cheng *et al.*, 1997). Compared to IT, much less work has been done in monkeys to explore the visual properties of neurons in the entorhinal cortex, hippocampus, amygdala and parahippocampal gyrus. There are however some notable exceptions that have shown that MTL neurons can also show selective responses to complex visual stimuli (see for example Suzuki *et al.*, 1997, Rolls, 1984, Leonard *et al.*, 1985, Rolls *et al.*, 1989, Miyashita *et al.*, 1989, Nishijo *et al.*, 1988, Brothers and Ring, 1993). A comparison between our data and the monkey MTL electrophysiological responses is discussed at the end of this Chapter.

The strong input from visual areas, data from functional imaging, scalp and intracranial electroencephalographic field evoked potentials and some single unit studies in humans, together with the existence of patients with severe visual deficits after damage to the temporal lobe suggested to us that neurons in the human MTL might show responses to the presentation of different types of visual stimuli. In this Chapter I will summarize our efforts to attempt to better understand the visual description of the world in the human MTL. I will illustrate some of the neuronal responses to complex visual stimuli that we have observed showing that some units change their firing rate selectively upon presentation of specific categories of stimuli while others respond more specifically to individual stimuli. Stimuli were drawn from a collection of over a thousand pictures that included different types of faces, household objects, spatial scenes, animals and foodstuffs (see Figure 3-1). I will start by giving several examples of the types of visually selective responses as well as non-selective responses both in the MTL and in the supplementary motor area. Then I will describe the properties of these visual neurons including the distribution of selective responses to different types of stimuli and possible segregation, their latencies and durations, the responses to repeated presentation of the

same stimuli and bursting. I will finally compare our results to those obtained in animal electrophysiology experiments and the observations using distinct techniques in humans such as functional magnetic imaging, evoked potentials. I will also discuss the relationship between our results and neurological cases of patients with visual object recognition deficits. A brief summary of some parts of this Chapter has been reported previously (Kreiman *et al.*, 2000a).

3.2 Introduction

The neuronal responses shown in the current Chapter were obtained while subjects were looking at complex visual stimuli on a monitor screen. Stimuli were presented for 1000 ms and subtended a visual angle of approximately 5 degrees³¹.

Throughout the course of my thesis work, I performed three different experiments where visual stimuli were presented. One was "presenter stimuli," the second one was "visual imagery" and the third one was "flash suppression." While there were differences in the three tasks, common to all of them is the fact that subjects observed a picture in a monitor for 1000 ms. In presenting the results, one possibility would be to describe the visual responses during each of these experiments separately. Since several of the analyses and main conclusions are quite similar, this would involve a fair amount of repetition. Alternatively, one could pool all the data from the visual responses and just

³¹ Stimuli were always presented in the center of the screen and subjects were looking straight at the stimuli so that the pictures encompassed the fovea. In most of the experiments there was no fixation point (see comments on eye movements in Section 3.11.2). In some of the experiments stimuli subtended a visual angle of approximately 3 degrees instead of 5 (see Chapter 6). There was some variation in this angle given that the patient's head was not fixed in any way; we estimate this range to be from 4.2 to 6 degrees from one experiment to another in those cases where the target angle was 5 degrees. Repetitions in which subjects made significant moves of their heads were discarded so that there was not much variation within a given experiment. The size of the images was large enough that they were easily recognizable even when presented in the minimum boundaries of this range of visual angles. The receptive fields of IT neurons in monkeys are very large and typically extend into both visual half fields; they almost always include the center of gaze. Extensive data from monkey electrophysiology experiments show that at least the responses of neurons in inferotemporal cortex are largely invariant to large changes in size (Tanaka, 1993, Logothetis and Sheinberg, 1996, Gross, 1994). For example, Schwartz and colleagues showed that a change in stimulus size from 13 to 28 degrees of visual angle did not significantly alter the responses of neurons in IT (Schwartz *et al.*, 1983). To my knowledge, several of these questions have not been studied in detail in the MTL. Here I assume that the MTL neurons also have large receptive fields and are mostly invariant to large changes in stimulus size.

point out the differences among the different experiments. I have opted for the latter possibility. One of the experiments, "presenter stimuli," only involved the binocular presentation of visual stimuli for 1000 ms. Immediately after each picture disappeared, an auditory tone was presented to request the subject to respond by pressing a button indicating whether the picture included a human face or not. As in the other experiments, the order of presentation of the stimuli was randomized subject to the constraint that the same picture or different pictures from the same category of stimuli (see below for the definition and description of the different categories that were shown) were not presented more than three times in a row. In the second experiment (visual imagery), there were two stimuli per repetition, each was repeated five times and the two stimuli were randomly chosen with the only constraint that they could not belong to the same category (see Chapter 4 for more details). Subjects also had to perform the same human face discrimination response by pressing a button after an auditory tone was presented³². The third experiment was flash suppression. Here a stimulus was presented monocularly³³ and then another stimulus was flashed on the contralateral eye. The subject's task was related to the flash suppression experiment (see Chapter 6 for details).

It could be claimed that the fact that subjects would have to imagine the stimuli afterwards has an influence on the type of visual responses that we observed. It could also be argued that the fact that a different stimulus would be flashed in the flash suppression experiment would change the preceding neuronal response to the monocular stimulus. Overall, we did not observe any clear trend of this sort. In most cases, because of time constraints the same subject did not perform the three types of tests. Therefore, even if there were any differences in the visual responses, it is unclear that these could be directly attributed to the posterior task rather than to other important variations across subjects (see Section 3.11.8). I will point out throughout the description of the data whenever we did observe any clear differences between these three different experiments.

³² There were two auditory frequencies (high and low) for the tones; these were randomly chosen for each presentation of a visual picture. These tones were also used during the imagery test (see Chapter 6).

³³ It is quite unlikely that neurons in such high visual areas would respond differently to monocular compared to binocular stimulation (Wandell, 1995), but we have not directly tested this. Overall, the percentage of responsive neurons in the three tasks was roughly compared (see text). Data from monkeys suggest that neurons in inferotemporal cortex seem to show large receptive fields typically involving more than one visual hemifield and respond to stimuli presented through either eye.

3.2.1 Types of stimuli

A careful selection of which stimuli to present could be fundamental in the success of this enterprise. However, this has turned out to be a crucial issue since we lack a sufficient understanding of visual object recognition to design an algorithmically efficient and systematic method of trying to identify what kind of images the neurons respond to. In V1, one could present bars within the receptive field and systematically vary their orientation, contrast, and spatial and temporal frequency in order to maximize the neuronal response. Most of the work in the macaque higher visual areas such as inferior temporal (IT) cortex or the superior temporal sulcus (STS) has been based on a brute force approach in which a large database of pictures is shown in the hope that the neuron will respond to one or a subset of those images. This is the approach that we have followed so far.

We have compiled a database of approximately 1,000 images drawn from the following groups of stimuli: faces by unknown actors denoting emotional expressions (known as Ekman faces, Ekman, 1976), household objects, spatial layouts, animals, cars, drawings and photographs of famous faces, foodstuffs and abstract patterns. A sample of the type of stimuli is shown in Figure 3-1, and a more detailed description of some of the characteristics of each group of stimuli is given in Table 3-1. Previous work recording the activity of single neurons in the MTL had shown that some units could differentiate between Ekman faces and objects (Fried *et al.*, 1997). In addition, the putative role of the amygdala in processing expressions, particularly fear (LeDoux, 2000), led us to conjecture that these emotional expressions could efficiently drive the activity of neurons in the amygdala. The collection of Ekman faces includes neutral expressions and faces denoting fear, surprise, disgust and happiness by both male and female unknown actors. In several of the studies in monkeys, animals were highly overtrained for specific stimuli or groups of stimuli (Logothetis and Pauls, 1995, Logothetis *et al.*, 1995, Miyashita and Chang, 1988, Sakai and Miyashita, 1991, Yakovlev *et al.*, 1998). It has been shown, for example, that individual neurons can respond vigorously to complex fractal patterns (Miyashita and Chang, 1988, Yakovlev *et al.*, 1998) and to specific conformations and views of paperclips (Logothetis *et al.*, 1995, Logothetis and Pauls, 1995). Furthermore, it has been suggested that familiarity could be important in modulating the responses of

neurons in the anterior temporal lobe, particularly in the perirhinal cortex (Brown and Aggleton, 2001). Therefore, we decided to include famous people in the hope that some neurons would respond to these previously seen (and quite possibly over-trained) stimuli. Cars constitute another category of stimuli where there are several individual elements with slight differences, and many people are quite well trained at fine discriminations among cars. Furthermore, it could be argued that cars belong to a group of objects that serve a specific purpose and there are reports in the literature of patients who show specific deficits in the recognition of manipulable objects or visual stimuli with specific functions (Humphreys and Riddoch, 1993, Forde and Humphreys, 1999). We included spatial layouts because of the important role of the hippocampus in navigation (O'Keefe and Dostrovsky, 1971, Maguire *et al.*, 1998, Wilson and McNaughton, 1993, Muller, 1996, Kahana *et al.*, 1999) and because functional imaging data had suggested the activation of parahippocampal areas when human observers viewed pictures of scenes and landscapes (Epstein and Kanwisher, 1998, Epstein *et al.*, 1999, Maguire *et al.*, 1998, Aguirre *et al.*, 1996, Gorno-Tempini *et al.*, 2000). There had also been some reports that neurons in the amygdala could respond to foodstuffs (Rolls *et al.*, 1977, Rolls *et al.*, 1979). We also presented patterns including sinusoidal gratings and more complex abstract shapes because some neurons in monkeys were found to respond to specific fractal patterns after training (Miyashita and Chang, 1988). These stimuli are also interesting because they cannot be argued to belong to natural groups of pictures. There are also interesting observations from neurological studies where patients lose the capability of recognizing quite specific groups of stimuli including several of the categories that we have used (see for example McCarthy and Warrington, 1994, Heilman and Valenstein, 1993, Forde and Humphreys, 1999, Damasio, 1990) and discussion in Section 3.12).

It can be argued that the above division of categories is quite arbitrary³⁴. For example, in what sense is a car not an object? Cars were typically embedded in a natural scene and therefore one could further ask in what sense are cars separable from the other natural scenes? Given that after repeating the pictures of the unknown actors subjects became familiar with their faces, one could ask to what extent these faces do constitute a

³⁴ *Aspirat primo fortuna labori.*

separate category from the famous faces one. We also analyzed separately the responses to drawings and photographs of famous faces and it is not evident either to what extent these constitute two independent groups. Furthermore, it can be argued that not all stimuli that were presented constitute "main-entry" level categories in contrast to subordinate categories of larger groups (Logothetis and Sheinberg, 1996). For example, to what extent do houses and natural scenes belong to sub-groups of a main category of spatial layouts and not two different categories? These questions may depend on the criteria utilized to define the categories. These points are well taken. Indeed, several neurons answered some of these questions for us. For example, many neurons responded both to drawings and photographs of famous people (see for example Figure 3-4). Other neurons responded to both objects and cars, or cars and natural scenes. While these divisions were coarsely based on neurological, functional imaging and electrophysiological data as we have briefly discussed above, this does not necessarily imply that they constitute the boundaries that neurons use to discriminate stimuli into separate groups. I think this is a very important and fascinating question and I will come back to this point in Section 3.12 after showing the data.

It can also be argued the list of categories is very far from complete. For example, there were no faces of children, there were no familiar faces that were not famous, there were no pictures of plants, there was no separate category for kitchen utensils and so on. Although there were some neutral faces within the Ekman set, all of these unknown actors also appeared in other images while depicting emotional expressions. There is no claim to completeness of any sort in the number of categories in our stimulus set. Furthermore, we do not make any claims to completeness within each category either (for example, there were plenty of animals or groups of animals that were not shown; the same is true about the other categories).

3.2.2 Behavior: visual recognition

For most of the data presented in the current Chapter, subjects were asked to make a very simple discrimination indicating whether the picture showed a human face

or not³⁵. This was done to engage their attention and make sure they had seen the picture. All pictures were clearly visible, easily recognizable and presented for a relatively long time so that the task did not pose any difficulty to the subjects. This is evidenced by their performance in this task, which was above 95% (i.e., at ceiling). While this does not necessarily imply that subjects did recognize every single picture, we feel confident that this was the case. During both the visual imagery and flash suppression tasks we debriefed the subjects asking them to describe the stimuli and this made it clear that they could easily identify all the stimuli. To what extent the identification of the stimuli influenced the neuronal response or not is unclear. But there is data recorded from monkeys showing that familiarity as well as visual associations can modulate the activity of single neurons in the inferior and medial temporal lobes (Logothetis *et al.*, 1995, Miyashita and Chang, 1988, Sakai and Miyashita, 1991, Brown and Aggleton, 2001).

Trials in which subjects made a mistake or in which there was significant movement of the eyes, head or body were discarded from analysis. These movements were discriminated by the investigator (G.K.). However, we did not accurately monitor eye movements (see discussion about eye movements in Section 3.11). While it may be interesting in general to study the error trials, there were too few in our sample to perform a meaningful analysis.

3.2.3 Neuronal recordings

The methods for acquiring and processing the neuronal data were described in Chapter 2. In this Chapter, I will describe the responses of 1044 neurons in 22 patients. Based on MRI confirmation (see Chapter 2), the localization of these units is shown in Table 3-2. Given that in some cases we recorded from the same microwire in two different days, it is important to distinguish whether the units obtained from these microwires were the same or not; this is discussed in Section 3.11.7³⁶. Stimuli were

³⁵ The exception is the flash suppression experiment, described in Chapter 6, where subjects did not have to make any response after the monocular picture but had to indicate their percept by pressing a button after the flash (see Chapter 6 for details).

³⁶ The numbers reported throughout the text correspond to the total number of units according to the criteria described in Section 3.11.7 without any normalization. If we recorded from a neuron on two separate days but the change in the waveform shape or amplitude was beyond the threshold, these were counted as two

presented on a monitor (Sony CPD-G200 Trinitron) attached to a laptop computer (Dell Inspiron 5000, Delaware) through the VGA port. Behavioral responses were acquired through custom-built push buttons attached to the laptop. The time of presentation of the stimuli and the timestamps of the behavioral responses were recorded on the same computer that acquired the neuronal data.

3.2.4 Distribution of the number of presentations

Ideally, one would like to present a large number of groups of stimuli or categories (n_{cat}) to be able to detect neurons that respond to a specific group of stimuli, a large number of individual stimuli (n_{ind}) because neurons may respond specifically to some individual pictures, a large total number of repetitions per category ($catn_{rep}$) and per individual stimulus ($indn_{rep}$). Furthermore, to address the question of whether there are differences between individual stimuli within a group, one would like to maximize as well the number of distinct pictures per group $n_{ind_per_cat}$. Unfortunately, given the time constraints that we have in our experiments (see comments in Chapter 2), it is not feasible to maximize all these variables at the same time. It turns out that this apparently trivial empirical issue may turn out to be relevant for the collection of data, the types of analyses that are possible and the interpretation of the results. This is very easy to see in the extreme situations. Let us say that we only present pictures of faces and nothing else ($n_{cat}=1$). Then, of course, we would not be able to answer the question of whether the unit responded differently to faces than to objects, or other types of stimuli. But we could potentially present a large number of faces and then carefully quantify how the unit responds to different faces. In another extreme, one could present each individual stimulus only once but present several stimuli from many categories. Then, one could compare the responses among categories, but we would be unable to study the potentially different responses to individual stimuli within a category. The data that we have collected involved a compromise among all these and other extremes. We actually shifted the values of these parameters in the three experiments to be described in Chapters 3, 4 and 6. Therefore, we will here describe these parameters separately for each experiment.

separate units. This analysis was applied uniformly to all the data before analyzing the firing rates, visual selectivity or any other properties of the neurons (Section 3.11.7).

During the presenter stimuli test, visual stimuli were presented while subjects indicated the presence or absence of a face. These data are only described in the current Chapter. The total number of presentations ranged from 132 to 700 (428 ± 220 , the distribution is shown in Figure 3-2a, left). The overall number of individual stimuli presented per experiment ranged from 25 to 223 (Figure 3-2b, left inset), but if we only consider those stimuli presented at least four times, the range was from 25 to 92 (66 ± 19 , the distribution is shown in Figure 3-2b, left). For these, the number of repetitions per individual stimulus was 6.1 ± 2.1 (Figure 3-2c, left). We presented stimuli from nine different categories (except for the first two patients where stimuli were drawn from only three different categories³⁷). The number of repetitions for each category of stimuli ranged from 5 to 167 (67 ± 37 , the distribution is shown in Figure 3-2d, left). The number of individual stimuli per category was 8 ± 10 (Figure 3-2e, left). The imagery test is described in detail in Chapter 4. It clearly required much more time and therefore the total number of presentations was appreciably smaller ranging from 62 to 400 (200 ± 82 , the distribution is shown in Figure 3-2a, middle), less than half of the number that was shown in the presenter stimuli test. The average number of different individual stimuli was 30 ± 11 . Each stimulus was repeated five times, but these five repetitions were only interspersed by one other stimulus and not randomly distributed throughout the whole experiment (see Chapter 4). The number of categories was the same as in the presenter stimuli paradigm, and there were a total of 4 to 86 (25 ± 16) presentations per category. There were on average 4.1 ± 2.7 individual stimuli per category. Finally, during flash suppression (described in detail in Chapter 6), the total number of presentations ranged from 100 to 520 (319 ± 123 , the distribution is shown in Figure 3-2a, right). The number of individual stimuli was 36 ± 16 and each individual stimulus was repeated 7.7 ± 3.6 times. In most of the experiments, we only presented five categories of stimuli, but in some experiments we presented stimuli from three or nine categories. The total number of presentations per category was 67 ± 36 , and the number of individual stimuli per category was 12.5 ± 10 .

³⁷ Ekman faces, household objects and spatial layouts.

The total number of presentations was not uniformly distributed among all the categories of stimuli. In the first two patients, only stimuli from three categories were used, and in the latest six patients, stimuli were drawn from five categories. In addition to this, even in those experiments where stimuli were drawn from the nine different categories, the stimuli were not uniformly distributed. We biased the number of presentations because, coarsely based on the monkey electrophysiology literature, we *a priori* assumed that there would be more selective responses to some types of stimuli than to others. Thus, over all the experiments reported in the current Chapter, the proportion of presentations per category were 19.9% for Ekman faces, 10.6% for objects, 18.8% for spatial layouts, 13.1% for animals, 2.2% for cars, 8.4% for drawings of famous people, 11.7% for photos of famous people, 2.8% for foodstuffs and 12.5% for patterns.

3.3 Visual selectivity, definitions and examples

3.3.1 Category specificity, criteria

Several interesting questions arise in the study of the responses of the neurons to the diverse set of visual stimuli. One particularly intriguing question is whether the neurons responded selectively to some specific stimuli. While the seemingly innocent word “responded” may seem self-explanatory to some readers, it is clearly not so. The problem of the type of codes neurons use to represent information constitutes a very provocative and active area of research (see for example Rieke *et al.*, 1997, Koch and Laurent, 1999, Kreiman *et al.*, 2000c, Krahe *et al.*, submitted, Hopfield, 1995, Softky, 1995, Shadlen and Newsome, 1994, deCharms and Zador, 2000); see also the discussion in Sections 3.10 and 3.12.6). At this point it is still unclear what the best definition of neuronal response should be; it could very well turn out to be that there is no one single "best" definition and that neurons can use different ways of encoding information in different areas and networks or even under different circumstances. A neuron could change its activity in many different ways upon presentation of different visual stimuli. We have started by using a very simple definition of response in which activity was

integrated over a 900 ms window starting 100 ms after stimulus presentation. The 100 ms shift was introduced in order to account for the neuronal latency that is estimated to be approximately 100 ms in the monkey inferior temporal cortex (Schmolesky *et al.*, 1998, Tamura and Tanaka, 2001, Desimone *et al.*, 1984, Perrett *et al.*, 1982, Rolls *et al.*, 1982). This is by no means the only definition of “response” and I will briefly discuss this towards the end of this Chapter and give a glimpse of other possibilities that we have started to explore (see Sections 3.10 and 3.12.6). However, we have observed very interesting results with this naïve and conventional definition of a response; therefore, most of the results in the current Chapter will be based on this simple spike count criterion. Once we have settled into this definition of response, one can compare the neuronal spike count responses to all individual stimuli and we have actually performed this comparison (see Section 3.3.5). However, if a unit responds in a broader manner to several stimuli, however, its response could be missed from this kind of analysis. Let us qualitatively see why this is so. Let us assume that there is a unit that responds strongly to all the pictures of animals but not to any other image. This will lead to a large variance in some of the among groups responses but a small variance in the comparison across different pictures of animals. This could easily lead to a non-significant result in the comparison of an animal against all other stimuli because of the responses to the other animals. Several lines of evidence had suggested that there might be specialized responses to broad groups of stimuli (see brief discussion above). We therefore compared the neuronal responses to the nine groups or categories of stimuli divided as indicated above (see Table 3-1 and Figure 3-1). It should be emphasized that this division of stimuli into different categories was fixed and it was done before studying any of the neuronal responses.

A unit was considered to be selective to a particular category i ($i=1,\dots,9$) of stimuli if the following conditions were met:

(i) an analysis of variance (ANOVA) showed that the variability in the responses among categories was significantly greater than the variability within categories (Keeping, 1995, Scheffe, 1953).

(ii) the responses to category i were significantly different from the baseline based on a paired t test (Keeping, 1995). The baseline was computed in the $[-1000,0)$ ms interval with respect to stimulus presentation.

(iii) the responses to category i were significantly different from those to other categories based on Scheffe's multiple comparison procedure (Scheffe, 1953).

(iv) the variability in the responses to different individual stimuli belonging to category i was not significantly different from the variability to repeated presentations of the individual stimuli based on an ANOVA.

3.3.2 Some relevant technical issues about the analysis

There are several technical questions that we should discuss about these analyses. First of all, the total number of presentations per stimulus is quite important. The distribution in the number of presentations per stimulus was discussed above (see Section 3.2 and Figure 3-2). There were several pictures that were presented less than four times. While it is possible to do a statistical analysis with very small repetition numbers, one has to be extremely careful about the interpretation of these results. We therefore imposed a limit of four presentations to include a stimulus in our analysis in criterion (iv) above. For criteria (i)-(iii) we performed the analysis both including every stimulus and excluding those that were not presented at least four times. Since the number of repetitions per category was typically quite large (Figure 3-2), there was no difference between these two alternatives. The cut-off of four presentations is arbitrary and stems from practical constraints. While we would like to have ten presentations per picture, this would impose severe constraints on the total number of pictures that we could present given the limited amount of recording time that we typically have (see Chapter 2).

Secondly, the classical analysis of variance and the Student's t tests assume a normal distribution of the firing rate in each group (Keeping, 1995). This is indeed far from clear in our sample. There is a natural left wall at 0 spikes/sec in the distribution of firing rates. As discussed in the previous Chapter, many of the neurons show a very low firing rate. Therefore, the left wall can make the distribution highly asymmetric and non-gaussian (the left wall is much less relevant at high firing rates). While there is some robustness to deviations from normality in several statistical tests including the ANOVA

and t tests (Keeping, 1995, Scheffe, 1953), we also performed non-parametric analyses. A Wilcoxon rank-sum test was used to compare two groups of values (Lehmann, 1975). A non-parametric analysis of variance was used in the comparison of multiple groups. I will briefly discuss the non-parametric analysis of variance used in criterion (iv). Let c_{ij} represent the firing rate of the neuron in response to repetition j of individual stimulus i . $i=1, \dots, k$ where k is the total number of different individual stimuli within the selective category. $j=1, \dots, n_i$ where n_i is the number of repetitions of stimulus i . The ANOVA tests whether the variability in the response of the neuron to different individual stimuli is larger than the variability in the responses to different repetitions of the same stimulus (Keeping, 1995). Let S_w represent the sum-of-squares *within-groups*

$$S_w = \sum_{i=1}^k \sum_{j=1}^{n_i} (c_{ij} - \bar{c}_i)^2$$

where \bar{c}_i is the mean firing rate for stimulus i . Let S_b denote the sum-of-squares *between groups*

$$S_b = \sum_{i=1}^k n_i (\bar{c}_i - \bar{c})^2$$

where \bar{c} is the overall mean ($\frac{1}{N} \sum_{i=1}^k \sum_{j=1}^{n_i} c_{ij}$) with $N = \sum_i n_i$.

Note that S_w measures the variability in repeated presentations of the same stimulus while S_b measures the variability in the responses to different stimuli. Our statistic parameter will be the ratio between the two estimators of the standard deviation³⁸:

$$t^* = \frac{S_b / (k - 1)}{S_w / (N - k)}$$

We estimated the actual distribution of t^* by using a bootstrap procedure (Efron and Tibshirani, 1993). For each iteration with a different shuffled order of the stimuli, we computed the sum-of-squares values and the statistic t^* (the * denotes the fact that this was computed after shuffling the stimuli indices). Under the null hypothesis, H_0 , there is no difference among the responses to distinct stimuli within the selective category. If the

³⁸ Under the assumption of a normal distribution of firing rates in each group, t follows an F distribution with $k-1$ and $N-k$ degrees of freedom. This is the classical ANOVA (Keeping, 1995, Scheffe, 1953).

neuron does indeed respond more strongly to some stimuli than to others, then the observed value of the statistic \hat{t} computed from the actual order should be separated from most of the random shuffles. The probability that this happens under the null distribution can be estimated from the number of iterations in which the statistic t^* is larger than the observed value \hat{t} :

$$p = \text{prob}_{H_0} (t^* \geq \hat{t}) \approx \#\{t^* \geq \hat{t}\} / n_{\Pi}$$

where n_{Π} represents the number of different permutations that were analyzed and $\#$ denotes the cardinal or number of elements in the set. We used $n_{\Pi} = 1000$. This gives an accurate estimation of p (Efron and Tibshirani, 1993). The results of the non-parametric analysis of variance for one particular example are illustrated in Figure 3-13 which shows the distribution of values of t^* for all the different permutations. The arrow indicates the value obtained with the actual order of presentations and the shaded area is proportional to the p value. A low p value indicates that it is unlikely that the null hypothesis holds; that is, it supports the claim that the neuron shows a distinct response to different individual stimuli. This method was also applied for the non-parametric analysis of variance used in criterion number (i) above.

The analysis of variance (criterion (i)) does not specify which stimulus the neuron is selective to. For that purpose, a post-hoc multiple comparison analysis is required. In our case, it does not appear practical to perform all possible pairwise comparisons. While this is feasible, the number of comparisons is too large (on the order of a thousand or more in the case of individual stimuli, see below) and therefore one must be extremely careful in protecting the overall experiment error (Scheffe, 1953). Instead we compared each stimulus to the rest of the stimuli; this still requires a multiple comparison procedure but the number of comparisons is the number of different groups of stimuli that were presented. We used the Scheffe multiple comparison procedure (Scheffe, 1953). Briefly, using the same nomenclature from above, we want to test the null hypothesis H_0 given by

$\sum_{i=1}^k \alpha_i \bar{c}_i = 0$ where the coefficients are subject to the constraint that $\sum_{i=1}^k \alpha_i = 0$. The

Scheffe method involves computing the statistic $\sum_{i=1}^k \alpha_i \bar{c}_i / \sqrt{\sum_{i=1}^k \alpha_i^2 \frac{MS_e}{r}}$ where MS_e is the

mean square error and is divided by the number of repetitions per stimulus. This is compared with the critical value given by $\sqrt{(k-1)F_{1-\alpha, k-1, N-k}}$ (Scheffe, 1953). This gives the lowest amount of false alarm (type II) errors.

Finally, criterion number (ii) also requires some careful consideration. The baseline response of each neuron can be computed by averaging over all stimuli or by averaging only the activity prior to the occurrence of each individual stimulus. If there is an overall activity change before a specific stimulus is shown (independent of the stimulus itself) and we compare against the overall baseline, the neuron may appear to show a strong response to the stimulus when it did not. On the other hand, if we only compare to the specific baseline for each stimulus, significance can also arise artificially due to a random change in the baseline but not during stimulus presentation. We computed the baseline in both ways; we did not observe any overall difference between the results from either procedure.

3.3.3 Category specificity, examples

I would like to illustrate our findings now with a few examples. An example of a unit that responded selectively to only one category of stimuli is shown in Figure 3-3. This unit was recorded from an electrode located in the right entorhinal cortex. This unit responded selectively to pictures of animals according to the criteria postulated in the previous Sections. The activity of this neuron during the [100;1000) ms interval after stimulus onset was different from baseline for animal stimuli ($p < 10^{-4}$), but not for the other stimulus categories ($p > 0.1$). A one-way ANOVA comparing firing rates between categories yielded $p < 0.001$, and comparing the activity for animals to all other categories yielded statistically significant differences ($p < 0.001$). The latency of response for this neuron was 219 ms, and the duration of increased response over baseline was 752 ms. How specific was the response of the neuron within the selective category? If the average increased response to animals were due to an enhanced firing for only a few pictures of animals, one might expect to observe a bimodal (or multimodal) distribution of firing rates. However, the distribution of firing rates for this neuron during presentation of different pictures of animals did not show any clear signs of

multimodality (Figure 3-3**b**). Although there was variability in the response of the neuron to individual instances of animals, the neuron responded above baseline for all pictures of animals (Figure 3-3**c**; $p < 0.005$). We also compared the variability for different presentations of the same animal to the variability across different pictures of animals using an analysis of variance (ANOVA). Both parametric and non-parametric ANOVA tests failed to show differences between individual stimuli ($p > 0.4$).

Another example of a category-selective unit is shown in Figure 3-4. This unit was recorded from an electrode located in the right anterior hippocampus. This neuron showed an increased firing rate over baseline in response to drawings of famous people as well as, to a lesser degree, to photos of famous people ($p < 0.001$). A one-way ANOVA yielded $p < 0.001$, and subsequent across-categories, pair-wise comparisons also showed that the activity during stimulus presentation was significantly higher for these two categories. Although the peak response was larger for drawings than for photographs (13.9 spikes/s versus 9.6 spikes/s), the average activity was not significantly different (Wilcoxon rank-sum test, $p > 0.05$). This neuron did not respond to faces *per se*, as indicated by the lack of change in the activity for emotional faces of unknown actors. The distribution of firing rates for photos of famous people for this neuron did not show clear signs of displaying more than one mode (Figure 3-4**b**). Variability in the responses to distinct individual photos (Figure 3-4**c**) was not higher than variability across different presentations of the same photograph (ANOVA, $p > 0.2$). Similar results held for drawings of famous people. The response to all individual stimuli within the selective categories was significantly different from baseline ($p < 0.01$). This neuron belongs to the same patient as the one in the previous example.

3.3.4 Responsive but non-selective units

We also observed some neurons that showed an enhanced response beyond the baseline during the presentation of visual stimuli but did not show any preference to any category or individual stimulus. If the ANOVA analysis (criterion (i) above) was not significant but the response in 75% of the categories was significantly different from baseline (criterion (ii) above), the neuron was defined to be responsive but non-selective. An example of such a neuron in the medial temporal lobe is shown in Figure 3-5. This

neuron was located in the parahippocampal gyrus. Note that the neuron fired almost exclusively during the presentation of a visual picture and not during the preceding or posterior time period. There was no clear difference in the responses to different types of stimuli. While there was some degree of variability in the number of spikes in each category, no statistical significant difference was found in the ANOVA analysis ($p > 0.2$).

This type of non-selective neuronal response was most prominent in the supplementary motor area. One example of a strongly responsive but non-selective neuron is shown in Figure 3-6. This unit increased its activity above baseline for every type of visual stimulus that was presented. It is quite striking how the neuron was almost completely silent in the absence of the visually presented stimulus. We should also point out the trivial fact that the subject kept his eyes open during the entire experiment and therefore there was a visual input outside the vertical dashed lines shown in the figure; however, assuming that a large fraction of the time he was staring at the screen, the monitor was blank except in the presence of a stimulus. Whether the neuron fired when the subject looked elsewhere I do not know. Some colleagues have suggested that this response could be related to the fact that subjects were engaged in a behavioral task. Although the task was completely trivial, it was a task nonetheless. The location of this unit, supplementary motor area, immediately prompts us to think about the movement itself (including possible eye movements although these are typically controlled in several other brain areas at least in the macaque brain, Kandel *et al.*, 2000.) During the picture presentation itself (which is when the neuron responded) there was no movement. I actually tried to eliminate as much as I could any artifact due to movements. True, I could have missed some small movement, but it is not possible that I missed several hundreds of overt movements. The subject's response occurred more than 1000 ms after the onset of neuronal response. Could the neuronal response be a preparation for movement? It seems plausible although unclear. There was no correlation between the response of the neuron and whether the subject would press one button or the other (that is, whether he had seen a face or not). Half way through the experiment, we switched the buttons the subject had to press in order to indicate his response. This did not affect the neuronal activity at all. One possibility is that the neuron was involved in signaling the presence of a sudden change in the visual world. Consistent with this hypothesis is our

observation that during the flash suppression task, we observed some neurons that fired non-selectively upon presentation of both the monocular stimulus and the flashed stimulus (see Chapter 6). Visual responses in motor areas have been described in the monkey pre-cortex (see for example Murata *et al.*, 1997, Graziano *et al.*, 1997, Rizzolatti *et al.*, 1997.) Rizzolatti and colleagues suggested that these neurons behave as “mirror neurons” and could be involved in processes of motor learning (Rizzolatti *et al.*, 1997). Of course, our experiments were not specifically designed to study the visual responses of the SMA and therefore much more extensive testing would be required in this realm.

3.3.5 Visual responses to individual stimuli, criteria

If a neuron responded in a very specific manner to only one individual picture, it would be easy to miss the change in firing rate in the previous analysis. This is because that highly selective activity would be diluted among all the other non-preferred stimuli from the same category over which the results were averaged. We therefore studied the responses of each unit to the 1000-ms presentation of the visual stimuli by using the same count in the [100,1000) interval described above. A neuron was considered to be visually selective to a specific stimulus S if the following three criteria were met:

(i) an analysis of variance (1-way ANOVA) yielded $p < 0.05$. This indicates that the variability in the responses to different stimuli was greater than the variability to repeated presentations of the same individual stimulus.

(ii) The response of the neuron to stimulus S was significantly different from that to the rest of stimuli (multiple comparison by Scheffe's method Scheffe, 1953.)

(iii) The response of the neuron to stimulus S was significantly different from the baseline activity (interval [-1000,0) ms with respect to stimulus onset at $t=0$ ms).

Several technicalities with respect to the methodology were discussed above with respect to category-selective units (see Section 3.3.2).

3.3.6 Visual responses to individual stimuli, examples

Following the scientific tradition of showing nice examples, I will illustrate our findings by showing as examples the responses of two neurons that displayed “nice”

selective responses. The danger in this approach is that the reader may get the impression that these are the typical responses that we observed in the more than 1000 neurons that we recorded from when they are not. The summary of the results presented later in this Chapter (see also Table 3-2) will convince the reader that most neurons did not respond selectively to either individual stimuli or categories of stimuli. Within those neurons that did respond selectively, some units showed stronger responses than others. The apparent strength of the responses correlates quite well with the probability of misclassification p_e obtained in the ROC analysis and we show the distribution of p_e values and therefore the quality of the selective responses for all the selective units in Figure 3-11 (see also Section 3.7).

A striking example of a selective response is illustrated in Figure 3-7. This unit was located in the amygdala, in a different subject than the ones from the previous examples. The response to all stimuli that were presented at least four times is shown in the figure. There was an increase in the response to only three out of the total of more than fifty stimuli. These three preferred stimuli were a black and white drawing of the face of ex-president Clinton laughing heartedly, a color photograph showing the face of ex-president Clinton smiling and a color photograph of ex-president Clinton, standing next to Hillary Clinton and greeting an actress³⁹. The unit reached peak activity rates of over 20 spikes/sec in bins of 200 ms only for these three pictures. It is quite clear that the neuron does not respond to all faces. It does not respond to all famous faces either (there were several other photographs and drawings of famous characters in the set including, for example, those of Batman and Michael Jordan). It does not respond to just any president (see arrows in the figure pointing to drawings or photographs of George Bush senior, George Washington and John F. Kennedy). The patient did not specifically manifest any specific love or hatred towards ex-president Clinton. The three pictures could hardly differ more in terms of their image properties (at the pixel level). One is black and white suggesting that there is a strong invariance to the color of the stimulus.

³⁹ This last picture was actually presented by mistake. All the photographs and drawings of famous people that were presented corresponded to faces of only one person per picture. I specifically tried to avoid presenting more than one person because I thought that this would make the interpretation even more complicated and the stimulus set even more inhomogeneous. I had changed the proportion of stimuli that were being presented the previous night and this picture had escaped the rigorous control. Ah, mighty serendipity.

Two of them show a front view of the face while the last one shows a full body view including a small side view of his face in a different position within the picture. There is basically no distracting stimulus in two of the pictures whereas there are several other foci of attention in the last picture. The degree of invariance of the response of this neuron is simply astounding. Whether the neuron would also respond to the word Clinton, to Monica Lewinski, to Hillary Clinton alone, upon the patient imagining Clinton, etc. I do not know.

Another example of a different unit in the amygdala is shown in Figure 3-8. This neuron responded selectively to only one of the pictures in the stimulus set that we presented. This was a black and white drawing of the face of Curly, one of the characters of a famous TV series called "The three stooges." I would like to emphasize that I am not claiming here that this is the only possible stimulus in the world that the neuron would respond to. That, I do not know. The claim is much more modest; among the 47 pictures that were presented at least four times⁴⁰, the neuron only showed an increased firing rate for Curly. I do not know whether the neuron would also respond to the other two characters of the three stooges, or to any other TV comedian. To be exact, I do not even know if the neuron would respond or not to a picture of the planet Jupiter. I am not claiming either that this is the only neuron in the whole brain that would respond to Curly. The representation in the brain should be somewhat robust and therefore it is unlikely that a single neuron in the whole brain would represent the face of Curly. What would happen in that case if for some reason that particular neuron were injured or lost? Would the subject stop recognizing Curly? This situation seems very unlikely. In spite of the fact that there are more questions that I cannot answer than the number of questions to which I can give a data-based reply, the degree of selectivity of these units is quite astounding and strikingly reminds us of the "grandmother neuron" theory of visual object recognition (Barlow, 1972, Poggio, 1990, Gross, 1994, Kandel *et al.*, 2000.)

⁴⁰ No response was apparent to any of the other stimuli that were presented less than four times either. But it is hard to make statistically meaningful claims with such low numbers.

3.4 Summary of visual responses

3.4.1 Distribution of category-specific responses

We recorded from a total of over 1000 units (see Table 3-2). Over 900 of these were in the medial temporal lobe. This is mostly due to the fact that more electrodes were implanted in the medial temporal lobe as described in Chapter 2 but also to our bias in selecting electrodes from the MTL to record from. Most of the visually selective units (both units responding to categories and to individual stimuli) were located in the MTL. There was a clear difference in the distribution of the responses in the MTL compared to the SMA (and the other non-temporal areas). In the former, most of the neurons that showed some statistically significant response were selective either to a category of stimuli or to specific individual stimuli. In contrast, in the latter, most of the responsive neurons were non-selective. Only very weak category-selective responses were observed in the SMA and AC. None of the SMA neurons, and only two of the extra-temporal units, showed selective responses to individual stimuli.

Upon analyzing a large number of neurons, one might expect to find some units that show up as statistically significant purely by chance. This constitutes an important concern when the number of significant neurons is rather small as in our samples. It should be noted that we are reporting the proportions here over the total number of neurons that we recorded from without eliminating any units from the analysis. This is quite different from approaches taken in other electrophysiological reports. One typical bias is introduced by analyzing only neurons that pass a threshold firing rate, typically set between 1 and 5 spikes/sec. In our case, we are including a large number of units with very low firing rates (see Chapter 2). It is interesting to note that the percentages of selective neurons would increase significantly if we were to exclude those units with low firing rates⁴¹. In other words, the number of selective neurons is not distributed uniformly between those with higher and lower firing rates. This is in part due to the fact that it is more difficult to detect statistically significant small changes in the spike counts when the

⁴¹ For example, as reported in Chapter 2, 21% of the neurons had firing rates below 0.5 spikes/sec (see Section 2.6.1). Only three of these neurons were selective according to the criteria specified above.

numbers of spikes are very low. Back to the question of attempting to assess whether the number of selective neurons could be obtained by chance or not, a common mistake is to compare the percentages with some magical number like 5%. It is not true that random firing would lead to 5% of neurons being significant using the criteria defined above for selectivity (Section 3.3.1 and 3.3.5). This was verified by simulating for each unit a Poisson process with the same firing rate. Yet another way to evaluate this question is to compare the distribution of p values expected by chance with the actual distribution of p values. These two distributions can be compared using a χ^2 test (Kreiman *et al.*, 2000a, Keeping, 1995). Both of these methods clearly show that the number of non-selective units, category-selective units, and units selective to individual stimuli in the MTL cannot be due to chance ($p < 10^{-4}$ for the comparison with a Poisson process and $p < 10^{-3}$ for the χ^2 test). In the frontal lobe, the number of non-selective neurons could not be due to chance according to both tests ($p < 10^{-5}$ for the comparison with a Poisson process and $p < 10^{-5}$ for the χ^2 test). However, we cannot reject the hypothesis that the number of selective neurons in the frontal lobe is due to chance ($p > 0.05$ for both tests).

Overall, 12% of the units in the MTL showed selectivity to one or a few categories of stimuli (Table 3-2). The distribution of the selectivity of the neurons in each location to the different categories of stimuli is shown in Figure 3-16. The proportions of units were computed from the total number of visually responsive units in each area. The responses are clearly not uniformly distributed among the different types of stimuli for the hippocampus and the parahippocampal gyrus. We quantified this by comparing the actual distribution with a uniform distribution using a χ^2 goodness of fit test. The null hypothesis of a uniform distribution could not be discarded at the 0.05 level for the amygdala and entorhinal cortex. However, the null hypothesis could be rejected at the 0.001 level for both the hippocampus and the parahippocampal gyrus. There were more responses in these two areas to spatial layouts and very few to some of other categories of stimuli. The hippocampal network has been implicated in navigation and the processing of spatial information. In rodents, O'Keefe and colleagues have described the existence of very specific "place cells" that are strongly and selectively activated when the animal is in a specific position within a maze or environment (O'Keefe and Dostrovsky, 1971, Muller, 1996). Different neurons show a preference for different positions within a

specific environment layout (Wilson and McNaughton, 1993). The trajectory of the animal can be reconstructed quite accurately from the activity of several tens of these neurons (Wilson and McNaughton, 1993). Specific oscillations and modulation of these oscillations have been shown in local field potentials in humans during navigation (Kahana *et al.*, 1999). Lesions in the hippocampus (Murray *et al.*, 1998) or knock-out of specific genes required for normal synaptic plasticity in the hippocampus (McHugh *et al.*, 1996) lead to severe navigation deficits. In humans, the parahippocampal area seems to be activated according to functional imaging evidence during the observation of houses and other spatial layouts (Epstein *et al.*, 1999, Epstein and Kanwisher, 1998, Maguire *et al.*, 1998). The navigational deficits and disorientation in the elder have been associated with hippocampal deficits (Kandel *et al.*, 2000). Are some of the processes involved in actual navigation similar to the ones involved in recognizing a spatial layout or a scene? To what extent does looking at a picture of a house or a landscape require the involvement of networks used for navigation? Our task did not involve movement, imagination of movement, or the ability to understand the disposition or location of different parts of the environment in the picture. Subjects were just looking at the pictures in a monitor. This is similar to some of the experiments done in fMRI. It should be noted that the pictures of houses did include a small entrance or garden and did not just represent the façade of the house. To what extent this gives the impression of an “environment” is questionable. We therefore performed a post-hoc analysis in which we separated the “spatial layouts” stimuli into houses and natural scenes. We did not observe any statistically significant differences for the neurons selective to spatial layouts in the direct comparison of the responses to these two groups. It should also be noted that the activation of the hippocampus may be related to a more general role in association of information than to navigation alone (Eichenbaum, 1996, Eichenbaum, 1997).

It was somewhat surprising to us that we did not observe any differences in the proportion of units selective to these different categories in the amygdala. We had expected that more neurons would respond to the Ekman faces than to other faces. Perhaps the set of stimuli represented by the Ekman faces did not really arouse any emotions in the subjects nor did they invoke specialized centers for the recognition of

emotional expressions. Alternatively, it is conceivable that the famous faces (and perhaps even some other stimuli) also gave rise to similar emotions that are processed in the amygdala. Indeed, we observed a large number of neurons that responded selectively to faces (both to the categories containing faces but also to individual faces; see Figure 3-17). It should also be noted that the amygdala encompasses a very large brain area comprising several different nuclei and sub-nuclei that are different at the anatomical, physiological and molecular level (LeDoux, 2000, Cheng *et al.*, 1997, Pikkarainen *et al.*, 1999, Zirlinger *et al.*, 2001). It is likely that most of the neurons were in the basolateral part of the amygdala but the resolution of the MRI localization of electrodes makes it difficult to precisely pinpoint which sub-nuclei the microwire was in (see Chapter 2). Pooling together the neuronal responses from different areas within the amygdala might have obscured potentially important differences in the visual selectivity of each sub-region.

Some of the neurons showed changes in firing rate in response to more than one of the categories (see for example Figure 3-4). The proportion of neurons, relative to the number of selective neurons, that responded to one category or more than one category is illustrated in Figure 3-17. Most neurons only responded to only one group of stimuli. Of the neurons that responded to two categories, most responded to two “related” categories such as drawings and photos of famous people or emotional faces. Some also responded to cars and spatial layouts (most of the cars were embedded in some scene in the background; see Figure 3-1).

As noted above, we also observed neurons that showed significant but non-selective changes in firing rate during stimulus presentation (such as the examples shown in Figure 3-5 and Figure 3-6). The proportion of non-selective neurons in each region is indicated in Table 3-2. Overall, 7% of the units showed a significant but non-selective visual response.

Although most of the visually selective neurons showed an increase in the firing rate upon presentation of visual stimuli, some cells had reduced firing rate from baseline (four neurons in the amygdala, four in the entorhinal cortex, five in the hippocampus and none in the parahippocampal gyrus). This corresponds to approximately 1% of the total number of neurons and 13% of the total number of selective responses. These values are

somewhat lower than the proportion of neurons that show an inhibitory response in the monkey inferior temporal cortex where some studies report that up to 25% of the selective responses corresponded to a decreased firing from baseline (Tamura and Tanaka, 2001). Given the low firing rates (see the distribution of firing rates in Chapter 2), it is typically difficult to detect a reliable decrease in firing rate. Decreased activity from baseline was also observed in the visually responsive but non-selective neurons (one neuron in the amygdala, five in entorhinal cortex, four in hippocampus and none in the parahippocampal gyrus).

Most of the neurons responded during the stimulus-presentation period, but there were some that responded when the stimulus was removed. To address this, we computed for all neurons in the presenter stimuli or imagery tests, within the 2000 ms after stimulus onset, the number of spikes in a 600-ms interval centered on the peak of the response and statistically analyzed the responses as described above. There were eight selective neurons (three in hippocampus, three in entorhinal cortex and two in the amygdala) that showed a statistically significant late response and were not detected with the previous analysis⁴². This corresponds to less than 1% of the total number of neurons and 7% of the number of selective units. It has been previously reported that neurons can respond to visual stimuli with a long delay even after the stimulus disappears (Yakovlev *et al.*, 1998, Miyashita and Chang, 1988). It is likely that in those cases the activity was related to the nature of the match to sample task in which the monkeys had to retain some visual information in the inter-trial periods in order to perform the task correctly and thus get a reward. The neuronal activity during these delay periods could represent a neuronal correlate to visual imagery similar to what we describe in Chapter 4. In the absence of such a task as we report here, we did not find many neurons that responded beyond the stimulus presentation period.

⁴² The data presented in the current Chapter from the flash suppression experiment only includes the monocular presentation. Immediately after the 1000 ms monocular picture, a different picture was flashed and therefore it is not possible to clearly study delayed responses to the first stimulus.

3.4.2 Distribution of responses to individual stimuli

Overall, we observed that 6% ($n = 59$) of the total number of MTL neurons showed selectivity to one or a few individual stimuli but not to all stimuli within a category. It should be noted that this group of neurons does not show any overlap with the category-selective neurons that we have just described. A neuron was classified either as category-selective or selective to individual stimuli (or non-selective) but not within both groups (see Sections 3.3.1 and 3.3.5). We observed more neurons selective to individual stimuli in the flash suppression experiment than in the previous two experiments; this is probably due to the fact that in that experiment we presented a smaller number of categories and more individual stimuli per category and a higher number of repetitions per stimulus (see Figure 3-2). Although the total number of units selective to individual stimuli was less than the category-selective units it was clear these highly selective responses could not be explained by random firing; this was assessed both by comparing to the number of expected neurons by chance using a χ^2 test ($p < 10^{-3}$) and by comparison to Poisson neurons as explained in the previous Section ($p < 10^{-4}$). In order to address whether there were differences in the distribution of neurons selective to different types of individual stimuli among the different MTL locations, we compared the proportion of selective neurons in each area to a uniform distribution using a χ^2 test (see Figure 3-17)⁴³. The null hypothesis of a uniform distribution of selective neurons across the nine categories could not be discarded for the selective neurons in the EC and PHG. However, both in the amygdala and in the hippocampus, the uniform distribution could be discarded at the 0.005 level. In the amygdala, most neurons responded selectively to human faces (particularly Ekman faces), there were also several units that responded to animals but none or almost none to individual objects, spatial layouts, cars, foodstuffs or patterns. In the hippocampus, we observed a predominance of responses to individual spatial layouts but also to emotional and famous faces.

⁴³ In Figure 3-17 we show the proportion of those units that were selective to individual stimuli within each of the nine categories that were presented. One could ask about the distribution of responses to each of the total of approximately 1000 different pictures that were presented to these 22 patients. We did not present the same pictures to all the patients or from one experiment to another. Therefore, it is complicated to compare whether there were more neurons that were selective to Batman than to Bush or to a tiger.

Most of the units that were selective but not category-selective responded to only one stimulus within the set of stimuli that we presented (Figure 3-18; see for example Figure 3-8). Responses to more than one individual stimulus were mostly within a category. In other words, there were very few instances of a unit that responded, for example, to a famous face and a calculator and a car. To address whether this could be due to chance, let n_i be the total number of stimuli within category i ($i=1, \dots, k$), let N be the total number of presentations $N = \sum_{i=1}^k n_i$, and let s be the total number of individual stimuli that the unit responded to according to the criteria defined in Section 3.3.5. The total number of combinations of s stimuli out of the total of N is simply given by

$$\binom{N}{s} = \frac{N!}{(N-s)!s!}$$

Now let us compute how many of these correspond to all s stimuli belonging to one category (any category). Repeating the same computation we just did, the number of combinations of s stimuli out of the total of n_i stimuli in category i is simply given by

$$\binom{n_i}{s} = \frac{n_i!}{(n_i-s)!s!}$$

where we define the value $\binom{n_i}{s}$ to be 0 if $s > n_i$. We can therefore compute the probability

P of obtaining by chance all individual stimuli from the same category:

$$P = \frac{\sum_{i=1}^k \binom{n_i}{s}}{\binom{N}{s}}$$

For several units that fire independently, the overall probability is the product of the individual probabilities assuming independent firing. When we compared these values to the ones that we have actually observed, it seems evident that the neurons that responded specifically to more than one individual stimulus preferred stimuli belonging to the same category. It could be argued that this is simply due to the fact that stimuli from the same category were more similar from a physical point of view than stimuli from different categories. Pictures from the same category could be more likely to share

features like color, size, textures, number of salient aspects, organization of features and so on. Indeed, as we will discuss below, this is clearly one possible interpretation of the origin category-selective neuronal responses that we have observed. In several cases, this is far from clear, though. For example, we emphasized before the high degree of dissimilarity from a physical point of view of the three pictures of Clinton that the neuron responded to (Figure 3-7). In other cases, when a neuron responded to different animals for example, it is not quite self-evident to me that a white eagle in a blue sky resembles the face of a tiger (from a strictly physical/structural viewpoint) more than a picture of a gray car in a blue background. Alternatively, the common response to several individual stimuli within a category (or to all stimuli within a category) may be due to the representation of other features shared by pictures in the same category beyond the strictly physical or structural ones. Further experimentation is required for us to be able to fully address this question. Interestingly, some studies have shown that monkeys can be trained to categorize visual stimuli and that neurons in IT and frontal areas can show selective responses that seem to respect the behaviorally trained categorical boundaries (Freedman *et al.*, 2001, Vogels, 1999b, Vogels, 1999a).

3.5 Is there any clustering of selective units?

Several networks in the nervous system show an astounding and marvelous organization. For example, a topographical map of auditory space can be found in the owl's brain equivalent of the inferior colliculus (Konishi, 1995, Konishi, 1993, Konishi, 1991). The visual world is topographically mapped onto the primary visual cortex (Wandell, 1995). Furthermore, neurons that have a similar orientation preference are localized within the same columnar structure (Kandel *et al.*, 2000, Wandell, 1995). Tanaka and colleagues suggested in a marvelous and intriguing paper that there might also be a columnar organization in the complex preferences of neurons in IT (Tanaka, 1993, Tanaka, 1996, Fujita *et al.*, 1992). It is therefore interesting to ask whether we observed any pattern or systematic organization in the location of the responsive neurons.

Unfortunately, this is a complicated question to address given our clinical constraints. The work of Tanaka and colleagues was carried out in monkeys by moving the recording electrode. This is something that we cannot do in our patients where the microwires were fixed. Furthermore, our data are pooled from different microwires with a separation larger than the inter-columnar separation described in monkeys (see (Kreiman *et al.*, 2000a, Fried *et al.*, 1999) and Chapter 2). The error in our estimation of the location of the microwires can clearly span more than a cortical column. Furthermore, the responsive neurons are distributed among all the patients that we have studied and it is quite complicated to draw any conclusions about detailed topography by comparing across patients.

Given these constraints, it is unclear to what extent we would be able to detect any degree of spatial clustering of selective responses even if there were a very strong topography. We can attempt to compare the selectivity of neighboring neurons in our recordings. There are two practical ways to define the term ‘neighboring’ here. In one, neurons recorded in the same microwire are considered to be neighbors. The distance between separate neurons recorded in the same microwire can be on the order of $<200 \mu\text{m}$ (Johnston and Wu, 1995); see also for example the simulations of extracellular potentials in (Holt and Koch, 1999). Alternatively, we can consider all neurons recorded within the same electrode (even those in separate microwires) to be neighbors. Here, the distance can be up to several mm (see Chapter 2). Let us consider first those neurons recorded in the same microwire. It should be noted that in several cases we only observed one neuron per microwire (see Chapter 2 and Appendix 1). From the total of 104 category-selective units in the MTL (see Table 3-2), 26 were recorded from microwires in which we did not observe any other neurons. Therefore, we cannot analyze the question of whether there is any clustering of selective units for those within approximately $200 \mu\text{m}$. The remaining 78 neurons had 89 neighboring units. Assuming as a null hypothesis that there is no clustering of responses, we would expect that the probability of observing a selective unit in those neighboring units is not affected at all by the fact that there was a neighbor unit that was selective. That is, we would expect approximately 11 out of those 89 neurons to be selective. We observed 19 (almost twice as many as expected by chance) of those units to be selective. This would argue that there

could be some degree of clustering. Of course, it is quite possible that there is a much stronger clustering of selective neurons that we have actually been able to estimate. Maybe several of the other neighboring units are also selective and we simply failed to find the right stimulus. A counterargument is that there could be some flaw in the spike sorting procedure. In one extreme scenario, we could imagine that there is only one neuron and our spike sorting mislabels them as two independent units. Then we would observe two selective units in the same microwire but this would simply be an artifact of the spike sorting process. The observation that not all of these neighboring units had the same selectivity shows some evidence against this interpretation. Forty-two percent of the neighboring units showed a different selectivity.

The second definition of "neighboring" considers all the units within the same electrode to be neighbors (i.e., from all the microwires from the same probe; see Chapter 2 for a description of the electrodes). According to this definition, there were 7 units that were selective and there were no other units recorded in the same electrodes. In the remaining cases, there were a total of 217 neighboring units. The expected number of selective units among those given the overall proportion in Table 3-2 would be 26 and the observed value was 29. It is therefore quite unclear that this could be taken as evidence for the existence of some topographical organization among the selective units at the level of different microwires from the same electrode. It should be emphasized that units from separate microwires could be spread apart by as much as a few millimeters; this probably spans several columns of cortical tissue. These values suggest that there is no particular organization of visually selective units that spans several mm in the MTL. Neurons in V1 show yet another level of organization by connecting neurons that have similar selectivity in distant columns by horizontal connections. Whether there are any such connections or not in the MTL is not possible to assess given our data.

We should remark that Tanaka *et al.* searched for the optimum stimulus of the cell by monitoring the activity on-line of a single neuron and moving the electrode. The clinical constraints and off-line nature of multiple neuron recording prevented us from searching for the optimum stimulus. This may also have a strong impact in the responses of neighboring neurons. If a neuron shows a weak selectivity to a particular stimulus or stimulus category, it is possible that an adjacent neuron could show a slightly different

stimulus preference. Perhaps this slight change in preference is enough for us to miss this type of weak selectivity.

Segregation of selectivity in humans is strongly suggested by the neurological literature (see for example Nachson, 1995, Tranel *et al.*, 1990, Warrington and Shallice, 1984, Warrington and Mc Carthy, 1983, Humphreys and Riddoch, 1993, Humphreys, 1996.) There are several descriptions of patients with circumscribed damage to the brain that show very specific deficits. A particular example is the case of prosopagnosia where subjects lose the ability to visually recognize faces (Heilman and Valenstein, 1993, Tranel *et al.*, 1990, Milders and Perrett, 1993, Dixon *et al.*, 1998). This suggests that there could be some modularity in the representation of information required for visual object discrimination in the brain. If selective neurons were organized in a completely random fashion, then one would expect more general visual object recognition deficits that would not be specific to visual or semantic categories of stimuli. The localization resolution for the damaged areas in the neurological studies showing category-specific deficits is much coarser than the level of single neurons. It is also possible that in several of these patients, the lesions occur in areas (such as inferior temporal cortex and fusiform gyrus) that directly project to the neurons we recorded from. Information from segregated and modular structures could thus converge on the MTL.

3.6 Latencies and durations

We were interested in estimating how fast and for how long selective neurons in the MTL responded to their preferred stimuli. The latency and duration were computed from the spike density function (*sdf*). We computed the *sdf* by convolving the spike train with a gaussian of 100 ms fixed width and then averaging over all the repetitions of stimuli that the neuron responded to (Richmond *et al.*, 1990, Sheinberg and Logothetis, 1997). This involved averaging over all stimuli for responsive but non-selective neurons, averaging over all repetitions within the selective category/categories for the category-selective neurons and averaging over all repetitions of the selective stimuli for those neurons that responded to individual stimuli. The spike trains as well as the gaussian

were binned in 1 ms bins before the convolution operation. An example of the *sdf* superimposed on the PSTH is shown in Figure 3-9a. The latency was defined as the first time point at which the *sdf* deviated by more than two standard deviations from the mean firing rate in at least five consecutive bins (first arrow in Figure 3-9a). The end of the response was similarly defined as the first time point at which the *sdf* returned to the mean firing rate within two standard deviations for at least five consecutive bins (second arrow in Figure 3-9a). The response duration was computed by subtracting the latency from the end of the response.

There was no significant difference in the latencies and durations of the responses among the three experiments (one-way ANOVA, $p > 0.15$ and $p > 0.1$ respectively)⁴⁴. Therefore, the data from all experiments were pooled in the following analysis and plots. The latency of the responses for those units that were non-selective or selective to specific categories ranged from 29 ms to 757 ms (208 ± 173 ms, see Figure 3-9b)⁴⁵. There was no overall significant difference in the latencies between those units that were category-selective and those that were non-selective. We observed that there was a significant difference in the latencies among the different recording locations (ANOVA, $p < 0.05$). The latencies in the SMA were significantly shorter than those in the MTL and other locations. The distribution of the latencies for the SMA and MTL are plotted separately in the insets of Figure 3-9b; the mean latency in the SMA was over 50 ms shorter than that in the MTL (167 ± 162 ms and 224 ± 174 ms respectively). Together with the observation that most SMA units were non-selective (see above and Table 3-2), this reinforces our suggestion that the nature of the visual responses in these two areas is quite different. There was no significant difference in the latencies of the units in different regions of the MTL. We also computed the latency of the responses for those units that showed selective responses to individual stimuli but not to a whole category (Figure 3-9c). The mean latency for these was 146 ± 90 ms and was significantly shorter than the

⁴⁴ Just to make sure this is clear, this comparison refers to the visual responses in the presenter stimuli experiment, the visual responses during the visual imagery experiment and the monocular presentation during the flash suppression experiment. The latencies and durations during visual imagery are described in Chapter 4; the ones for the flash period during flash suppression are discussed in Chapter 6.

⁴⁵ No value was removed to compute this average. In comparing with other reports of latencies, it should be noted that several reports show the mean latency after imposing some upper boundary in the data to eliminate some abnormally high latencies. The same is true about the response durations described below.

values for the category-responsive units (t test, $p < 0.01$). As discussed before, most of these units were in the MTL (and none of these were in the SMA, see Table 3-2). The responses to individual stimuli were generally stronger and this makes it easier to detect the onset of the response. The broader selectivity to categories of stimuli with a smaller number of spikes yields a longer and more variable estimation of the latency. The broad response to a group of stimuli is an average of the responses to individual stimuli; while there was no significant difference in the overall spike count in the time intervals of analysis (see above), it is still possible that neurons could distinguish between the stimuli based on a differential latency to the distinct pictures within a category. It is not easy, however, to address this question precisely. It is possible to estimate the latency for each individual picture within each category; in order to perform any serious comparison one would like to estimate the errors in this estimation of the latency; the simplest way to do this would be to compute the variability in the latency for a given picture and compare this value with the variability in the latency for different pictures. It is not realistic, however, to precisely estimate the latency of the response to each individual repetition of a given picture given the large variability in the responses and the low firing rates. Instead, we used a Monte Carlo procedure to attempt to address this question (the analysis is similar to the non-parametric analysis of variance described in Section 3.3). Briefly, we computed the standard deviation in the latencies to different individual stimuli within each category for the actual presentation order and random shuffles of the pictures in each presentation (we used $n=10,000$ random shuffles). There was a significant difference among the latencies to different pictures within a category in only 9% of the category-selective neurons.

The results shown in Figure 3-9 illustrate the average latencies for different types of stimuli. We addressed the question of whether there were any overall differences in the latencies to specific categories of stimuli⁴⁶. There were no overall differences in the response onset latency to the different categories of stimuli (one-way ANOVA, $p > 0.2$). It should be noted that several factors are averaged in this analysis; one of them is the

⁴⁶ It is not simple to address this question in the case of the responses to individual stimuli. For most of the individual stimuli, we did not observe any selective neurons in any of the patients. Some stimuli were repeated (from one patient to another) much more than others. Only in very few occasions did we obtain several neurons in different patients responding to the same stimulus.

location of the recordings. A more precise analysis would require a two-way (or multiple-way) analysis of variance since there could very well be interactions in the data (for example, hippocampal neurons may show shorter latencies for spatial layouts whereas amygdala neurons may show shorter latencies for faces). The problem with these multiple-factor analyses is that we quickly run out of data. As we collect more and more data on the responses of these neurons, we hope to be able to address these questions. We also studied the variability in the time to the first spike instead of the latency as a putative way of encoding information about different visual stimuli; these data are discussed in Section 3.12.6.

The duration of the responses for those units that were non-selective or selective to specific categories ranged from 42 ms to 1190 ms (452 ± 260 ms, see Figure 3-10a). There was no overall significant difference in the durations between those units that were category-selective and those that were non-selective. We did not observe either a significant difference in the response durations among the different recording locations (ANOVA, $p > 0.05$). The mean duration for those units that responded to individual stimuli was 689 ± 199 ms and was significantly longer than the value of 451.55 ± 260 ms for the category-responsive units (t test, $p < 0.01$). As reported for the latency, we did not observe clear differences for the response durations to different individual stimuli within the selective categories. Finally, we did not find a statistically significant difference among the response durations to different categories of stimuli (1-way ANOVA, $p > 0.1$).

3.7 Guessing which stimulus was presented

Our attempts to record the activity of individual neurons during different perceptual or behavioral tasks implicitly assume that the activity of individual neurons is somehow correlated with the perceptual or behavioral state of the subject. By carefully studying the activity of neurons that are directly involved in representing the processes involved in visual object recognition, for example, it should be possible at least in principle to understand the subject's percept. We therefore posed the following question: "Can we guess the stimulus identity or the stimulus category that was presented by

analyzing the firing pattern of a single visually selective neuron?" Given the large number of neurons in the brain, it is usually thought that rather large ensembles of neurons may be required in order to encode any information about stimuli. However, in some cases, investigators have suggested a sparser code where important information can be retrieved from small groups of neurons, even from single neurons. For example, Young *et al.* have suggested that faces are sparsely represented in the macaque inferior temporal cortex (Young and Yamane, 1992) and a similar proposal was put forward by Logothetis upon observing the specific encoding of stimuli by individual neurons in the same area (Logothetis and Pauls, 1995, Logothetis *et al.*, 1995). In motion area V5⁴⁷, the behavior of the monkey in determining the direction of motion of a random dot stimulus display can be quite accurately predicted based on the spike counts of individual neurons (Movshon and Newsome, 1992, Celebrini and Newsome, 1994, Newsome *et al.*, 1989). A careful information-theoretic quantitative analysis pioneered by Bialek and colleagues has revealed that substantial stimulus-specific information can be conveyed by single neurons (Bialek *et al.*, 1991, Rieke *et al.*, 1997, Kreiman *et al.*, 2000c, Wessel *et al.*, 1996).

It is therefore of interest to study whether it would be possible to predict the stimulus that was presented based on the firing activity of a single neuron. We addressed this in a quantitative manner by performing a receiver operating characteristic (ROC) analysis as used in signal detection theory and psychophysics (Green and Swets, 1966, Gabbiani and Koch, 1998). For those units that showed visual selectivity, we addressed the question of how well an ideal observer could estimate whether the preferred stimulus (or preferred stimuli) had been shown or not by observing the firing rate of the unit during the presentation of the pictures. For each visually selective unit, we computed the distribution of firing rates for the preferred stimuli (the ones for which it showed a differential response as discussed above) and the non-preferred stimuli (the remaining stimuli) as shown in Figure 3-11a. From the distribution of firing rates we evaluated, by sliding a threshold T over the whole range of firing rates, the probability of correct detection (P_{CD}) and the probability of false alarm (P_{FA}). Assuming chance performance,

⁴⁷ This area is also called MT (Allman and McGuinness, 1988); I use here the European terminology V5 so that it does not get confused with the areas that we are recording from in the human medial temporal lobe (MTL).

$P_{CD}=P_{FA}$ (Figure 3-11b, dashed line). The departure from the diagonal shows the possibility of discriminating between the preferred and non-preferred stimuli based on the firing rate of the single neuron. The plot of P_{CD} as a function of P_{FA} is called a receiver-operating characteristic curve. The probability of misclassification plotted against the probability of false alarm can be defined as

$$p_{error}(T) = 1/2 P_{FA}(T) + 1/2 (1 - P_{CD}(T)).$$

The overall probability of error, p_e , is then defined as the minimum value of this function (arrow in Figure 3-11c). A value of $p_e = 1/2$ corresponds to chance performance while a value of $p_e = 0$ indicates that it is possible to predict with perfect accuracy based on the number of spikes whether the preferred stimulus was presented or not.

The p_e values ranged from 0.01 to 0.45. There was no significant difference among the different regions in the MTL and therefore the data were pooled. There was a significant difference between those units that were selective to categories and those that responded specifically to individual stimuli. The distribution of p_e values for the former is shown in see Figure 3-11d; the mean value was 0.30 ± 0.08 . The distribution of p_e values for those units selective to individual stimuli is shown in see Figure 3-11e; the mean value was 0.21 ± 0.10 . One possible explanation for the difference in the p_e values for neurons selective to categories or to individual stimuli is that in the latter case we have found a few individual stimuli to which the units are more sharply tuned to whereas in the former case we have not. A neuron may respond to one category, say animals, more strongly than to the other ones. But it may be that if we had shown an image of a particular animal (for example, we did not include photos of owls in our sample), the neuron would have fired even much more strongly and we would have been able to detect an enhanced response to an individual animal. In some cases we found the equivalent of the owl and in some we did not. There may be some complex multidimensional tuning curve (although I do not know what exactly the axes of this tuning curve would be). In some cases we found stimuli that were very close to the peak (or much closer to the peak) and in some cases we found several pictures that were on some of the not so steep slopes. We did not find a significant difference in the p_e values for different types of stimuli. As we have discussed before, it would be interesting to look at the joint distribution of p_e for location and type of stimuli but we easily run out of data for this analysis.

In summary, in spite of the fact that there are several millions of neurons in the human MTL, it is quite striking that we can actually guess with relatively high accuracy whether the preferred stimulus was present or not by observing the firing rate of a single neuron. The probability of error in this analysis can be lower than 0.05 in the best cases. It should be reminded that this only takes into account the small proportion of selective neurons and not the majority of neurons that did not show any selectivity. However, as we have argued, the lack of selectivity in those neurons could simply be due to our failure to find the preferred stimulus within the very small set of stimuli that were presented. By combining the activity of a very small number of neurons it might be possible to predict the presence or absence of specific stimuli with a very high degree of accuracy, suggesting a sparse representation for the visual recognition of objects.

3.8 Change in neuronal response with time

Neurons, in particular those in lower visual areas, show strong adaptation upon maintenance of a constant visual stimulus. It has been argued that several parts of the nervous system are more sensitive to changes in the environment than to steady stimulation. Indeed, we have discussed above that most of the selective neuronal responses that we have observed are transient and do not last for the whole period of presentation of the stimulus. I think this is indeed a crucial issue that has been underestimated in the studies of perception and I will discuss it in more detail in the last Chapter.

A somewhat related question (although potentially very different from a mechanistic point of view) is the change in the response upon repeated but not necessarily consecutive presentations of the same stimulus. One could conceive the existence of neurons specialized in detecting novelty, that would respond exclusively when a particular stimulus is presented for the first time (or the first few times). The response of such a neuron would decay strongly with repeated presentations. One could also speculate about the existence of neurons that signal familiarity. These units would fail to respond to an unknown stimulus and only respond to what was déjà vu. The

activity of such a unit would increase strongly with repeated presentations. Evidence for changes in firing rate upon repeated presentation of identical stimuli has been observed in the more anterior parts of the monkey IT cortex and in the monkey MTL (see (Brown and Aggleton, 2001) and Section 3.12.5). These data have been taken to indicate that neurons in the MTL could be involved in the detection of stimulus familiarity in addition to recognition of visual pictures.

It could be argued that the neuron illustrated in Figure 3-4 constitutes such an example. This hippocampal unit fired selectively to the visual appearance of drawings and photographs of famous (and therefore familiar) people. Debriefing indicated that the subject did recognize the stimuli; the unit responded to all stimuli within this category (it would have been quite interesting to study the responses of the neuron for famous characters that were not recognized by the subject). We did not present pictures of other familiar stimuli such as relatives, or a familiar house or car, etc. Therefore, we cannot truly distinguish between a selective neuronal activation to familiar stimuli in general or to famous people in particular.

The above comments about Figure 3-4 relied on previous presentation of those stimuli (through watching TV, reading newspapers, etc.). To what extent do the neurons adapt to or enhance their response to repeated presentations of the same stimulus (or group of stimuli)? All the data analysis presented above averaged the neuronal activity for successive presentations of the same stimulus or stimulus group without taking into account the order of the presentations that were assumed to be equivalent. To address the possible effects of adaptation or enhancement in the response we studied the change in the firing rate with the repetition number of identical stimuli.

As a first approach, we computed the slope and correlation coefficient for the change in firing rate in the [100;1000) ms with repetition number for those units that showed selective responses⁴⁸. A change in firing rate upon repeated presentation of the preferred stimulus could be simply due to a general overall trend of firing rate with time (because of some artifact due to problems with the equipment, to slight movement of the

⁴⁸ As a first approach we are investigating here a linear change. There is no clear reason why the change in firing rate should follow a linear trend (for example, there could be a steep exponential drop). If there is a large general effect in the population, however, these preliminary results should be apparent in spite of this assumption, justifying a more refined analysis.

electrodes, changes in the extracellular milieu, changes in the neuronal excitability and so on, Brody, 1999). In order to control for possible non-specific effects of time, we also computed the trend for firing rate with time for all the responses. The distribution of the slopes for the category selective units is shown in Figure 3-12a. When we computed the change of firing rate with time for all type of stimuli, we observed a very narrow distribution of slopes centered on 0 (inset). The slopes for the selective stimuli were also distributed around 0 with a slightly larger variability (Figure 3-12b). The same result was true for the correlation coefficients (Figure 3-12c-d). We obtained the same general observations for the individual stimuli but the variation around 0 was even larger. There were some isolated individual cases where there did seem to be a correlation or anti-correlation of firing rate and presentation order. However, we observed as many examples where there was an increase as cases where there was a decrease. Some examples are illustrated in Figure 3-12e-h. Figure 3-12 e and f show two examples of selective neurons (one selective to a category and another one selective to an individual stimulus) where there was a decrease of firing rate with repetition number. This could be due to the neuron being adapted to the stimulus; it is interesting to note that in the last repetitions, the neuronal response is barely distinguishable from baseline. In contrast, in Figure 3-12g and h we observe two examples of neurons that seemed to show an increase in firing rate with repetition number. However, in the overall average we did not observe any clear effect of changes in firing rate with repetition number.

3.9 Bursts and visual responses

As we have showed in the previous Chapter several of the units showed two peaks in the interspike interval distribution, one sharp peak at short interspike interval (ISI) values, and another broad peak at long ISIs. This is characteristic of neurons that fire in bursts. Bursts have been hypothesized to play an important role in the transmission of information in the nervous system given their higher probability of eliciting an EPSP in the postsynaptic neuron and their lower chance of failures (Lisman, 1997, Koch, 1999). Several investigators have shown that in very different systems bursts of spikes can

convey more information than isolated spikes (see for example the work of Reinagel *et al.*, 1999, Lisman, 1997, Gabbiani *et al.*, 1996, Martinez-Conde *et al.*, 2000, Krahe *et al.*, *submitted*). We therefore explored whether bursts of spikes could convey a more efficient or different signal than isolated spikes in our data.

Units were divided into bursting or not bursting according to the criteria described in Chapter 2. We observed that 22% of the category-selective neurons and 19% of the neurons that responded to individual stimuli were bursting units. These percentages are quite similar to the ones obtained in the whole population (see Chapter 2). For this subgroup of bursting units, we computed the probability of misclassification for the preferred stimuli from the ROC analysis for the spikes occurring within bursts as in Section 3.7. A similar analysis in a separate study had shown that bursts perform better at a feature extraction task in pyramidal cells of the weakly electric fish (Krahe *et al.*, *submitted*, Gabbiani *et al.*, 1996). However, within these data we did not observe a statistically significant difference for the probability of misclassification between the entire spike trains or only those spikes that occurred within bursts for either the category-selective units (2-tailed t test $p > 0.1$) or the units selective to individual stimuli (2-tailed t test, $p > 0.1$). In our study of stimulus encoding by pyramidal cells in the weakly electric fish *Eigenmannia*, it is worth mentioning that the difference between bursts and all spikes is clearly significant but in absolute terms it is small, approximately a difference of 5% in p_e . In our data set, several of the recordings had lower *snr* values than in the electric fish data. As we discuss in Chapter 2 and in Appendix 1, bursting is probably one of the most difficult problems for spike sorting of data from single electrodes (the problem is somewhat alleviated with the usage of tetrodes). This is because it may be very hard to distinguish between two spikes from the same neuron occurring within a burst and two spikes from adjacent neurons occurring in close temporal synchrony. We have addressed this problem by considering the refractory period as well as the temporal lag and its variance in the auto-correlogram and cross-correlogram. However, it is possible that some of the second spikes in the burst could be misclassified (this is also true about the third and subsequent spikes in a burst particularly if they show a very different amplitude). It is possible that the fact that we did not detect any changes in p_e when

comparing only those spikes within bursts to isolated spikes is related to this methodological difficulty.

3.10 Neuronal representation: rate codes, timing codes and other codes

There has been much controversy and discussion over the past decade about the type of code that neurons use to represent information. Two widely discussed limits are the rate code (where the only important variable is the count of spikes in windows of several hundreds of ms to seconds) or a timing code (where the precise timing of spikes at the ms level carries relevant information). The issue can be posed quantitatively by asking what are the time windows relevant for information processing in the nervous system in different neural networks. That is, if a neuron responds to a given picture presented during 1000 ms with, say, 10 spikes, does the detailed pattern of the response matter? Figure 3-22 shows three theoretical examples (these are not real data) that illustrate the point. The format is the same as that illustrated in the previous histograms in this Chapter. Following the analysis outlined in the previous sections in this Chapter, there would be no apparent response to the presentation of the "different stimuli." The neuron responds to all three stimuli with approximately 10 spikes in the 1000 ms of stimulus presentation. However, upon closer examination, we can observe in the response to the second (middle) and third (right) stimulus that there are 3 spikes that occur in a rather specific and quite precise pattern during each of the ten repetitions. This is not the case for the first stimulus (left). That is, the neuron could be "encoding" the second stimulus not by the total number of spikes, but using the precise timing of some of the spikes. And the third stimulus could be represented by the same neuron by using a different temporal code. Furthermore, it is possible that some form of combined activity of several neurons could be relevant to encode information. The number of possibilities is immense and fascinating.

I have done some preliminary investigations of some of these questions including the relevance of spike timing in information encoding in the MTL and the capability of information encoding by pairs of neurons. But time and space constraints preclude from expanding on the subject in here. If time permits, I will describe our preliminary observations in the last Appendix. For the moment, we should keep in mind that spike counts are only part of the story.

3.11 Other questions and technical issues

3.11.1 Seizure focus

Given that some of the neurons that we have recorded from were within the seizure focus, it is important to evaluate whether some of the results that we have described could be due to some brain abnormality caused by the seizures. The information recorded during seizures from the depth electrodes was used to localize the seizure focus (see Chapter 2)⁴⁹. Eighty-five percent of recorded neurons were outside the clinically determined zone of seizure onset (that is, either in the other hemisphere or in a different brain area on the same side). Ninety-four percent of responsive neurons were outside the seizure focus. Because we did not observe any differences in the waveforms, firing rates, interspike interval distributions or response properties of neurons inside or outside the seizure focus, all neurons were included in Table 3-2. The percentage of responsive neurons would increase from 20% to 22% if we excluded neurons within the seizure focus.

3.11.2 Eye movements and blinks

We make frequent eye movements of varying magnitudes, from small microsaccades to large saccades spanning several degrees of visual angle. Interestingly, to a large degree, we are not aware of these eye movements. In spite of the fact that the

⁴⁹ UCLA has not yet given me the information corresponding to the seizure focus in 45% of the patients described in this Chapter. Therefore, all the data regarding seizures described in this Section correspond to the 12 patients where I have the clinically determined seizure focus.

visual field changes quite dramatically (luminance temporarily goes to 0 during a blink and the whole scene is rapidly swapped by a new one during the course of a saccade), the whole world seems to remain quite constant for us. In fact, most naïve subjects vastly underestimate the number of blinks or saccades made. Therefore, one could speculate that the neurons whose activity correlates with our perceptions should be quite oblivious to eye movements. While we have not carefully explored the neuronal activity in the MTL during eye movements in our tests, most of the evidence suggests that eye movements could not influence our results. I think it is worth briefly discussing the evidence.

Ringo *et al.* (Ringo *et al.*, 1994) were among the first to explore the problem in the anterior and medial temporal lobe in monkeys. In their experiments, fixation was controlled before task start while recording the activity of temporal lobe neurons (fixation was controlled within a relatively coarse limit of 10°). They defined a saccade as an eye movement that began with a fixation (no movement $> 1^\circ$ in previous 200 ms) and traveled $> 4^\circ$ within 50 ms. Up, down, left and right movements were tested separately. They report that 25% of hippocampal cells, 52% of parahippocampal cells and 23% of IT cells showed saccadic modulation. Interestingly, no correlation was found between those cells with significant visual responsiveness and those with significant saccadic modulation. It should be noted that these were spontaneous saccades. Maybe saccades occurred due to some motivational or attentional change and so did the neuronal response. In this scenario, both saccades and neuronal responses would be correlated (i.e., occur roughly at the same time) but there would be no causal link between the two. In a follow up study they studied the responses to compulsory saccades of 10° to the left, right, up, down of the fixation cross (Sobotka *et al.*, 1997). In this study they report that in the anterior hippocampus, 12.5% of neurons respond to cued saccades and 12.5% (n=1) to spontaneous saccades. In anterior IT, 45% responded to cued saccades and 9% (n=1) to spontaneous saccades while in the parahippocampus, 48% responded to cued and 26% (n=7) to spontaneous saccades. These percentages were computed over the number of units that did not respond to the dot itself. Several of the responses during the saccades were direction specific. In the hippocampus and parahippocampus this may be related to the spatial responses of these units. On the other hand, a recent detailed study of IT found that free viewing has virtually no effect on the IT neuronal representations

that likely support object recognition (DiCarlo and Maunsell, 2000). They further claimed that the mild excitatory and inhibitory neuronal responses reported by Ringo's group could be due to the effects of large eye movements made in complete darkness.

In the studies of Ringo's group, only a small proportion of neurons respond to saccades (only neurons with firing rates > 2 spikes/sec were considered; this left out almost half of their population; therefore, compared to our samples where we analyzed all the data, the percentages should be divided by 2). These responses were most prominent in the parahippocampal gyrus and quite weak in the anterior hippocampus. For example, in the latter work, only one neuron seemed to respond to cued and spontaneous saccades. In those cases where there was a response, the activity was related to large eye movements. The cued saccades were more than 10° ! And the spontaneous ones were around 5° .

In our experiments, care was taken to discard large head or eye movements. The method was admittedly very crude. The experimenter (G.K.) marked each record where a significant movement occurred and this was discarded. It is likely that several small saccades were missed. But most of the large eye movements and blinks were probably detected⁵⁰. Furthermore, in nine subjects we performed a very short and crude eye movement test in which subjects were asked to fixate and direct their eyes to targets located 10° to the right, left, up or down (Figure 3-20a). An example of the neuronal response of a selective neuron during this eye movement task is shown in Figure 3-20b; there was no clear modulation in the activity of this neuron during the cued saccades to any of the four locations. None of the selective neurons showed a differential response to eye movements in this task. The task was admittedly crude. To begin with it was rather short, with at most 10 saccades to each location. While the experimenter verified that the subject was indeed making the saccades, we did not use any quantitative and accurate device to monitor fixation and eye movements.

Ringo's report claimed that there is no interaction between responses to saccades and to visual stimuli. Assuming independence, if 10% of the neurons respond to saccades in the hippocampus and 12% respond selectively to visual stimuli, then less than 2% of

⁵⁰ This does not apply to the flash suppression and binocular rivalry experiments where subjects wear goggles and I cannot clearly see their eyes (see Chapter 6).

the cells would be responding in the two situations. Several of the responses observed by Ringo *et al.* were directionally selective. For these neurons, the only confounding factor would be if the subjects moved their eyes in a specific direction if and only if the selective group of stimuli was presented (for example, if subjects made an eye movement to the right every time a famous face was presented but not upon presentation of other stimuli). All these data strongly support the idea that eye movements do not significantly alter the visual representation in the MTL.

3.11.3 Behavioral task

In most of the data reported here, subjects were instructed to perform a very simple discrimination task by distinguishing those pictures that showed human faces from those that did not (this was done for the presenter stimuli test and the imagery test but not during the flash suppression test; the task in that case is described in detail in Chapter 6). This was designed merely to engage their attention and make sure they had observed the pictures. This raises the concern that some of the selective neuronal firing could be due to the task and not to the visual responses *per se*⁵¹. We directly compared the neuronal activity during presentation of human faces versus other pictures by averaging the responses to all human faces or other stimuli. Thirty-eight units showed a significantly different activity in response to faces or non-faces according to this analysis. Given the biological relevance of faces for humans and other primates, a human face has been hypothesized to represent a very particular type of visual stimulus. Therefore, these responses may not be related to the behavioral discrimination task but to the importance of faces. Furthermore, of those 38 units, 22 were selective to one or two of the three categories of stimuli that included human faces and are reported in Table 3-2. These units were selective to particular types of faces and not to all faces and therefore it is not possible to claim that the response of these neurons was related to the task. Of the remaining 16 units, a post-hoc multiple comparison test also showed differential

⁵¹ Furthermore, faces have been suggested to constitute a very special type of visual stimulus for primates in general and for humans in particular. Several investigators have described the existence of areas in the brain that respond specifically to faces (see for example Tong *et al.*, 2000, Kanwisher *et al.*, 1997, Fried *et al.*, 1997, Rolls, 1984, Leonard *et al.*, 1985, Allison *et al.*, 1994, Adolphs *et al.*, 1994, Desimone, 1991; see however Gauthier *et al.*, 1999, Gauthier and Logothetis, 2000, Gauthier and Nelson, 2001).

responses to different stimuli and therefore it is unlikely that those responses would be related to the task. Responses to faces were also observed in a previous study in the absence of a behavioral task (Fried *et al.*, 1997). Finally, we observed also several units that responded vigorously to faces (or subsets of faces) in the flash suppression experiment (see Chapter 6). In this task, subjects did not have to make a discrimination of human face versus non-face.

We also studied whether there were units that showed a modulation of their activity related to the motor response itself. For this purpose, we performed two additional analyses. First, we computed histograms of all the neuronal responses locked to the time at which the subjects pressed the button by averaging over all stimuli. We then evaluated whether there was a significant difference in the response in the interval 300 ms before and 300 ms after the button was pressed. This interval was chosen so that it would not overlap on average with the periods of visual presentation. Among the MTL units, we found seven neurons (three in the amygdala, one in the hippocampus, two in EC and one in PHG) for which the response before the button press was significantly different from that after the button press and also from that during the baseline interval (Wilcoxon test, $p < 0.05$). One of these neurons was visually responsive but non-selective. We also found four neurons that showed an increased activity in a 600 ms window around the responses compared to baseline (Wilcoxon test, $p < 0.05$). None of these neurons was visually responsive.

3.11.4 Color versus black and white and other inhomogeneities

As described above, pictures were drawn from a large collection and different sources (see Table 3-1 and Section 3.2.1). While they were all presented at the same size, they were not normalized or modified in any other way. The pictures differed in contrast, color, texture and several other variables. To what extent this influences the neuronal responses in the MTL has not been explored to our knowledge. However, in inferior temporal cortex in the macaque brain (particularly in anterior regions within IT), there is a large degree of invariance in the neuronal responses to several different types of physical changes in the visual stimuli. Investigators have determined that, in general, relatively large changes in size, position, contrast and color (Tanaka, 1996, Gross, 1994,

Logothetis and Sheinberg, 1996) lead to only small changes in neuronal activity. Furthermore, Sary *et al.* have shown that IT neurons can respond to a specific shape defined by very different cues such as luminance, motion or texture (Sary *et al.*, 1993). Based on these data, we speculated (and hoped) that we would obtain visual responses without having to fine tune or explore for each picture a large range of colors, sizes, contrasts and so on. The severe time constraints in our experiments preclude us from even trying to present all pictures in a combinatorial array of parameters. If a neuron responded exclusively to a picture of a specific house shown in a particular size, orientation, color, position and so on, we would probably miss its response. This is true about the electrophysiological investigations in monkeys as well. In some sense, therefore, it is easier to identify neuronal responses that show some degree of invariance. Therefore, it is possible that such highly specific neurons that are picky for a particular combination of features do exist and we simply miss it in our time constrained electrophysiological investigations.

From the set of pictures that were presented, one could still ask whether the units responded not to a specific category of stimuli or a specific individual picture but to some physical aspect of the images such as their color, texture or contrast. It is not particularly trivial how to derive a quantitative measure of texture for the collection of inhomogeneous images that we have used. We did compare the responses to all color pictures versus all black and white pictures. Overall, there were 45 units that yielded a significant value in this analysis at the 0.05 level (2-tailed t test and also non-parametric Wilcoxon test). Of these, 24 units were category selective according to the criteria defined above and therefore responded differently to some colored stimuli than to others. Of the remaining 21 units, a post-hoc multiple comparison test also showed differential responses to different stimuli for 15 of them and therefore it is unlikely that those responses would be related to the color of the stimuli *per se*. The ideal test of course would be to present exactly the same stimulus under different physical transformations but it should be reminded that all the data were analyzed off-line and we could not present all stimuli under different possible transformations without severely reducing the stimulus set.

Yet another dimension along which stimuli could be classified is in the general category of “living things” versus “non-living things.” It has been suggested by neurological studies that this could be an important functional dissociation (see for example the work of Forde and Humphreys, 1999, Tranel *et al.*, 1990, Damasio, 1990, Warrington and Shallice, 1984, Warrington and McCarthy, 1994). We directly compared the neuronal responses to stimuli within these two groups, and we found that most of the neurons that showed a selective response to living or non-living things actually showed more specific changes in activity. Thus, selective neurons typically responded to only some categories or individual stimuli within the general group of “living” or “non-living” stimuli. Some of the histograms illustrated in this Chapter demonstrate this point (see for example Figure 3-3 and Figure 3-4).

It is conceivable that selective neurons responded not to one specific simple stimulus feature such as color but to a specific conjunction of features. As we will discuss in Sections 3.12.7 and 3.12.8) one of the explanations that has been suggested for the category-selective responses that we have observed is that stimuli within a category share specific physical or structural attributes. Whether a category is defined by a conjunction of structural features or by a more semantic distinction based on function, experience, emotional significance or other parameters is a fascinating question that deserved further investigation.

3.11.5 Error trials

The task that subjects had to perform required discriminating whether the picture included a human face or not (except for the flash suppression experiment where the task is described in Chapter 6) and was entirely trivial. There was no occlusion, faces were large and easily recognized, no tricks. The number of errors was consequently very small (the percentage correct ranged from 93 to 99). It is in general quite interesting to explore the neuronal activity during error trials. In particular, if the subject made an error, would the neuron also “make a mistake” and fail to increase its firing rate or not? This leads to the question of whether the trial-by-trial neuronal activity directly correlates to the behavioral response or the perception. The distinction between these two possibilities is important. In one case, the subject really perceives the face and makes a mistake in

choosing which button to press. In the other case, the subject actually fails to discriminate the face, that is, there is a “perceptual mistake.” As I said, the task was so simple in our case that I am skeptical to believe that there would be any true perceptual mistake where subjects could confuse a face for a hat or vice-versa (we did not deal with prosopagnosic patients!). Therefore, I suppose most of the mistakes were related to a simple mistake in choosing which button to press. Indeed, in a large percentage of the mistakes, subjects explicitly manifested this immediately after pressing the button. If we restrict our attention to the selective units and their preferred stimuli, there were very few error trials in each experiment to make any quantitative analysis. We directly compared the neuronal responses before the error trials to those in the correct trials by pooling together all the error trials regardless of what the stimulus was. Over the whole populations there were 14 units that showed a statistically significant difference in error and correct trials. None of the visually selective units showed any differential activity during error trials. Since I am arguing that there was no perceptual mistake, the neuronal activity therefore follows the perception and is not dissociated from the percept during the mistakes.

3.11.6 Right versus left

In 1861 Broca described patients with severe impairments in the production of language after they had sustained specific damage to their brains (Broca, 1861). In addition to being one of the first pieces of evidence for cerebral localization, it constitutes one of the best cases to this day where a strong difference can be drawn between the right and left hemispheres. Is there any difference between the right and left hemispheres in terms of the type and properties of the neuronal responses? As we have discussed in Chapter 2, we recorded from more neurons in the right hemisphere than in the left hemisphere (see Table 3-2). Overall, the ratio of the number of units in the right versus that on the left hemispheres was approximately 1.7. This ratio ranged from 0.48 to 8 in different regions (the value of 8 corresponds to the occipital units where the total numbers are very small; see Table 3-2). We are still not sure about the origin of this bias (see Chapter 2 for discussion).

Although there were more category selective units in the right hemisphere, the overall ratio of selective units could be explained by the larger number of recorded units.

After normalizing by the number of units, the overall ratio of category selective units in the right to those in the left was 1.02. Within the MTL, this ratio ranged from 0.41 in the parahippocampal gyrus to 1.93 in the amygdala. The ratio of the number of units selective to individual stimuli in the right to that on the left hemisphere after normalizing was 0.84. This ranged from 0.14 in the PHG to 2.62 in the amygdala. The overall number of responsive but non-selective units in both hemispheres was also similar after normalizing by the number of recorded units (right/left = 0.98). Within the MTL, this ratio was 0.86.

Hemispheric differences could be manifested only for specific types of stimuli. We therefore compared, for each category of stimuli, the number of selective units in the right and left hemispheres. The ratio of selective units in the right to that in the left hemisphere after appropriate normalization was 1.15 for emotional faces, 0.49 for objects, 1.75 for special layouts, 1.3 for animals, 0.39 for cars, 1.92 and 3.51 for drawings and photos of famous faces, 1.56 for foodstuffs and 1.42 for abstract patterns⁵³. It is interesting to observe that in functional imaging evidence some investigators have reported a predominance for the right hemisphere in the response to faces in the fusiform face area (Kanwisher and Moscovitch, 2000, Tong *et al.*, 2000).

Thus, our results do not show a clear overall site and stimulus independent difference between the proportion of responsive units (category selective, selective to individual stimuli or non-selective) after correcting for the number of units recorded in each hemisphere. There is still a larger absolute number of recorded units in the right hemisphere. It should be noted that the selection of the units to be recorded from is done before showing any stimulus in any experiment. Therefore, it is not possible that our experiments or the type of pictures in our experiments would be responsible for this bias⁵². When studying in more detail the hemispheric differences for specific areas or specific types of stimuli⁵³, there does seem to be some differences that could potentially

⁵² Of course, while we scrutinized each channel to choose the set of units to record from, subjects had their eyes open and were doing something. We cannot rule out that this visual (or non-visual) activity (including sometimes watching TV, reading but also talking, etc.) could engage the two hemispheres differentially. Other possible causes for this difference are discussed in Chapter 2.

⁵³ There could very well be some interaction between these two variables (recording location and stimulus category). For example, there seems to be in our preliminary data more neurons responding to famous faces in the right than in the left amygdala. However, we quickly run into very small numbers (even smaller than the ones we are reporting here) when trying to study these interactions. There are still other variables that

be interesting and relevant. For example, both for units responding to categories or individual stimuli, there seems to be a predominance for the right hemisphere in the amygdala and the left hemisphere for the PHG. We observed more units selective to faces (particularly famous faces) in the right than in the left hemispheres. To what extent this is due to random variation or to a true differentiation between the two hemispheres is unclear given the low number of units that we are working with. Hemispheric differences in the human brain (and also in animals) have been suggested based on neurological, EEG and functional imaging studies. The most robust differentiation concerns the predominance of the left hemisphere for language in most right-handed subjects. But differences in visual capabilities and imagery, typically describing predominance in the right hemisphere, have also been suggested. Not all studies agree on this matter, however. As we collect more data, we hope to be able to shed more light on this issue.

3.11.7 Stability of recordings: Monday/Tuesday neurons

In several cases, we have recorded from the same microwire on subsequent days. For the analysis and accurate report of the number of neurons that we recorded from and the ones that were visually selective, we would like to know whether these are the same neurons or not. In other words, what is the exact correspondence between, say, a recording on Monday from a given channel and on Tuesday from the same channel? This turns out to be a complicated question to study quantitatively and carefully. Other than qualitative reports, I do not know of a careful report addressing the problem in the animal literature either. This is probably because at least up to now, several investigators insert an electrode on one day, remove it after the end of the experiments and then start all over the next day. The argument is that the probability of hitting the same neuron is negligibly low. However, several of the modern multi-unit recording approaches involve the chronic implantation of a set of recording electrodes, an idea more akin to what we do in our

could potentially be very important. One of them is whether the subjects are right, left-handed or ambidextrous (as assessed both behaviorally and by the Wada test). For example, there might be more units selective to faces in the right hemisphere of right-hemisphere dominant subjects and in the left hemisphere of left-hemisphere dominant subjects. Again, the problem with adding these variables into a joint analysis (which would be the correct way of analyzing these data) is that we run into very low numbers. Therefore, all the results described in this section should be deemed as a partial look at conditional distributions where we do not know whether variables are independent or not.

patients (see for example Siapas and Wilson, 1998, Nicolelis *et al.*, 1997, Nicolelis *et al.*, 1999, Wilson and McNaughton, 1993). In these cases, the question arises as to how stable the recordings are across days. There is one particular application for which this question may turn out to be crucial. If the electrodes are implanted for clinical reasons, for example, in order to be able to extract neural information, decode it and guide a prosthetic limb, it will become important to understand how stable the recordings and the representations are across time.

I have not yet developed a fully satisfactory quantitative way to address this problem (at least to me), and I hope to explain the difficulties in this Section. The analysis that follows corresponds to approximately 50% of the total data (a total of 447 units). Let x_i represent a waveform within a given cluster from a certain microwire recorded on day 1 (Monday from here on). Let the total number of waveforms recorded on Monday be n_x ($i=1, \dots, n_x$). $x_i \in \mathcal{R}^d$ where d is the number of bins per waveform that depends on the sampling rate and the time length of each waveform. Similarly, let y_i represent a waveform recorded from the same microwire on Tuesday.

Our first observation is that the number of clusters obtained in both days is not necessarily the same. Actually, in 54% of the cases, there was a difference in the number of clusters in both days. For those cases where there was a difference in the number of clusters, the mean difference was 1.4 ± 0.6 clusters. Overall, the mean difference was 0.9 ± 0.4 clusters. Assuming we trust our estimation of the number of clusters and unless we claim that there is a lazy Monday neuron that only wakes up on Tuesday, it is hard to explain this fact without assuming some movement of the electrode or some basic change between one day and the next. It is unclear whether the electrode itself would move or the microwires would move with respect to the electrode. The electrode is sealed to the skull specifically to avoid movements since it could potentially be dangerous for the subject. But this does not necessarily imply that there cannot be a very tiny displacement, say of less than one micron, in the electrode. And perhaps such a small displacement could give rise to a considerable shift in the microwires. The usual rigid body assumption in physics does not seem to apply here. The microwire could easily move with respect to the electrode as well. Unless the microwire is very close to the soma, it does not take much more than a few tens of microns to change the recording completely (see for example

simulations of extracellular recordings at different distances from the soma by Gary Holt, (Holt and Koch, 1999). Indeed, there has been data indicating that electrodes implanted in epileptic patients can move substantially (Wellmer *et al.*, 2000). Elger's group reports that the shift in electrode contact position from implantation to removal is in the range of 2 mm!⁵⁴

One would like to develop a quantitative statistical criterion to accurately determine what is the probability that x_i and y_i actually come from the same neuronal source and the differences are just due to random noise. This is difficult because of the dimensionality curse (Duda and Hart, 1973, Bishop, 1995). The question is complicated even for small multidimensional problems and in our case d was typically at least 20 (2 ms sampled at 10 kHz) and in several cases > 30 . Indeed, solving this problem constitutes an important part of developing an efficient spike-sorting algorithm (see Appendix 1).

We have used a heuristic classification scheme here instead. We compute the mean sources for each day, \bar{x} and \bar{y} . We computed the Euclidian distance between \bar{x} and \bar{y} divided by the estimation of the noise (see Chapter 2 and Appendix 1). The waveforms were considered to be indistinguishable if this normalized distance was less than a threshold of 2. According to this criterion, we obtained that 87% of the comparisons were deemed to be different. One of the main problems with this claim is that this could very simply be due to a change in the impedance of the recordings from one day to the next. A change in the impedance could arise as a consequence of salt deposition in the microwires, gliosis near the electrodes and also electrode shifts. It is also likely that other factors change in addition to the impedance. To begin with, it is only on rare occasions that we go on Tuesday morning and obtain the same beautiful units on the oscilloscope that we had on Monday. We typically spend several hours on Tuesday until we get everything working again⁵⁵. Even in the absence of impedance changes or electrode

⁵⁴ It should be noted, however, that this is not the same type of electrodes that we have used. The above value corresponds to grid electrodes; the electrodes that we have used are described in Chapter 2. The shift in the position of the microwires is likely to be much less. However, even a shift of less than 100 μm could significantly alter the extracellular recordings.

⁵⁵ In addition to these difficulties, it is not uncommon for the different investigators to change the settings in every possible piece of equipment from one day to the next. Incidentally, this is quite different from

movements, it is possible that some other external factors influence the level of noise in the recordings as well.

In order to attempt to model a linear change in impedance, we simply scaled the waveforms on Tuesday. It should be remarked that this additional degree of freedom drops the most important variable in spike sorting. Monkey electrophysiologists have used and still use dual window discriminators to sort spikes for several decades now. Normalizing to equate the heights of the waveforms therefore removes the most fundamental cue for the separation of distinct units recorded in the same microwire. Using Datawave's manual cluster cutting procedure, again, the height turns out to be the first variable that one looks at in order to separate waveforms. We scaled the waveforms on Tuesday so that the Euclidian distance between the two waveforms is minimized:

$(\bar{x} - \alpha\bar{y}) \cdot \bar{y} = 0$. After determining α ($\alpha = \frac{\bar{x} \cdot \bar{y}}{\bar{y} \cdot \bar{y}}$), we computed the normalized Euclidian

difference as before.

This is shown for one example in Figure 3-14. Figure 3-14a illustrates an example of a recording on two consecutive days. The circles and crosses correspond to the measured values while the continuous lines represent a cubic spline interpolation with an upsampling factor of 10. Visual inspection suggests that the neurons have different heights but similar shapes. Upon normalization as indicated above, the two waveforms appear very similar (Figure 3-14b, the continuous line indicates the normalized Monday recording while the symbols show the original Tuesday recording). The inset shows the difference between the two.

This transformation dramatically increases the number of waveforms that are considered to be the same. However, upon closer examination it turns out that many of the clusters from the same microwire recorded on the same day (separated by cluster-cutting) would also turn out to be equivalent after removing the height with the above procedure. It could be argued that there is a specific reason to do this when comparing Monday and Tuesday neurons, namely, possible changes in impedance from one day to the next (see above.) However, this normalization procedure can lead us to the following

several experiments in an electrophysiology lab where an investigator is usually quite reluctant to let other people randomly change all the settings after he spent several hours, days or weeks fine-tuning them.

problem: it turns out that in many cases, cluster 1 from Monday is at the same time equivalent to cluster 1 *and* cluster 2 on Tuesday. That is cluster 1 on Tuesday looks very much like cluster 1 on Monday after a normalization factor > 1 (that is, after normalization the Euclidian distance between the two clusters is less than the established threshold.) However, upon cluster 2 from Tuesday is also similar to cluster 1 from Monday after using a factor < 1 . Upon directly comparing cluster 1 and 2 on Tuesday they appear to be clearly different (it should be noted that there is no reason to apply any normalization upon comparing two clusters recorded in the same experiment.) Therefore, we cannot claim that cluster 1 on Monday is at the same time equivalent to both neurons recorded on Tuesday. This is illustrated in a clear example in Figure 3-14c. In this figure the mean waveforms recorded for all clusters recorded in the same electrode on two consecutive days are shown. It seems clear that we can claim that there are two distinct neurons even though they were recorded from the same microwire on the second day. After normalization, however, the Monday recording appears rather similar to both clusters (with different normalization factors). This is illustrated in Figure 3-14c and Figure 3-14d. In one case, $\alpha > 1$ whereas in the other $\alpha < 1$. In other cases, we obtained quite stable recordings across several days. Two examples of this are shown in Figure 3-14e and f where the waveforms from two units recorded on three different days are shown. Without any normalization the waveforms appear to be very similar and cannot be distinguished according to the criteria described above.

In order to quantify roughly the number of units that are the same after normalization, we arbitrarily choose one of the two to solve the problem in the previous paragraph (i.e., we count the number of units in the above paragraph as 2). We systematically chose the one that yielded the smaller distance after normalization. We allowed each comparison between different clusters to have its own separate normalization factor. While there may be some factors that could influence all electrodes simultaneously, a single value of α for all the channels would have yielded rather inaccurate results. Alternatively, it could be assumed that there is a separate normalization factor per electrode (common to all clusters within that channel). We tested this in a few cases and the results appear very similar. With this criterion, we observed

that 67% of the units that were being compared turned out to be the same (this is to be compared to the 13% obtained previously without normalization).

In summary, we have described three different procedures for comparison of recordings across days: (1) if two neurons are recorded from the same microwire on two consecutive days, they are assumed to correspond to the same neuron; (2) if the Euclidian difference between the two waveforms after allowing for an additional degree of freedom that linearly multiplies one of the waveforms to account for impedance changes is less than threshold, the two neurons are assumed to be the same; (3) if the Euclidian difference between the two waveforms is less than a threshold, the two neurons are assumed to correspond to the same unit (no normalization). The order in which they are presented indicates the order of how conservative they are to postulate that the neuron is the same. Throughout the text we have opted for criterion (3). We directly compared these three possibilities in a subsample of 447 units. The total number of units recorded from the same microwire on separate days was 148 (33%). If we decide to lump these 148 units that are "counted twice," we are left with 299 units that were not recorded on different days on the same microwire (criterion (1).) The total number of units assumed to be the same with the Euclidian comparison allowing an additional degree of freedom for scaling was 99; the total number of units after lumping together these 99 units would be 348 (criterion (2).) Of the 148 units, the total number of units assumed to be the same with the Euclidian comparison but with no scaling was 20 yielding a total of 427 units (criterion (3).)

Perhaps a more important question is what happens with the selectivity of the units from one day to the next. However, in our case, we did not present exactly the same set of pictures on subsequent days. Therefore, several of the changes in selectivity to be described below could be easily explained by the changes in the stimulus. There were 61 category-selective units in this sub-sample of the data. From this, there were, according to the scaled Euclidian comparison above, 36 that were not recorded from on two separate days. Of the remaining 25 units, 8 were not selective on Monday, but they were selective on Tuesday. Nine units were selective on Monday but not on Tuesday. Six units showed the same selectivity on Monday and Tuesday. Four units had different selectivity on Monday and Tuesday. In principle, even if we were absolutely sure by an independent

measure that we were recording from the same neuron on two separate days, it is not necessarily true that the selectivity should remain the same. One would assume *a priori* that a neuron that represents faces on Monday would still represent faces on Tuesday. To my knowledge, this has not been proved clearly in monkeys in the temporal cortex (Nicollelis, while recording in a different brain area, has claimed that some neurons retain some of their properties for months, but it is unclear how to extrapolate from these results to the visual selectivity in the human MTL). An example of a unit that showed very similar selective responses on two different recording days is shown in Figure 3-20. This unit was located in the amygdala and increased its firing rate selectively upon presentation of four different faces. In this case, the same four faces were presented among several other pictures in two separate experiments (with approximately 24 hours difference between the two) and the neuron responded selectively only to these four faces within the set of stimuli that were presented.

3.11.8 Variability across subjects

All the data shown in this and all other Chapters of this thesis (with the exception of the specific examples, of course) correspond to averages across subjects. We have observed large changes from one subject to another. Although one may expect to observe in the long run statistically somewhat comparable results from one subject to the next, it is difficult to perform a detailed comparison across subjects. Let us take the proportion of selective units for the sake of discussion (similar difficulties and comments apply to other variables). In some subjects we did not observe a single selective unit whereas in other subjects there were many (the values shown in Table 3-2 are the total values for 22 subjects). But there was also a large variation even in the number of recorded units from one subject to another (see Chapter 2).

Human beings are fortunately very different from one another, and these differences are likely to be reflected in the properties of their neurons and neuronal networks. However, one could expect that the photoreceptor cells in the retinae of different subjects with normal vision would behave more or less similarly. It is still unclear what kind of individual differences could be expected at the neuronal level in the MTL of different subjects. But several other changes occur from one patient to another

that make these comparisons extremely difficult. First of all, although in several cases we are recording from similar structures, most likely we are recording from very different neurons. For example, it is quite possible that what we call hippocampus in one subject corresponds to a pyramidal CA1 cell in one subject and a granule cell in the dentate gyrus in another. As discussed above, we also recorded from different hemispheres. Even in monkey electrophysiological experiments where it is possible to control somewhat more accurately the general area of recording, there are cases where important variations are reported between different individual animals⁵⁶. The electrodes themselves could also be slightly different from one patient to another (there can be up to a two- to four- fold change in the impedances for example). The subject's age, gender, educational background and several other variables differ as well. In Chapter 2, we argued that we did not observe clear differences in the proportion of units of subjects with different ages, gender, IQ and so on, but these variables could still contribute to differences in the selectivity of the neurons. The focus of the seizures can also vary from one subject to the next but again it is not clear whether this would influence the selectivity of the neurons, particularly because most of our data come from outside the seizure focus. It has been suggested that the subject's experience could modify the size of a brain area. For example, pianists could have an expanded representation of the brain areas responsible for fine finger movements. Would a subject who is an expert in cars have more neurons responding to cars and a subject that watches TV continuously have a larger representation for famous actors? While this is highly oversimplified, these are interesting but difficult questions given all the important changes that occur from one patient to another.

3.12 Summary and discussion

We have found individual neurons in the MTL in the human brain that show selective responses to complex visual stimuli. The neurons were very diverse in their

⁵⁶ Even in electrophysiological experiments where it is possible to record from individually identifiable neurons, there are some variations (although not nearly as dramatic as what we observe) from one subject to the next.

preference for different types of visual stimuli in all the areas within the medial temporal lobe that we have explored, particularly in the entorhinal cortex and amygdala. Increasingly complex stimulus attributes are represented from the retina to the higher visual areas. Evidence from neurology, functional brain imaging and evoked-potential studies in humans, and from single-neuron electrophysiology and lesions in monkeys, suggests a fundamental role for the medial temporal lobe in the recognition and storage of visual information.

3.12.1 Lesions and neurological studies

Object recognition deficits have been extensively described in the neurological literature. A large fraction of these deficits are very specific to particular groups or categories of stimuli. One particular well-known case is prosopagnosia in which patients show deficits in the visual recognition of faces (Tranel *et al.*, 1990, Dixon *et al.*, 1998, Nachson, 1995, Milders and Perrett, 1993, Heilman and Valenstein, 1993). The deficit can be quite striking as illustrated by Oliver Sacks in his book “The man who mistook his wife for a hat” (Sacks, 1998). Prosopagnosic patients can typically recognize people by other traits such as their gait or voice but not by visual inspection of their faces. Recognition deficits are not restricted to the identification of faces. Nielsen originally reported on the examination of several patients who showed a differential impairment for recognition of living things compared to non-living things (Forde and Humphreys, 1999); this was later observed in several other cases and reported in the seminal work of Warrington and McCarthy (Warrington and Shallice, 1984). Warrington and McCarthy also described two patients who had a much more severe impairment for recognition of “objects” compared to foodstuffs or living things (Warrington and Mc. Carthy, 1987, Warrington and Mc Carthy, 1983). Warrington and Shallice suggested a double dissociation between the two groups by reporting about 4 patients who had exactly the opposite type of impairment, with an almost intact ability for objects and incapable of recognizing foodstuffs and living things (Warrington and Shallice, 1984). While the number of clinical cases is too vast to attempt to give an overview here, it is important to emphasize that these category-specific impairments do not seem to be due to “low-level” visual impairments. Evidence for this comes from the observations that the deficits can be

quite specific, subjects perform normally on several other visual/perceptual processing tasks such as matching objects presented in different views. Interestingly, in some but not all of these cases, subjects are unable to access information about the poorly recognized stimuli from other modalities in addition to vision.

The localization resolution from several of these neurological studies is unfortunately quite coarse⁵⁷. Furthermore, it should be noted that the damage does not necessarily respect anatomical or physiological boundaries. But in several cases these lesions seem to occur in areas that project directly to the MTL structures that we have studied. Specific recognition deficits occur also in some cases after damage in the MTL (Adolphs *et al.*, 1994, LeDoux, 2000, Buffalo *et al.*, 1998). Another case worth discussing is that of patients with Alzheimer's disease. This is a prototypical neurodegenerative disease characterized by several structural abnormalities in the brain, particularly in the MTL (including the entorhinal cortex, the amygdala and the hippocampus). The dysfunction and neuronal death is associated with cytoskeletal abnormalities. The most notable deficits in people with Alzheimer include memory problems. Some patients with severe Alzheimer also develop other cognitive deficits and dementia. Spatial orientation problems are not uncommon in people with Alzheimer. This constitutes yet another link for the role of the hippocampus and adjacent areas in memory processes (particularly anterograde associative memory and spatial memory).

What defines the boundaries between categories and therefore the type of deficits that are observed is still unclear. Several alternative variables have been suggested to play a crucial role in accounting for the type of selective impairments that are observed in these patients. Since one of the most commonly reported deficits involves the dissociation between living and non-living things, it has been proposed that there is a separation given by whether functional or sensory information is used to represent a stimulus. Assuming within the nine categories of stimuli that we have presented that faces, animals and foodstuffs correspond to "living things" whereas the others do not, we directly compared the neuronal responses between these two groups and we observed that this could not account for the responses. Neurons responded in a more specific manner to

⁵⁷ Several of these studies were done before the advent of structural MR imaging and therefore only CT information is available (and in some cases the CT is not provided either).

certain categories or stimuli and not to all living or non-living things (see Section 3.11.4). Another variable that has been suggested to play an important role is the extent to which a stimulus can be easily decomposed into several parts and this is related to the degree of complexity and difficulty in recognizing particular groups of image. According to this view, much more visual processing and information is required to distinguish between two faces than to distinguish between a television set and a sofa. Members of a category of stimuli can share specific sets of physical attributes that group them together. Thus, structural similarity has been suggested to account at least partly for the category-recognition deficits. Physical similarity among stimuli within a category could also at least partly account for some of the selective responses to categories of stimuli. In other cases, however, it is not entirely obvious that the different stimuli that a neuron responded to shared more physical features than stimuli from different categories. Finally, the familiarity with a particular stimulus or category of stimuli, the frequency with which the stimulus is encountered, the age of acquisition of the object's name and function could also have some weight in the neuronal representation and therefore on how easily a subject could lose the capacity for recognition. Recent neuropsychological experiments have accounted for these variables by using regression analysis methods or by carefully designing and balancing the type of stimuli that were presented. While several of these variables can be important for object recognition it seems clear that any single one of these variables cannot account for all the wide range of deficits that have been observed (Warrington and Shallice, 1984, Humphreys and Riddoch, 1993, Forde and Humphreys, 1999, Damasio, 1990, Tranel *et al.*, 1990).

In monkeys, there have been several studies that show severe deficits in object recognition upon damage in the anterior and posterior inferior temporal cortical areas and also in the medial temporal lobe (Meunier *et al.*, 1996, Murray, 1996, Buckley and Gaffan, 1997, Meunier *et al.*, 1993, Buckley and Gaffan, 1998, Kluver and Bucy, 1939, Gross, 1994, Mishkin, 1982). Bilateral lesions to the inferior temporal cortex in monkeys have been shown to produce significant impairments in the capacity to visually recognize complex objects (Mishkin, 1982, Tanaka, 1993, Gross, 1994). In contrast to lesions in earlier visual areas, IT ablations leave intact several “basic” visual capabilities such as sensitivity to light intensity, contrast or motion. In the MTL, it was originally thought that

combined lesions in the hippocampus and amygdala could lead to visual recognition deficits (Mishkin, 1978). However, it was later shown that damage to the rhinal cortex and not the amygdala and hippocampus was responsible for the object recognition impairments (Murray and Mishkin, 1998, Meunier *et al.*, 1993, Meunier *et al.*, 1996). Within the rhinal cortex, damage to the entorhinal area alone seems to produce significant but milder visual deficits than ablation of the perirhinal cortex (Meunier *et al.*, 1993). Lesions to the rhinal cortex mediate recognition in more than one sensory modality. Amygdalectomy produces a significant disruption of food preference learning and both learned and unconditioned fear responses (Gaffan, 1994, Murray, 1996, LeDoux, 2000). Animals with damage to the hippocampus can have an intact immediate memory and abnormally rapid forgetting. These animals show a temporally graded memory loss similar to that observed in human patients where memories acquired shortly before damage are lost whereas information acquired much earlier is spared. A typical behavior where rats with lesions in the hippocampus fail is the Morris water maze task. Here, rats or mice learn to avoid the water and swimming by finding a platform in a pool that is located just beneath the surface. Training involves different starting positions and animals with hippocampal lesions fail in this spatial task (Morris *et al.*, 1982). Hippocampal lesions also impair other type of associative declarative memory tasks (Eichenbaum, 1997, Eichenbaum, 1996, Murray, 1996).

While it is difficult to directly compare the results of neurological or lesion studies with our recordings from single neurons given the different spatial and temporal resolutions, our brief overview of these investigations shows that the deficits that are observed are compatible with an important role of the MTL in the identification and storage of visual information.

3.12.2 Comparison with fMRI and ERP studies in humans

Several functional imaging studies show that the activation of different areas in the human temporal lobe correlates with subjects' observing pictures belonging to different categories of stimuli. In particular, areas specialized for faces, spatial layouts, objects and animals have been described in the literature. In one of these experiments, naming pictures of animals and tools was associated with bilateral activation of the

ventral temporal lobes and Broca's area. While this could be associated with the language requirement of the task, it is of interest to note that some of the earliest stages of visual processing including the left medial occipital lobe were also activated selectively during naming of animals while naming tools was associated with an area in the left middle temporal gyrus (Martin *et al.*, 1996). Ishai *et al.* (Ishai *et al.*, 1999, Ishai *et al.*, 2000) have observed that several areas of the visual extrastriate cortex are activated to a larger or lesser degree during the visual identification of objects. Based on this evidence they have suggested a distributed representation of objects across the visual extrastriate cortex (Ishai *et al.*, 2000, Haxby *et al.*, 2000). In contrast, other functional imaging reports suggest a more local representation. For example, the fusiform gyrus in the human brain has been suggested to be specialized to represent information about faces (Tong *et al.*, 2000, Kanwisher *et al.*, 1997, Kanwisher and Moscovitch, 2000, Kanwisher, 2000, Treisman and Kanwisher, 1998). Another area that has been described to be specialized in the visual recognition of specific stimuli is the parahippocampal area which is activated when subjects observed pictures of spatial layouts similar to the ones used in our study (Epstein and Kanwisher, 1998); some investigators have suggested that this activation is more related to the explicit or implicit navigational requirements of the tasks involved (Aguirre *et al.*, 1996, Aguirre and D'Esposito, 1997, Maguire *et al.*, 1998; see however Epstein *et al.*, 1999). Spatial-topographical processing can activate not only the parahippocampal cortices but also occipito-parietal regions (Gorno-Tempini *et al.*, 2000).

Specific changes in activity on presentation of faces, objects and letter strings has also been observed in evoked potentials studies in humans (Puce *et al.*, 1999, Allison *et al.*, 1999, McCarthy *et al.*, 1999, Allison *et al.*, 1994). Allison *et al.* (1999a,b,c) performed a very extensive study of 98 epileptic patients in whom electrodes were placed directly upon the cortical surface. They used a strip of electrodes that were placed subdurally. They observed short-latency evoked response potentials (ERPs) in striate and extrastriate areas and a series of longer-latency face specific responses. A lateral surface negative peak with a latency of approximately 200 ms was recorded from the middle temporal gyrus and longer latency potentials were observed in posterior ventral occipitotemporal and both posterior and anterior lateral temporal areas. Specific responses to object and gratings were also observed in extrastriate areas. It is somewhat

hard to attempt to directly compare these results with ours given the large differences in spatial resolution (single neuron versus several mm^2 of cortex, see Chapter 2 and Logothetis *et al.*, 2001) and the extent of the recordings (limited to the MTL in our case versus a more global and comprehensive coverage in several of these other studies). However, it is interesting to note the general agreement in the latencies of the responses and the existence of responses to complex stimuli in the MTL in both cases. Interestingly, after characterizing some of the face-responsive areas in a previous study, electrical stimulation in the fusiform and inferior temporal gyri produced a temporary inability to name familiar faces (Allison *et al.*, 1994).

The resolution of functional imaging and electrical evoked potentials does not allow us to address the nature of the detailed neuronal representation of objects in different areas of the visual system. While we and others (Heit *et al.*, 1988) have observed the existence of single neurons that show selectivity to individual stimuli in the human brain, it is still unclear how these give rise to the type of signals observed in the more global types of studies described in this Section. Single neuron changes in neuronal activity upon presentation of faces have been described in epileptic patients before (Heit *et al.*, 1988, Ojemann *et al.*, 1992, Fried *et al.*, 1997). The studies of Heit *et al.* and Ojemann *et al.* differ from the studies that we have done in that theirs were conducted during surgery. I have argued about the difficulties and constraints of single neuron recordings in human patients (see Chapter 2). Recording while patients are undergoing neurosurgery evidently imposes even more severe limitations; this makes the amounts of data even more limited but the attempts are truly heroic. Selective responses were observed not only to faces but also to words (Heit *et al.*, 1988, Ojemann and J., 1999). Holmes *et al.* also observed activity in the right lateral temporal cortex correlated with short-term visuo-spatial memory processes and perception (Holmes *et al.*, 1996).

3.12.3 Comparison with monkey electrophysiological experiments

Single inferotemporal cortex (IT) neurons in monkeys respond to complex visual stimuli including faces, objects and abstract patterns (Gross, 1994, Desimone *et al.*, 1984, Perrett *et al.*, 1982, Rolls *et al.*, 1982). Interestingly, only a small fraction of the neurons explored neurophysiologically typically respond selectively to these and other complex

visual stimuli. This is also true in our data set; most of the neurons that we have recorded from did not seem to respond selectively to any of the stimuli that we have presented (see Section 3.4). This could be due to several reasons and it is interesting to discuss some of the possibilities. One possible explanation for this observation is that we simply failed to find the right stimulus for these neurons. As we discuss in this Chapter, the number of possible different visual stimuli is simply immense (one could even claim that such a space is not finite if we consider all possible variations of colors, shapes, angles, and combinations of different stimuli; but some units, particularly in these high visual areas, can show strong invariance properties to some of these variables). Even in the most heroic experiments we only end up exploring a very small fraction of such a possible universe. In the best cases, we have presented up to one hundred stimuli. Maybe one of the neurons that did not respond selectively would have fired very strongly if we had shown the subject a picture of a pair of shoes (we did not). From this standpoint, it is actually more astounding and miraculous that monkey electrophysiologists and we actually observe any selective responses than the great majority of cases in which we do not. The probability of finding a selective response in this approach in which we randomly present a series of complex stimuli probably depends on the width of the tuning curve of neurons for different stimuli, an issue that I will discuss later. It should also be noted that we are looking at a very simple and gross measure of neuronal response, that is, overall changes in spike count in large time windows. Maybe neurons did respond selectively in some cases and we simply did not read the information correctly (see for example Figure 3-22). Maybe in addition to not knowing how to decode the information from a single neuron in the most accurate way, the representation actually relies on the combined activity of several neurons. Thus, it is possible that several units that we labeled as non-selective actually gave very strong responses when looked at in the right way. An alternative explanation is that there are indeed neurons that do not show visual responses in the MTL. Extracellular recordings typically monitor the activity of different types of neurons; there are several distinct types of neurons based on morphological studies and investigations of specific molecular markers (Ziringer *et al.*, 2001, Kandel *et al.*, 2000). Maybe some of them show visually selective responses whereas others play modulatory roles and they would not respond in a specific way to any type of visual

stimuli even if we had the opportunity of showing millions of pictures. It may be conceivable that pyramidal cells show visually selective responses whereas interneurons do not (and these are just two examples of the large variety of neuronal types that one may record from). Perhaps in the MTL, some neurons are involved in representing information from other modalities like audition. Or perhaps some neurons respond selectively to the integration or conjunction of specific visual and auditory cues but not to either one alone.

Information from the neocortical neurons in the inferior temporal cortex of macaques is conveyed to polymodal association areas in the medial temporal lobe⁵⁸. Compared to IT, less work has been done in monkeys to explore the visual properties of neurons in the entorhinal cortex, hippocampus, parahippocampal gyrus or amygdala. Overall, approximately 11% of the neurons that we recorded from in the MTL showed a visually selective response to categories of stimuli and approximately 6% showed selectivity to one or more individual stimuli but not to a category. These percentages are comparable to those reported in monkeys. In the entorhinal cortex, 11% of the neurons show selective visual responses (Suzuki *et al.*, 1997), compared to 14% in our study. It is interesting to note that in the work of Suzuki *et al.*, each cell was only presented initially with 6 to 18 stimuli and then 6 of those were selected for a delayed match to sample task. Thus, the above percentage of selective neurons is based on only six different pictures. Suzuki *et al.* observed an approximately equal number of selective and non-selective neurons; this also agrees with our observations. Some of the neurons also responded specifically to a given cued location in a place memory task in the study of Suzuki *et al.* (we did not test this in our case), and the authors therefore suggested that entorhinal cortex neurons process information both about objects and locations.

In the monkey amygdala, out of 1000 neurons, approximately 12% showed visual responses. About 33% of those were selective for faces (Leonard *et al.*, 1985, Rolls, 1984). Although several neurons responded to other non-face stimuli, the main goal of the study of Rolls' group was to explore the activity during recognition of faces. Most of these neurons did not typically respond to only one face but to several or all of them. In

⁵⁸ Information is sent to frontal cortex and feedback is provided to lower visual areas as well. A detailed description of the anatomical connectivity of area IT is given in the reviews of Tanaka (Tanaka, 1996) and Logothetis (Logothetis and Sheinberg, 1996).

several cases, investigators could observe a difference in the responses to some faces compared to others. These results are also very similar to the ones that we have observed in our recordings in the amygdala both in terms of the proportions of selective units as well as in their selectivity. In our data, although we observed several units selective to either emotional faces or famous faces, the distribution of selective responses over all the categories that we tested could not be discriminated from a uniform distribution for the category-selective units (in other words, there were selective responses to objects, foodstuffs and other stimuli as well). For those amygdala units selective to individual stimuli, more neurons responded to faces than to other stimuli (see the distribution in Figure 3-17.)

Studies of categorical representation at the level of single neurons have also been performed in monkeys but not in the medial temporal lobe to our knowledge. Monkeys can be trained to categorize stimuli (see for example Vogels, 1999a). There is evidence both in inferior temporal cortex as well as in frontal cortex that individual neurons can show changes in the firing rate that are category dependent (Vogels, 1999b, Freedman *et al.*, 2001). The study in IT showed category-specific modulation of single neuron activity and the authors interpreted the results as evidence that a small group of IT neurons is necessary and sufficient to provide accurate categorical information (Thomas *et al.*, 2001). The study of Freedman *et al.* showed evidence that these neuronal responses are due to training in the monkey and can be modified according to experience and re-training. While we did not specifically train our subjects in discriminating any of the categories (nor did we ever tell them that stimuli were drawn from specific categories), people have extensive natural training in their every day lives in separating visual stimuli into different groups. All the categories that we have used fit well into highly trained and easily recognizable groups of stimuli.

We showed that in the relatively small matrix of hippocampus, entorhinal cortex, parahippocampal gyrus and amygdala, there is a remarkable degree of segregation of categories at the level of single neurons. Neurons in these regions show visual object discrimination among at least nine stimulus categories. Based on the firing rate of individual neurons, it was possible to predict with a mean probability of error of 0.30 whether the preferred stimulus category was presented or not and with a mean probability

of error of 0.21 whether the preferred individual stimuli were presented or not (Figure 3-11). This, by itself, shows a striking degree of visually selective firing on a trial-by-trial basis. By combining the activity of multiple neurons, an even higher level of accuracy can be achieved. Such category-specific processing may be important not only in object recognition, but also in the representation and retrieval processes that have been closely linked with the medial temporal lobe (Zola-Morgan and Squire, 1993, Squire and Zola-Morgan, 1991).

3.12.4 Speed of processing in the visual system

Another important piece of information in understanding how visual information is represented is the study of how fast subjects can recognize or discriminate a visual stimulus. The distinction between these two processes is important since categorization or discrimination could involve different neuronal mechanisms and therefore different speeds than recognition of a specific instance of a stimulus. Some psychophysical studies have used rapid serial visual presentation paradigms where unrelated pictures are presented at high rates on the same screen location (Potter and Levy, 1969). Under these conditions, subjects can detect or remember specific images even at presentation rates of up to 72 pictures *per second* (Keysers *et al.*, 2001). Interestingly, a large proportion of neurons in superior temporal sulcus in the monkey still responded selectively to the effective stimulus at these high presentation rates. In rapid two-alternative categorization tasks, human subjects can give a correct response under stringent conditions as quickly as approximately 250 ms. Psychophysical measures of reaction time include both the visual processing speed and the response execution time. Several experiments have therefore estimated the speed of processing visual information in the visual system by using scalp electroencephalographic (EEG) recordings. The lowest boundary set based on these evoked potential responses yielded a value of 150 ms (Thorpe *et al.*, 1996). This was obtained while subjects were performing a “go”/“no go” rapid categorization task for pictures of animals and distractors. EEG data from subdural electrodes has yielded a negative peak at a latency of 200 ms arising from the MTL (McCarthy *et al.*, 1999, Allison *et al.*, 1994).

Response latencies of single units have been studied in several monkey electrophysiology experiments in the visual system. Schmolesky *et al.* performed a comparison of different areas under the same visual stimulation conditions and reported a latency of 82 ± 21 ms in V2, a value approximately 20 ms longer than the ones in V1⁵⁹ (Schmolesky *et al.*, 1998). V4 cells showed latencies of 104 ± 23 ms. They do not report any values for IT areas. There is an approximate 15 to 20 ms delay in the average latency from LGN to V1, V1 to V2 and V2 to V4 paths. Extrapolating from this, one would predict a mean latency of approximately 120 ms for IT neurons. The exact latency of IT responses varies from one study to another but a safe boundary region would be between 100 and 150 ms. For example, Tamura and Tanaka report a latency for the initial neuronal response of approximately 130 ms. Perrett *et al.* report a range of latencies from 80 to 220 ms with a mean of approximately 150 ms (Perrett *et al.*, 1982, Rolls *et al.*, 1982) whereas Leonard *et al.* report values ranging from 70 to 150 ms in the superior temporal sulcus (Leonard *et al.*, 1985). This may depend to some extent on the type of pictures that were presented and the tasks and degree of overtraining in the task. Recent work has shown that the responses of some neurons in IT may show two components, an earlier one with a latency somewhat less than 100 ms and encoding global information and a later component encoding fine information with a latency 50 ms longer.

The MTL shows significantly longer latencies. Most responses in entorhinal cortex show latencies between 100 and 300 ms with a mean of 181 ms (Suzuki *et al.*, 1997). In the amygdala, Leonard *et al.* reported that most of the neurons showed latencies in the 110 to 200 ms range (Leonard *et al.*, 1985, Rolls, 1984) and a separate study showed latencies ranging from 60 to 300 ms (Nakamura *et al.*, 1992). In the hippocampus, Miyashita *et al.* reported latencies ranging from 50 ms to more than 500 ms. They report the mean value only for those neurons with latencies ≤ 260 ms and this is 154 ± 44 ms (Miyashita *et al.*, 1989). Another study showed latencies in the hippocampus ranging from 80 to more than 340 ms (Rolls *et al.*, 1989) with a mean of 168 ms after excluding the values beyond 260 ms. It should be noted that estimating the latency is difficult and inaccurate for neurons that show low firing rates and are only poorly

⁵⁹ Significant differences were observed between different layers in V1 and it is also possible that different neuron types in V2 also show different latencies (Schmolesky *et al.*, 1998).

selective; furthermore, some neurons showed more than one peak in the PSTH. For that reason these studies report the latency for the majority of neurons with stronger responses, but several longer latency values are excluded from the averages. In our case we showed all the data (Figure 3-9) and this included some neuron with very poor responses that had very long latencies and slightly shifted the average towards longer latencies. It should also be considered that latencies in monkeys could be slightly shorter due to overtraining and to the smaller conduction distances. However, by and large the estimations of latency in the monkey MTL are quite close to the ones that we have obtained and show a clear increase compared to the data from IT and STS.

If it were possible to fully identify a stimulus in 100-200 ms, then it is possible that neurons that show latencies above 300 ms may not contribute directly to the perceptual process. Instead, one can hypothesize that those units may be involved in some other process such as storing the information or in the neuronal correlates of consciousness. It should be noted, however, that it is still difficult to put a strict and sharp bound on the amount of time required to fully identify a stimulus. Even if a particular picture can be recognized or remembered at a presentation rate of 72 pictures/sec, this does not mean that 14 ms are enough for recognition (since the process of identification or recognition of a picture could easily occur afterwards during presentation of a separate stimulus). Studying the speed of neuronal processing can yield important information regarding the locus of correlations between neuronal activity and perception.

3.12.5 Recognition, memory and familiarity

A well-known characteristic of the classical anterograde amnesic syndrome shown in the seminal study of patient H.M. (Milner, 1972, Penfield and Milner, 1958) is the loss of recognition memory. Interestingly, electrophysiological data from monkeys performing recognition memory tasks show that some neurons respond differently to subsequent presentations of a visual stimulus that has been seen previously (Brown and Aggleton, 2001, Li *et al.*, 1993). In most of these cases, investigators have reported a reduction in the neuronal response; enhancement is more rare except in specific tasks⁶⁰.

⁶⁰ In one study, for example, enhancements in the neuronal response were observed when the monkeys were rewarded for repetitions of a target stimulus but not of non-target stimuli.

Interestingly, this property seems to be prevalent in anterior temporal cortex, particularly in the perirhinal cortex. But adaptation or reduction of the neuronal response seems to be very rare in IT/TEO and the hippocampus and parahippocampal areas (Riches *et al.*, 1991). While perirhinal neurons can show single-trial learning, short latencies and storage across more than 24 hours, very few hippocampal units changed their activity upon repeated presentation of the same stimulus, these responses show much longer latencies than the ones in perirhinal cortex and typically do not persist for 24 hours (Brown and Aggleton, 2001). Evidence from immediate-early-gene expression imaging in rats has also shown activation by novel stimuli in the perirhinal cortex but not in the hippocampal neurons. These observations seem to be concordant with the data that we have described in Section 3.8 whereby most selective neurons in the MTL do not significantly alter their response upon repeated presentation of the same stimulus. It is interesting to remark that our ability to recognize and identify a given percept is not deteriorated by repeated presentation of the same stimulus. Therefore, the recognition of stimuli, like some of the neurons that we have recorded from do not seem to show significant changes in the response with repeated presentation.

3.12.6 Neural codes

The data that I have presented is based on the assumption that the spike counts integrated over a relatively large window of several hundreds of ms carry some meaning for the nervous system. This is based on several experiments like the ones we have shown where interesting and clear changes in firing rates are observed in single neuron recordings. A large fraction of the animal electrophysiology literature studies the neuronal responses based on this assumption. It should be noted, however, that this is by no means the only possibility of encoding information in the nervous system. Several other possibilities have been proposed but it is beyond the scope of the current work to exhaustively discuss the different alternatives and the evidence for and against these.

It has become apparent in several systems that neurons can be very reliable (Mainen and Sejnowski, 1995, Kreiman *et al.*, 2000c) and that the precise timing of spikes can be very important in carrying information about the stimulus (Rieke *et al.*, 1997, Bialek *et al.*, 1991, Gabbiani and Koch, 1998, Kreiman *et al.*, 2000c, Wessel *et al.*,

1996). One way of characterizing the trial-by-trial variability in the responses to identical stimuli has been the comparison of the spike count variance to the mean spike count (Koch, 1999). This has shown that in several systems the variance is very close to the mean as expected for a Poisson process (Koch, 1999, Gabbiani and Koch, 1998, Shadlen and Newsome, 1994, Shadlen and Newsome, 1998). In contrast, some neurons can show a remarkably low variance, in fact, the variance can be as low as the minimum possible variance for a given spike count (Berry *et al.*, 1997, Berry and Meister, 1998, Bair and Koch, 1996, van Steveninck *et al.*, 1997, Kreiman *et al.*, 2000c). We computed the variance and mean in the spike counts to repeated presentation of the preferred stimuli of the selective neurons. The data, shown in Figure 3-21a-b, show that the values are very close to those obtained in several cortical recordings in monkeys and cats and quite close to what would be expected from a Poisson process. This matches our observations from Chapter 2 (see also Chapter 5), where we described that the coefficients of variation of the interspike interval distribution are very close to 1. The slope of the variance versus mean plot was 1.08 and 1.19 for category-selective units and units selective to individual stimuli respectively.

One elegant and simple proposal in which the timing of spikes could be relevant to encode information is that the time to the first spike carries information about the visual stimuli (vanRullen and Thorpe, *In preparation*, Thorpe *et al.*, 1996). In our data, there is a large variability in the latency to the first spike as shown in Figure 3-21c. Over the whole data set, there was no significant difference in the time to the first spike for each of the different categories of stimuli (Figure 3-21d, one-way ANOVA, $p > 0.25$). We evaluated for each individual unit (regardless of whether it was selective or not with our previous analysis) whether the time to the first spike was sufficient to discriminate between different categories of stimuli and/or different individual stimuli. Over the whole data set of over 1000 neurons, we observed 16 neurons that yielded a significant difference in the latency to the first spike for different categories and 18 neurons that yielded a significant difference for distinct individual stimuli. One such example is shown in Figure 3-21e-f. In spite of the significant difference in latency to first spikes (one-way ANOVA, $p < 0.01$), there is considerable variability in the occurrence of the first spikes. One possibility is that the first wave of activity (represented perhaps by the first

burst of action potentials or the first set of synchronized spikes) rather than the first spike in itself carries information about the visual world (vanRullen and Thorpe, In preparation).

There are several other possibilities in the encoding of information by the timing of spikes and a more detailed analysis of spike timing would be quite interesting (see for example Kreiman *et al.*, 2000c). A theoretical example of how important the timing of spikes could be and what we could miss by counting spikes in large time windows is depicted in Figure 3-22. The data in this figure are from a simulation and do not correspond to empirical data. We show a case where the spike counts do not reveal any differences among the responses of the neuron to three different stimuli. However, upon closer inspection of the raster plots, it becomes evident that there is a striking pattern of spikes that occurs for the second and third stimulus (Figure 3-22, middle and right plots). Furthermore, it is possible to distinguish between the second and third stimulus by the precise timing of spikes within these triplets. The numbers of possibilities are fascinating and immense. If time permits, I will show some preliminary data regarding these questions in one of the appendices to this thesis.

It is very likely that the combined activity of multiple neurons can yield much more information about the world than the firing of an isolated unit. I am being deliberately vague when I say "combined" activity because, again, there are multiple possibilities including population codes and precise synchronous interactions (see for example Krahe *et al.*, submitted, Dan *et al.*, 1998, Abeles, 1991, Fitzpatrick *et al.*, 1997, Warland *et al.*, 1997, deCharms and Zador, 2000, Singer and Gray, 1995, Singer, 1999, Shadlen and Newsome, 1994, Shadlen and Newsome, 1998). If time permits, I will also discuss some studies of our data addressing whether we can extract and decode information from multiple neurons given the empirical constraints in our recording setup.

3.12.7 What do these neurons really “like”?

There are a large number of questions about the selectivity of these neurons that our preliminary study has not addressed. I enclose here all these issues within the general title implied by the first question in this paragraph. For example, let us assume that we show there is a neuron that responds to a picture like that of Clinton as shown in Figure

3-7. Would this neuron respond to any blond person? Any U.S. (ex)-president? Any person who is famous? Any person whose last name starts with “C”? Any presentation of Clinton? The word “Clinton”? Any person who was involved in a “sex” scandal? The emotions that arouse in the subject upon thinking of Clinton? And so on. Just by chance, some of these questions can be partially addressed from our data (for example, this particular neuron did not seem to respond to other U.S. presidents). We could pose a similar set of questions for any of the other responses. For the example illustrated in Figure 3-3 where a neuron responds to some animals: Does the unit respond to large animals? Small animals? Specific kingdoms? Specific conjunctions of animals and colors? Other living non-human species? Objects that resemble animals? Cookies with the shape of an animal? And so on. As briefly mentioned in the introduction, we still do not understand what the relevant variables are in the representation of stimuli in these higher visual areas. All of these questions are fascinating in themselves. But addressing them would require a whole experiment in and by itself (if not a full Ph.D.). It was not the original goal of our tests to address all of these questions. Much more extensive work will be needed before we can provide a detail answer to all of them. Given the severe time constraints that we have in our experiments, it may be better to study some of these questions in detail in recordings in macaques. On the other hand, it is conceivable that this could be an area where cultural differences could influence the results and important distinctions could be found between humans and monkeys. For example, Clinton may not be particularly relevant in the monkey world. Several colleagues have also suggested that some neurons could respond to a more “cognitive” category rather than a visual category, the distinction being that non-visual cues can influence the former but not the latter. For example, it is possible that the Clinton neuron would also respond to Monica Lewinski and/or to the White House. If a neuron responds to a baseball (see for example Chapter 4, Figure 1), it is conceivable that the neuron could also respond to the hotdog or beer that people typically eat while watching baseball. Again, the answer to these questions is that I do not know.

Even after constraining the “natural world” based on specific cognitive categories, or using some constraints based on the frequency content of natural scenes, or smoothness constraints, etc., the total number of stimuli to explore is just too large to

attempt an exhaustive study (in monkeys or in humans). What is needed is a serious systematic study based on a parametric variation of stimuli. Some studies have attempted to do this in the monkey inferior temporal cortex (see for example Yamane *et al.*, 1988). In our data, it is not entirely obvious that neurons are responding to a combination of physical attributes. By physical attributes I am specifically referring to information that can be gathered even by a computer about the picture itself. These include the color, size, texture and many others. As we have argued before, the structural features of the three pictures of Clinton in one of our examples are strikingly different. A combination of cognitive or biological features could also be quite relevant in determining the activity of these neurons. These could include familiarity, function, emotional aspects, biographical aspects and many more. Structural and cognitive attributes need not be represented separately and it is possible that neurons can respond to specific combinations of these features.

Most of the category-selective neurons responded to only a single stimulus category, rather than weakly responding to a large fraction of all stimuli (Figure 3-16). Our data thus support the existence of sparse coding in the medial temporal lobe. A sparsely coded representation has been suggested for information processing in the rodent and primate hippocampus (McNaughton and Nadel, 1990, Alvarez and Squire, 1994, Treves and Rolls, 1994) and for processing of faces and objects in IT (Young and Yamane, 1992, Logothetis *et al.*, 1995). For example, the studies of Logothetis *et al.* have shown that a sparse representation for specific views of paperclips in different shapes could be observed in the monkey inferior temporal cortex after extensive training (Logothetis *et al.*, 1995, Logothetis and Pauls, 1995). A sparse representation implies that it may not be necessary to record the activity of several millions of neurons simultaneously to be able to understand the neuronal representation of objects in the temporal lobe. The probability of misclassification that we have reported is quite poor compared to the behavior and recognition capacity of the subject; but it is quite striking that so much information can be conveyed by the spike count of a single neuron.

Whereas a significant proportion of neurons are selective for faces in the superior temporal sulcus in the monkey, responses in the IT cortex in general are much more varied and strong activation has been reported for different types of stimuli. Selective

responses were also diverse in our sample of MTL units. Category-selective responses to all the different groups of stimuli were found in the entorhinal cortex and amygdala neurons (Figure 3-16). However, in the hippocampus and parahippocampal gyrus, we observed more responses to images showing spatial layouts, including houses, natural scenes and interiors. The rat hippocampus contains place cells that respond selectively to the position of a rat while it is navigating in a maze (O'Keefe and Dostrovsky, 1971, Wilson and McNaughton, 1993, Muller, 1996). Neurons in the monkey hippocampus respond selectively depending on the position of the stimulus in a conditioned spatial response task (Rolls *et al.*, 1989, Miyashita *et al.*, 1989, Rolls *et al.*, 1993). Functional MRI studies report parahippocampal and hippocampal activation associated with navigational tasks as well as while observing images similar to the ones shown in our study (Epstein and Kanwisher, 1998, Aguirre *et al.*, 1996, Aguirre and D'Esposito, 1997). This area is posterior to the medial temporal recording sites in our study, and is likely to project to the hippocampus. In the responses to individual stimuli, there were more units in the amygdala that responded to faces (both Ekman faces and famous faces) than to other types of stimuli.

Data from very different experiments and using distinct techniques are converging to show an important role for the human medial temporal lobe in visual object recognition. Our data establish that single neurons in humans explicitly respond to specific complex stimuli, which may be relevant to the representation and retrieval of visual information.

3.12.8 Categories, exemplars, both or none?

A fascinating and complicated problem in the object recognition research field involves trying to understand the type of models that are compatible with the way we recognize objects and how neurons could implement such models. Categorization is one of the most fundamental cognitive tasks in which the visual system is involved. Presumably, from an evolutionary point of view⁶¹, it is crucial to distinguish at least some basic categories of stimuli. Is this food or is it a poison? Is this a prey or predator? Does

⁶¹ A famous Caltech Professor always says that one should always hold a glass of wine while discussing evolution in these terms.

this way lead home or to the enemy's mouth? Categorization involves a strong degree of invariance to the properties of the stimulus. There are stimuli that are physically very dissimilar (in color, texture, etc.) and belong to the same category (a red apple and a yellow banana) while there may be stimuli that are very similar and belong to different categories (a red tennis ball and a red apple).

One of the underlying assumptions in several models is that there is some stored prototype and neurons would measure some form of distance to this prototype for identification and also for classification. A fundamental question concerns the nature of this prototype. In one scenario, this prototype consists of individual exemplars (for example, one for a particular side view of Clinton, another for a red apple, etc.). Alternatively, there could be a more general prototype (such as a general face prototype). It is not the scope of our work here to discuss the evidence for and against these different models. Stimulus categorization can be achieved from the selective responses of neurons that are not perfectly category-selective (Thomas *et al.*, 2001). Furthermore, Vogels *et al.* present a model with a self-organizing Kohonen map, suggesting that only one processing stage beyond IT is necessary for categorization based on their IT data. Based on this model, both frontal cortex and the MTL constitute excellent potential targets for categorization based on the neuronal responses that have been observed and on the direct input from IT.

To what degree do the two types of analysis that we have performed (category-selective and responses to individual stimuli) uncover two dissociable groups of neurons? As emphasized several times, above, we did not exhaustively explore the space of visual stimuli. It is therefore possible that some of the so-called category-selective neurons actually respond in a much more specific way to some other stimuli that were not presented. For instance, maybe the neuron illustrated in Figure 3-3 shows a general modulation for pictures of animals but also responds very strongly to a photograph of an elephant (that we did not present in that particular experiment) and would actually thus have been able to differentiate between different animals if we had presented the "right" ones. The reverse possibility is also conceivable. Perhaps some of the neurons with a remarkable fine-tuning of specificity such as the one in Figure 3-8 do not respond only to one or a few stimuli but to a whole category of stimuli that we did not present.

3.12.9 Final remarks

Categorization requires the selection between different stimuli that are physically similar but belong to different groups and the abstraction to group stimuli that do not look the same. This is an astounding level of invariance required for a nervous system. Being able to recognize a physically complex merge of features is a daunting task in itself. We still do not understand how selectivity to bars of different orientation in V1 (Livingstone and Hubel, 1988, Wandell, 1995) or simple shapes (Hedge and Van Essen, 2000) can lead to the recognition of a face like Curly's at different sizes, colors or positions such as the responses of several inferotemporal neurons or the neuron in the amygdala that we have presented here.

The ability to respond to a category poses an even greater challenge to the brain. After somehow extracting several specific features, some of that information has to be transformed and much discarded to reach the possibility of generalizing across wildly different pictures in the same group. From a behavioral point of view, this is indeed not a trivial feat and monkeys typically require extensive training. In our tests, we have had the enormous advantage of studying highly trained subjects because categorization is part of our everyday visual recognition capacity. Whether invariant neuronal responses require training or not is a question that still has to be addressed.

This work also opens many new interesting routes for further exploration. As we have emphasized above it would be fascinating to understand what specific combination of structural and cognitive features neurons are selective to. Are some of these indeed "cognitive neurons" that respond to specific conjunctions of information that we have acquired about function, biographical interest, emotional significance, historical relevance and other complex non-visual features of the stimuli?

We have observed fascinating and strongly selective visual responses that suggest the possible existence of a sparse representation of the visual world in the human temporal lobe. The selective responses to complex stimuli in the MTL could be very important both for the identification and recognition of visual stimuli but also for the storage of information for subsequent retrieval.

3.13 Figure legends

Figure 3-1: Sample of stimuli

Sample of the type of stimuli that we have used. Pictures were pseudo-randomly drawn from a relatively large collection of over 1000 pictures (but not all pictures were presented in any one given experiment; see Figure 3-2). Stimuli were drawn from nine different categories including photographs of faces of unknown actors denoting emotional expressions (Ekman faces), photographs of household objects, photographs of spatial layouts including houses and natural scenes, photographs of animals, photographs of cars, black and white drawings of famous people (Bill Clinton in this example), color photographs of famous people (actress Sharon Stone in this example), photographs of foodstuffs and abstract patterns. A description of the stimuli is given in Table 3-1. Further examples of the type of stimuli that we used were given in (Kreiman *et al.*, 2000a) and (Gross, 2000).

Figure 3-2: Distribution of the number of presentations per experiment

The distribution of the number of presented stimuli in each experiment is shown for the visual part of the three experiments described in the current Chapter: presenter stimuli (left), visual imagery (middle) and flash suppression (right). **a)** Total number of presentations per experiment. Bin size = 50. **b)** Number of individual stimuli that were presented more than $n=4$ times. The inset shows the distribution of the total number of individual stimuli that were presented. Bin size = 2. **c)** Number of repetitions per individual stimulus. Bin size = 1. **d)** Total number of presentations in each category. Note that these presentations were not uniformly distributed among all categories (see for example the number of presentations in each category in Figure 3-3). Bin size = 10. **e)** Number of individual stimuli presented per category. Bin size = 2.

Figure 3-3: Selectivity to categories, example 1

(a) Post-stimulus time histogram (PSTH) of the responses of a neuron in the right entorhinal cortex. The rasters and histograms are aligned to the onset of the stimulus. The stimulus was presented between $t = 0$ and $t = 1000$ ms (indicated by dashed vertical lines in each histogram). Responses were averaged for all stimuli within a given category using a bin size of 200 ms (this bin size was used to display the PSTHs but the statistical analysis used a window of 900 ms; see Section 3.3.1). The dashed horizontal line indicates the mean firing rate over the whole experiment (10.3 spikes/s). The category and the number of stimuli presented in each category are indicated at the top of each histogram. The firing rate in the [100;1000) ms interval upon presentation of a picture of an animal was significantly different from that in the -1000 to 0 ms baseline preceding the stimulus onset ($p < 10^{-4}$). The probability of error (p_e) from the ROC analysis (see Section 3.7 and Figure 3-11) was 0.21. There were only five presentations of food items in this experiment, and they were not included in the graph. The neuron did not respond to food items based on these five repetitions. (b) Distribution of firing rates during presentation of pictures of animals; the histogram shows the distribution of mean firing rates during the visual response in each trial, bin size of 1.5 spikes/s. There is no clear sign of bimodality in the distribution. (c) PSTHs showing the responses of this neuron to each individual stimulus within the category of animals. Although the responses vary from one stimulus to another, the neuron responds to all stimuli within this category (comparison with baseline, $p < 0.01$). An analysis of variance comparing the responses to different individual animals did not yield significance ($p > 0.4$). The scale and conventions are the same as in (a). The size in this reproduction may not be large enough for the reader to discern the stimuli; from left to right they correspond to rabbit, snake, tiger, scorpion, eagle, dolphins and deer. The activity of this unit was recorded during the “presenter stimuli” test.

Figure 3-4: Selectivity to categories, example 2

(a) PSTH of the responses of a neuron in the right anterior hippocampus. The notation and conventions are the same as in the previous figure. The firing rate in the [100;1000)

ms interval on presentation of a drawing or a photo of a famous face was significantly different from that in the -1000 to 0 ms baseline preceding stimulus onset ($p < 0.001$). Note that this neuron did not respond to just any face, as it failed to change its activity to the unknown actors depicting emotional expressions (top, left histogram). The mean firing rate over the whole experiment was 3.1 spikes/s, and the p_e was 0.19 . There were only five presentations of food items in this experiment, and they were not included in the graph. The neuron did not respond to food items based on these five repetitions. **(b)** Distribution of firing rates during presentation of famous faces. Histogram distribution of mean firing rate responses in each trial, bin size of 1.5 spikes/s. There is no clear sign of bimodality in the distribution. **(c)** PSTHs showing the responses of this neuron to each individual stimulus within the category of photos of famous faces. Although the responses vary from one stimulus to another, the neuron responds to all stimuli within this category (comparison with baseline, $p < 0.05$). An ANOVA comparing the responses to different individual famous faces did not yield significance ($p > 0.2$). The scale and conventions are the same as in **(a)**. The subject recognized all these pictures according to the posterior debriefing. The size may not be large enough to discern the pictures; from left to right, they correspond to Arafat, ex-president George Bush, Julia Roberts, Bill Clinton, Mick Jagger, Nicole Kidman, Jim Carter, Rod Stewart, Barbara Streisand, O.J. Simpson and Paul McCartney. The activity of this neuron was recorded during the “presenter stimuli” test.

Figure 3-5: Visually responsive but non-selective unit, example in the MTL

PSTH of the responses of a neuron in the left parahippocampal gyrus. The notation and symbols are the same as in the previous figure. The firing rate in the $[100;1000)$ ms interval upon presentation of stimuli from any of the categories was significantly different from that in the -1000 to 0 ms baseline preceding stimulus onset ($p < 0.01$). Although the response to some of the categories (e.g., cars) was slightly weaker than that to others (e.g., animals), the variability in the responses to individual presentations within the categories was as large as that across different categories, yielding a p value in the ANOVA > 0.1 . This neuron was therefore labeled responsive but non-selective. Note that the mean firing rate shown by the horizontal dashed line is computed over the whole

experimental session (including rest intervals and so on.) Thus, at low firing rates, in some cases such as this one, it may not give the best indication of a mean activity and the PSTH during stimulus presentation is not far beyond this mean. The neuron hardly responded during the 1000 ms period before or after stimulus presentation. The activity of this neuron was recorded during the “imagery” experiment (see Chapter 4 for details about this experiment).

Figure 3-6: Neuronal responses outside the medial temporal lobe

PSTH of the responses of a neuron in the right supplementary motor area. The notation and symbols are the same as in the previous figures. The firing rate in the [100;1000) ms interval on presentation of any of the categories was significantly different from that in the -1000 to 0 ms baseline preceding stimulus onset ($p < 10^{-8}$). Although the response to all the categories was not identical, the variability in the responses to individual presentation within the categories was as large as that across different categories, yielding a p value in the ANOVA >0.2 . This neuron was thus labeled as non-selective. Note that the neuron hardly responded before or after stimulus presentation. Note also the rather precise latency of the response with respect to the onset of the visual presentation. The activity of this neuron was recorded during the “presenter stimuli” experiment.

Figure 3-7: Selectivity to individual stimuli, example 1 (Clinton neuron)

Post-stimulus time histogram (PSTH) of the responses of a neuron in the right amygdala. The rasters and histograms are aligned to the onset of the stimuli. Here we show the responses to every individual stimulus that was presented at least four times. The stimulus was presented between $t = 0$ and $t = 1000$ ms (indicated by dashed vertical lines in each histogram.) Bin size = 200 ms. The dashed horizontal line indicates the mean firing rate over the whole experimental session (2.78 spikes/s.) The number of repetitions of each stimulus is indicated at the top right corner of each PSTH. Note that the order shown here is arbitrary; the actual presentation order was completely randomized. The firing rate in the [100;1000) ms interval upon presentation of any of the three different

pictures of Clinton (purple boxes, mean value of 15.1 spikes/s) was significantly different from that in the -1000 to 0 ms baseline preceding the stimulus onset ($p < 10^{-4}$) and significantly different from the responses to any of the other stimuli ($p < 0.01$). The activity of this neuron was recorded during the flash suppression experiment. The three pictures where Clinton was presented are shown enlarged at the left corner. Note that the neuron did not respond to other pictures of famous people (pictures around the Clinton ones). The three pictures of Clinton have very different physical properties. The unit did not respond to other pictures of American presidents (indicated by arrows).

Figure 3-8: Selectivity to individual stimuli, example 2 (Curly neuron)

Post-stimulus time histogram (PSTH) of the responses of a neuron in the right amygdala (in a different patient from the one shown in the previous figure). The format and conventions are the same as in the previous figure. The firing rate in the $[100;1000)$ ms interval upon presentation of the drawing of Curly (one of the characters of “The Three Stooges,” a famous U.S. TV series, mean value of 7.9 spikes/sec) was significantly different from that in the -1000 to 0 ms baseline preceding the stimulus onset (1.7 spikes/s during the baseline, two-tailed t test $p < 10^{-3}$) and also different from the response to the other stimuli ($p < 0.01$). The activity of this neuron was recorded during the flash suppression experiment. The picture of Curly and the neuronal response to its presentation are shown enlarged at the right bottom corner of the figure. Note that the neuron did not respond to other black and white drawings of famous characters.

Figure 3-9: Distribution of neuronal response latencies

(a) Computation of the response latency and duration. The spike density function (*sdf*) was computed by convolving the spike train with a gaussian of 100 ms fixed width and then averaging over all the stimuli the neuron responded to. Here we show a PSTH of the response of a neuron to its selective category (animals in this particular example). The superposed black trace corresponds to the estimation of the *sdf*. The spike train was binned with a 1 ms bin size before the convolution. The stimulus was presented between

$t=0$ and $t=1000$ ms (indicated here by dashed vertical lines). The dashed horizontal line shows the mean response of the neuron over the whole experiment. The dash-dot lines show two standard deviations in the estimation of the baseline response. The latency was defined as the first time point at which the *sdf* exceeded the two standard deviations limit in five consecutive bins of 1 ms width (indicated here by the first arrow). Similarly, the end of the response was defined by the time point at which the *sdf* went back to baseline within two standard deviations for five consecutive bins (indicated here by the second arrow). **(b)** Distribution of the latency of the responses for all units that were category selective or responsive but non-selective (see Table 3-2.) Bin size = 50 ms. The arrow indicates the mean value. The insets show the distribution of latencies for all the responsive units in the MTL (top) and SMA (bottom). **(c)** Distribution of the latency of the responses for all units that were selective to individual stimuli (see Table 3-2.) Bin size = 50 ms. The arrow indicates the mean value. All these units were in the MTL.

Figure 3-10: Distribution of neuronal response durations

(a) Distribution of the duration of the responses for all units that were category selective or responsive but non-selective. Bin size = 50 ms. The arrow indicates the mean value. **(b)** Distribution of the duration of the responses for all units that were selective to individual stimuli. Bin size = 50 ms. The arrow indicates the mean value. All these units were in the MTL.

Figure 3-11: ROC analysis and p_e distribution

(a) Distribution of firing rates for the neuron shown in Figure 3-3 for the preferred category (top, animals) and non-preferred category (bottom, all other categories). Bin size = 1.5 spikes/s. These represent the conditional probability distributions $P(f | stim \in preferred\ category)$ and $P(f | stim \notin preferred\ category)$, where f denotes the firing response computed in the [100;1000) ms interval and $stim$ indicates the stimulus. **(b)** ROC analysis. Each stimulus is classified into the preferred category if the firing rate during the [100;1000) ms interval is above a given threshold T . The probability of

correctly classifying a stimulus (P_{CD}) is plotted as a function of the probability of making a false alarm (P_{FA}). This was calculated for successive values of the threshold T by integrating the tails of the two distributions: $P(f > T | stim \in preferred\ cat.)$ and $P(f > T | stim \notin preferred\ cat.)$. The dashed line indicates chance performance ($P_{CD} = P_{FA}$). The dots indicate the actual data using numerical integration, whereas the continuous line indicates the values after assuming a Gaussian distribution of firing rates. (c) The overall probability of error, $p_{error}(T) = 1/2 P_{FA}(T) + 1/2 (1 - P_{CD}(T))$, is plotted as a function of the probability of false alarm. The classification performance of the neuron based on the firing rate was characterized by the minimum in this curve, p_e (indicated by an arrow, 0.21 in this particular example.) (d) Distribution of p_e for category selective neurons; data from all MTL regions are averaged in this plot. Bin size, 0.025. The arrow indicates the mean value. (e) Distribution of p_e for neurons selective to individual stimuli but not to a category; data from all MTL regions are averaged in this plot. Bin size, 0.025. The arrow indicates the mean value.

Figure 3-12: Change in firing rate upon repeated presentation of stimuli

(a, left) Distribution of the slope of firing rate versus presentation time for the preferred stimuli of category-selective units. Bin size = 0.05 Hz/presentation. The inset shows the corresponding distribution of slopes for all stimuli. (a, right) Mean slope of firing rate versus presentation time for the preferred stimuli of category-selective units in each location (red, amygdala; green, entorhinal cortex; blue, hippocampus; yellow, parahippocampal gyrus). Values are in Hz/presentation. The inset shows the corresponding mean values for all stimuli. (b, left) Distribution of the slope of firing rate versus presentation time for the preferred stimuli of units selective to individual stimuli. Bin size = 0.05 hz/presentation. The inset shows the corresponding distribution of slopes for all stimuli. (b, right) Mean slope of firing rate versus presentation time for the preferred stimuli of category-selective units in each location. Values are in hz/presentation. The inset shows the corresponding mean values for all stimuli. (c, left) Distribution of the correlation coefficient of firing rate versus presentation time for the preferred stimuli of category-selective units. Bin size = 0.05. The inset shows the

corresponding distribution of correlation coefficients for all stimuli. **(c, right)** Mean correlation coefficient of firing rate versus presentation time for the preferred stimuli of category-selective units in each location. The inset shows the corresponding mean values for all stimuli. **(d, left)** Distribution of the correlation coefficient of firing rate versus presentation time for the preferred stimuli of units that responded to individual stimuli. Bin size = 0.05. The inset shows the corresponding distribution of correlation coefficients for all stimuli. **(d, right)** Mean correlation coefficient of firing rate versus presentation time for the preferred stimuli of units selective to individual stimuli in each location. The inset shows the corresponding mean values for all stimuli. **(e)** Example of a category-selective unit that showed a significant decrease in firing rate with presentation time. The activity of the unit during presentation of the preferred stimuli is shown in green while the baseline previous to these presentations is shown in blue. The green and blue dashed lines show the corresponding linear fits to the data. The black horizontal line shows the average activity of the unit over the whole experiment. The cyan dashed line shows the fit for the change in firing rate with time for all presentations. **(f)** Example of a unit selective to an individual stimulus that showed a decrease in firing rate with presentation time. The notations and symbols are the same as in part **(e)**. **(g)** Example of a category-selective unit that showed an increase in activity upon repeated presentations of stimuli from the same category. The notations and conventions are as in part **(e)**. **(h)** Example of a unit selective to an individual stimulus that showed an increase in firing rate upon presentation of repeated occasions of the same picture. The notations and symbols are as in part **(e)**.

Figure 3-13: Non-parametric analysis of variance

This figure illustrates the probability distribution of the statistic t^* from the bootstrap statistical analysis for 1000 iterations over the data from the response to photos of famous faces illustrated in Figure 3-4. The statistic t^* measures the ratio of the within group to the between group variance (see Section 3.3.2). The arrow indicates the observed value for the actual order (\hat{t}). The probability of obtaining such an extreme value is the integral of the curve from the arrow to infinity. In this case $p=0.19$. This argues against the

possibility of discriminating between different photos of famous people based on the firing rate of this neuron.

Figure 3-14: Stability of recordings (Monday and Tuesday neurons)

(a) Average waveforms of two units recorded on two consecutive days from the same microwire. Each trace shows the 2 ms stretch of the action potential that we recorded and analyzed. The y axis indicates the amplitude in A/D mV after amplification (the distribution of amplitudes of waveforms was shown in Chapter 2). The points mark the actual data; the continuous line shows interpolated values after fitting a cubic spline with an upsampling factor of 10. The blue trace corresponds to data from the first recording day; the black trace shows data recorded on the next day. (b) Comparison of the original waveform recorded on day 1 (crosses) with the mean waveform from the second day after allowing for a linear scaling factor (continuous line). The inset shows the difference between the two curves and the dashed line indicates the noise estimation (see Chapter 2 and Appendix 1). (c) Average waveforms of one unit recorded on day 1 (blue circles) and two units recorded in the same microwire on day 2 (black squares and crosses). The continuous line shows values interpolated with a cubic spline as in (a). (d) Comparison of the original recording with cluster 1 of the recording from the second day after normalization by linear scaling (top) and cluster 2 of the recording from the second day after linear scaling (bottom). (e) Example of one unit recorded on three consecutive days from the same microwire. (f) Example of another unit recorded on three consecutive days from the same microwire (corresponding to another cluster from the same microwire illustrated in (e)).

Figure 3-15: Distribution of category-specific responses

Distribution of category-selective neurons for each category in each location. The percentage of neurons selective for each category of stimuli is shown for each location in the medial temporal lobe (red, amygdala; green, entorhinal cortex; white, hippocampus; blue, parahippocampal gyrus). The percentages are based on the total number of category

selective and non-selective units in each location. Data from the right and left hemispheres have been pooled. Note that the values represented here are not normalized in any way for the number of presentations in each category. The criteria to place a unit in one of these graphs are discussed in the text. If a neuron responded to more than one category but not to all of them (see Figure 3-17 for the proportions), it was added in all the groups it was selective to in this figure.

Figure 3-16: Number of categories that neurons were selective to

Percentage of neurons that responded selectively to 1, 2, ..., 5 categories (neurons that responded to more than five categories were considered to be responsive but non-selective; see text for details). The percentage is computed over the total number of category selective units. The criteria to determine whether a neuron was selective or not are described in the text (Section 3.3.1). The great majority of selective units responded to only one category.

Figure 3-17: Distribution of selective responses to individual stimuli per stimulus category and location

Distribution of neurons selective to individual stimuli within the indicated categories. The percentage of neurons selective to individual stimuli in each category is shown for each location within the medial temporal lobe (red, amygdala; green, entorhinal cortex; white, hippocampus; blue, parahippocampal gyrus). The percentages are based on the total number of selective neurons in each location. Data from the right and left hemispheres have been pooled. Note that the values represented here are not normalized in any way for the number of presentations in each category. The criteria to consider a unit to be selective to individual stimuli are discussed in the text (Section 3.3.5). If a neuron responded to individual stimuli from more than one category (see Figure 3-18), it was added in all the corresponding groups in this figure.

Figure 3-18: Number of individual stimuli that neurons were selective to

(a) Percentage of neurons selective to 1, 2, ..., 7 different individual stimuli (there were no units that were selective to more than 6 discrete individual images according to the criteria described in the text). The percentages are based on the total number of neurons selective to individual stimuli. (b) For the neurons that responded to more than one individual stimulus, percentage of the responses that were to stimuli in 1, 2, 3 or 4 categories. For those cases where the neuron responded to two stimuli in two categories, the red bar shows the proportion where one response was to an emotional face and another response was to a famous face.

Figure 3-19: Eye movements

In 40% of the subjects, we performed a very simple eye movement control task to explore whether there was any modulation in the neuronal activity due to eye saccades. The subject was instructed to fixate for 1000 ms on a small cross placed the middle of the screen (a, left) and then make a saccade to a target which was randomly displayed in one of four possible locations to the right, left, above or below the fixation cross (a, right). The target was on for 1000 ms. The experimenter (G.K.) coarsely verified the saccades, but we did not perform any accurate measurement of eye movements. An example of the neuronal activity of the visually selective unit shown in Figure 3-8 during the saccades is illustrated in part b. Histograms are aligned to the time the saccade was instructed by the appearance of the target as in the right of part a (indicated by the dashed vertical line). Bin size = 200 ms. The horizontal line indicates the mean firing rate of the unit over the whole experiment. This unit did not show any differential activity due to saccades in general nor any direction-specific saccade activity.

Figure 3-20: Stability of recordings: visual selectivity across days

Example of a unit in the amygdala that was recorded from on two consecutive days and where some of the same stimuli were presented in the two experiments. The unit responded selectively to four different faces: a black and white photograph of an

unknown actress denoting surprise, a black and white drawing of ex-president Clinton laughing, a color photograph of singer Paul McCartney and a black and white drawing of actor Sylvester Stallone. The left column corresponds to the PSTHs for the activity recorded on day 1, and the right column shows the PSTHs for the responses recorded on the following day (approximately 24 hours of difference between the two tests.) It is unclear why the neuron would respond selectively to these four pictures. Maybe some special movement or position of the mouth is common to all four pictures. Or maybe there is some other common element to these four pictures.

Figure 3-21: Variability in spike counts and first spike times

Spike count variance versus mean spike count for all neurons selective to a category (**a**) or individual stimuli (**b**). The dashed line shows the $y=x$ line. A linear fit to the data yielded a slope of 1.09 and 1.19 and correlation coefficients of 0.92 and 0.89 for (**a**) and (**b**) respectively. Counts were obtained in the [100,1000) ms interval with respect to stimulus presentation (see Section 3.3.1 for discussion). The distribution of the latency from stimulus presentation to the occurrence of the first spike is shown in (**c**) after averaging over all pictures. Bin size = 20 ms. The arrow indicates the mean value. Only those stimulus presentations where a spike occurred before 600 ms were included in this distribution. The corresponding distribution for each category of stimuli is shown in (**d**). The distribution of the time to the first spike in a neuron that showed a significant differential response among stimuli (according to the criteria specified in the text) is shown in (**e**) and (**f**). Part (**e**) corresponds to the distribution of the latency to the first spike for each category of stimuli while the part (**f**) illustrates the distribution of time to first spikes for each individual stimulus that was presented at least four times. Bin size = 20 ms. Arrows indicate the mean values.

Figure 3-22: Spike timing. A glimpse of what we are missing and what's ahead.

Modeled data showing just one of many other possibilities of information encoding that would be overlooked in our analysis of spike counts. In these simulated data, rasters and

PSTHs are shown in the same format as in the previous figures in this Chapter. No strong and evident response can be observed from these histograms and a statistical analysis based on spike counts suggests that there were no differences between the three responses. A more detailed observation, however, shows the occurrence of a rather precise pattern of spikes in the middle and right raster plots but not in the first one. The spikes involved in such a pattern are marked in red in the middle histogram and in green in the right histogram. Note that the precise timing of spikes within this pattern in the middle and right rasters is different and that this is only poorly noticed in the PSTHs due to the background firing rate of the unit.

3.14 Table legends

Table 3-1: Description of stimuli

Name	Description	Color	Examples
Ekman faces	Faces (photos) by unknown actors denoting emotional expressions	B&W	woman smiling; angry man, etc.
Household objects	Common household objects (photos)	B&W / color	CD, cup, baseball, hairbrush
Spatial layouts	Indoors and outdoors pictures of houses and landscapes (photos)	Color	house, mountain, beach, etc.
Animals	Animals (typically one per picture), (photos)	Color	tiger, snake, spider, elephant
Cars	Different types of cars (photos)	Color	yellow 1940s car, red modern 2door car
Drawings of famous faces	Famous actors and characters, typically with exaggerated expressions (drawings)	B&W	Clinton, Michael Jordan, Frank Sinatra
Photographs of famous faces	Famous actors and characters (photos)	Color	Clinton, Paul Mc Cartney, Nicole Kidman
Foodstuffs	Different sorts of foodstuffs (photos)	Color	Eggs, pizza, orange juice
Patterns	Abstract lines and other patterns (photos)	B&W	Vertical grating, quilt-like pattern

Description and example of the type of stimuli used within each of the nine categories that were presented. Figure 3-1 shows a sample of one picture from each category.

Table 3-2: Total number of units and number of selective units

	Medial temporal lobe									Occipital/Parietal					Frontal						
	Amy		Hip		EC		PHG		Total	Occ		Par		Total	AC		Fr		SMA		Total
	L	R	L	R	L	R	L	R		L	R	L	R		L	R	L	R	L	R	
units	77	186	68	145	137	224	43	30	910	1	8	4	13	26	31	15	5	9	19	29	108
non-sel	7	13	9	10	0	14	3	2	58 (6%)	0	1	0	0	1 (4%)	1	2	0	0	9	7	19 (18%)
sel	6	28	14	13	10	24	7	2	104 (11%)	0	1	0	0	1 (4%)	2	0	0	2	2	1	7 (6%)
ind	3	19	3	4	8	11	10	1	59 (6%)	0	0	1	0	1 (4%)	0	1	0	0	0	0	1 (1%)

Total number of units, responsive but non-selective units, category-selective responses and units that responded to individual stimuli. The criteria to define the unit in any of these groups are described in the text. The data are distributed into the Medial temporal lobe, occipital/parietal lobes and frontal lobes. Note that occipital does not mean V1; most of those neurons were in more anterior parts of occipital cortex, presumably in an area analogous to V5 (C. Wilson, personal communication). Within the frontal lobe, the units were located in the anterior cingulate, supplementary motor area or orbitofrontal and frontal cortex (these two pulled together with the label Fr.). Our resolution did not allow us to identify which field in the hippocampus we recorded from, which layers we recorded from in the EC, or whether the microwires were in the basolateral amygdala or other regions within the amygdala (Ziringer *et al.*, 2001). The percentages are based on the total number of units in each region.

4 Visual Imagery

4.1 Introduction

Vivid visual percepts can be voluntarily generated in our minds in the absence of concomitant visual input. While trying to count the number of flowers in Van Gogh's sunflowers, understanding a description or recalling a path, subjects report forming an image in their "mind's eye." The feeling is quite different from the one experienced upon answering a more abstract question such as "What is the name of the capital of the United States?" or "What is the definition of liberty?" We use imagination to solve problems, to navigate, to understand what we read, to create art. Whether this process is accomplished by the same neuronal mechanisms as visual perception has been a matter of debate for a long time. Evidence from functional imaging, psychophysics, neurological studies and monkey electrophysiology suggests a common processing mechanism. However, there are patients with deficits in one but not the other. We directly investigated the neuronal substrates of visual recall by recording from single neurons in the human medial temporal lobe while the subjects were asked to imagine previously viewed images. We found single neurons in the hippocampus, amygdala, entorhinal cortex and parahippocampal gyrus that selectively altered their firing rates depending on the stimulus the subjects were imagining. Of the neurons that fired selectively during both vision and imagery, the majority had identical selectivity. Our study reveals single neuron correlates of volitional visual imagery in humans and suggests a common substrate for the processing of incoming visual information and visual recall. A summary of this chapter has been reported previously (Kreiman *et al.*, 2000b).

4.2 Brief historical remarks

Visual imagery is one of our most astounding and pervasive capacities. In the absence of any input from the world, we can generate a visual percept in our minds. Imagery is not restricted to the visual modality but it is the representation of the visual world that I have been mostly concerned with for my thesis. Tracing the history of ideas and hypothesis regarding visual imagery is by itself a daunting task. I will therefore just limit the description to some of the key historical thoughts. As in many other areas of Neuroscience, complete empiricists could fiercely argue that not much understanding has been gained from these philosophical arguments and counter-arguments⁶². Even if this statement were correct, I still find several of the philosophical ideas about images and imagination quite picturesque and worth a brief discussion.

Plato was perhaps one of the first philosophers to dedicate some serious thought to the matter⁶³. He argued that our impressions are like images painted in our minds (Plato, 1999). Several centuries later, Descartes proposed that perceptions are a matter of the soul but he argued that the soul converses with the body through a special area of the brain called the pineal gland⁶⁴. Certain philosophers have assigned to imagination a central role in the power of the mind to represent any reality. Kant and Hume, for example, appealed to imagination to explain how thoughts were possible. Several influential thinkers including William James (James, 1890) and Sigmund Freud (Freud, 1966) pondered on the relationship between seeing and imagining. The former argued that mental images lack the pungency and tang of the real world. However, an elegant experiment carried out in 1910 by Perky (Perky, 1910) showed that it is not always trivial to discriminate between the two. By projecting faint images onto a screen she managed to

⁶² One particular example related to the questions addressed in the current chapter is the famous and hotly debated issue of whether it is possible or not to imagine a triangle which is neither equilateral, scalenum or isosceles, neither oblique nor rectangular but all and none of these at once (Berkeley, 1709, Honderich, 1995).

⁶³ As always; it would be more precise to say that Plato's ideas are among the first documented records of thoughts about the matter. It is quite likely that many others also debated and had very original ideas which were not set in papyrus. Publish or perish.

⁶⁴ As we have commented on before, in the case of Descartes, one should be cautious in attempting to distinguish his ideas from what he wrote. The latter were heavily monitored by the Inquisition and the consequences of a rejection were much more severe than when Nature sends a manuscript back without review.

show that the generation of a mental image can interfere with seeing faint and fine visual details.

The era of behavioral psychology, however, largely ignored the problem. Watson directly attacked the concept of imagery (Kosslyn, 1994). Imagination was a subjective state not amenable to study experimentally. Some even claimed that there is no such thing as imagination. I would like to ask the reader to engage in a very simple exercise: imagine the letter K, place a square on the left side, and a circle inside the square; finally, rotate the whole figure 90 degrees to the left. What object can you “see”? Or try to explain someone over the phone the directions to get to your house. Chess players typically enjoy impressing novices and friends by playing “blindly,” that is, using only their mind’s eye. It seems evident that imagery constitutes a central form of cognition and there are several examples from even trivial tasks like the ones just illustrated where visual imagination is used. A visual image in the mind in the absence of direct input is formed to solve many every day simple tasks even without explicitly noticing this.

Psychologists started reconsidering the problem of imagination in the last decades of the twentieth century. Roger Shepard and Stephen Kosslyn demonstrated that complex aspects of visual imagery could be studied by using carefully the tools of psychophysical examination. They showed that there is an astounding similarity in the structural information obtained in perception and the one contained in a mental image and that the functional processes that take place during imagination can play an important role in perception (Kosslyn, 1994, Shepard, 1987). An important piece of information came from the work of neuropsychologists Edoardo Bisiach and Claudio Luzzatti by studying patients with neglect syndrome. Upon showing these patients a picture of the Piazza Milano, the subjects could only name the buildings to the right of the point of view and ignored or failed to name those on the left side. Interestingly, when they were asked to close their eyes and imagine being in the Piazza looking in a particular direction they ignored the left side of the world exactly as they did while looking at the pictures (Bisiach and Luzzatti, 1978).

The advent of functional imaging techniques and scalp electrophysiological event related potential (ERP) recordings allowed psychologists and cognitive scientists to pry into the processes inside the brain in order to attempt to more precisely localize the

neurophysiological activity changes that occur during imagination. Accordingly, the last decade of the past millennium has seen also a wealth of new information from careful experiments in delayed match to sample tests in monkeys, neurological studies of patients with imagery deficits in addition to fMRI, psychophysics and ERPs. The results of these experiments will be discussed later in the chapter and compared to what we have observed.

While several of the reports obtained from these techniques suggested the possibility of a common representation for vision and imagery, the debate was still open (Coslett, 1997, Behrmann *et al.*, 1992, Farah, 1988, Kosslyn, 1994). This was due primarily to two observations: (1) Studies of neurological patients had suggested that in most patients, visual deficits were accompanied by concomitant visual imagery deficits. However, a double dissociation was suggested based on the observation of some patients that showed difficulties in one but not the other process. (2) Some (but not all) functional imaging observations had suggested activation of brain areas exclusively during one process or the other in addition to areas activated during both.

We had observed that neurons in the human MTL can respond in a selective manner to complex visual stimuli (see Chapter 3). Given the fundamental role of the MTL in memory processes (Zola-Morgan and Squire, 1993, Squire and Zola-Morgan, 1991) and the observations that electrical stimulation in the human brain could elicit visual percepts or interfere with visual processes (Penfield and Perot, 1963, Penfield and Jasper, 1954, Fried *et al.*, 1982), we conjectured that the activity of these neurons might also be modulated in a very specific manner depending on the type of stimuli that subjects imagined. We had the exciting opportunity of investigating visual imagery at an unprecedented spatial and temporal resolution in human subjects.

4.3 Methods

4.3.1 Tasks

Although the question may seem quite trivial at first glance, there are several possibilities in the design of an experiment to explore the neuronal correlates of visual imagination and compare the neuronal activity during imagery and vision. There are advantages and disadvantages to each of them. Given infinite time, one would wish to compare the results from all or at least several of them. Reality constraints force us to choose one path.

One of the first questions that arise is how to make sure that subjects are indeed forming a visual image in their minds. This is indeed a very difficult question. As we mentioned previously in this chapter, this problem is actually sufficiently complicated and challenging to force the majority of the “hard” scientists to completely abandon the problem of visual imagery. This is clearly one alternative, although one that would not contribute much towards understanding. One way to address this problem is to force the subjects to perform some task that depends crucially on the formation of such images. Some examples are given by the simple tasks proposed in the previous section. Thus, we could directly ask the subject to estimate the number of doors in his house, the number of Van Gogh sunflowers, the eye color of Madonna and so on. We probably would not want to have subjects give a one paragraph response to each question because of possible language confounds and because of time limitations. Therefore, it would be easier to have a yes/no question for each case. One would then change the above questions appropriately. For example, we could ask whether there are more than five doors in his house or not, or whether Madonna’s eye color is green or not, etc. There are a number of drawbacks to this approach. First, there would be several confounds such as the comprehension of the question, the evaluation of the response and the response itself. Second, it could actually be argued to what extent the answer to such questions really does invoke visual imagery more than the tasks we have actually used. Third, since we wish to perform statistics, if we wanted to present each picture five times, we would have to formulate five different non-overlapping questions for each picture (such as, “is the

number of Van Gogh sunflowers more than 10?”, “are most flowers pointing to the left or right?”, “is there any green color in the painting?”, “were there more than four wilted flowers?”, “were the flowers inside a flowerpot?”). Note that it may be complicated to come up with five non-overlapping questions that are not too simple or too complicated. Also, after answering one or two, maybe the subject can answer the next question without forming a visual image. Fourth, it is also possible that the subject could answer some of the questions “by heart,” particularly, after asking several questions on the same picture. Finally, this approach requires much more time than the approach we actually pursued.

The task was divided into two parts: a vision and an imagery task. During vision, two images were separately shown for 1000 ms (Figure 4-1**a-b**). After each picture, a tone reminded the subjects to press a button indicating whether the picture was a human face or not. This was done to engage the subject’s attention and make sure that they had seen the picture. The task itself was trivial as evidenced by the performance above 97% correct. Subsequently, subjects closed their eyes and imagined one or the other picture upon listening to high and low tones alternated every 3000 ms (Figure 4-1**d**). This was repeated for approximately 30 different pairs of images during each session (ranging from 8 to 41), 1-4 sessions per patient, depending on clinical constraints. Imagery was verified by debriefing following each repetition by requesting detailed descriptions of the pictures and asking whether they could form a visual image or not. If subjects reported that they could not visualize the images, the trials were discarded. We also discarded cases in which subjects could not remember the images while describing them. This task depended on the reliability of the subjects to indicate their behavior (overall, less than 5% of the total number of trials were discarded). Each individual image appeared in only one pair. In the first two patients, only faces, objects and spatial layouts were presented. In the remaining patients, stimuli were chosen among the full set of stimulus categories (see Chapter 3, Figure 1 for a sample of the figure and Table 1 for a description of the different categories of stimuli).

The time periods chosen for visual presentation and visual imagery were different as described above. A time interval of 1000 ms was too short for visual imagery. We can shift our gaze at a much faster pace and scenes can change quite rapidly both in the real world and in a film. Yet, when I informally tried asking some colleagues to switch their

mind percept every 1000 ms, everybody reported it was too fast and confusing. On the other hand, a visual presentation period of 3000 ms would have been too boring and would have reduced the number of stimuli that we could present. It does not seem likely that our main conclusions could be related to this difference; however, the differences observed in the latencies and durations could be related to this (this is discussed in further detail later).

Several functional imaging experiments are fond of comparing the activity during a specific task to that during a “blank” state. I would like to briefly discuss some of these experiments. Mellet *et al.* (Mellet *et al.*, 1996) asked subjects to mentally build complex objects based on verbal instructions. The vocabulary in the verbal instructions was reduced to six words to minimize possible lexico-semantic confounding factors. Subjects had to visualize the object thus built in memory for 5 seconds and then "... delete it from their minds..." There was no on-line behavioral control. The authors used debriefing and testing after the experiment was finished to verify behavioral performance during the 5 seconds of data acquisition. Fletcher *et al.* (Fletcher *et al.*, 1995) compared retrieval of "highly imageable words" such as 'car' and 'truck' and "non-imageable words" such as 'come' and 'go'. A subjective rating of how much they used imagery as a mnemonic rule for retrieving the associate pairs was acquired after the test. Kosslyn *et al.* (Kosslyn *et al.*, 1999) showed very simple stimuli presented simultaneously (bars of different orientation, width and size) and then asked subjects to make a comparison between them. This requires subjects to visualize the stimuli in their minds. While I find this to be a more controlled paradigm, the stimuli they used are probably not very interesting to higher visual areas (very few neurons responded to patterns like gratings in our sample; see Chapter 3). In addition, multiple stimuli are presented and imagined simultaneously so that temporal and stimulus specificity are lost. Kosslyn *et al.* (Kosslyn *et al.*, 1995) use a rest condition in which subjects are asked to "... close their eyes, relax and to have it black in front of their mind's eye..." They were interested in comparing presumed activation of V1 when subjects imagined objects of different sizes (the crucial point to argue topography). They therefore "... urge subjects to visualize the objects at the correct size..." The results suggest a topographical activation with different areas activated with small and large images. Mellet *et al.* (Mellet *et al.*, 1998) compared brain activation upon

listening to dictionary definitions of concrete (bottle, guitar) and abstract (theory, grammar) words. During the former, they "... explicitly encouraged subjects to form visual images" while during the latter they "... instructed subjects not to force themselves to produce mental images..." They also use a "rest control" with eyes closed and no particular instructions other than not to move. Kosslyn *et al.* (Kosslyn, 1994, Kosslyn *et al.*, 1995) asked subjects to "...empty their minds of all other thoughts..." and asked subjects to "... view the stimuli without trying to make sense of them or make any connections between them..." Subjects were debriefed after the experiments, and they all reported not having thought about or made connections between the stimuli. None reported having visualized any pattern in the grid. D'Esposito *et al.* (D'Esposito *et al.*, 1997) used concrete words (e.g., 'apple') versus abstract words (e.g., 'freedom'). Subjects were instructed to imagine the appearance of the named object during the concrete condition and to listen passively during the abstract condition. Ishai and Sagi (Ishai and Sagi, 1995) performed a very elegant psychophysical experiment showing that visual imagery can affect perception in a similar way that perception does. There is no clear control over what subjects are imagining other than the strong results of the psychophysics experiments themselves. The bottom line is that if subjects had not followed the instructions, it would be hard to explain the magnitude and parameter-dependency of the results. In our experiment, we specifically tried to avoid asking subjects to "have their minds blank." Therefore, there was no blank space between the periods in which subjects had to imagine one or the other stimulus. Because of this, during visual imagery we did not have a baseline prior to each 3000 ms interval. Instead, we used the 1000 ms before the first imagery period for each pair as a baseline. It could be argued, *bien sure*, that subjects were still forming some images during this period. However, when we directly compared the baseline during imagery to that during vision and to the spontaneous activity over the whole experiment, we did not observe any significant difference (see Figure 4-4); this is discussed in more detail below (see Section 4.4).

4.3.2 Electrode implantation, recordings and stimulus presentation

The procedures for electrode implantation, data acquisition, spike sorting and so on were described in detail in Chapter 2. The current chapter describes data from a total of 37 experiments in 15 patients (10 right handed, 7 male, 21 to 44 years old). Visual stimuli were presented on a monitor (Micron 500 FGx, Micron Electronics, Nampa or Sony CPD-G200 Trinitron) situated approximately 50 cm away from the subject attached to a laptop computer (Dell Inspiron 5000, Dell Computer Corporation, Round Rock) as in the previous chapter and auditory stimuli were presented through speakers (Harman/kardon, Harman Multimedia, Northridge). Responses to visual stimuli were acquired through a pair of custom-built push buttons attached to the laptop. The timestamp of these responses as well as the visual and auditory stimuli marker timestamps were sent to the neuronal data acquisition computer and recorded in the same computer as the electrophysiological data. Subjects debriefing responses were marked down in the same laptop computer by the experimenter (G.K. in 95% of the experiments).

4.3.3 Data analysis

We used the same criteria described in detail in the previous chapter to define whether a unit was selective to a particular category of stimuli or not. Briefly, a neuron was *selective* to a stimulus group if: *i*) the firing rate during stimulus presentation was significantly different from the preceding baseline, *ii*) an analysis of variance and multiple pairwise comparisons using Scheffe's method addressing whether there were differences among the stimulus groups yielded $p < 0.05$ and *iii*) an ANOVA comparing the variability to distinct stimuli within the selective category to the variability to repeated presentations of the same stimulus showed $p > 0.05$. The interval of analysis for the visual presentation was [100;1000) ms with respect to the picture onset as described in the previous chapter. During visual imagery we used the interval [100;3000) ms with respect to the tones. In both cases, spike counts were normalized by the interval duration. We observed neurons selective to faces, objects, spatial layouts and other stimuli. The neuronal responses during the visual presentation part of this test were described in more detail in the preceding chapter. If the across-groups comparisons were not significant but

the activity was different from baseline, the neuron was defined as responsive but non-selective. To take into account any effects due to the different intervals for counting spikes, we also compared the responses in a 600 ms window centered on the peak firing rate. All the results described in this chapter were very similar using this shorter interval, and therefore we only give the values for the [100;1000) and [100;3000) ms intervals for vision and imagery respectively.

In the previous chapter we have also described the existence of neurons that showed selectivity for a few individual stimuli within a category but did not respond to all the stimuli that were presented within the category. In the current experiment, the five presentations of each individual stimulus were not completely randomized with respect to the total set of individual stimuli that were presented (a pair of stimuli were presented five times each according to the scheme described above and in Figure 4-1). This presents an important concern for the analysis of the responses to individual stimuli. The individual pictures and the categories in one pair and the next one were independent (with each individual picture shown only once), justifying the analysis for categories of stimuli as described above and in Chapter 3.

The peak, latency and duration were estimated from the spike density function (Richmond *et al.*, 1990, Sheinberg and Logothetis, 1997) as described in Chapter 3. For the selective neurons we computed the probability of error, p_e , for classifying the stimulus as belonging to the preferred stimulus category or not (Gabbiani and Koch, 1998, Green and Swets, 1966) as we described in Chapter 3.

4.4 Neuronal responses during visual imagery

We directly studied the neuronal correlates of visual imagery by recording from 429 single neurons in the human medial temporal lobe. The location of these units, as assessed from the structural MRI information (Chapter 2) is indicated in Table 4-1.

4.4.1 Comparison of baselines

The criteria used to define a unit as selective were described above and in Chapter 3. One of the criteria requires the comparison of the activity of the neuron during visual presentation or imagination with the baseline. The baseline during vision was computed in the $[-1000,0)$ ms interval before each presentation. During visual imagery, we avoided comparing to a baseline between tones when subjects could still imagine the stimuli (see discussion above); instead, we used the $[-1000,0)$ ms interval before the first tone. Neuronal activity during the baseline constitutes a potential concern in visual imagery experiments (Kosslyn, 1994, O'Craven and Kanwisher, 2000, Binder *et al.*, 1999) (Kosslyn and Ochsner, 1994, Kosslyn *et al.*, 1995, Kosslyn, 1994). We therefore directly compared the activity in the baseline during vision to that during imagery (Figure 4-4). There was no significant difference in the firing rate in the vision and imagery baselines (t test, $p > 0.2$). There was a strong correlation between the two baselines ($r^2 = 0.97$). The same results hold when considering only those units in the MTL ($r^2 = 0.96$). We also directly compared the baselines for those units that were selective during both vision and imagery ($r^2 = 0.91$). Furthermore, both baselines were strongly correlated with the average firing rate of the unit computed over the whole experimental session. Neither the baseline during vision nor that during imagery was statistically different from the spontaneous activity in the whole experiment (t tests, $p > 0.4$ and $p > 0.3$ respectively). The correlation coefficients were 0.99 and 0.98 for vision and imagery respectively. These data suggest that our baseline was a rather accurate assessment of spontaneous activity.

4.4.2 Selectivity during vision and imagery: examples

We found that some of the neurons showed selective changes in their firing rate while subjects viewed the figures and when they were visually recalling the images with closed eyes. Figure 4-2a shows an example of a neuron in the entorhinal cortex that increased its firing rate selectively when the subject viewed pictures of objects. The mean firing rate during the $[100,1000)$ ms interval after stimulus onset for objects was 16.8 ± 3.6 (mean \pm s.d.) spikes/s. This was significantly higher than the baseline and also higher than the activity for all other types of stimuli (ANOVA and pairwise comparisons, $p <$

10^{-3}). The same neuron also increased its firing rate when the patient recalled the objects with closed eyes (Figure 4-2b), but not when the subject imagined other stimuli. There was no significant difference in the responses to distinct objects ($p > 0.2$).

Another example of a neuron that showed selective changes in firing rate during vision and visual recall is shown in Figure 4-3. This example corresponds to a neuron located in the amygdala in a different patient from the neuron illustrated in the previous figure. This unit showed an increased firing rate when the subject saw pictures of animals and when she formed mental images of the same pictures, but not during vision or recall of other stimuli. The neuronal activity for pictures of animals was statistically significant ($p < 10^{-3}$) from the response to all other types of stimuli that were presented.

As described in the previous chapter, some neurons showed a non-selective increase in firing rate upon visual presentation of different pictures. For these neurons, the firing activity did not distinguish among the different categories of stimuli or different individual stimuli. An example of the responses of such a neuron to visual presentation is illustrated in Figure 4-7a. This neuron was recorded from the left parahippocampal gyrus. Note that there were very few spikes from this neuron outside the visual presentation interval; the baseline activity was 0.5 ± 0.9 spikes/s. The overall activity during visual presentation was 3.3 ± 1.9 spikes/s. During visual imagery, the unit did not show any modulation of the activity either (Figure 4-7b). The firing rate during visual imagery seemed to be maintained at a relatively constant level of 2.6 ± 1.4 spikes/s, higher than the visual imagery baseline of 0.3 ± 0.7 spikes/s. None of the units that were responsive but non-selective during vision showed selective activity modulation during imagery.

4.4.3 Selectivity during vision and imagery: summary

We found a total of 54 neurons (13% of all recorded neurons) that showed selective changes in firing rate to one or two of the categories of stimuli during visual presentation (Table 4.1). There were 31 neurons (7% of all recorded neurons) that showed selective changes in firing rate while subjects visually imagined the same stimuli. Of these 31 neurons, 9 (29%) were activated exclusively during imagery and 22 (71%) were also selective during vision. The striking observation was that, out of these 22

neurons, 20 (91%) showed the *same* selectivity during vision and visual imagery. While the significance criterion was set to 0.05, most of the p values (vision: 70%; imagery: 73%) were below 0.01. Assuming a null hypothesis of independence between vision and imagery, one would expect at the 0.01 level to obtain approximately one neuron in 90,000 showing the same selectivity during both vision and imagery by chance. The statistical significance of our results is discussed in more detail below.

As discussed in the previous Chapter, we do not know whether all the units that did not show a selective response are indeed neurons that are not interested in any visual stimuli or we simply could not find within the very small set of stimuli that we explored what the units preferred. This applies to visual imagery as well and therefore the observation that a neuron was not modulated during a visual imagery experiment does not mean that it does not generally respond to imagery of other stimuli that were not tested. It is conceivable that subjects were equally engaged in the visual imagery task in all of the trials. The above results constitute averages over all neurons and the proportions may be different for different types of neurons. Furthermore, the results that we have reported here correspond to a very simple measure of neuronal response obtained by counting spikes in relatively large time windows and it is possible that precise patterns of spike timing and/or the combined activity of multiple neurons can yield more information about the percept. Therefore, the above results estimating the proportion of neurons whose activity is modulated during imagination are likely to represent a lower bound.

A summary of the responses during vision and imagery for those units that were selective during both processes is shown in Figure 4-9. The average response was obtained by averaging the *sdf* to the effective (red) and ineffective (blue) stimulus categories after normalizing by the peak of the response and subtracting the mean firing rate. The figure shows a selective increase of the response for the preferred stimulus both during vision and imagery. Interestingly, the average neuronal response is transient and does not last for the entire presentation or imagery period. The average duration of the responses are described below.

4.5 Comparison of neuronal activity during vision and imagery

We have shown that there are neurons that were activated both during vision and imagery and had a similar selectivity in the two processes. Here we directly compare the properties of the responses of those neurons.

4.5.1 Response latencies and durations during vision and imagery

We estimated the response duration and latency from the spike density function for both vision and imagery. The details of the computation were described in the previous Chapter. The neuronal responses during visual imagery were longer during vision ($p < 0.05$, see Figure 4-8a). The durations during vision ranged from 299 to 1207 ms (522 ± 246 ms) whereas during imagery the range was much broader going from 207 to 1947 ms (906 ± 593 ms). The longer durations and larger variability in the response durations could be simply due to the differences in the perception time intervals during the two tasks, but it should be pointed out that the durations were on average less than 50% longer during imagery (while the time interval of 3000 ms was three times longer than the 1000 ms of visual presentation). Most of the visual responses were transient and significantly shorter than the whole period of visual stimulation as described in our previous Chapter. The same was true for visual imagery where the mean response duration was approximately a third of the whole period in which we requested subjects to form a mental image. In the case of imagery, it can readily be argued that subjects did not maintain the mental picture during 3000 ms as an explanation for this. But in the case of visual responses, subjects did see the picture during 1000 ms and the duration of the neuronal response was still approximately half of this interval.

The latencies (Figure 4-8b) as well as the peak times (Figure 4-8c) were longer and more variable during imagery compared to vision. The latencies during vision ranged from 71 to 640 ms (257 ± 265 ms) whereas those during imagery ranged from 95 to 1200 ms (461 ± 385 ms). Again, these differences could be due to the different time intervals in

the two tasks. Perhaps subjects, knowing in advance that they had 3000 ms to imagine the stimuli, deliberately started forming a mental picture later on. The actual image formation time could be more variable from trial to trial, thus creating a longer more variable latency. It is also possible that it does indeed take longer to develop a mental picture than to interpret an incoming visual signal. The larger variability in the latency and duration of the neuronal responses could represent a correlate of the more spatially and temporally diffuse nature of imagery, the lack of pungency and tang that William James described.

4.5.2 Comparison of firing rates

Is the firing rate of the neurons during vision different than during imagery of the same stimuli? To address this question, we computed for the 20 neurons with the same selectivity the firing rate over the whole stimulus period for the selective stimuli and divided it by the baseline activity. There was a strong correlation between this normalized firing response for vision and visual imagery ($r^2=0.90$, see Figure 4-5a). The slope between the activity during vision and imagery was 0.76 indicating that the firing rates were about 24% higher during vision. The baselines in the two tasks were very similar as discussed in Section 4.4; therefore, this correlation was not due to this normalization procedure. The correlation was similarly very strong between the absolute firing rates (Figure 4-5b) and also between the firing rates after subtracting the baseline (Figure 4-5c). Since the durations of the selective responses were typically shorter than the whole stimulus period (see previous section), this firing rate is an underestimation of the response of the neuron. Furthermore, the intervals in the two tasks were different. We therefore computed the firing rates in a 600 ms period centered on the peak firing rate. The correlation coefficient between the firing rate in this interval during vision and imagery was 0.85 and the slope was 0.76 (Figure 4-5d).

There is a potentially important confound in the comparison of the selectivity for categories during vision and imagery. To see why this is so, let us suppose that we presented stimuli s_1 , s_2 and s_3 within a given category. Furthermore, let us assume that the neuron responded very strongly to s_2 but yielded almost no response to the other two stimuli during vision and that the unit showed a vigorous response to s_3 and almost none to s_1 and s_2 during imagery. Upon averaging, we may see a significant response to this

category both during vision and during imagery, but it is quite clear that there is no real shared selectivity between the two processes. This situation, however, does not seem to apply to our data. First, one of the criteria for selectivity to a category was that there should be no difference to distinct stimuli within the category (see Section 4.3 and Chapter 3). Therefore, for all the units reported as selective in Table 4.1 there was no significant difference between the responses to different individual stimuli within the selective category both during vision and during visual imagery. Therefore, we can exclude the above assumption that the unit would respond significantly different to only one stimulus within the group. Furthermore, we directly computed the correlation coefficient between firing rates for all individual stimuli within the categories the neurons were selective to (Figure 4-6). The correlation of firing rates between vision and imagery was also high for individual figures ($r^2=0.94$, slope=0.91; see for example Figure 4-1). Therefore, the correlation that we observed in the number of selective neurons and their firing responses between the process of vision and imagination does not rely on the issue of category selectivity.

4.5.3 Guessing the stimulus from a single neuron's activity

Is it possible for an ideal observer to predict the stimulus group that the subject was viewing or imagining based on the firing rate of a single neuron? Figure 4-1 shows a particularly clear example of selective firing in a single repetition of the neuron whose average activity was depicted in Figure 4-2. The data is shown again in Figure 4-10a but without the labels of which stimulus (the face or the baseball) was shown in each period. By observing the activity of this neuron, it is possible to predict with rather high accuracy what the subject was viewing (Figure 4-10a) or imagining (Figure 4-10b). Indeed, very few mistakes are made by setting a threshold such as the one shown as a red horizontal dashed line in Figure 4-10a-b and claiming that the baseball was presented or that the subject was imagining the baseball in those time periods where the *sdf* exceeded this value.

We addressed this question quantitatively by carrying out a receiver operating characteristic (ROC) analysis as described in Chapter 3 (see also (Gabbiani and Koch, 1998, Green and Swets, 1966, Kreiman *et al.*, 2000a). This analysis estimates the position

in which the threshold must be set in order to yield the minimum probability of error, p_e , in classifying the stimulus as belonging to the preferred category or not based on the average firing rate (see Chapter 3 for the details of the computation). $p_e=0$ corresponds to perfect classification and $p_e=0.5$ to chance performance. p_e ranged from 0.10 to 0.44 for vision (0.30 ± 0.08 , see Figure 4-10c) and from 0.08 to 0.46 for visual imagery (0.31 ± 0.08 , see Figure 4-10d). This means that in approximately 7 out of 10 trials, an ideal observer would be correct in guessing whether the subject saw or imagined the neuron's preferred stimulus by counting the number of spikes. Our perception and our imagination are clearly more accurate than this 30% error rate. Presumably, a much higher accuracy can be achieved by using further information such as the combined activity of multiple neurons. However, it is quite remarkable that such a relatively low error rate can be obtained from the firing rate of one individual unit.

There was a strong correlation in the p_e values for the neurons selective during both processes (Figure 4-10d). The correlation coefficient between the two was 0.80 and the slope was 0.71. That is, the more likely for an ideal observer to accurately determine whether the preferred stimulus was present or not during vision, the higher the chances he will also be correct during visual imagery.

4.6 Other questions and technical issues

4.6.1 Tones and auditory responses

The imagery percept change was instructed by presenting the subjects with an auditory tone (see Section 4.3.1). Is it possible that the selective responses that we observed were due to this tone? While the MTL does receive input from auditory areas of cortex (Nolte, 1998, Suzuki, 1996, Insausti *et al.*, 1997, Insausti *et al.*, 1995), this does not seem to be the case. To address this issue, we directly compared the average neuronal responses after high and low tones. This was analyzed by comparing the activity in a window of 500 ms before and after each type of tone. A neuron was considered to selectively respond to a high (low) tone if the activity in the 500 ms after high (low) tones

was significantly different from the response in the 500 ms before the tone and from the 500 ms of activity after the low (high) tone. Note that all the different visual stimuli that we presented were averaged in this analysis. None of the visual or imagery selective units showed a significant difference in this comparison ($p > 0.05$). It should be reminded that the high and low tone were randomly assigned to the different stimuli. Therefore, for a visually selective neuron such as the ones illustrated here (see Figure 4-2 and Figure 4-3), the tones preceding the different stimuli within each category were typically different. Furthermore, there were several other stimuli that shared the same tone and to which the neuron did not respond. Two neurons (neither selective) showed a different activity to high and low tones *per se* at the $p < 0.01$ level. These results do not mean that neurons in the MTL do not respond to auditory stimuli. It is possible, for example, that some neurons respond to more complex auditory stimuli. The visual world analogous of the two tones would perhaps be to test the neurons only with two gratings of fixed frequency and contrast, one horizontal and one vertical. Most likely, we would not have observed many visual responses with only these two stimuli. The fixed-frequency tones were not particularly interesting and were repeated hundreds of times during the course of a test. These observations suggest that the selective neuronal responses that we have observed are not caused by or correlated with the auditory tones⁶⁵.

4.6.2 Motor responses during vision

In addition to the processing of incoming visual input and the recall of information, the visual presentation and imagery periods of the task differed in the requirement that subjects were involved in making a visual discrimination after the presentation of each picture but not during imagery. Pictures were presented in the monitor during 1000 ms and at an appropriate size and resolution so that the task was evidently trivial (as assessed also by the performance, see Section 4.3.1). As we have discussed in the previous Chapter, it is very unlikely that any of the selective responses that we have observed were due to this task. We compared the activity before versus after

⁶⁵ One could also ask the question of whether it is also true in other modalities that there is a common process to perception and imagination. For example, would the same areas be active upon listening to Beethoven's 9th symphony compared to imagining the sound?

the behavioral response and the peri-response activity to the baseline (see Chapter 3 for details). None of the visually or imagery neurons showed statistically significant changes in firing activity that correlated with pressing the button or making the human face / not human face discrimination. Selective responses were also observed in a previous study where subjects were not required to give any behavioral response to the presentation of different pictures (Fried *et al.*, 1997). Therefore, we infer that it is unlikely that the conclusions that we have drawn from this work of a common representation for vision and imagery are caused by the simple visual discrimination task in which subjects were involved while observing the pictures.

4.6.3 Right versus left

Was there any difference in the visual imagery processes between the right and left hemispheres? Unfortunately, the amount of data that we have makes it very difficult to carefully address this question. As we have discussed in Chapters 2 and 3, we recorded from more neurons in the right hemisphere than in the left hemisphere (see Chapter 3 and Table 4-1). However, the overall number of visually selective neurons in both hemispheres was similar after normalizing for the total number of recorded neurons. Of the 31 neurons that showed selective responses during imagery, 15 were in the right hemisphere and 16 were in the left hemisphere. This corresponds to 5% and 11% of the total number of units in each hemisphere respectively. Given the low numbers it is unclear whether this indicates a real predominance of the left hemisphere during imagery or is simply due to random variations. The low numbers also preclude from drawing any clear conclusion about the differences between hemispheres specifically in each location within the MTL or for specific types of stimuli. There could very well be an interaction between location, stimulus specificity and hemisphere (for example, there may be more neurons activated during imagery of faces in the right amygdala and more neurons activated during imagery of spatial layouts in the left parahippocampal gyrus; I am not claiming there is any difference, I am just giving a hypothetical example). Of the 20 neurons that showed the same selectivity during vision and imagery, 8 were in the right hemisphere. Again, to what extent this is or is not due to random variations is unfortunately unclear. The discussion that we will present in Section 4.7.4 regarding the

study of differences between the four distinct areas that we have recorded from in the MTL and extrapolations from our current data *set* also applies here. Unless the differences between the two hemispheres were very pronounced, it would be necessary to record from a large number of neurons to be able to make a statistically meaningful claim about these differences. Hemispheric differences in the human brain during visual imagery have been suggested based on neurological, EEG and functional imaging studies. Some of these differences, however, could be related to the language processing requirements of the tasks. We should definitively keep our eyes open to possible differences between the two hemispheres as we collect more data.

4.6.4 Statistical significance of the results

It could be argued that some units could be selective both during vision and imagery by chance. It is therefore important to estimate the probability of obtaining 20 neurons that show the same selectivity during vision and imagery. We performed three different tests to address this question. First, we assume that by chance one would obtain $p_v=5/100$ (1/100) neurons which are selective during vision and $p_i=5/100$ (1/100) neurons which are selective during imagery at the 95% (99%) confidence level. Assuming as the null hypothesis that vision and imagery are independent process, the probability that a neuron will be selective during both vision and imagery will be $p_v*p_i=25/10000$ (1/10000). But in most cases, this neuron would show different selectivity during vision and imagery. Given $n_{cat}=9$ categories, there are n_{cat}^2 different possible combinations. Out of these, n_{cat} will show the same selectivity. Therefore, the probability that the neuron will show the same selectivity during both processes is $1/n_{cat}$. The overall probability of finding a neuron with the same selectivity during vision and imagery by chance is therefore $p_v*p_i/n_{cat} = 25/90000$ (1/90000).

In the above test we assumed that the probability of finding a neuron that is selective by chance is 5/100 (1/100). But we actually found a percentage of neurons that were selective which was somewhat larger than that (see Chapter 3 and Table 4-1). Then, if we now already believe that the neurons are not firing randomly, we acknowledge that some of them will be selective with a probability p_v during vision and a probability p_i during imagery. From our data, let us estimate $p_v = 54/429 = 0.13$ and $p_i = 31/429 = 0.07$

at the 95% confidence level (see Table 4-1). According to the null hypothesis of independence between vision and imagery the probability of finding a neuron which is selective during both processes is $p_v * p_i = 0.009$ (in order to compute the expected value of how many one would obtain after repeating the experiment $n=429$ times, we would have $429 * 0.009 = 3.9$). If we further request that the neurons should show the same selectivity during both processes, the probability per neuron becomes: $p_v * p_i / n_{cat} = 0.001$ (and the expected value 0.43 for $n=429$ neurons).

In this test, we assumed as a null hypothesis that vision and imagery are independent processes. But let us say that we already acknowledge that there is an interaction between them. Then, we need a measure of these interactions in order to do any computation. Let us assume that out of 429 neurons, we already know that we will get 22 neurons that will be selective during both processes (see Table 4-1). Then the next question is what is the probability that they will show the same selectivity. As a null hypothesis, let us assume that the selectivity of the neuron during vision and imagery are independent. Then, as discussed in the first test above, the probability for each neuron to show the same selectivity is $1/n_{cat}$ (there are n_{cat}^2 possibilities and n_{cat} show the same selectivity). Then the expected value is 2.4 neurons with the same selectivity during both vision and imagery. I only showed the expected values in this discussion, in all these tests the standard deviations can be easily computed from the expression for a binomial distribution (Keeping, 1995).

I ran test 2 and test 3 in a bootstrap procedure with 100,000 simulations as well. During the second test, the number of neurons selective during both vision and imagery was 3.89 ± 1.86 (range = 0 to 25). The number of neurons selective during both vision and imagery and with the same selectivity was 0.45 ± 0.69 (range = 0 to 10). For the third test, the number of neurons that were selective in both processes with the same selectivity was 2.47 ± 1.49 (range = 0 to 12). These statistical tests and simulations clearly show that the results that we have observed could not be due to a neuron firing randomly that passes the criteria for selectivity for the same stimuli both during vision and imagery by chance.

4.7 Summary and conclusions

4.7.1 Is visual imagery indeed visual?

Images pervade our mental processing of all types of information. They appear guided by the sensory input to the retinae in our every moment visual experience but they are also prominent in the absence of retinal activity during thoughts and imagination. It has also been argued that imagination and the retrieval of stored representations play a crucial role in visual object recognition.

There is converging evidence from several different studies using very distinct techniques that suggests that there may be a common representation in the brain that is used and required for recognizing visual stimuli and to imagine them. These techniques include psychophysical and psychological experiments, non-invasive imaging and scalp electrical recordings in normal subjects, electrophysiological experiments in monkeys, the study of neurological patients as well as lesions in monkeys and electrical stimulation of the brain in epileptic human patients. While a gross overview of these techniques agrees in the general finding that vision and imagery are closely related, quite often they can differ in the details and it is therefore important to consider them carefully.

In the last decade there have been several investigations exploring the activity in the human brain using non-invasive scalp recordings and functional imaging. Most studies agree that there are similar areas that are activated during vision and imagery using very different tasks and exploring different brain regions (Kosslyn *et al.*, 1997, D'Esposito *et al.*, 1997, O'Craven and Kanwisher, 2000, Goebel *et al.*, 1998, Mellet *et al.*, 1998, Mellet *et al.*, 1996, Ishai *et al.*, 1999, Roland and Gulyas, 1994, Roland and Gulyas, 1995, Frith and Dolan, 1997, Buckner *et al.*, 1996, Wheeler *et al.*, 2000; see, however, Howard *et al.*, 1998). It should be emphasized that not all areas are co-activated during both processes in the above studies. One of the points where several studies diverge is on whether occipital cortex is activated during visual imagery or not in functional imaging studies⁶⁶. Unfortunately, we typically cannot record from V1 because

⁶⁶ It is interesting to observe that V1 is also an area of hot debate in other experiments such as binocular rivalry (see Chapter 6) and attentional modulation.

this is not where the doctors implant the electrodes; therefore, I will not discuss the role of V1 in visual imagery here (for discussion see Kosslyn *et al.*, 1999, Kosslyn *et al.*, 1995, Miyashita, 1995, Ishai and Sagi, 1995, Roland and Gulyas, 1994, Kosslyn and Ochsner, 1994, Kosslyn, 1994).

Anterior areas in the temporal lobe seem to be engaged in the process of visual imagery according to functional imaging evidence in several of the studies from very different groups. Kosslyn *et al.* found joint activation of the left occipito-temporal junction, the associative area defined by Brodmann's area 19 and middle temporal gyrus (Kosslyn *et al.*, 1997). D'Esposito *et al.* report the activation of the left inferior temporal lobe (Brodmann's area 37) in most subjects and occipital association cortex (BA 19) in some subjects during mental image generation. O'Craven and Kanwisher report activation of the fusiform gyrus and parahippocampal area during imagery of faces and places respectively (O'Craven and Kanwisher, 2000). Mellet *et al.* report the activation of the inferior temporal cortex, superior temporal sulcus and fusiform gyrus (Mellet *et al.*, 1998) as well as parahippocampal gyrus (Mellet *et al.*, 1996) for spatial imagery. Several studies also report activity in the MTL during imagination (Mellet *et al.*, 1996, Ghaem *et al.*, 1997, Roland and Gulyas, 1994, Roland and Gulyas, 1995).

The medial temporal lobe is also activated according to functional imaging evidence in other situations in which there are strong experiential visual phenomena in the absence of retinal input. One such situation is the case of dreams to be discussed more extensively in the next Chapter and where the MTL is activated (Braun and Herscovitch, 1998, Maquet *et al.*, 1997, Maquet *et al.*, 2000, Maquet and Franck, 1996). Another condition is during hallucinations in schizophrenics (Silbersweig *et al.*, 1995) or in patients with Charles Bonnet syndrome (ffytche *et al.*, 1998). The content of hallucinations seems to correlate with the functional specialization of each region activated in fMRI. Strong activation of the MTL was observed in these studies during the perceptual phenomena that occur during the hallucination period.

The activity of neurons in monkey extrastriate cortex is correlated with the contents of the pictorial short-term memory (Miyashita and Chang, 1988, Miyashita, 1993, Yakovlev *et al.*, 1998). Neuronal correlates of visual recall and prospective coding have been described in the monkey frontal cortex areas (Rainer *et al.*, 1999, Miyashita

and Chang, 1988, Tomita *et al.*, 1999). Furthermore, these studies show top-down interactions between prefrontal and temporal cortex during recall. In typical experiments in which stimuli were shown to monkeys, the neuronal activity ceased within a short time of the disappearance of the stimulus. The seminal work of Miyashita's group showed that neurons in inferior temporal cortex continue to fire in a selective fashion in the absence of any visual stimulus in studies involving delay match to sample tasks. This neuronal activity during the delay period was highly specific and showed the invariance properties to several physical attributes of the objects such as their size, orientation, color and position (Miyashita and Chang, 1988). During the delay period, monkeys need to remember the stimulus they have just seen in order to be able to make a correct decision at the end of each trial and thus get a reward. It is suggested that monkeys do this in the same way that most humans would, that is, by keeping a mental image of the stimulus in their minds during the delay. Therefore, the activity during this period has been interpreted to represent a neuronal correlate of visual imagery (Miyashita, 1988, Miyashita and Chang, 1988, Sakai and Miyashita, 1991, Miyashita, 1993, Sakai and Miyashita, 1994a)⁶⁷. In subsequent work, Sakai and Miyashita showed that the activity of some neurons in the anterior inferior temporal cortex correlates with retrieval processes from long-term memory as well. The work of Rainer *et al.* later showed that early on in the delay period, activity in frontal cortex represents the visual input whereas during the last period of the delay, the neuronal response better correlates with the anticipated target. The top-down signal from frontal cortex to the temporal lobe was suggested to play a crucial role in the control of memory retrieval (Tomita *et al.*, 1999). This could be due to

⁶⁷ When we first submitted the manuscript to Nature, we proudly announced that we were the first to study the relationship between vision and imagery at the single neuron level. For sure, there have been several philosophical treatises on the matter for several centuries. Extensive data has been accumulated from psychological and psychophysical experiments as well. Moreover, in the last decade, a wealth of data from functional imaging experiments have become available as we have described in this chapter. But, to the best of our knowledge, we were the first to look at single neurons. Claiming to be the first is surely a nice strategy to get other scientists annoyed. And, sure enough, Earl Miller rejected this by claiming that several of the studies in monkeys reported here had already shown this. I am personally extremely fond of the monkey electrophysiology literature. The work of Miyashita's group as well as that of Miller's group has clearly illuminated (and will continue to do so) our understanding of the functioning of the temporal lobe. But I think it is not entirely trivial that the fascinating responses obtained during delayed match to sample tasks in monkeys correspond univocally to visual imagery. Let us then rephrase our claim to be more precise. We have shown for the first time that the volition-induced imagined percepts share a common representation in the medial temporal lobe in humans at the level of single neurons.

direct interaction of the frontal lobe with the anterior inferotemporal cortical areas or through the MTL.

Important information comes from the study of neurological patients that show specific deficits in the recognition of stimuli. Neuropsychological studies have shown that most of the patients that have perceptual deficits also display a deficiency in the possibility of imagining the same stimuli (Young *et al.*, 1994, Farah *et al.*, 1992, Farah, 1988). There are, however, exceptional reports of a few cases where patients show visual recognition deficits but keep a normal capacity for visual imagery (Behrmann *et al.*, 1992, Bartolomeo *et al.*, 1998, Coslett, 1997, Young *et al.*, 1994). Conversely, a double dissociation has been suggested on the basis of a few other patients that show imagery deficits but keep a normal capacity of visual discrimination (Coslett, 1997). Unfortunately, the anatomical location of the lesions in these patients are not as well delimited as one would like in order to fully understand whether there are indeed two separate representations, one used during vision and a different system used during imagery. Going over the literature, it seems that the number of patients that show a deficit in only one of the processes but not in the other is much less than the number of patients that show a common difficulty both in recognition and imagination. While this may be to some degree related to the different ways of testing the patients and controlling for recognition and imagery impairments, it might also be indicative of the way the two processes are represented in the brain. In this light, several regions would be involved in a common representation whereas only a few specific areas would be exclusively required by either vision or imagery. The general comparably poor localization resolution of these studies makes it difficult to discern which areas could be required only for imagery or vision.

Electrical stimulation in the human temporal lobe can interfere with (Fried *et al.*, 1982, Ojemann and Mateer, 1979, Ojemann *et al.*, 1998) or elicit visual recall (Penfield and Perot, 1963, Penfield and Jasper, 1954). It should be noted that while several of these observations are more anecdotal rather than quantitative, they constitute fascinating evidence about the possible link between neuronal activity in the human temporal lobe and perception. These observations were made during surgery in epileptic patients who were perfectly conscious during the tests. The patient cannot feel when the brain is

touched; therefore, the surgeon can check the reliability of the responses. Out of a total of 520 surgeries in one of the most comprehensive studies where electrical stimulation in the temporal lobe was performed⁶⁸, experiential responses were observed in experiential responses in 40 patients (7.7%; Penfield and Perot, 1963, Penfield and Jasper, 1954). It is interesting to note that this percentage is roughly comparable to the percentage of units that showed a selective response during visual imagery. These experiments are quite difficult and heroic. In some of the subject debriefing responses, there was a poor reproducibility after repeated stimulation separated by several minutes. However, it should be noted that the electrode could move considerably in the course of several minutes. Therefore, the lack of reproducibility could be simply due to the fact that the stimulation site shifted to a different location. In some other cases, there is a remarkable degree of similarity in the description of the experiential phenomena described by the patients. Eliciting an effect (e.g., experiential phenomena) by stimulating area X does not necessarily imply that area X causes the effect. Stimulation may be sent from X to Y and Y might actually represent the neuronal causal correlates of the effect. Thus, most of the experiential responses were obtained by stimulating in neocortical targets (see however Halgren *et al.*, 1978) in the temporal lobe rather than in the MTL but this does not necessarily imply that the results of stimulation are not conveyed and spread to other nearby areas. It should also be noted that the spatial resolution of electrical stimulation in the human brain can be rather coarse compared to the level of single neurons. It is likely that electrical stimulation in these reports changed the activity of a large number of neurons. It seems that a completely random arrangement of neuronal selective responses would not necessarily give rise to the complex experiential phenomena that the patients reported. In the careful stimulation experiments performed in the monkey V5/MT visual cortical area (Salzman *et al.*, 1992, Salzman *et al.*, 1990), the columnar organization of direction selective neurons seems to have been crucial for the effective manipulation of the monkey's percept. Interestingly, Tanaka and others have suggested a columnar organization of selective responses in the monkey anterior inferior temporal cortex (Miyashita, 1993, Fujita *et al.*, 1992, Tanaka, 1993). It is also interesting to point out that

⁶⁸ 612 other areas were exposed in other surgeries. No experiential responses were observed by stimulation outside temporal lobe.

visual and/or auditory hallucinations sometimes come to patients with seizures (Penfield and Jasper, 1954, Fried *et al.*, 2000). In several cases, the reports obtained from electrical stimulation were concordant with the typical experiential phenomena that the patient described during seizures. Experiential phenomena during seizures, due to electrical stimulation, hallucinations, imagery and dreams seem to share several interesting properties including the common activation of anterior and medial temporal lobe areas.

We observed three different types of selective neurons. Some neurons responded during processing of incoming visual information but not during imagery (32 out of 54). There were also neurons activated only during visual recall (9 out of 31); these units might be involved in retrieval mechanisms dissociated from vision (Behrmann *et al.*, 1992, Bartolomeo *et al.*, 1998). Finally, some neurons responded selectively during both vision and imagery (22 out of 31). Of these 22 neurons, 20 showed identical selectivity during the two processes. The large proportion of neurons with similar selectivity during the two processes suggests the existence of a common processing system for the encoding and retrieval of visual information in the MTL. We did not find regional segregation of neurons selective during vision and imagery. We found neurons from all these three groups in the four regions that we have studied within the MTL (Table 4-1). While we have not observed any obvious difference in the number of selective units during vision and imagery in these different locations, the total number of units may not be enough yet to detect these differences. This is further discussed in Section 4.7.4. Interestingly, we have not found two units from different groups on the same microelectrode. Given that the results reported in Table 4-1 constitute an average over several patients, it is not clear whether the neurons from these three different groups were arranged in any special way within each structure.

Our results provide a rare opportunity to directly observe the activity of the human brain in the absence of external visual stimulation. The evidence that we have observed suggests a common neuronal mechanism for the interpretation or creation of images from retinal input or memory. The selective neuronal activity in the MTL during imagination may represent the retrieval of the picture information from memory or the maintenance of the visual percept during imagination. The firing of these neurons could represent a correlate of the percept common to vision and imagery. Given the prominent

role of the medial temporal lobe in declarative memory, it also seems possible that these neurons could be activated during storage of incoming visual inputs and later reactivated during the mnemonic retrieval process required for imagery.

4.7.2 The lack of pungency and tang in imagination

Given the weaker percept during imagery, it may seem surprising that we have observed such a small difference in the firing responses during both processes. For those neurons that were selectively activated in both processes, there was approximately 25% less activity during imagery overall (see Section 4.5.2 and Figure 4-5). This difference in firing activity does not seem to correspond to the "...lack of pungency and tang..." that William James ascribed to imagination (James, 1890). Our introspection tells us that mental images seem to be much more diffuse, ephemeral and dream-like than our representation of the world during visual input. One possibility is that this difference is related to our observation that there were fewer neurons recruited during visual imagery compared to vision (Table 4-1). Alternatively, it could be that the difference is related to the existence of more dramatic changes in neuronal activity in other brain areas. It is also conceivable that there is some other variable dependent on the neuronal response other than the spike counts that we have computed; perhaps there is a larger difference between the strength of the response in vision and imagery when comparing, for example, the synchrony in the firing between neurons.

There are several reports from functional imaging that show that within those areas that are co-activated by same-modality perception and imagery, the response is weaker during imagination (Porro *et al.*, 1996, O'Craven *et al.*, 1997, Goebel *et al.*, 1998, Chen *et al.*, 1998) but there is also evidence that in some cases the imagery signal can be greater than the one observed during perception (Binder *et al.*, 1999, Kosslyn *et al.*, 1993). Kosslyn *et al.* proposed that the increase in brain activity during imagery is due to a requirement for more processing to generate an internal mental image than to merely process one that comes in the input. Recently, O'Craven *et al.* reported that higher visual areas in the temporal lobe that are selective for faces and places are correspondingly activated during visual imagery. The strength of the signal obtained during imagery was

approximately 60% of the one obtained during vision of the same stimuli (O'Craven and Kanwisher, 2000).

In monkeys performing delay match to sample tasks, the activity during the delay period was consistently weaker than that during perception (Miyashita and Chang, 1988, Sakai and Miyashita, 1994b, Tomita *et al.*, 1999, Rainer *et al.*, 1999). Following the interpretation of Miyashita and colleagues (see above), this suggests that activity in the inferior temporal cortex as well as in the pre-frontal cortex is lower during imagery compared to perception.

4.7.3 Alternative interpretations

Even if the temporal aspect in our paradigm is more diffuse (so that if a subject has to imagine a face and a car, they do not imagine the stimuli for the whole duration of the 3000 ms and they imagine sometimes the other stimulus as well), the fact that there are selective responses would be hard to explain if subjects were not doing different things while they are imagining distinct visual stimuli. Furthermore, the fact that the selectivity is similar in the imagery and visual responses in most of the cases argues for the fact that they were indeed imagining the right stimuli. As in many of the experiments above, the posterior debriefing (by asking them to indicate in which trials subjects did not form a mental picture and to describe the images) together with the physiological results confirm that most of the subjects were indeed forming a specific mental image in their minds during the imagination part of the test. There are, however, several alternative interpretations and issues that are worth discussing.

One concern that was raised in several of my talks was the question of language. This actually involves several related questions. Is it possible that the subjects were not imagining the pictures but were actually forming a mental image of words related to the pictures? For example, instead of forming a visual picture of the Van Gogh sunflowers, maybe they were actually mentally forming the word "S-u-n-f-l-o-w-e-r-s." The use of tones was an attempt to avoid the usage of words in the instructions. When subjects were debriefed after each session, they claimed to be imagining the pictures themselves and not the words. Then, in order to maintain the word-imagery hypothesis, one would have to admit that subjects were imagining words and not pictures but, for some reason, they

claimed to imagine pictures during debriefing. In addition to that, as discussed in the results, we observed neuronal responses that were similar during imagery and vision. It is still possible, of course, that during vision subjects were still creating words in their minds for each picture they saw. But the hypothesis starts getting somewhat complicated. Maybe during vision the responses were indeed visual and they were related to language during imagery. This would require the activation of the medial temporal for both vision and language related processing. Along these lines, however, a more complex but also perhaps more feasible hypothesis can be formulated. Maybe these neurons respond to complex cognitive concepts. Do not ask me, please, to precisely define what a complex cognitive concept is. The semantic information about the stimuli could then contribute or lead to the selective neuronal response. The semantic information would also be common to vision and imagery. According to this interpretation, the neurons were activated while observing and imagining a visual picture not necessarily because of the common mental image but because of the common process of interpreting, recognizing, understanding and assigning a meaning to the picture or object. An example may clarify this point. Maybe the neuron responding to the baseball (see Figure 4-1) was selective to the general concept of a ball used during sports or to the specific shape, feeling, speed and other attributes of the spherical object used to play baseball. Perhaps the semantic concept the neuron was selective to is even broader. Thus, the neuron may have been equally well aroused by anything related to a baseball game including the visual presentation or imagination of a baseball, a baseball t-shirt or even while imagining eating hotdogs in a baseball court in the absence of the baseball itself. This point is quite related to our discussion in Chapter 3 of what the neurons are actually selective for. Given the limited time in our recordings, the number of pictures that were presented (and the number of different visual images the patients had to form in their minds) was quite limited. These clearly constitute fascinating questions for future research (see Section 4.8).

It has been claimed that there are strong variations in the individual capacity to form visual images (James, 1890, Charlot *et al.*, 1992, Kosslyn *et al.*, 1996, Finke, 1985). It should be noted, however, that this is not an easy question to address experimentally. But let us go along with it. Does this translate into individual differences in the neuronal responses? Indeed, we have observed large changes in the comparison of responses from

one subject to another. Such differences were also prominent in some of the functional imaging reports commented on before (see for example D'Esposito *et al.*, 1997, O'Craven *et al.*, 1997). In most reports, these activation variations were attributed to individual differences in the imagery capacity. One should be extremely careful about this type of interpretation in our data. For example, in some subjects we did not observe a single neuron responsive during imagery. Should we conclude from this that the subject lacks the capacity to form mental images in his or her mind? I think not. It is obviously possible that we simply did not record from the neurons that were selective to the stimuli the subjects were imagining. Let us go back to the visual responses where the problem is exactly equivalent but easier to understand. In some subjects, as discussed in the previous Chapter, we did not observe a single neuron that was visually selective. Should we conclude from this observation that the subject was blind? The claim is obviously ridiculous. The same applies to the imagery data. In addition, one should keep in mind that there could be several other differences among subjects. Among them could be the fact that separate subjects formed different mental images and engaged differently upon recalling the same stimulus. Also, the electrode locations changed from subject to subject (we were not recording from the same neurons). Therefore, while we did observe subject-to-subject variability, it is unclear that we can directly relate this to the possible individual differences in their mental imagery abilities.

As we have discussed in Section 4.3, there were several other ways to design the imagery experiment. Our empirical verification of visual imagery was based on debriefing the subjects after each session. It is quite conceivable that the subjects were distracted in some trials and that they failed to form visual images in their minds in others. In contrast to the visual presentation paradigm where we have a robust idea of the few trials in which subjects did not see the images (because they looked elsewhere or they were distracted or another reason) and we could discard those trials from the average, the data during visual imagery is much more noisy in this sense. Even when we did ask subjects explicitly to indicate trials in which they were distracted or they could not imagine the pictures, it is conceivable that subjects did not want to admit that they were distracted in some trials. Even if they did form visual images in every single trial, it is more than possible that the latency to the image formation was different from trial to

trial. Indeed, we obtained rather large variations in the latencies during imagery. The changes in latency could translate into missing a unit that was only weakly selective⁶⁹. Therefore, I believe that our results set a lower bound on the actual responses in the human brain during visual imagery.

In our experiment, subjects were shown a picture and asked to imagine it shortly afterwards. This may involve the activation of a long stored representation but it could also recruit the activity of areas specialized for short-term memory (here practically defined to be roughly on the order of 5-60 seconds). Another question that we did not address is whether these neurons would also be activated in processes that require retrieval from long-term memory. An alternative explanation of the selective activation observed here is that there could be specific neuronal mechanisms involved in the retrieval of information. Activity during vision and imagery could thus be correlated to the storage and recall of information. The line between vision and imagery or perception and storage is not necessarily a sharp one. Similarly, there could be a continuum rather than a dissociation between the processing of visual input, its storage and its recall. Thus, imagery has been suggested to play a crucial role in the normal processes of object recognition. The identification of a visual stimulus may require the recall and matching of stored information after appropriate scaling, rotations and other transformations.

In our report published in *Nature* (Kreiman *et al.*, 2000b), the low numbers precluded from making any clear conclusions about possible differences between the four areas that we have studied in the MTL. For this reason, I attempted to continue the experiments and record more data after the report had been accepted for publication. Unfortunately, in spite of continuing the recordings for an additional year and several patients, the number of neurons that we have obtained is still too low and we cannot yet make any clear claims regarding regional specificity (see Table 4-1). In order to get robust results on this matter, one would need probably much more data (see the discussion about extrapolation in the next Section).

⁶⁹ For example, let us say that a unit shows a 20% increase in firing rate during imagery for a duration of 300 ms. If the latencies in the 5 trials are separated by several hundreds of milliseconds, then basically, the overall firing rate change after averaging would be close to 4% and then it is very likely to miss the response.

4.7.4 Regional differences: how much longer will it take?

The brain is not an homogeneous ensemble of neurons. Therefore, it would be very interesting to directly compare the different regions, at least those in the medial temporal lobe. For example, one would like to pose questions such as whether there are more imagery selective neurons in the entorhinal cortex than in the other regions. However, it is not easy to give a statistically meaningful answer to this question given the low number of neurons. But let us make some assumptions and at least try to get some feeling for the magnitudes involved.

As a null hypothesis, let us assume that the selective neurons are uniformly distributed among the different regions in the medial temporal lobe. For the sake of argument, we will take $k=4$ regions, namely the amygdala, the entorhinal cortex, the hippocampus and the parahippocampal gyrus. Then, our null hypothesis says that for a total number N of selective neurons, we should expect approximately $0.25N$ in each region. In order to compare the two distributions, let us use a simple χ^2 test (Keeping, 1995)⁷⁰. This involves simply computing the following quantity

$$\chi_s^2 = N \sum_{i=1}^k \frac{(p_i - \pi_i)^2}{\pi_i}$$

where p_i is the proportion of selective units observed in region i ($i=1, \dots, k$) and π_i corresponds to the expected proportion in that region. For the null hypothesis of a uniform distribution, $\pi_i=0.25$ for every i . This quantity has (in the limit of large N) a χ^2 distribution with $k-1$ degrees of freedom. Note that the closer the proportions are to the uniform distribution, the smaller the value of χ_s^2 will be and therefore the less likely that we can reject the null hypothesis.

We would like to know how large N should be in order to be able to observe a statistically significant difference for a given deviation from the uniform distribution. Let us take as an example the case where one region has more selective units than the other regions (this is of course not the only possibility, but the other ones can be analyzed in a

⁷⁰ This test actually requires us to make some strong parametric assumptions about the distribution of the data, such as a normal distribution of the frequencies around their mean. This assumption is actually not necessary, and I have actually also used a non-parametric goodness of fit test. While the results are not identical, the details of the non-parametric test would require some further explanation and the main points are well illustrated with the more commonly used χ^2 test.

similar way). Figure 4-11 shows the required N as a function of the ratio of the percentage of selective neurons in this region beyond that in the other areas. For the actual observed total number of selective neurons, we would only be able to detect a significant difference if the proportion of selective neurons in one area were more than 200% of that in the other areas. For a 50% increase in the proportion, at the 0.05 significance level, we would need above 200 neurons in order to make a strong claim of regional specificity. Linearly extrapolating from the rate of data acquisition⁷¹ and selectivity that we have observed, this would require approximately 50 years of time.

This analysis simply counts the total number of selective neurons in each area. There could, however, be an interaction between the type of selective responses and the area. For example, there could be more neurons responding selectively during imagery of faces in the amygdala while there would be more neurons that are specifically activated during imagery of spatial scenes in the hippocampus. The study of these interactions requires even more data.

4.7.5 An image is worth a thousand words

Our results should not be taken to imply that there is an underlying image-like representation in our brains of everything. It is not clear to me at this moment that images could represent all concepts and ideas. A long-standing tradition in philosophy, championed by Locke among others, tried to argue that we cannot think of anything that we have not seen before or that is not part of what we have seen before. Abstract concepts and ideas such as “liberty” or “men are derived from animals by natural evolution” are hard to paste in the form of images alone. To address this problem, Berkeley went even farther to claim that there is no such thing as a truly abstract idea.

Some of the neurons that we have studied showed a common activation when the brain was stimulated from the outside world through the retina and when it was stimulated internally. But all the images were easily imaginable stimuli; furthermore, subjects had actually observed the stimuli just prior to the imagery task. It remains to be investigated whether there is any specific activation upon thinking of more abstract

⁷¹ It should be noted that this linear extrapolation is a very bad assumption of course.

concepts. Interestingly, several investigators have actually used these abstract nouns as non-imagery controls in their functional imaging experiments (see for example Mellet *et al.*, 1998). However, the results should be interpreted with caution given that they specifically emphasize that they encourage subjects not to form visual images upon thinking of abstract nouns.

4.7.6 Cause or correlation?

Neurophysiological data as well as other measurement techniques provide a correlation between a particular behavior or sensation and the neuronal activity. In our case, there was a correlation between the changes in firing rate and the specific stimulus that the subjects were seeing or imagining. Such a correlation is quite interesting in itself. But it does not directly address the more fundamental question of causality. That is, are the activity changes that we have observed causally related to the percept? In other words, does the percept occur if and only if there is activation of the selective neurons in the MTL?

There are several scenarios in which the activity changes do not indeed represent a causal link to perception. In the simplest case, the causal changes could be taking place pre- or post-synaptically to the units that we recorded from. The feed-forward or feed-back to the unit we were monitoring would immediately cause the change in firing rate but the percept would have occurred regardless of what happened to our unit as long as there were concomitant changes in activity elsewhere. For example, the unit could be activated during visual input due to direct and strong connections from inferior temporal cortex in order to initiate the process of memory storage. During visual imagery, the unit could be activated due to feedback (from frontal cortex or also from inferior temporal cortex) involved in the retrieval of information.

Other experiments and other sources of data would be required to address this issue. One important experiment in this regard would be the study of the effects of electrical stimulation in the neighborhood of the imagery selective units (see Chapter 7 for a more detailed description of this experiment). Another useful source of information is the study of patients with lesions circumscribed to the area of interest. A relatively

recent technique allows to stimulate human brain and cause reversible lesions in a non-invasive way by using transcranial magnetic stimulation (see Appendix 3). The resolution of these techniques is however very coarse compared to the level of single units that we have studied. Given that the question of correlation versus causality seems to be crucial in the interpretation of all the studies shown in this thesis, I will discuss this in more detail in Chapter 7.

4.8 Outlook and open questions

My job in enumerating a long list of problems that are fascinating and still open to research has been facilitated by the excellent feedback that I have obtained from scientists and non-scientists. The number of outstanding problems in visual imagery is too large to even attempt to make any exhaustive list; I will just focus on a few that I think are amenable to single unit research and are of particular interest to our endeavors.

In our experiments, we have always presented the pictures in the monitor a few seconds (up to 15 seconds) before the imagery period. It could be argued that this would activate neuronal systems involved in short term memory retrieval. It is possible that retrieval from long-term memory could invoke different mechanisms and therefore yield different results. It should be noted that I do not necessarily mean years when I talk about long term here. While the retrieval of information from the distant past is also interesting, I am mostly referring to several hours interval between the presentation and the imagination. Interestingly, there is substantial evidence pointing to a very distinct role of the hippocampus in the processing of retro and antero-grade information. The classical example of patient H.M. (Milner, 1972, Penfield and Milner, 1958) showed that he had severe deficits in the retrieval of recently stored information while he could perfectly recall information that had been acquired long before the surgery. Thus, imagining the sunflowers after observing them in a monitor could be very different from imagining the painting if the subject has not seen it for several years.

Another interesting question is the issue of whether we can imagine something that we have never seen. Berkeley and several others have argued (Honderich, 1995) that

we can only re-create what we have already experienced. Thus, we cannot think of an animal that we have never seen and which is not a combination of animals that we have seen. I am not sure that this is correct but it has interesting implications for the question of creativity. We all like to believe that we originally generate art or science. But perhaps we are simply putting together different pieces that were already there to some extent. This does not mean that the particular combination is not new and fascinating in itself. The process of recalling information from memory to form a visual image could potentially involve different neuronal mechanisms than the process of creating a new visual image.

A somewhat related question is how neurons would respond in the case where there is a discrepancy between the input and the imagery. Perky's experiment (Perky, 1910) suggested that one can confuse a mental image for a real one. What would happen at the neuronal level if we see one stimulus while imagining another? My bias would be towards predicting that vision will win since imagery lacks the pungent and tang as William James put it. This scenario is not necessarily as strange as it may seem. It is conceivable that the visual input and the stored information do not match in the recognition of a visual stimulus that has recently experienced some changes (e.g., when a friend has a hair-cut or a given product line changes its container).

Several people have also asked me about the relationship (or lack thereof) between imagery and language. Would the neuron illustrated in Figure 4-1 also respond to the word "baseball"? I am not sure how to neatly dissociate the two. That is, even if we present the word "baseball" and the neuron responds, it could be argued that the word elicits visual imagery in the subject. As I have discussed below, it could be argued whether the neuronal activation that we have observed is not due to the access of semantic information about the stimuli. There are several important questions regarding whether there is a dissociation at the neuronal level between visual and semantic information and whether the neurons in the MTL represent one or the other.

A fundamental missing link is the question addressed in the previous Section of whether there is a causal link between neuronal activity and perception. I would very much like to address or see the answer to this issue. Would electrical stimulation of the selective neurons give rise to a percept? If we recorded from a neuron that responded

selectively upon presentation of a baseball, would the subject "see" a baseball if the neuron (and adjacent neurons) were externally activated? It should be noted that the techniques for electrical stimulation used in epileptic patients by Ojemann, Penfield and others cannot reach the resolution of a single neuron. Probably several thousands of neurons (if not more) are actually activated during the process.

4.9 Figure legends

Figure 4-1: Visual imagery. Single trial example and experiment design.

Two images (**a**) were shown separately during 1000 ms. Each image was repeated five times and the order of presentation of the two pictures was randomized. The black and white horizontal bars indicate the visual presentation of one or the other picture in (**b**). After each picture, subjects pressed a button indicating if the picture was a human face and another button if it was not. (**c**) Activity from a neuron in the entorhinal cortex; the continuous line shows the spike density function while each individual tick represents a spike. After the total of ten visual presentations, subjects closed their eyes and imagined one or the other picture upon hearing a high or low tone. Tones were alternated every 3000 ms (**d**). (**e**) Data from the same neuron during visual imagery. This neuron showed an increased firing rate for pictures of objects (including the baseball illustrated here and two other objects) during both vision and imagery ($p < 10^{-3}$).

Figure 4-2: Selective neuron, example 1

Responses of the same neuron shown in Figure 4-1 during vision (**a**) and visual imagery (**b**). The post-stimulus time histograms were computed by averaging activity for all stimuli within each stimulus group using a bin size of 200 ms. The total number of presentations in each category is indicated in parentheses above the PSTH. The horizontal dashed line shows the average firing rate over the entire experimental session. In (**a**), the dashed vertical lines delimit the stimulus presentation time while in (**b**) the imagery period was from $t=0$ to $t=3000$ ms. This neuron increased its firing rate over the baseline ($p < 10^{-4}$) upon visually presenting objects but not other stimuli. There were three different objects that were presented including the baseball illustrated in Figure 4-1. An ANOVA and pairwise comparisons indicated that the response to objects was significantly different from that to other stimuli ($p < 10^{-3}$; see text for details). Interestingly, the neuron also increased its activity when the subject recalled the same

objects in her mind with her eyes closed (comparison with baseline, ANOVA and pairwise comparisons $p < 0.001$) but not during imagery of other stimuli. There was no significant difference in the responses to distinct objects (vision: $p > 0.2$; imagery: $p > 0.2$).

Figure 4-3: Selective neuron example 2

Responses of a selective neuron in the left amygdala of a different subject during vision **(a)** and visual imagery **(b)**. The format and conventions are the same as in the previous Figure. This neuron increased its firing rate over the baseline upon visually presenting animals ($p < 10^{-5}$) and also during imagery of animals ($p < 10^{-4}$) but not to other stimuli. ANOVA and pairwise comparisons also showed that the response of this neuron was highly selective during both vision ($p < 10^{-3}$) and visual imagery ($p < 10^{-3}$). There was no significant difference in the responses to distinct animals (vision: $p > 0.15$; imagery: $p > 0.3$).

Figure 4-4: Correlation of firing rates during the baseline

Baseline activity during vision (black circles) and during imagery (black squares) is shown as a function of the firing rate computed over the whole experiment for all the units studied in this Chapter. The dashed diagonal indicates the $y=x$ line. The values for the units that were selective during both vision and imagery are shown in red. The main plot was cut at 10 spikes/sec for clarity (this comprises $>90\%$ of the data). All the data are shown in the inset in the top right of the figure. The total number of points in this figure was $n=509$ and included units both in medial temporal lobe regions as well as in other areas. There was no significant difference between the three methods of estimating the baseline.

Figure 4-5: Comparison of activity during vision and imagery

(a) Correlation of normalized firing rates between vision and visual imagery for the neurons that showed the same selectivity during both processes (Table 4-1). The firing rate was computed in the [100,1000) ms and [100,3000) ms interval for vision and

imagery respectively and was normalized by dividing by the baseline (see text for details). The solid line shows a linear fit to the data (slope=0.76; correlation coefficient=0.90). The two examples from Figure 4-2 and Figure 4-3 are marked in red. **(b)** Correlation of absolute firing rates between vision and imagery. The intervals for spike counts are the same as in **(a)**. The dashed line shows $y=x$ and the solid line shows a linear fit (slope = 0.85 and $r^2=0.96$). **(c)** Correlation of firing rates after subtracting the baselines. The linear fit yields a slope of 0.72 and $r^2=0.91$. **(d)** Correlation of normalized firing rates between vision and imagery where spikes were counted in a 600 ms interval centered on the peak of the response. The linear fit yields a slope of 0.76 and a correlation coefficient of 0.85. The dashed diagonal indicates $y=x$. The error bars correspond to s.d.

Figure 4-6: Comparison of firing rates for individual stimuli

Correlation of firing rates for each individual stimulus within the selective categories for all the neurons that were selective during both vision and imagery (see Table 4-1). The dashed diagonal shows the $y=x$ line. The solid line shows a linear fit (slope = 0.91, correlation coefficient 0.94). The time intervals where spikes were counted are described in the previous figure.

Figure 4-7: Non-selective unit, example

Responses of a non-selective neuron in the left parahippocampal gyrus of a different subject during vision **(a)** and visual imagery **(b)**. The format and conventions are the same as in the previous examples in this Chapter. This neuron increased its firing rate over the baseline upon visually presenting pictures from all different categories ($p<0.01$). The activity during imagery was similar to that during visual presentation and larger than the baseline activity ($p<0.05$).

Figure 4-8: Distribution of latencies and response durations for vision and imagery

The latency and duration of the responses were computed from the spike density function (*sdf*). The *sdf* was estimated by convolving the binned spike train (bin size=1ms) with a gaussian of 100 ms width and then averaging across trials (see Chapter 3). The latency was defined as the first time at which five consecutive bins deviated more than two standard deviations from the mean firing rate. The end of the response was similarly defined as the first time point at which five consecutive points go back to the mean firing activity within two standard deviations. **(a)** Mean response duration for vision and imagery (left) and frequency distributions for vision (center) and imagery (right). Bin size =100 ms. **(b)** Mean response latency for vision and imagery (left) and frequency distributions for vision (center) and imagery (right). Bin size = 50 ms. **(c)** Mean peak time for vision and imagery (left) and frequency distributions for vision (center) and imagery (right). Bin size =100 ms.

Figure 4-9: Summary comparison of vision and imagery

Mean normalized activity for preferred stimuli (red) and non-preferred stimuli (blue) during vision **(a)** and imagery **(b)**. The mean normalized response was obtained by normalizing the spike density function by the maximum of the response after subtracting the mean firing rate and then averaging across all neurons. The shaded regions correspond to the s.e.m. The visual stimulus was presented between $t=0$ and $t=1000$ ms (dashed vertical lines in **(a)**). The imagery period was from $t=0$ to $t=3000$ ms (dashed vertical lines in **(b)**).

Figure 4-10: ROC analysis

(a) Would it be possible to guess beyond chance which stimulus was presented if we had lost the labels? Here we show the same data from Figure 4-1 where the labels for the two stimuli were removed. By setting a threshold as indicated by the red horizontal dashed line, and guessing that a baseball was presented whenever activity crossed that threshold, one can determine the presence or absence of the selective neuron with very high

accuracy. A higher accuracy could be obtained in this case by setting a slightly higher threshold. This is addressed quantitatively by using an ROC analysis in parts **b-d**.

(b) Distribution of p_e for all visually selective neurons. Bin size = 0.025. The arrow indicates the mean value.

(c) Distribution of p_e for all imagery selective neurons. Bin size = 0.025. The arrow indicates the mean value.

(d) Probability of error (p_e) in predicting the percept based on the firing rate for imagery (y axis) versus vision (x axis). The p_e ranges from 0 (perfect classification) to 0.5 (chance performance); see text and previous chapter for details. The dashed line shows $y=x$. The solid line indicates the linear fit to the data (slope=0.71; $r^2=0.80$). The examples from the previous figures are indicated by squares.

Figure 4-11: Regional differences, extrapolation

Estimation of the total number of selective neurons required to make a claim about differences between regions for different assumptions regarding the percentage in excess of selective neurons in any one given region. The significance level for the computation was 0.05 (blue circles) or 0.01 (red circles). The horizontal dashed line shows the actual number of neurons selective during both vision and imagery (see Table 1).

4.10 Tables

Table 4-1: Number of selective neurons in the medial temporal lobe

	Amy	EC	Hip	PHG	Total
<i>N</i>	93	168	123	45	429
Vis. Selective	9 (10%)	19 (11%)	20 (16%)	6 (13%)	54 (13%)
Im. Selective	5 (5%)	10 (6%)	11 (9%)	5 (11%)	31 (7%)
Vis. & Im.	3 (60%)	7 (70%)	8 (72%)	4 (80%)	22 (71%)
Vis. & Im. same	2 (67%)	7 (100%)	7 (88%)	4 (100%)	20 (91%)

Number of neurons recorded in each location (*n*) and number of selective neurons. The percentages for the number of selective neurons correspond to the number of recorded neurons in each location. "Vis. & Im." indicates those neurons selective during vision and imagery. For these, the percentages are based on the total number of imagery selective neurons. "Vis. & Im. same" indicates the number of neurons that showed the same selectivity during vision and imagery. For these, the percentages are based on the "Vis. & Im." row. Locations: Amy, amygdala; EC, entorhinal cortex; Hip, hippocampus; PHG, parahippocampal gyrus. Of the 429 neurons, 66% were in the right hemisphere (67% in the Amy, 63% in the EC, 77% in the Hip, 44% in the PHG). The differences in selectivity between the right and left hemisphere are discussed in the text.

5 Neuronal activity during sleep

5.1 Introduction

Although we spend approximately a third of our lives sleeping⁷², we still do not understand why we need to sleep. This has long been an outstanding question for scientists and philosophers. Sleeping is not clearly controlled by our volition. Even in circumstances where it is extremely dangerous to fall asleep, for example while driving, people do fall asleep. The prevalent view of sleep (even today) is that it consists of a resting state: after thinking and working for a whole day, the brain needs to take some time to save energy and be prepared for the following day. In modern day terminology, this would be analogous to suspending the computer when it is not being used or shutting it down at night. This view of sleep as a passive recess period has ancient roots. According to Greek mythology Hypnos⁷³, the God of sleep, was the brother of Thanatos, the God of Death. His sons were the "bringers of dreams": Morpheus (people), Icelus (animals) and Phantasus (inanimate things). According to many religions, it is God's will to decide whether we will wake up after sleep or "continue sleeping." In 1834, in one of the few books on sleep of those times, it was stated that sleep is an intermediate state between wakefulness and death (MacNish, 1834.)

This view of sleep has changed radically over the past 50 years due to multidisciplinary research on the psychological, electrophysiological, behavioral and functional aspects of sleep. One of the most important changes in sleep research was the introduction of the scalp electroencephalogram (EEG) to measure in a non-invasive way the activity of the brain during the sleep period (Berger, 1930). Nathaniel Kleitman and

⁷² This is an approximate average value and the reader should not be concerned if he thinks that he is far below or above average.

⁷³ Somnus in the roman literature.

colleagues showed the existence of stereotyped EEG characteristics that are repeated in a cyclical fashion throughout the night (Dement and Kleitman, 1957a, Kryger *et al.*, 1994). One of the most important observations was the distinction between slow-wave sleep (SWS) and rapid-eye-movement (REM) sleep. The former is characterized by low frequencies and high amplitudes in the EEG whereas high frequencies and low amplitudes are prevalent in the latter. In addition, as the name implies, REM is characterized by rapid movement of the eyes as observed in the electro-oculogram (EOG) and a flat muscle tone evidenced in the electro-myogram (EMG) (Aserinsky and Kleitman, 1953, Dement and Kleitman, 1957a). Based on the amplitude and frequency of the EEG, four different stages of non-REM sleep have been described. These are typically referred to as Stage 1 through 4 and are characterized by increasing amplitudes and lower frequencies. The EEG activity during REM is indistinguishable from that during the wake state (Figure 5-1). The seminal work of Aserinsky, Dement and Kleitman showed that there is a strong correlation between REM sleep and the phenomenon of dreaming (Aserinsky and Kleitman, 1953, Dement and Kleitman, 1957b, Hobson, 1988, Hobson, 1995). When subjects are woken up during a slow wave sleep period they may relate some kind of thought or discussion on what they were thinking about but typically they do not report the vivid aspects characteristic of dreams. When the same subjects are woken up during a REM sleep period, in about 80% of the cases, they typically give upon debriefing descriptions of vivid visual hallucinatory dreams. The EEG, in combination with the EOG and the EMG has become one of the most fundamental tools to study sleep.

The biophysical origin of the EEG signals is still not fully understood. The EEG signal provides an average of cerebral activity over relatively large regions. It is thought to arise as a consequence of the synchronous activity of large ensembles of neurons (Kandel *et al.*, 2000). According to this view, the correlated firing of a large group of neurons (on the order of several millions) would be evidenced at the scalp level by weak but clear electrical signals measured in the EEG. This does not necessarily mean that the electrical signal is some form of average of the spikes within the neuronal network. The scalp signal could very well be composed of a weighted average of the corresponding excitatory post-synaptic potentials (EPSPs) or dendro-somatic components.

The vivid visual nature of dreams⁷⁴ constitutes yet another clear example in which what we perceive is dissociated from the physical input to the retina. In most cases, subjects sleep with the eyes closed and there is no incoming visual input whatsoever during the period of dreams (even in cases where people have their eyes open or semi-open during sleep, dreams are only weakly affected by the visual input). Therefore, dreams represent a fascinating opportunity (albeit an extremely difficult one) to study the neuronal representation of visual perception in the absence of concomitant retinal activity.

We therefore decided to commence by investigating the activity of ensembles of single neurons in the human brain during different wake and sleep periods. In this Chapter, I will describe our quantitative study of the spontaneous firing activity of individual neurons and the correlations among them. First, I will report the firing rates and coefficients of variation of individual neurons during different epochs of the sleep-wake cycle. Then, I will describe the correlations between pairs and triplets of neurons and the modulation in synchronous activity during sleep. I will finally compare our results to those obtained by electrophysiological exploration of animals and other techniques used to study the human brain and speculate on the possible implications and interpretations of the data. At the time of writing this Chapter, a summary of these results has been submitted for publication (Kreiman *et al.*, *submitted*).

5.2 Methods

5.2.1 Polysomnographic studies

We directly investigated the neuronal processes during sleep in the human brain by recording the activity of multiple single neurons. Subjects were patients with pharmacologically intractable epilepsy who were implanted with chronic depth electrodes

⁷⁴ Different people remember their dreams more often than others. But most normal people experience the phenomenon of dreaming. The exceptions to these are congenitally blind people or patients with specific neurological disorders (Solms, 1996). It is interesting to note that most of the neurological patients that do not experience visual dreams also show deficits in visual imagery (Solms, 1996).

in order to localize the seizure focus for possible surgical resection as described in Chapter 2. Prior to depth electrode implantation, patients gave their informed consent for participation in these studies under the approval of the UCLA Internal Review Board. All the data described in the current Chapter were acquired by the laborious endeavor of Rick Staba and Charles Wilson from UCLA. The analysis of the data was performed together with a summer research student, Jonathan Lin.

Polysomnographic sleep studies were conducted on the hospital ward within each subject's room. Neuronal recordings were acquired 48-72 hours following surgery and typically began between the hours of 10pm and midnight and ended at 6am the following morning. Sleep staging was carried out according to the criteria of Rechtschaffen and Kales (Rechtschaffen and Kales, 1968). The sleep record consisted of two electro-oculogram (EOG) leads recording eye movements, two electro-myogram (EMG) leads placed on the chin to record sub-mental muscle tone, and two O Flexion scalp leads placed during surgery at "10-20" positions C3 and C4 referred to the contralateral ear to record cortical EEG activity (Kryger *et al.*, 1994). Sleep stages were categorized as waking, drowsy, non-REM sleep stages 1 through 4, and rapid-eye movement sleep based on the EEG/EOG/EMG activity. Stage scoring was performed by Rick Staba and Charles Wilson.

5.2.2 Single neuron recordings in the human brain during sleep

Single units recorded during the stages defined as waking, non-REM stages 3 and 4 (from herein referred to as slow wave sleep), and REM sleep were analyzed. States of drowsiness, non-REM stage 1 and non-REM stage 2 were not included for analyses. Neuronal data for analysis were acquired during movement-free epochs of ten-minute approximate duration in well identified awareness states. These states were: quiet wakefulness (wake), slow wave sleep (SWS) and rapid eye movement sleep (REM). The *wake* period was recorded either before falling asleep or immediately upon awakening. Typically, only one epoch was recording during each state. In one patient (R.Z.) there were no clear signs of slow wave sleep activity and therefore only the wake and REM periods were recorded. Single neurons were isolated from the multi-unit activity based on the peak, width and other parameters of the waveforms. The spike sorting procedure was

performed by Rick Staba. More details about the spike sorting procedure are given in Chapter 2 and also in Appendix 1.

We recorded from a total of 143 neurons in 14 patients (9 male, 36 ± 11 years old ranging from 19 to 51). The location of the electrodes was assessed by structural magnetic resonance imaging obtained at 1.5 Tesla before removal of the probes (see Chapter 2). According to this estimation of the position of the microwires, the location of the units is indicated in Table 5-1. The number of neurons recorded per patient ranged from 7 to 15 (10.2 ± 2.6 , mean \pm s.d.). The number of neurons may slightly differ from one epoch to another because in some cases the recording was lost between epochs. There were 122, 115 and 130 units recorded during wake, SWS and REM respectively.

5.2.3 Study of significance in correlations

Several of the data analysis methods that we use in this Chapter were described in previous Chapters (see in particular, Chapter 2). In this Chapter we also study the degree of coincident firing by pairs and triplets of neurons. A pair of neurons firing independently will, by chance, show some spikes that occur at approximately the same time. It is therefore essential to understand to what degree the correlations that we measure between neurons can be due to chance. This entails carefully assessing the probability of coincident firing.

Let \mathbf{w} , \mathbf{x} , \mathbf{y} represent three spike trains typically represented as $x(t) = \sum_i \delta(t - t_i)$

where the t_i indicate the times at which the neuron fired an action potential. We will here write the spike train in the binned form $\mathbf{x} = \{X_1, \dots, X_n\}$. $X_i = 1$ if and only if the neuron fired a spike in bin i ($i = 1, \dots, n$) where n is the total number of bins in the spike train. n depends on the bin size and the total recording time. We used bin sizes of 10, 20 and 40 ms. Each recording epoch lasted approximately 600 seconds (674 ± 252 secs; range = 359 to 1547 secs). The raw cross-correlation between two neurons can be computed as

$$R_{xy}(\tau) = \sum_i X_i Y_{(i+\tau)} \equiv X \circ Y \quad 1.$$

and represents a value proportional to the probability that neuron \mathbf{y} fires τ bins after neuron \mathbf{x} . This can be normalized to indicate the number of coincidences per second

(there are several other ways of normalizing the cross-correlogram but this is the one we will use in the plots). The extension to the three-point correlation requires a new lag parameter:

$$R_{wxy}(\tau_1, \tau_2) = \sum_i W_i X_{(i+\tau_1)} Y_{(i+\tau_2)} \quad 2.$$

Let us assume that we have two neurons, each firing at an unknown rate. We record the activity in an interval of time (with n bins) and we count the number of spikes in each neuron: $X = \sum_{i=1}^n X_i$ and $Y = \sum_{i=1}^n Y_i$. Let $X=k$ and $Y=l$. We measure also the number of coincident spikes C ($C = \sum_{i=1}^n X_i Y_i$) and obtain $C=m$. In order to assess the significance of the measure C we need to evaluate the probability of obtaining the count m given the individual counts k and l in each neuron in a total of n bins. Assuming independent firing, this probability follows a hypergeometrical distribution given by (Aersten *et al.*, 1989, Perkel *et al.*, 1967, Palm *et al.*, 1988)

$$pr[C = m \mid X = k, Y = l] = \frac{\binom{l}{m} \binom{n-l}{k-m}}{\binom{n}{k}} \quad 3.$$

We can easily evaluate the first two moments of this distribution

$$E(C \mid X = k, Y = l) = \frac{kl}{n} \quad 4.$$

$$V(C \mid X = k, Y = l) = \frac{1}{(n-1)} k(1 - \frac{k}{n}) l(1 - \frac{l}{n}) \quad 5.$$

This can be expressed in terms of the estimated rates $\kappa=k/n$ and $\lambda=l/n$:

$$E(C \mid \kappa, \lambda) = n\kappa\lambda \quad 6.$$

$$V(C \mid \kappa, \lambda) = \frac{n^2}{n-1} \kappa(1-\kappa)\lambda(1-\lambda) \quad 7.$$

If the firing probabilities are known ($pr[X_i = 1] = p$ and $pr[Y_i = 1] = q$ for every i), then we can estimate the probabilities of coincident counts by writing:

$$pr[C = m] = \sum_{i=1}^n pr[C = m | X = k] pr[X = k] \quad 8.$$

The conditional probabilities can be computed by assuming without loss of generality that $X_1, \dots, X_k = 1$ and $X_{k+1}, \dots, X_n = 0$, therefore, $C = \sum_{i=1}^k Y_i$. Assuming a binomial distribution:

$$pr[C = m | X = k] = b(m; k, q) = \binom{k}{m} q^m (1-q)^{k-m} \quad 9.$$

and therefore,

$$pr[C = m] = \sum_{k=1}^n b(m; k, q) b(k; n, p) \quad 10.$$

and the first two moments are

$$E[C] = npq = np_{coinc} \quad 11.$$

$$V[C] = n(pq - p^2q^2) = np_{coinc}(1 - p_{coinc}) \quad 12.$$

Adding a third neuron, with a spike count $W = \sum_{i=1}^n W_i$ and letting $W=j$ equation (3)

above gets converted into

$$pr[C = m | W = j, X = k, Y = l] = \frac{\binom{k}{m} \binom{n-k}{l-m} \binom{n-m}{j-m}}{\binom{n}{l} \binom{n}{j}} \quad 13.$$

from which we can compute the first two moments:

$$E[C | W = j, X = k, Y = l] = \frac{klj}{n^2} \quad 14.$$

$$V[C | W = j, X = k, Y = l] = k(k-1)l(l-1)j(j-1)/(n^2(n-1)^2) \quad 15.$$

In terms of the estimated firing probabilities

$$E[C | \iota, \kappa, \lambda] = n\iota\kappa\lambda \quad 16.$$

$$V[C | \iota, \kappa, \lambda] = \iota(\iota n - 1)\kappa(\kappa n - 1)\lambda(\lambda n - 1) \frac{n}{(n-1)^2} \quad 17.$$

If the firing probabilities are known ($pr[W_i = 1] = o$, $pr[X_i = 1] = p$ and $pr[Y_i = 1] = q$ for every i), then the probability of coincidence (assuming independence) in any given bin is given by

$$p_{coinc} = opq \quad 18.$$

and the first two moments for the number of coincidences in n bins is simply:

$$E[C] = nopq = np_{coinc} \quad 19.$$

$$V[C] = nopq(1 - opq) = np_{coinc}(1 - p_{coinc}) \quad 20.$$

We also compared the results from the analytical equations above to a simulation of Poisson neurons (Gabbiani and Koch, 1998). For each neuron we fitted a homogeneous Poisson process with the rate given by the mean interspike interval of the unit. For each pair and triplet of neurons we ran 10,000 simulations of 10 seconds duration for each of the neurons. Spikes were binned as before and we computed for each simulation the number of coincidences between pairs of neurons (C_2) or triplets of neurons (C_3) from equations 5.1 and 5.2. As expected, there is a strong correlation between the values of the first two moments as obtained from the simulations or the analytical equations (both for two-point correlations and three-point correlations).

The above two methods require strong assumptions about the structure of the spike trains. Separate bins are assumed to be independent and it is unlikely that this is true at least at short time intervals in the nervous system. The assumption that interspike intervals are exponentially distributed (as in a Poisson process) also poses strong restrictions on the neuronal firing. For a large number of neurons, an exponential distribution can very well fit the interspike intervals (see Chapter 2). However, a large fraction of neurons show bursts of spikes that give rise to a bimodal ISI distribution and cannot be accounted for by a homogeneous Poisson process. We therefore implemented a model-free procedure to assess the significance of the coincident firing. Assuming that there are no long-term⁷⁵ correlations between the spike trains, we divided the spike trains into segments of T seconds each. We used the following values of T : 5, 10 and 20

⁷⁵ The exact meaning of “long-term” requires further clarification. As described below, we compared segments of >5 seconds duration. Hence, this procedure requires assuming the absence of precise correlations in the spike train beyond 5 seconds. This does not mean that there cannot be neuronal processes that occur with a periodicity of approximately 0.2 Hz. But we do assume that whatever those processes are, they cannot cause a correlation of <10 ms width between the spike trains.

seconds. If the total length of the spike train is L , this gives us $S=L/T$ segments that we label $\mathbf{x}^1, \dots, \mathbf{x}^S$ and $\mathbf{y}^1, \dots, \mathbf{y}^S$. Computing the cross-correlation as indicated in equation (5.1) is similar to computing

$$R(\tau) = 1/S \left(\sum_{i=1}^S R_{x^i y^i}(\tau) \right) = \langle x^i \circ y^i \rangle \quad 21.$$

where $\langle \rangle$ represents the average across segments. This was compared to the distribution of all possible permutations⁷⁶: $R_{x^i y^j}(\tau)$ by computing the deviations in the correlogram after subtracting these permutations. The quantitative analysis requires the following additional definitions:

$$P_x(t) = \langle x^i(t) \rangle \quad P_y(t) = \langle y^i(t) \rangle \quad 22.$$

$$\sigma_x^2(t) = E(x^2) - (E(x))^2 \quad \text{and} \quad \sigma_y^2(t) = E(y^2) - (E(y))^2 \quad 23.$$

Note that P_x and P_y are equivalent to a PSTH (post-stimulus time histogram) in the more traditional case where each segment corresponds to a repetition in the presentation of a stimulus (see Chapter 3). The expected value of R as well as the degree of departure expected by chance can be easily evaluated. The stimulus-induced effects are typically subtracted in either of two related ways by subtracting the shuffle corrector or the shift predictor. The shuffle-corrector can be defined as

$$K = \langle x^i \rangle \circ \langle y^i \rangle = P_x \circ P_y \quad 24.$$

while the shift predictor can be defined as

$$D = \langle x^i \circ y^{\Pi(i)} \rangle \quad 25.$$

where $\Pi(i)$ is a permutation of the order of the presentations. The average here was computed over all possible permutations. The shuffle-corrected cross-correlogram can be expressed as⁷⁷

$$V = \langle (x_i - P_x) \circ (y_i - P_y) \rangle \quad 26.$$

If \mathbf{x} and \mathbf{y} independent, then the expected value of V is zero. If there are departures from zero, how significant are they? Assuming that \mathbf{x} is independent of \mathbf{y} , independence

⁷⁶ There are a total of $S(S-1)$ different permutations excluding the diagonal terms and S^2 total pairs.

⁷⁷ Note that $V = \langle x_i \circ y_i \rangle - \langle P_x - P_y \rangle = R - K$. Analogously, one can define $V_D = R - D$. Assuming that different segments are independent, then $E(V) = E(V_D)$ (Brody, 1999).

between different trials, and independence in different bins within each trial, the variance in the null hypothesis of V is (Brody, 1997):

$$\sigma_V^2 = \frac{1}{S} (\sigma_x^2 \circ \sigma_y^2 + \sigma_x^2 \circ P_y^2 + \sigma_y^2 \circ P_x^2) \quad 27.$$

The same procedure can be easily extended to the correlations between triplets of neurons (although with a non-trivial increase in the computational cost). Significance in departures from $V=0$ have typically been assessed by comparing V to the standard deviation in the null hypothesis σ_V . One problem with this approach is the strong assumption of independence between different bins within the same repetition. In order to address this issue, we also compared these results with those obtained using a Fisher permutation procedure to evaluate the significance in the departures from the null hypothesis (Efron and Tibshirani, 1993, Steinmetz *et al.*, 2000).

We used bin sizes of 10, 20 and 40 ms and threshold of 2, 3 and 4 sigma values. Overall, the general findings to be reported below were not dependent on the specific method used to analyze the level of correlations or the different parameters involved. We directly compared all the methods in neuronal simulations. By and large, all these methods yield similar boundaries of rejection of the null hypothesis of independent firing. An example of the boundaries and results from each of these methods is illustrated in Figure 5-5. The analytical method shows the expected value and two standard deviations for the number of coincidences in Figure 5-5a (given by equations 5.6 and 5.7). The cross-correlogram after subtracting the shuffled correlograms using segments of 10-second duration is shown in Figure 5-5b. An example of a correlogram of two Poisson neurons with firing rates fitted to the two neurons is shown in Figure 5-5c. Finally, the distribution of all the peaks of the cross-correlograms from 10000 iterations of the Poisson neurons is shown in Figure 5-5d together with the actual value of the peak (indicated by *). The synchronous activity of this pair of neuron was detected by all methods and parameters. There were, however, some weaker correlations where the degree of significance did vary from one statistical assessment procedure to another. It is not practical to give all the results for every possible combination of methods and parameters. I will argue, however, towards the end of this Chapter that the main conclusions that we draw were robust to alterations in these methods and variables.

In order to compare the correlations between two neurons obtained during different epochs, we computed the peak, width and area of the cross-correlograms. The properties of the correlograms were computed after interpolating by fitting a cubic spline (Dierckx, 1993) with an upsampling factor of 10. The width was computed at half the peak height after subtracting the baseline value. The area was computed by numerically integrating the correlogram between -100 and 100 ms (Press *et al.*, 1996). The strength of the correlogram was measured as the ratio of the peak value to the baseline.

5.3 Single neuron activity during sleep

We first sought to determine if there was any clear large-scale modulation in the mean activity of the neurons during the dramatic changes in the state of awareness that occur during sleep. Figure 5-2a-c shows the overall distribution of firing rates during each epoch. The median values of the firing rates are indicated in Figure 5-2d. There was no significant difference in the global firing rates among wake, SWS and REM sleep periods (one-way ANOVA, $p > 0.2$). It is conceivable that there could be differences in firing rates only in specific locations. Figure 5-2e shows the mean firing rate within 5 regions for which the number of neurons was large enough to perform a meaningful comparison ($n \geq 5$ in all epochs). Again, the analysis of variance failed to yield any significant difference in the firing rates among the three states in any of the regions (one-way ANOVA, $p > 0.1$). There seems to be a slight decrease in the firing rate of SMA units during the wake state but the small number of units ($n = 5, 8$ and 9 in each epoch respectively) precludes from obtaining a statistically significant difference.

We next studied in further detail the temporal structure of the neuronal activity of single units. We computed the coefficient of variation, CV (Gabbiani and Koch, 1998), for every unit; the distribution for each epoch is shown in Figure 5-3a-c. There was a small but marginally significant difference in the CV across epochs (one-way ANOVA, $p < 0.05$); the variation in interspike intervals was slightly larger during SWS (Figure 5-3g). This was particularly evident in the hippocampus but it was also statistically significant in the entorhinal cortex (Figure 5-3i). We also computed the CV_2 , a measure

of within-trial variation that is more robust to large changes in instantaneous firing rate and bursting activity (Holt *et al.*, 1996). The distribution of CV_2 values is shown in Figure 5-3**d-f**. There was no statistically significant difference among the CV_2 during the three epochs. The average value of CV_2 was very close to 1 (as expected for a Poisson process) for all the regions that we studied Figure 5-3**j**.

Given these small differences in the temporal structure of the spike trains as evidenced by the differences in CV but not in CV_2 , we speculated that there could be a different proportion of bursting units or bursting activity during the three different sleep epochs. Several investigators have described that neurons in the MTL, particularly in the hippocampus, can fire strong bursts of activity (see for example (Csicsvari *et al.*, 2000, Buzsaki, 1989); see also Chapter 2). The interspike interval distribution of some neurons could be very well fitted by an exponential distribution (Chapter 2, Figure 10**a**) after shifting to account for the refractory period. In several other cases, however, an exponential distribution gave only a very poor description of the distribution of ISIs (Chapter 2, Figure 10**b**). A bimodal distribution was evident in most of those cases showing a strong sharp peak at short intervals and broader and weaker peak at longer intervals. This is typical in neurons firing in bursts (Gabbiani *et al.*, 1996, Bastian and Nguyenkim, 2001). In order to evaluate quantitatively whether the neuron fired in bursts or not we computed the autocorrelogram for the neuronal spike train. For a Poisson process, the autocorrelogram is flat and equal to the mean firing rate of the neuron (that is, the probability of firing another spike τ ms after a spike is constant, regardless of the value of τ). This is the case for the neuron whose autocorrelogram is shown in Figure 10**c** in Chapter 2. The decrease at low time lag values near zero is due to the refractory period. We computed the 99.9% confidence intervals for the autocorrelograms under the assumption of a Poisson process (Abeles, 1982). The bursty neuron deviated significantly from this assumption (Chapter 2, Figure 10**d**). The maximum interspike interval corresponding to a burst was computed from the time at which the peak in the autocorrelogram returned within the confidence interval limit.

The percentage of units that showed bursting behavior was 33% (40/122), 57% (65/115) and 38% (49/130) during wake, SWS and REM sleep (see Figure 5-4). The

maximum interspike interval for bursts were 38 ± 18 , 40 ± 20 , 39 ± 18 ms⁷⁸ (ranging from 8 ms to 89 ms) and there was no significant difference among the three epochs. Within those cells that showed bursting behavior, we computed the total proportion of spikes that occurred in bursts. Approximately half of the spikes in bursting cells occurred within bursts regardless of the sleep epoch.

5.4 Synchronous activity during sleep

It is quite striking that one can record any electrical activity at the level of the scalp by using EEG. Since its inception, this was taken to imply the coordinated activity of large number of neurons. The localization resolution of EEG, however, does not permit to assess accurately which neuronal networks in the brain give rise to the scalp signal. Single unit studies in animals had shown that pairs of neurons could show synchronized activity during the sleep period, particularly after the animal had practiced some specific tasks (Wilson and McNaughton, 1994, Skaggs and McNaughton, 1996, McNaughton, 1998, Louie and Wilson, 2001). The relevance of these experiments to memory processing will be discussed later in this Chapter (see Section 5.5). We therefore explored the degree to which pairs and triplets of neurons in the human brain fired in close temporal synchrony during the different periods of the wake-sleep cycle.

5.4.1 Increased synchrony during slow wave sleep

An example of the neuronal activity of three separate neurons in one of the patients in each of the three epochs is illustrated in Figure 5-6. These neurons were

⁷⁸ This is typically too long for a normal Ca^{2+} burst (Kandel *et al.*, 2000). Intracellular recordings typically show bursts of less than 10 ms duration. The maximum interspike interval for bursts in pyramidal cells of the weakly electric fish are typically less than 10-20 ms. The low firing rates of these neurons make it more difficult to precisely estimate the duration of the bursts. Upon considering only neurons with firing rates beyond 1 spike/sec and excluding those bursts longer than 50 ms the vast majority of the bursts were less than 20 ms suggesting that the longer values here reported are due to anomalous long "bursts" of slowly firing units.

recorded simultaneously from 3 different microwires located in the right hippocampus⁷⁹. The neurons tended to fire coincidentally within small time windows of <50 ms more often during the slow wave sleep epoch (Figure 5-6b) than during the other two periods (Figure 5-6 a, c).

We quantitatively evaluated the degree to which pairs and triplets of neurons fired in close temporal synchrony. We analyzed a total of more than 400 pairs and more than 1000 triplets of neurons. The total number of neuronal pairs and triplets analyzed during each epoch is indicated in Table 5-2. Cross-correlograms were computed between any possible pair or triplet of neurons recorded simultaneously within each subject. The cross-correlograms for the same units shown in Figure 5-6 are shown in Figure 5-7. The two-point correlograms show that the three pairs of neurons enhanced their synchronous activity during SWS (Figure 5-7c) compared to both the wake state (Figure 5-7a) and REM sleep (Figure 5-7e). The peak in the correlograms during SWS were 0.20, 0.33 and 0.93 coincidences/sec and occurred at time lags of -1, -2 and +1 ms for the three pairs respectively. The peak-to-baseline ratios were 2.9, 5 and 5.3. This was highly significant ($p < 0.001$) and not due to random coincident firing as assessed by comparing to the null hypothesis of independent firing responses (see Section 5.2.3). The peak to baseline ratios during the wake state or REM sleep were less than 1.1 and were not significant ($p > 0.2$). Furthermore, the three neurons tended to fire simultaneously more often during SWS (Figure 5-7d) than during the wake state (Figure 5-7b) and REM sleep (Figure 5-7f). The peak values in the three point correlograms were 0.02, 0.11 and 0.01 coincidences per second during wake, SWS and REM sleep. The coincident firing observed during SWS was highly significant ($p < 10^{-4}$) whereas during the wake or REM sleep states the number of coincident spikes could not be distinguished from random simultaneous firing ($p > 0.2$). The peak during SWS was centered near zero, achieving the maximum at time lags of 3 and -10 ms.

It should be noted that synchrony does not follow the transitive property. In other words, it is conceivable that a neuron A fires in synchrony with another neuron B, B fires in synchrony with C but A does not fire at the same time than C. This is due to the fact

⁷⁹ Two microwires within the same electrode could be up to a few mm apart (see Chapter 2 for a description of the electrodes). It is unclear how to set a lower bound on how close two microwires could be. The distance between separate electrodes was larger, encompassing one to several centimeters.

that when two neurons are synchronized not all spikes occur in close temporal alignment. We have shown this in very simple neuronal models (not shown here). If all spikes were temporally correlated, then synchrony would necessarily follow the transitive property. The non-transitivity implies that three-point correlations are not necessarily a consequence of two-point correlations⁸⁰. In our data, however, in most of the cases where we observed synchronous activity between three pairs of neurons, the corresponding triplet *also* showed synchronous activity.

Figure 5-8 illustrates all the neuronal pairs that showed significantly enhanced simultaneous firing for one patient during each of the three epochs. Only those neurons that were recorded during the three epochs are shown in this figure. The number of pairs exhibiting synchronized activity was larger during SWS. Furthermore, in those cases where the same pair of units showed significantly synchronized activity during two epochs, correlations were stronger during slow wave sleep (indicated in the figure by the color code of the lines joining two units). It is interesting to note that most of the pairs of neurons that showed correlated activity were recorded from two microwires within the same electrode; this is discussed in further detail below.

There was a large degree of variability across subjects. As we have discussed in previous Chapters, there were several variables that changed from one subject to another (see in particular the discussion about subject-to-subject variability in Chapter 3). The number of pairs recorded in all three epochs also varied significantly from one subject to another and this is one of the factors that make across subject comparisons difficult. Of course, the total number of pairs that show synchrony depends on the number of pairs that can be recorded from. The proportion of pairs that are synchronized, however, can be artificially high or low in cases where the number of pairs is very low.

Overall, synchronous activity was larger during SWS than during the other epochs and the wake state showed an intermediate value both for pairs (Figure 5-9a) and triplets of neurons (Figure 5-9b). The proportions of pairs that showed significant synchronized firing were 11.1%, 20.2% and 7.8% during wake, SWS and REM sleep states

⁸⁰ The non-transitive character of synchrony could potentially prove to be very important in the encoding of information. One of the roles that has been suggested for synchrony is in solving the so-called binding problem (Singer and Gray, 1995, Singer, 1999). A single neuron A could participate in representing information about two different stimuli by firing some spikes in close temporal synchrony to another neuron B and other spikes whenever a third neuron C fires.

respectively and the proportions of triplets that showed significant correlated activity were 2%, 8% and 3% during wake, SWS and REM sleep states respectively. There were very few cases of inhibitory correlograms (that is, where there was a significant decrease in the probability of simultaneous firing). The average proportion of inhibitory correlograms was $1\% \pm 0.3\%$ for pairs of neurons, less than 1% for triplets of neurons and there was no clear difference across epochs. The overall firing rates were 2.4 ± 4.5 , 2.6 ± 6.9 and 2.3 ± 6.6 (median \pm s.d.) for wake, SWS and REM respectively. Given the low firing rates of the neurons, it is statistically very difficult to detect a decrease in the probability of simultaneous firing.

Due to experimental constraints, in some cases we could not record from the same neuron in the three epochs. Eighty-four units (out of the total of 143 neurons) were recorded in all three epochs. Of the total number of pairs recorded in all three epochs, ninety-one pairs (32%) showed coherent activation during at least one of the three epochs. Of these, 16% showed significantly increased synchrony in all three epochs, and 38% showed synchrony only during SWS (Figure 5-9c). We directly compared the strength of the correlations for those pairs that showed synchrony in more than one epoch (see Section 5.2.3). We observed that correlations were significantly stronger during SWS (Figure 5-9d). For the neurons synchronized in all 3 epochs, the peak to baseline ratios of the correlograms were 1.9 ± 1.1 , 2.6 ± 1.0 and 1.4 ± 0.4 for wake, SWS and REM sleep. There was a statistically significant difference among the three epochs (one-way ANOVA, $p < 0.05$). For those pairs synchronized only during wake and SWS the ratios were 1.9 ± 0.7 and 2.5 ± 1.3 respectively (t test, $p < 0.05$). For those pairs synchronized only during SWS and REM the ratios were 2.4 ± 1.0 and 1.5 ± 0.4 respectively (t test, $p < 0.05$). Thus, not only more pairs and triplets show synchronous activity during SWS but also the coincident firing is stronger during SWS compared to the wake and REM sleep states.

The distribution of the widths of the correlograms (see Section 5.2.3) is shown in Figure 5-9e. The mean overall value for the width was 28 ± 19 ms. There was no significant difference among the three epochs (one-way ANOVA, $p > 0.1$). Therefore, correlations occur within the same temporal scale in the three epochs and the stronger synchrony is due to a larger number of coincident spikes within similar time windows.

It should be noted that we recorded neuronal activity from areas both outside and within the seizure focus. Therefore, caution should be taken in extrapolating the results. It is particularly important to address this question given that hypersynchrony has been associated with absence seizures. The location of the seizure focus was determined for each patient by the clinical team based on the data acquired from the depth electrodes as well as other clinical measurements and tests (see Chapter 2). Sixty-two out of the one 143 neurons (51 in the temporal lobe, 11 in the frontal lobe) were inside the clinically determined seizure focus. We repeated our analyses considering only those neurons outside the seizure focus. Here we define a neuron to be outside the seizure focus by considering those units recorded either in the same hemisphere but within a different brain lobe or in the contralateral hemisphere. In two of the patients, all the recorded units were within the seizure focus; therefore they were excluded from this analysis. The overall percentage of synchronized pairs of units were 10%, 17% and 6% and the overall percentage of synchronized triplets of units were 2%, 6% and <1% for wake, SWS and REM sleep states. The observation that there was no major change in the percentages of synchronized pairs or triplets suggest that seizure activity cannot account for the results that we have described in this Chapter. Some of the pairs and triplets of neurons that showed correlated activity in our overall summary in Figure 5-9 were within the seizure focus. However, the stronger correlations during SWS remain after eliminating all the units inside the epileptic focus.

Activity during the wake state was recorded after “lights out” before subjects fell asleep or in the morning immediately after awakening. It should be noted that this is quite different from several of the active exploration tasks studied in rats (see for example Wilson and McNaughton, 1994, Skaggs and McNaughton, 1996, McNaughton, 1998, Sutherland and McNaughton, 2000, Nadasdy *et al.*, 1999). In several of those studies, it has been found that pairs of neurons that show synchronous activity during the task in the wake state show a similar pattern of activation during the sleep period. In our case, synchronous activity was reduced during the wake state; this may be due to the lack of an attentionally demanding active task. We compare in more detail our results to those obtained in animal electrophysiology experiments in Section 5.5.

Most of the synchronized pairs of neurons corresponded to local interactions within the same brain area. Less than 7% of the synchronized pairs included units in different hemispheres (Figure 5-10a). However, 18% of the synchronized pairs (20% during the awake state, 17% during slow wave sleep and 17% during REM sleep) showed correlations between different electrodes (Figure 5-10c)⁸¹. Of these, the strongest interactions corresponded to those between the entorhinal cortex and the hippocampus, encompassing 75% of the total number of across-areas correlations⁸².

Most of the spikes ($76\pm 23\%$) that occurred coincidentally in two neurons appeared within short bursts of activity within the individual units (see Figure 5-4c-d). Bursts have been proposed to play an important role in information transmission. This is in due to the large efficiency they may achieve on transmitting a message across a synapse and eliciting an EPSP in a postsynaptic neuron. Synchronized bursts could have an even larger efficacy in information transmission. Therefore, it is interesting to observe that the majority of coincident events occurred between bursts of spikes and not isolated spikes.

5.4.2 Some technical issues about the analysis

As discussed above, the observed correlations were strongest among units that were on the same electrode. Since units in different electrodes were in different brain regions and separated by large distances, this observation could be related to a decrease in the possibility of detecting correlations across different brain regions. This could be due to an enhanced number of local connections between nearby neurons than between neurons in separate brain areas. One possible confounding factor, however, would be the existence of cross talk between microwires on the same electrode. Such an artifact, however, could not explain the differences that we observed across different awareness epochs and therefore seems unlikely. However, we took several precautions to address this question. First, an electrical artifact would lead to perfectly synchronous spikes

⁸¹ In most cases, this corresponds to different brain areas. In less than 5% of the cases were two electrodes targeted to anterior and posterior regions of the hippocampus. Except for these cases, two separate electrodes were always situated in different brain regions.

⁸² While the low number of pairs between frontal cortex and temporal lobe that were analyzed preclude from drawing clear conclusions, it is interesting to note that two out of eight pairs (25%), three out of ten pairs and none out of six pairs showed synchronous activation during the wake, SWS and REM sleep periods respectively (Figure 5-10b). The communication between the temporal lobe and frontal cortex has been suggested to play a prominent role in processes of memory consolidation (Siapas and Wilson, 1998).

(within the order of magnitude of the sampling frequency) between different microwires in the same electrode. These correlations should disappear after subtracting the perfectly matching spikes. We removed those synchronous spikes that occurred in different microwires within a time lag $-2\Delta t \leq \tau \leq 2\Delta t$ where Δt is the sampling step (0.1 ms in our case). The proportion (and identity) of the coincident pairs was basically the same and none of the conclusions presented before were modified. The perfectly synchronous spikes were removed from all the pairs and triplets of neurons claimed to show significant correlations in the current Chapter. We also computed the correlation between the raw signals before high-pass filtering (continuous signal, 0.5 Hz high passed). This was evaluated by computing the r^2 correlation coefficient between the two signals. I also computed the cross-correlogram; r^2 is proportional to the peak in the correlogram at $\tau = 0$ ms lag. This was done for signals between microwires within different electrodes and microwires within the same electrode. There was no difference between the correlation coefficients in the two cases (t test, $p > 0.1$). It should be noted, however, that this was done from data collected in RC Electronics. It is not the same data, not even the same patients. The correlation coefficients between wires were clearly positive (regardless of whether they came from the same or different electrode). This suggests that there is some common noise source such as 60 Hz or movement artifacts. Movement artifacts or 60 Hz noise are largely eliminated during the process of spike sorting and cannot account for the differences in the signals within and between electrodes. The same results were obtained after removing the spikes from the raw signal (given the sparseness of spikes, they do not contribute much to the correlation between the raw traces). Finally, the results were not altered when computing the correlations of the continuous signal after 300 Hz high pass filtering.

Some of the correlograms were computed between pairs of units recorded in the same microwire. Care should be taken in the interpretation of these correlograms. If the spike sorting procedure yields an inaccurate classification, the cross-correlograms could actually include pairs of spikes from the same unit. A unit that fires short bursts of activity, for example, could yield a peak at a short time lag in the correlogram that would be indicative of the bursting activity rather than any synchronous activation of two separate units. However, in our data, most (84%) of the 143 neurons that we recorded

from (see Table 5-1) correspond to units from separate microwires. Of the total number of pairs that showed statistically significant synchronization, less than 10% come from neuronal pairs recorded from the same microwire (Figure 5-10d). Therefore, our conclusions would not be altered if we excluded those pairs of neurons from the analysis.

Whether the coincident firing that was observed constitutes a significant deviation from what would be expected by chance constitutes a crucial issue when studying synchronous activity. We therefore used three different methods to assess the degree of departure from the null hypothesis of independence (see Section 5.2.3 and Figure 5-5). The first two assessment methods make strong assumptions about the structure of the neuronal spike train (in one case independence between bins and in the second a specific distribution for the interspike intervals is assumed; see Section 5.2.3 for details). The results that we obtained and the conclusions drawn were robust with respect to changes in the parameters and statistical assessment methods that we have used. Unfortunately, space constraints preclude from giving all the values in detail for every possible combination. The number of combinations of methods and parameters is simply too large. With the first assessment method, there were 3 bin sizes \times 3 significance levels \times 2 (with/without perfectly synchronous spikes) \times 2 (including/excluding units within seizure focus) = 36 combinations. With the second assessment method, there were 3 bin sizes \times 3 significance levels \times 2 (with/without perfectly synchronous spikes) \times 2 (Poisson/gamma) \times 2 (including/excluding units within seizure focus) = 72 combinations. Finally, with the third assessment method, there were 3 bin sizes \times 3 significance levels \times 2 (with/without perfectly synchronous spikes) \times 3 (values of T) \times 2 (including/excluding those units within the seizure focus) = 108 combinations. Let us just illustrate the qualitative robustness of our results in a somewhat strange way by referring to Figure 5-11. The figure shows the distribution of the ratios of percentages of significant pairs in different epochs for all possible combinations of parameters and significance assessment methods. The number of pairs that showed synchrony during SWS was always at least 25% larger (and up to 300% larger) than the corresponding number during the wake state. Similarly, the number of significant pairs during SWS was at least 45% larger than the corresponding value during REM and could be up to 400% larger.

5.5 Summary and discussion

5.5.1 Sleep and awareness

Before comparing our results with those obtained with other techniques, it is interesting to speculate on what could be learned about the exploration of the neuronal correlates of consciousness by studying the activity of individual neurons during different stages of sleep. It is therefore legitimate to ask why it may be of interest to our main endeavor to comprehend the neuronal processes that take place during sleep. Sleep constitutes a substantial portion of our lives; it seems to be essential since sleep-deprived rats die after approximately 17 days (Rechtschaffen *et al.*, 1983).

To begin with, sleep constitutes one of the major and perhaps simplest on/off switches to consciousness. Our level of awareness can dramatically change within a few minutes (possibly even faster than that) upon falling asleep. The investigation of the mechanisms that are involved in turning this switch on and off has been a very productive and interesting area of research over the last decades. Some brain nuclei, notably the brainstem and thalamic structures are involved in mediating these transitions and can modulate the level of activity of the rest of the brain by the massive release of neurotransmitters (Hobson, 1995, Kryger *et al.*, 1994). These are generally referred to as "enabling factors" for conscious sensation (Crick and Koch, 1998, Crick and Koch, 1990). It could be argued that this does not directly pertain to the more specific question of what are the neuronal correlates of the contents of visual consciousness. According to this view, the reticular-thalamic activating mechanisms reflect the analogous to the power cord in the computer. After all, if we unplug a computer (a desktop, not a laptop), the computer does not respond but still, we would not claim that the intelligence of the computer lies in its power cord. It seems quite clear nowadays that falling asleep does not correspond to shutting down a computer. A myriad of active processes occur in the brain during sleep. Furthermore, we can use these observations to significantly constrain our

hypothesis about the neuronal correlates of consciousness. If we hypothesize that a specific brain area, say X, directly represents the neuronal correlates of the contents of our visual percepts, then this area should not be activated during non-dream sleep⁸³. This is a rather simple idea and it probably does not require further argumentation. Yet, there are several important issues about the exact interpretation of this statement. For example, what do we mean by area X being activated or not and how could we test this in an experiment? I will not elaborate on this any further here. The claim is that non-dream sleep imposes important boundary conditions on our search for the neuronal representation of our percepts even if it does not lead directly to the neuronal correlates of specific contents of visual awareness.

Perhaps as importantly to our big goal, we must consider the period (or rather, periods) of dream sleep. Our visual experiences during dreams can be as intense as (or even in some sense more intense than) those during the wake state. We can have a strong feeling of qualia during our dreams. This is an extremely complicated issue and it is not my main purpose here to write an elaborate description of dreams or their contents. There are excellent treatises on this matter by many illustrious philosophers, psychologists and neuroscientists (Freud, 1955, Hobson, 1988, Kryger *et al.*, 1994). To be concrete, I suggest (I do not have any clear proof for the moment) that if an area X (with specific networks of neurons firing in a determined way) indeed represents a neuronal correlate to the contents of specific visual percepts, it will also be activated during those dreams that arise these visual percepts.

5.5.2 How are the specific contents of dreams represented?

A fascinating question that I would be very interesting in pursuing is to try to understand neuronal correlates of the contents of dreams. In other words, how is the visual information pertaining to the contents of dreams represented in the patterns of firing of the neurons in our brains? Let us say that we have found a neuron that is selectively activated upon seeing or imagining a specific stimulus, say Clinton for the sake of argument. Would this unit also respond selectively if the subject dreamt about

⁸³ I am assuming here that we are not visually conscious during non-dream sleep. This seems to be a well-accepted observation but we should keep in mind that this is an assumption for the moment.

Clinton? It is generally a very bad strategy in Science to claim that a particular problem will not be able to be answered. History typically proves scientists who make such claims wrong sooner or later. I will argue, however, that this specific question (exactly as I formulated it) will be extremely difficult to address in a quantitative and rigorous manner. Since I find this to be a rather fascinating question, I will devote a few paragraphs to discussing the feasibility of such an experiment. We could wake up the subject during REM sleep, debrief him of the contents of his dream and analyze the neuronal response surrounding such an event. The first question that arises is what are the chances that the subject will recall a dream upon awakening. There is a rather strong correlation between REM sleep and dreams (Hobson, 1995, Hobson, 1988, Horne, 1988, Dement and Kleitman, 1957b). It has been reported that in approximately 80% of awakenings from REM sleep, subjects report a vivid dream. When subjects are awakened from SWS, in most cases subjects do not report vivid dreams. If they do recall a dream, its contents are typically mental elaborations of plans or ideas and are usually deemed to be quite different from normal dreams and lack the hallucinatory and strong visual components characteristic of REM sleep dreams. It is therefore likely that we would be able to reliably obtain a dream report by awakening the subjects during REM sleep at the expense of disrupting their sleep. It should be noted that this does constitute a small experimental problem to begin with. While it may not be a fascinating scientific issue, it represents a practical problem. The patients stay in the hospital for a short period of time. The first nights after surgery they are typically very tired. Some nights they have to be sleep deprived (for clinical reasons). And in some other cases, they are soundly asleep due to the medications they receive. Therefore, there are not many nights in which the experiment could be carried out.

One of the main difficulties, however, is that we do not know whether the subject will be dreaming about the preferred stimuli of the neurons we are recording from (Clinton in our example). It is unclear how to evaluate with our current understanding of dreams the answer to this question. There is no clear way to induce or increase the probability that the subject will dream about this stimulus. Dreams are usually correlated with the diurnal experiences and activities. Perhaps by showing the picture of Clinton a large number of times during the day, we could increase the chances that the Clinton

would appear in the subject's dream. There are some controversial reports of lucid dreaming (LaBerge *et al.*, 1981, LaBerge, 1985, LaBerge, 1990) where the subject (typically the author himself or another person well-trained in lucid dreaming) can specifically select and direct the contents of his dreams. But, by and large, if the neuron is selective to Clinton or to dolphins, we cannot guarantee that the subject will be dreaming about those stimuli. Even if the subject does dream about Clinton or the dolphins, the stimulus will most likely not be isolated but in the context of a dream. It is not very clear at least at this point how this would affect the neuronal response (if at all).

Another major obstacle is the question of timing. Even if the subject tells us that he was indeed dreaming about Clinton, we cannot know within a ms precision when he started thinking about Clinton. Several of the neurons we recorded from showed a response duration of ~ 500 ms (see Chapter 3). If our precision of when the image of Clinton appeared in the dream is within a few minutes, we would not be able to perform any serious analysis. When a subject is woken up and reports a vivid visual dream, it does not mean that he was elaborating or experiencing immediately prior to awakening. Maybe the particular episodes in the dream that the subject narrates actually took place several minutes before awakening (may be even longer; this is still an open question in sleep research). Therefore, in contrast to our visual presentation (Chapter 2), visual imagery (Chapter 3) and flash suppression (Chapter 6) experiments, we do not know when exactly to look for the neuronal response to Clinton.

Finally, another major obstacle is that we need to perform some statistics. We lack the repeatability of the situation that is essential for statistical purposes. Even if the subject does dream about Clinton, we cannot be sure that he will do so in ten successive REM periods so that we can perform some statistical tests. Maybe Clinton will appear ten times in a single REM period but we would not know when exactly these ten occurrences took place in order to align the responses. Unfortunately, all of these factors make it unlikely that we will be able to address this issue in a rigorous manner soon.

5.5.3 Synchrony, EEG and oscillatory rhythms

Slow wave sleep is characterized at the scalp EEG level by low frequencies and high amplitudes whereas higher frequencies and lower amplitudes predominate in REM

sleep and the wake state (Figure 5-1). Here we have shown that there is a clear distinction also at the level of correlated activity between single neurons in the human medial temporal lobe among these distinct states of consciousness. Although the biophysical origin of the EEG signal is unclear, the traditional interpretation proposes that signals recorded at the level of the scalp are the result of massive synchronous activity in the brain (Kandel *et al.*, 2000). Typically, it is thought that the activity that can be detected at the scalp comes from cortical neurons and thus it is unclear to what extent deep structures like the hippocampus contribute directly to the large amplitude signals observed during SWS. It is possible, however, that another structure such as the thalamus, controls simultaneously the oscillatory activity in the hippocampus and cortical structures (Steriade *et al.*, 1993).

At the cellular level, there are specific rhythms in populations of thalamo-cortical and hippocampal neurons during sleep (Csicsvari *et al.*, 2000, Steriade *et al.*, 1993, Sejnowski and Destexhe, 2000). During SWS, cellular studies have described high frequency ripple activity in the hippocampus (Siapas and Wilson, 1998, Csicsvari *et al.*, 2000, Buzsaki, 1989) and spindles in neocortical local field potentials (Siapas and Wilson, 1998, Steriade *et al.*, 1993).

5.5.4 Sleep and memory

Several theories have proposed a prominent role for sleep in the consolidation of memories both during SWS (Wilson and McNaughton, 1994, Skaggs and McNaughton, 1996, McNaughton, 1998, Nadasdy *et al.*, 1999, Buzsaki, 1989, Sutherland and McNaughton, 2000) and during REM (Crick and Mitchison, 1983, Hopfield *et al.*, 1983, Poe *et al.*, 2000). According to this view, changes in the strength of synaptic connections between neurons would take place during sleep depending on the processes that occurred during the day.

Psychophysical evidence for a role of sleep in memory consolidation comes from several experiments showing an improvement on different visual discrimination and memory tasks after sleep. This enhancement in performance is sensible to sleep deprivation even after controlling for factors such as attention and fatigue (Karni *et al.*, 1994, Gais *et al.*, 2000, Stickgold *et al.*, 1999, Stickgold *et al.*, 2000).

There is also physiological evidence showing that neuronal activity during the wake period can influence the firing in subsequent sleep epochs (Pavrides and Winson, 1989). Sleep has been shown to induce robust changes in neuronal activity in neurons of the song motor control nucleus of zebra finches (Dave *et al.*, 1998). These changes have been hypothesized to allow the motor system to gain access to sensorimotor information required for song learning. In subsequent work Dave and Margoliash showed that the timing and structure of neuronal activity elicited by the playback of a song during sleep matches activity during daytime singing. Additionally, the spontaneous activity of these neurons during sleep matches their sensorimotor activity, a form of song "replay." Hippocampal networks in rats have been shown to also "replay" the activity experienced during the day in maze exploration tasks (Nadasdy *et al.*, 1999, Wilson and McNaughton, 1994, Skaggs and McNaughton, 1996, Sutherland and McNaughton, 2000). Neurons that fired together when the rat occupied specific locations within the environment exhibited an increased probability of firing together during the subsequent period of SWS (Wilson and McNaughton, 1994). Furthermore, the temporal patterns of neuronal firing sequences during sleep were found to reliably reflect the temporal order in which the cells fired during the navigation or exploratory behavior (Skaggs and McNaughton, 1996). Investigators have suggested that the information acquired during the wake experience is replayed during the subsequent sleep period. Nadasdy *et al.* suggested that the activity during sleep is a time compressed version of that during the wake period (Nadasdy *et al.*, 1999). Neuronal replay was also found during REM sleep, albeit with a substantially different time scale (Louie and Wilson, 2001).

The memory for temporal order has been suggested to be mediated by long term potentiation (LTP) changes in synaptic strength. Synapse strength in the central nervous system can be modified according to experience. One of the most extensively studied phenomena is long-term potentiation (Bliss and Lomo, 1973, Bliss and Collingridge, 1993, Madison *et al.*, 1991, Mayford and Kandel, 1999). The extent of synapse potentiation can depend on the pattern of stimulation; the precise timing of the input being essential in determining even the sign of the change in synapse strength (Zhang *et al.*, 1998, Markram *et al.*, 1997, Magee, 1997).

In humans, areas in the medial temporal lobe, in particular the parahippocampal gyrus, the hippocampus, entorhinal cortex and amygdala, are differentially activated throughout the sleep-wake cycle according to functional imaging evidence (Braun *et al.*, 1997, Maquet *et al.*, 2000, Hobson *et al.*, 1998, Maquet and Franck, 1996, Braun and Herscovitch, 1998, Hofle *et al.*, 1997, Maquet *et al.*, 1997). The medial temporal lobe is involved in the retrieval of experiences from memory (Kreiman *et al.*, 2000b). Furthermore, electrical stimulation can induce visual recall (Penfield and Jasper, 1954, Penfield and Perot, 1963, Halgren *et al.*, 1978, Fried *et al.*, 1982). Given the prominent role that the medial temporal lobe plays in memory processes in humans (Milner, 1972, Scoville and Milner, 1957, Zola-Morgan and Squire, 1993, Alvarez and Squire, 1994, Squire and Zola-Morgan, 1991, Penfield and Milner, 1958), it is interesting to speculate that the synchronous activity that we observed during slow wave sleep could be correlated with specific changes in the strength of neuronal synapses.

5.6 Figure legends

Figure 5-1: Schematic EEG characterization of the different sleep stages

Scalp electroencephalographic (EEG) recordings in combination with electro-oculogram (EOG) and electro-myogram (EMG) recordings reveal 6 different stages based on the amplitude and frequency components of the signals. Here we show schematic sample traces of the EEG signals obtained during these different stages. The wake state is characterized by low amplitude high frequency activity. As sleep progresses, low frequencies appear prominent and the amplitude of the EEG signals increase. The amplitude reaches its maximum values during stages 3 and 4 (slow wave sleep). Upon entering the rapid eye movement (REM) sleep stage, the EEG trace appears basically identical to that during the wake state. REM sleep can be distinguished from the wake state by the rapid eye movements evidenced in the EOG and the flat muscle tone observed in the EMG (not shown in this figure).

Figure 5-2: Firing rates of single units during different states

a-c Distribution of firing rates during wake state (**a**), SWS (**b**) and REM sleep (**c**). Bin size = 1 spike/sec. The arrow at the top shows the mean value. The x axis was cut at 20 spikes/sec for clarity. The numbers of units that showed firing rates beyond 20 spikes/sec were 4, 2 and 5 during wake, SWS and REM sleep respectively. The data from all the units recorded in each epoch were used to compute these distributions. **d**. Median firing rate of all single units during each epoch. We show the median value since the mean shows a larger value due to a few outliers with extremely high firing rates. Error bars show the s.e.m. The firing rate was computed over the entire epoch. **e**. Mean firing rate for each epoch (wake = black, SWS = gray, REM sleep = white) within each of five specific regions: Hipp = hippocampus, EC = entorhinal cortex, Fr = frontal cortex, SMA, supplementary motor area, Sub = subiculum (the other regions are not shown because there were less than 5 units recorded in all epochs).

Figure 5-3: Coefficient of variation during different states

a-c Distribution of coefficient of variation (CV) during the wake (**a**), SWS (**b**) and REM sleep (**c**) epochs (see text for the definition of CV). Bin size = 0.1. The arrow at the top shows the mean value. The x axis was cut at 2 for clarity. The numbers of units that showed a CV value beyond 2 were 15, 7 and 5 during wake, SWS and REM sleep respectively. **d-f** Distribution of CV_2 during the wake (**a**), SWS (**b**) and REM sleep (**c**) epochs (see text for the definition of CV_2). Bin size = 0.1. The arrow at the top shows the mean value. The data from all the units recorded in each epoch were used to compute the distributions in **a-f**. **g**. Mean CV of all single units during each epoch. Error bars show the s.e.m. **h**. Mean CV_2 of all single units during each epoch. Error bars show the s.e.m. **i**. Mean CV for each epoch (wake = black, SWS = gray, REM sleep = white) within each of five specific regions (see previous figure for location abbreviations). **j**. Mean CV_2 for each epoch within each five specific regions.

Figure 5-4: Spiking and bursting

a. Percentage of neurons that showed bursting behavior. The criteria to define a unit as bursting or not-bursting were described in the text (see Chapter 2). The percentages were computed over the total number of neurons recorded in each epoch. **b**. Percentage of the total number of spikes that occur in bursts. The burst duration limit was determined from the deviations from the expected values for a Poisson process in the autocorrelogram (see Chapter 2). Based on this duration, a spike was considered to be part of a burst if the corresponding ISI was shorter than the burst duration. **c**. Percentage of the total number of spikes that were synchronized within a time lag less than the cross-correlogram width for all pairs that showed significant correlations. The criteria to determine whether a pair of neurons showed synchronous activity or not are described in Section 5.2.3 (see also Figure 5-5). **d**. Percentage of the total number of spikes that were synchronized and in bursts. Only bursting pairs of neurons that showed synchrony were included in this computation. The percentages are based on the total number of spikes.

Figure 5-5: Assessment of statistical significance in cross-correlograms

a. Cross-correlogram between a pair of units showing the expected value (continuous line) as well as two standard deviations (dashed line) based on the null hypothesis of independence computed from equation 5.6 (Aersten *et al.*, 1989, Palm *et al.*, 1988.) The x axis shows the time lag between the spikes from the two neurons. The y axis shows the number of coincidences per second within each bin. Bin size = 10 ms. **b.** Cross-correlogram between the same pair of units after subtracting the shuffled-cross correlations of all segments of the spike train using an interval $T = 10\text{sec}$ (see text for details). The expected value assuming the null hypothesis of independence is 0 and the two standard deviations significance limit is indicated by the dashed line. The inset shows the mean shuffled-cross correlogram. **c.** Mean cross-correlogram obtained by simulating two Poisson neurons with the same firing rates as the real neurons in the previous correlograms. The number of iterations was 10000 and a refractory period of 2 ms was imposed on the modeled spike trains. Bin size = 10 ms. **d.** Distribution of the peak values of the cross-correlograms from the 10000 iterations of the Poisson neurons. The arrow indicates the mean value and the ticks show the two standard deviations limits. The * denotes the actual peak value obtained in the example shown in **a**.

Figure 5-6: Sample of spike trains

Example of neuronal spiking activity recorded from three neurons simultaneously during the wake state (**a**), slow wave sleep (**b**) and rapid eye movement sleep (**c**). Each tick corresponds to an action potential; each row within each epoch corresponds to a separate unit. The three neurons were recorded from separate microwires within the same electrode located in the right hippocampus. The shaded box encompasses 50 ms and distinguishes groups of spikes that appeared simultaneously within this time window in the three neurons. Scale bar = 1000 ms.

Figure 5-7: Examples of two-point and three-point cross-correlograms

Example of two- and three-point correlograms among the same neurons illustrated in Figure 5-6. The two-point cross-correlograms are shown at the left for each of the three possible pairs of neurons during the wake state (**a**), SWS (**c**) and REM sleep (**e**). The x axis represents the time lag between the two neurons ranging from -400 to 400 ms while the y axis shows the degree of synchrony in coincidences per second. The y scale varies from one pair to another but all epochs for a given pair are shown at the same scale. The time lags that showed significantly enhanced synchrony (> 3 s.d.) are shown in red. The three point correlograms are shown at the right for the corresponding epochs (**b**, **d** and **f**). The x and y axis represent the time lags between neuron 1 and neuron 2 and between neuron 2 and neuron 3 respectively. The z axis (in coincidences/sec) indicates the correlation function between the three neurons. The points that showed significantly enhanced synchrony (> 3 s.d.) are shown in red. Bin size = 10 ms both for the two-point and three-point correlograms.

Figure 5-8: Comparison of pairwise correlations in one subject

Summary of all the significantly correlated pairs of neurons in one subject during the wake state (**a**), slow wave sleep (**b**) and REM sleep (**c**). The data come from the same subject illustrated in the previous 2 figures. The vertices along each circle correspond to different units. The location of each unit is indicated next to each point (see legend to Table 5-1 for abbreviations). The small subindices indicate the specific microwire within the electrode. A line between two units indicates that the number of synchronized spikes in the corresponding pair exceeded the value expected by chance at the $p < 0.05$ level (see Section 5.2.3). The strength of the correlation was evaluated as the ratio of the peak to the baseline amplitude in the correlogram and is color-coded (see scale at the bottom of the figure). Only neurons recorded during all epochs are shown in this figure. According to several physiological and clinical criteria, the seizure focus for this patient was determined to reside in the left medial temporal lobe; therefore, the only unit illustrated here within the focus is the left entorhinal unit shown at the top. Note the increased number of synchronous pairs and the stronger correlations during slow wave sleep.

Figure 5-9: Summary of results

(a) Overall percentage of pairs that showed synchronized activity during each epoch (wake = black, SWS = gray, REM sleep = white). The data from all units (including units that were not recorded during all epochs and both units outside and inside the seizure focus) are pooled in this figure (see text for details). (b) Overall percentage of triplets that showed synchronized activity. The total number of pairs and triplets analyzed during each epoch is indicated in Table 5-2. A pair or triplet was considered to show synchronized activity only when the peak was beyond 2σ of the null hypothesis of independent firing (see Section 5.2.3). (c) Proportion of neuronal pairs that showed synchrony among the significant pairs that were recorded in all epochs. There were a total of 282 neuronal pairs that were recorded in all three epochs. Out of these, 91 (32%) showed significant peaks in the correlograms for at least one of the different epochs. all = pairs that showed synchronized activity in all epochs; ws = pairs that only showed synchrony during the wake state and SWS; wr = pairs that only showed synchrony between wake and REM; sr = pair that only showed synchrony during SWS and REM; w, s and r = pairs that only showed synchrony during the wake state, SWS and REM sleep respectively. All the proportions in this plot are based on the number of pairs that were recorded simultaneously in all three epochs. (d) Comparison of the strength of the correlations for those neuronal pairs that showed synchrony during SWS and at least one other epoch. The strength is measured as the ratio between the correlogram peak and the baseline (see Section 5.2.3). The abscissa always corresponds to the activity during SWS. Empty circles: neuronal pairs with temporally correlated activity during all epochs; activity during wake state. Filled circles: neuronal pairs correlated only during wake and SWS; activity during wake state. Empty squares: neuronal pairs synchronized during all epochs; activity during REM. Filled squares: neuronal pairs correlated only during wake and REM; activity during REM. The diagonal indicates the $y = x$ line. Note that most of the points fall below the diagonal. (e) Distribution of the width at half height of the cross-correlograms for all significant pairs during the wake (top), SWS (middle) and REM

sleep (bottom) states. Bin size = 5 ms. The vertical tip shows the mean value. Distribution plots were cut at 100 ms for clarity.

Figure 5-10: Anatomical location of synchronized pairs

Percentage of those pairs that showed synchronized activity that were in the same hemisphere (**a**), brain lobe (**b**), brain area (**c**) and microwire (**d**) during each epoch (wake = black, SWS = gray, REM sleep = white). Brain lobe corresponds to either the temporal lobe or frontal lobe since there were no units in other lobes (see Table 5-1). Brain area corresponds to one of the anatomical regions specified in Table 5-1. This corresponds to the percentage of neuronal pairs that were on the same electrode among those that showed synchrony.

Figure 5-11: Robustness of the results to parameter changes

For each possible combination of parameters and statistical assessment methods, we computed the ratio of the percentage of significantly synchronized pairs during SWS to the one during the wake state or REM sleep and also the ratio of the number of synchronized pairs during wake to those during REM. Here we show the distribution of the values for these ratios. (**a**) SWS/wake, (**b**) SWS/REM, (**c**) wake/REM. The arrow corresponds to the value reported in the text. Bin size = 0.2. The data average all possible combination of parameters including all the variants of the statistical significance assessment as described in Section 5.2.3, all the different possible bin sizes, thresholds and other parameters. The total number of parameters and combinations are given in the text.

5.7 Tables

Table 5-1: Number of units recorded in each location

	Amy	PHG	Hipp	EC	Temp	Sub	SMA	AC	Fr	Total
Left	2	4	29	15	2	0	0	2	3	57
Right	0	0	31	39	0	5	9	0	2	86
										143

Number of units recorded in each location (Amy = amygdala, PHG = parahippocampal gyrus, Hipp = hippocampus, EC = entorhinal cortex, Temp, anterior temporal cortex, Sub = subiculum, SMA = supplementary motor area, AC = anterior cingulate, Fr = frontal cortex). Within each location, we separate those units in the left and right hemispheres. There were 127 units in the temporal lobe (shaded) and 16 units in the frontal lobe.

Table 5-2: Number of pairs and triplets analyzed during each epoch

	Wake	SWS	REM
Number of pairs	484	441	544
Number of triplets	1288	1150	1555

Total number of neuronal pairs and triplets analyzed during the wake, SWS and REM sleep states.

6 Binocular Rivalry and Flash Suppression

6.1 Introduction

The brain can interpret several types of visual stimuli in more than one configuration. The famous artists Salvador Dalí (1904-1989) and Maurits Cornelis Escher (1898-1972) took advantage of this observation in several of their paintings. A classical example can be found in the Necker cube (Figure 6-1). The cube can be seen either in the configuration illustrated in Figure 6-1b or the one shown in Figure 6-1c. Interestingly, one can voluntarily switch between one and the other. This type of stimuli, generally called bistable percepts, are very interesting to neuroscientists in order to explore the changes that occur in the brain during one or the other interpretation. They are not limited to static stimuli (Bradley *et al.*, 1998). Binocular rivalry constitutes another example of a single stimulus that can be seen in two different ways. Normally, there is only a small difference between the views of the world obtained in the right and left eyes. This disparity is used by the brain as a fundamental cue to extract depth information, a process called stereopsis (Wandell, 1995, Wheatstone, 1838, von Helmholtz, 1866). If two dissimilar stimuli that cannot be fused are presented to corresponding areas of the two retinæ, subjects report observing a mixture of the two stimuli during brief periods of time. But, most of the time⁸⁴, the percept alternates between one and the other stimulus in seemingly random fashion (von Helmholtz, 1866, Dutour, 1760, Blake, In press, Levelt, 1968, Leopold and Logothetis, 1999). I provide a simple example of binocular rivalry in

⁸⁴ The proportion of time actually depends on several properties of the stimuli such as its size, frequency, content, etc. (Blake, In press, Blake, 1989, Levelt, 1968, Leopold and Logothetis, 1999). Dominance of stimulation to one eye can be complete; for example, it is interesting to note that if we close one eye, we do not see the world darker (nor do we see it disappear for several seconds!). Occlusion constraints can also affect the degree of rivalry or suppression between unpaired images (Shimojo and Nakayama, 1990, Nakayama *et al.*, 1995).

Figure 6-2. Upon fixating on the white cross through red/blue glasses, the horizontal and vertical gratings are projected onto one and the other retina⁸⁵. Perception alternates between the red horizontal grating and the blue vertical grating. Sometimes, one can see a patchy mixture of the two, but mostly one sees exclusively the blue grating or the red one. Even without glasses but with enough patience, a much weaker version of the phenomenon called monocular rivalry can be appreciated (Campbell and Howell, 1972, Andrews and Purves, 1997).

Bistable percepts constitute a fascinating tool for our endeavor of trying to understand at the neuronal level the dissociations between the visual input and our percepts (see Chapter 1). The same stimulation from the visual world can give rise to two different interpretations, that is, there are perceptual variations in spite of the absence of changes in the retinae⁸⁶. Binocular rivalry shares several properties with the Necker cube and other bistable stimuli. For the most part, we do not see a mixture of the two configurations during either binocular rivalry or examination of the Necker cube. Transitions between alternative percepts are stochastic. Furthermore, researchers have carefully measured the time interval between switching percepts and shown that this follows a gamma distribution of a low order in both cases (Leopold and Logothetis, 1999, Levelt, 1968, Borselino *et al.*, 1972). There are, however, some important differences: while it is easy to volitionally control the dominant configuration of the Necker cube, it is not particularly trivial to do so during binocular rivalry (Lack, 1978; see however von Helmholtz, 1866). In order to investigate the neuronal responses during these changes, it would be very interesting to be able to study neurons whose activity is correlated with the occurrence of one or the other percept. This is not easy to do with stimuli like the Necker cube or the Escher paintings. In contrast, binocular rivalry allows us to use basically any type of stimuli and therefore one can test binocular rivalry with the preferred stimuli of the neurons that are being recorded. Let us suppose that we found a neuron that shows a visually selective response to, say, a picture of Clinton. What would happen if we show

⁸⁵ This depends on the sharpness of the color filters and on the particular settings in the monitor or printer through which the image is rendered.

⁸⁶ To a very first approximation, and in the absence of any eye movements, the retinal image is constant (see description of the phenomenon later). Binocular rivalry can be strongly modulated by eye movements (Walker, 1978, Logothetis, 1998); however, they cannot be the cause in the generation of the perceptual alternations. One of the strongest pieces of evidence for this is the observation that binocular rivalry can occur with retinally stabilized images (Ditchburn and Protchard, 1960).

Clinton in one eye and a different stimulus in the other eye? Will the neuron be continuously active regardless of the fact that the percept disappears for several seconds from consciousness? Or will the response of the neuron be modulated depending on the perceptual state of the subject?

It is not easy to do these experiments in monkeys since it requires extensive training for the animal to be able to accurately report its percept. John Allman and colleagues demonstrated that monkeys could be trained to accurately report their changing percepts during binocular rivalry. Furthermore, their results showed that there is a strong similarity in binocular rivalry in macaque and human subjects (Myerson *et al.*, 1981). Nikos Logothetis and colleagues did several electrophysiological experiments to study the neuronal activity in monkeys during binocular rivalry in different areas of the visual cortex (Logothetis, 1998). They observed that in higher visual areas, almost all neurons follow the percept, that is, they respond when the monkey reported that it was seeing the stimulus the neuron preferred and not when the other stimulus was seen (Sheinberg and Logothetis, 1997). In lower visual areas, the activity of most neurons remains unaltered by the changing percept and more accurately reflected the constant visual stimulus (Leopold and Logothetis, 1996)⁸⁷. It seems quite clear that neurons in the lateral geniculate nucleus of trained monkeys respond constitutively regardless of the perceptual alterations (Lehky and Maunsell, 1996). The proportion of neurons that follow the percept increases from V1 to inferior temporal cortex in macaque recordings (Logothetis, 1998).

Here I will describe the activity of individual neurons particularly in the medial temporal lobe in the human brain while the subjects' perception was modulated in binocular rivalry and during a related phenomenon called flash suppression (to be described below). Our results show that the responses of most visually selective neurons are strongly modulated by these perceptual alternations. These neurons do not respond when the selective stimulus is suppressed from perception but strongly increase their firing rate when the preferred image comes into awareness. I will also compare our

⁸⁷ We will discuss later in this Chapter the evidence from psychophysical, functional imaging and MEG studies in humans.

results to those obtained in monkey electrophysiology in particular, but also to those from functional imaging and other techniques.

6.2 Brief historical remarks

6.2.1 Binocular rivalry

The function and mechanism of the combination of information from the two eyes has been extensively discussed for a very long time; for example, there are documented discussions on the topic in the work of Aristotle. It was Porta who first described in 1593 what can be interpreted as binocular rivalry while observing with each eye a separate page of a book using a partition between the eyes (Blake, In press, Alais *et al.*, 2000). Almost two centuries later Dutour (Dutour, 1760) noted alterations in color perception when the two eyes observed different colors. Both Dutour and Porta interpreted these observations as evidence for the “suppression theory” according to which we see through only one eye at a time. This is in contrast to the “fusion theory” which dates back to Aristotle that claims that we see through both eyes simultaneously. A more extensive and systematic characterization of binocular rivalry was performed by Sir Charles Wheatstone (Wheatstone, 1838). He observed the characteristic complete suppression of one of the two stimuli, the alternations between the two percepts and the fragmentation of the images during the transitions. Wheatstone also noted that slightly different views of the same stimulus could be fused to form a three-dimensional percept, a phenomenon now known as stereopsis. He thence revived the fusion theory and proposed that binocular rivalry occurs when fusion breaks down. Helmholtz considered the subject again in his famous *Treatise on Optics* (von Helmholtz, 1866). He attributed the phenomenon of binocular rivalry to alternations in selective attention. Interestingly, he claimed to be able to volitionally modulate the switch from one percept to another, a proposition that does not seem to be very well accepted nowadays⁸⁸. Helmholtz’

⁸⁸ There are important cognitive and attentional influences in binocular rivalry (Leopold and Logothetis, 1999, Lack, 1978) but perceptual transitions can occur spontaneously in the absence of any volitional

contemporary, Hering proposed that a competition at low levels of the visual system could account for the mechanism of binocular rivalry alternations. This debate of high versus low level mechanisms initiated by Helmholtz and Herring has not been fully resolved yet (Logothetis, 1998, Leopold and Logothetis, 1999, Blake, 1989, Blake, In press, Polonsky *et al.*, 2000, Tong and Engel, 2001). Several important properties of binocular rivalry were carefully studied in the doctoral dissertation of Levelt (Levelt, 1968). The second part of the twentieth century was fortunate to contemplate a wealth of new experiments on the psychophysics, electrophysiology and functional imaging of rivalry. Several of these more recent studies will be discussed later in this chapter.

6.2.2 Flash suppression

The phenomenon of flash suppression was introduced by Jeromy Wolfe (Wolfe, 1984). It consists of the perceptual suppression of a stimulus that was presented monocularly upon flashing a new stimulus to the contralateral eye while keeping the original stimulus in the ipsilateral eye (Figure 6-6). Given that when a stimulus is presented monocularly, say to the right eye, this eye is clearly dominant according to the theory of competition between two eyes (we do not see alternations with a black image upon closing one eye), Wolfe's prediction was that this eye would continue to be dominant during the flash period, and therefore the contralateral flash would be barely seen if at all. The suppression of the monocular stimulus is, however, very strong. It is not a complicated version of forward masking (the parameters are just wrong for what would be expected in forward masking). It is not due to light adaptation or any other mechanism that reduces the visibility of the ipsilateral flash (Wolfe, 1984). A short blank offset can be introduced between the monocular presentation and the flash without affecting the effect. However, the effect is not due to the offset reversing the dominance of the eyes because it can be elicited in the absence of any offsets.

As in the case of binocular rivalry, there are two extreme⁸⁹ alternative explanations that could account for the suppression. In one scenario, there could be a

control. Furthermore, changes from percept A to percept B occur even if the subject tries to exclusively maintain the percept of A.

⁸⁹ And several in between.

competition between the signals from the two eyes for perception. The new eye would win this competition due to either adaptation of the ipsilateral eye or to a larger change in its stimulation. Alternatively, the competition could be between the two patterns themselves. Here the new stimulus would win over the old stimulus. At the physiological level, the first hypothesis could be modeled by a competition between monocular channels, possibly at or before the binocular neurons of V1 in the visual pathway. The second hypothesis would be better modeled by the competition between neurons selective to the stimuli themselves in higher areas of the visual hierarchy. These two interpretations parallel the current controversy regarding the origin of alternations during binocular rivalry (Lee and Blake, 1999, Blake, In press, Logothetis *et al.*, 1996, Logothetis, 1998, Tong and Engel, 2001, Polonsky *et al.*, 2000). We have studied this issue using psychophysical experiments; the results of which are described in Appendix 2.

6.3 Experimental procedures

6.3.1 Neuronal recordings and stimulus presentation

The procedures for electrode implantation, data acquisition, spike sorting and so on were described in detail in Chapter 2. The current chapter describes data from a total of 11 experiments in six subjects (4 right handed, 3 male, 21 to 44 years old) for binocular rivalry and 30 experiments in 11 patients (8 right handed, 6 male, 24 to 48 years old) for flash suppression. There is an overlap of two subjects in these two groups where both the binocular rivalry and the flash suppression paradigm were tested.

Visual stimuli were presented on a monitor situated approximately 50 cm away from the subject connected to a laptop computer as in Chapters 3 and 4. Stimuli subtended a visual angle of approximately 3 degrees⁹⁰ and were presented separately to

⁹⁰ There was some degree of variation from patient to patient and even from experiment to experiment in the size of the images. It should also be noted that in contrast to experiments in monkeys where the head is somewhat fixed, subjects were free to move their head in our tests. However, I estimate that the range of visual angles is still not too large (see below) and I do not think this can affect the flash suppression effect (see Appendix 2; image size does influence the percentage of piecemeal rivalry during binocular rivalry, though). Patients typically rest their head on a pillow or on the bed. The distance to the monitor was approximately 70 cm. I let the patient somewhat adjust and fine-tune the distance to the monitor so that

the right and left eyes by means of a pair of liquid crystal glasses that transmit light to one eye or the other in interlaced fashion (Crystal Eyes, Stereographics, San Rafael, CA). The occlusion rate is controlled by an infrared synchronization signal originated by an emitter connected to the monitor. The system uses the field sequential technique with alternate fields presented to one or the other eye. The information from the infrared signal originates from the display system's video signal. Crystal Eyes uses the "above-and-below" method according to which two sub-fields are arranged above and below each other in a single standard field, rendering a 60 fields per second flicker-free image in each eye (Lipton and Meyer, 1984). The absence of flicker depends on a high-field rate output that is normally available in modern monitors. According to the manufacturer, there is a 1000-fold difference in transmittance between the two states of the goggles. Margaret Livingston has performed careful measurements with a photocell using the same brand of glasses that we have used and she reported a ratio value of 25:1 in the transmittance in the "open" to that in the "closed" state (Livingstone, 1996). The quality of the stereoscopic image depends not only on the low transmittance in the closed state but also on the capability of the display system to clear the field substantially before the next field is written. If this is not the case, subjects see a double or ghost image because of the cross-talk due to the left image being perceived also by the right eye and vice versa. This can occur due to leakage in the closed state from the goggles, to lags in the rise and fall time to the open and closed states in the goggles, and also other factors such as the phosphor persistence in the monitor. The measurements of Margaret Livingstone showed

they feel comfortable. Sometimes, they may feel that 70 cm is too far or too close and therefore this distance can vary by as much as 10-15 cm (this translates into a range of distances of 55 to 80 cm corresponding to 2.5 to 4 degrees of visual angle). One of the nice things about flash suppression is that it is quite strong and it is not affected by the size of the images (I have never tried sizes < 1 degree or beyond 10 degrees). Therefore, I prefer to be strict and firm with the patients in other things that I consider more important such as asking them to wear the goggles and not move or talk during the tests, etc rather than complain much about this issue. There is yet one more constraint: patients are video monitored 24 hours a day. Given the angle of the camera, if the monitor is too close to the patient it may partially occlude the view from this camera. Therefore, there is some variability (I estimate that this is up to 1 degree, see above) from patient to patient or sometimes from experiment to experiment in the same patient. Within an experiment, I think the variability is much less. Patients do move their heads to some extent. However, I try to discard as much as possible any trial where I observe a significant movement. There is basically only one degree of freedom in the movement that would alter the visual angle; this is moving towards the monitor. They cannot move away from the monitor since their head is against the bed. They could move to the sides but a 10 cm displacement to the sides would alter the visual angle by less than 0.1 degrees of visual angle. A 10 cm forward movement of their head would cause a change of 0.5 degrees of visual angle.

that each eye effectively sees 5% of the other eye's visual input⁹¹. However, the physiological (event related potentials) and perceptual effects of this cross-leakage seem to be negligible (Livingstone, 1996). Probably one of the best ways to separate the stimuli shown on one eye or the other is the stereoscope design of Wheatstone (Wheatstone, 1838). The simple mirror optics yields 0% cross-transmittance to the other eye. Indeed, we have manufactured and used this type of stereoscope successfully in several binocular rivalry and flash suppression experiments. The drawback is the relative difficulty to easily adapt and maneuver it in the clinical setup. Color filters have been used extensively in the past for stereo presentations and binocular rivalry experiments but they also suffer the problem of leakage.

Responses to visual stimuli were acquired through a pair of push buttons attached to the laptop. The time of these responses (see below for the description of the tasks) as well as the visual stimuli marker timestamps were recorded in the same data acquisition computer as the electrophysiological data.

6.3.2 Tasks

During binocular rivalry, two different images were presented to the two retinæ and subjects reported their percept by using two push buttons. They were instructed to press and hold one button upon perceiving one of the images, to push and hold the other button while perceiving the other image and to release both if they observed piecemeal rivalry defined as any mixture of the two stimuli (Leopold and Logothetis, 1996, Sheinberg and Logothetis, 1997). Each session lasted 40 seconds and included also 4-6 randomly inserted 1000 ms periods of monocular stimulation where only one stimulus was presented. This was done to verify that the subjects were reporting their percepts accurately and also to elicit flash suppression perceptual changes. In 3 subjects we also included before each rivalry session, a period where the two stimuli were smoothly merged into each other during binocular presentation. That is, we started by presenting one of the stimuli, say **A**, and we gradually added the second image, say **B**, while image **A** was disappearing. This was done to demonstrate the state of piecemeal rivalry and to

⁹¹ It should be noted that the monitor used for that evaluation (Silicon Graphics, Indigo XZ4000) was different than the one used in our tests (Sony CPD-G200 Trinitron).

coarsely mimic this state for comparison purposes. The merging was done linearly, that is, the image presented in frame i (I_i) had i/N of pixels from picture **A** and $(N-i)/N$ of pixels from **B** where N is the total number of frames. Frames were changed linearly with time and the pixels were chosen randomly. The linear change and the random pixel assignment constitute a poor approximation to the transitions during rivalry. The transitions in our perception seem to switch in a sigmoidal fashion (slowly near the two extremes). Patches in the images seem to switch together; as the name piecemeal rivalry indicates there are pieces of the image which may correspond to **A** and other pieces to **B**. These pieces are composed of patches of the images and not to individual random pixels. As discussed in the text below, by the time we were going to make these changes, we decided we were going to continue running flash suppression instead of binocular rivalry in order to acquire more data and therefore the merging was only done in this poor approximated fashion.

Figures were chosen from our collection of images containing different natural categories of stimuli (see Chapter 3 for a description of the images and Figure 3-1 for a sample). The two images were chosen to belong to different categories (with the natural categories defined as in Chapter 3, Table 1). During binocular rivalry, each individual image was used in only one session. Other than these constraints, the pictures were chosen pseudo-randomly.

During flash suppression, an image was presented monocularly for a period t_1 of 1000 ms⁹². Following the monocular presentation, the same monocular image was flashed onto the same eye while a different picture was flashed to the contralateral eye during a period t_3 of 500 ms⁹³ (Figure 6-6). Both images disappeared after the flash and this indicated the subject to give his response. Subjects were asked to report by pressing a button whether the second picture was the same as the first one or not. In contrast to the previous experiments, here we did not use any tone to request the response; the signal to respond was the disappearance of the images. We did not particularly ask subjects to

⁹² In the two first subjects t_1 was 1500 ms. A time of 1000 ms is enough to produce the flash suppression effect as reported by (Wolfe, 1984) and also verified in psychophysics experiments by ourselves (see Appendix 2). The shorter time allows us to present more stimuli.

⁹³ In 5 subjects, t_3 was 300 ms. 300 ms is more than enough to observe the flash suppression effect. However, some subjects perceived the images as being changed too fast and we adjusted t_3 to 500 ms. A flash period much longer than 500 ms causes the two images to start giving rise to binocular rivalry (Wolfe, 1984 and Kreiman *et al.*, unpublished observations) and we wanted to avoid that.

respond fast. Approximately 10% ($8 \pm 7\%$) of the trials were controls in which only one picture was presented during t_3 . The number of controls trials was kept very low, thus creating a clear bias in the proportion of responses. Thus, if subjects had indicated that there was always a change in the stimuli, they would have been correct approximately 90% of the time. This was designed in this way in order to maximize the amount of data obtained during flash suppression trials. We knew from extensive psychophysical testing (see Figure 6-7 and Appendix 2) that the suppression phenomenon is very strong and we did not need to control for the responses as carefully as in monkeys. Just to be sure, in 20% of the sessions we also asked subjects to report verbally what the first and second pictures were. This debriefing was also done after both pictures had disappeared from the screen. The results from this assessment were also very strong (Figure 6-7) suggesting that the subjects did indeed observe the suppression of the monocular stimulus. The monocular stimulus was randomly delivered to either the left or right eye. We did not ask subjects whether they could determine which eye received the monocular stimulus. Several psychophysical studies have shown that, in general, it is not possible to discriminate which through which eye a stimulus was observed if the experiment is carefully controlled. In several cases, subjects can learn to use other cues to determine the eye of origin but upon controlling for these cues, utricular discrimination seems to be at chance levels (Porac and Stanley, 1986, Ono and Barbeito, 1985, Blake and Cormack, 1979a, Templeton and Green, 1968, Blake and Cormack, 1979b). While it may be quite interesting to explore this question in V1 recordings, all neurons beyond V1 seem to respond to stimulation from either eye and it is therefore unlikely that this could influence the neuronal activity in the MTL.

As we have discussed in Chapter 3, it is not possible to maximize simultaneously the number of pictures, the number of repetitions per picture and the number of categories to be presented. The overall distribution of the number of stimuli presented during the monocular phase, during the flash period and the joint distribution of the number of presentations during both periods for each figure is shown in Figure 6-7. In some cases, there were very few presentations per picture. For the comparison of the responses to different individual stimuli, we only used those stimuli that had been presented at least 4 times as we have done in Chapter 3. Some stimuli were presented

enough times during the monocular period but not during the flash period due to the randomization of the presentation order. This marks the difference between the groups Cat, Ind and Cat*, Ind* in Table 6-2.

The experimenter (G.K.) attempted to distinguish and mark any trials in which subjects moved and could have thus affected the results. In the previous experiments, care was taken to attempt to mark as much as possible large saccades and blinks as well (see Chapter 3). Unfortunately, it is very difficult to detect these in the rivalry/suppression experiment due to the fact that the goggles are quite dark. While it is possible to see the patient's eyes through them, this requires much more attention than without the goggles. Since I am normally monitoring several things at once during the experiment it is likely that I have missed most of the saccades and blinks in this experiment. We did not attempt to independently measure eye movements with some other external device. To somewhat compensate for this, I insisted at the beginning in asking the subjects to fixate. However, these subjects were clearly not trained in psychophysical experiments and fixating and I did not monitor this because of the darkness of the goggles. There are two important questions related to blinks and eye movements. The first one is to what extent eye movements or blinks could directly influence the neuronal responses. As I have discussed in Chapter 3, it is unclear that blinks and eye movements would influence the neuronal responses in the MTL, particularly given the clear lack of interaction between visual selectivity and the few responses to eye movements that have been shown in the temporal lobe (DiCarlo and Maunsell, 2000, Ringo *et al.*, 1994, Sobotka *et al.*, 1997). The second question is also important: to what extent do blinks and eye movements influence flash suppression or rivalry? The answer for binocular rivalry is somewhat long and less relevant to our study because most of our data occurs during flash suppression but it seems clear that eye movements can influence the alterations that occur during binocular rivalry (Leopold and Logothetis, 1999). During flash suppression the situation is different. A 500 ms monocular presentation is enough to elicit the flash suppression effect (see (Wolfe, 1984) and Appendix 2). The duration of blinks and saccades is short enough that it would not change the flash suppression effect if they occur at the beginning of the monocular presentation assuming that the subjects do get 500 ms of monocular input. Furthermore, a

blank period t_2 can be inserted between the monocular input and the flash without altering the flash suppression effect (see (Wolfe, 1984) and Appendix 2). If the blink or saccade occurs near the end of the monocular period or at the beginning of the flash, this would be somewhat similar to this blank period. Inasmuch as the blink or saccade is similar to the blank interval, I think that the flash suppression would proceed normally. However, the influence of blinks and saccades in the flash suppression effect has not been directly tested to my knowledge. In a few subjects (see details in main text), however, t_3 was 300 ms. A blink could be problematic if it occurred during these 300 ms.

6.3.3 Data analysis

During binocular rivalry, raster plots and histograms were aligned to the times at which subjects reported a switch in their perception. The statistical analysis followed the lines of the methods defined in Chapter 3 with a few small modifications. From the reported perceptual switch times, a fixed interval length of 900 ms length was considered for statistical analysis⁹⁴. In contrast with the previous analysis, we did not allow here for a 100 ms latency; this was done to account for the behavioral delay in pressing the buttons⁹⁵. We estimated the baseline activity by measuring the number of spikes in the [-1000;-500) ms interval with respect to the time at which the subject pressed the button. While this interval is arbitrary, we tried to avoid using the period immediately preceding the button press because of the behavioral delay just mentioned. Similar results were obtained when considering only the 1000 ms before the start of each session.

The methods for data analysis during flash suppression used in this chapter have been defined and explained in Chapter 3. Visual selectivity was analyzed during the monocular presentation of the stimulus (the interval t_1 used for statistical analysis was

⁹⁴ In some cases, perceptual intervals were shorter than 900 ms (see for example the distribution in Figure 6-3). We also tried an interval of 600 ms centered on the peak of the response and the results were similar to the ones reported in the text.

⁹⁵ The delay from the percept to the behavioral response is most likely to be longer than 100 ms (Kandel *et al.*, 2000). Therefore, it is possible that the strongest neuronal response could occur even before the time at which the subjects pressed a button. To address this, we also tried shifting the data by multiples of 100 ms from 0 to 500 ms to account for the behavioral delay. It seems unlikely that this delay could be longer than 500 ms. The number of selective neurons that followed the percept for each of the time shift value ranged from 3 to 5. Therefore, it seems that our results are not due to missing a large number of selective units because of this time shift.

[100;1000) ms) as defined in Chapter 3. During the flash interval (t_3), we used two different periods for data analysis: [100;600) ms and [100;1000) ms with respect to the onset of the flash (corresponding to [1100;1600) ms and [1100;2000) ms with respect to the onset of the monocular stimulus). The results using the two intervals were similar and here we report the values corresponding to the first interval. A unit was said to “follow the percept” of a particular stimulus A (or a particular category A) if and only if:

- (i) its response during t_1 was selective to A (see criteria in Chapter 3)
- (ii) its response during t_3 when A had been presented monocularly and a different stimulus was flashed onto the contralateral eye was indistinguishable from baseline
- (iii) its response during t_3 was selective to A when another stimulus had been presented monocularly

6.4 Neuronal activity during binocular rivalry

6.4.1 Behavior

A common way of evaluating the behavioral responses during binocular rivalry is the shape of the normalized distribution of dominance periods. From the time intervals of dominance as reported by the subject (Figure 6-3b) one can compute the distribution of the duration of these intervals normalized by the mean dominance period. This has been shown to follow a gamma distribution of low order (Levelt, 1968) and is even used as one criterion to evaluate the training of macaques in binocular rivalry experiments (Myerson *et al.*, 1981, Leopold and Logothetis, 1999). This kind of distribution is also obtained in other kinds of bistable percepts (Borselino *et al.*, 1972). We show an example of the distribution of dominance periods fitted by a gamma curve in Figure 6-3c. This kind of distribution was obtained in all but one⁹⁶ of the subjects suggesting that they were observing and reporting normal binocular rivalry alterations.

⁹⁶ In this subject, we could only run a very short test with very few presentations. The subject did not seem to be able to clearly observe or clearly report the perceptual alternations during rivalry at the beginning. In the last two sessions, the subject was probably reporting somewhat more accurately her perception but that was too few data to result in the typical gamma distribution.

It should be noted that subjects were not trained in the binocular rivalry task before starting the tests⁹⁷. This is in stark contrast to the experiments performed in monkeys. The characteristics of the stimuli such as their saliency, color, contrast, frequencies and even their content and significance to the subject can strongly affect the mean dominance period of a particular stimulus (Blake, 1989, Blake, In press, Logothetis, 1998, Fahle, 1982, Levelt, 1968). However, because of time constraints, we did not attempt to fine-tune these parameters for each particular stimulus pair.

6.4.2 Neuronal selectivity during binocular rivalry

We analyzed the activity of a total of 119 units (67 in the medial temporal lobe) during the binocular rivalry test (see Table 1 for the location of these units). An example of the activity of a unit in the left amygdala during one session of the rivalry experiment is shown in Figure 6-4. This unit was selective to emotional faces. The stimulus that was presented to the subject is shown in part **a** and the subject's reported percept is indicated in part **c** of the figure. The neuron weakly followed the subject's percept as indicated by the *sdf* in part **b**, increasing its activity when the face was perceived. In this trial, there were several instances where a stimulus was presented monocularly and therefore the subject should have experienced flash suppression. One such example occurs towards the end when the house is shown monocularly and then, upon flashing the face, the subject perceives the face exclusively and there is a vigorous increase in the neuronal response. However, in the previous period where the house was presented monocularly, the subject did not report observing the face during the flash (at least not immediately) and there was no increase in the neuronal activity. The average activity of this neuron for all stimuli within the groups of stimuli that were presented is shown in Figure 6-5. Note that in contrast to all the previous histograms, these PSTHs are aligned to the perceptual switch as indicated by the subject's response while pressing the buttons (dashed vertical line). We can see an increase in the activity of this unit when the subject reported perceiving an emotional face compared to stimuli from the other categories. There seems to be a small inhibition in some of the other categories before the perceptual switch. This could be

⁹⁷ In two subjects, a short 5 minute rivalry session had been run on the same or previous day to the experiment itself.

related to the fact that during some of those periods a face was presented and was disappearing from perception.

Overall, as explained in detail below, the amount of data that we obtained during a binocular rivalry experiment was unfortunately very small. We obtained only 3 units that weakly followed the percept as the one just illustrated out of the total of 67 MTL units. We did not independently evaluate the visual selectivity of these neurons before the binocular rivalry experiment. This was done to attempt to acquire more data during the rivalry condition. Extrapolating from our previous results, we could expect to see approximately 9 units in the medial temporal lobe that would be visually selective. Therefore, it could be claimed that approximately 33% of the units we observed actually followed the percept. Extreme caution should be made however in this claim. First of all, the numbers are very small to accurately compute percentages; second, the number of 9 units was obtained by extrapolation and is not an actual measurement. Finally, we had much less data (in terms of number of stimuli as well as number of perceptual periods per stimuli, see below) to be able to detect the potential selective response of the units.

6.4.3 Amount of data obtained during binocular rivalry

The amount of recording time obtained in our approach⁹⁸ during binocular rivalry was very small. This is due to a number of practical and experimental constraints. Part of the neuronal data is recorded during piecemeal rivalry. While it may be quite interesting to analyze the neuronal activity during these transition periods, to a first approximation we only looked at the selective neuronal responses during the periods of dominance. Dominance periods can last up to 10 seconds (with typical durations on the order of 1-4 seconds). Based on our results described in Chapter 3 and the ones obtained by Sheinberg and Logothetis in the monkey inferior temporal cortex (Sheinberg and Logothetis, 1997), most neurons seem to show only transient responses to the stimuli. Therefore, if the

⁹⁸ Again, let me emphasize that we did not previously select which stimulus/stimuli the neuron was selective to as done in Logothetis' experiments (Leopold and Logothetis, 1996, Sheinberg and Logothetis, 1997, Logothetis and Schall, 1989). If we had known that the neuron responded strongly to a particular stimulus, we would have been able to present that stimulus extensively during the experiment. This goes back to our previous discussion of being able to evaluate the selectivity of the neurons on-line (see Chapter 3).

duration of the neuronal response is only 500 ms and the dominance period lasts 10 seconds, we are basically wasting 9.5 seconds of valuable data. On the other hand, some dominance periods can be very short (200-300 ms) and this again poses a problem in the analysis of the response because a fixed interval of, say, 600-900 ms as the ones used in the previous chapters, would encompass data corresponding to the perceptual period of piecemeal rivalry or dominance of the opposite stimulus. In the approach we have taken, we do not know a priori the selectivity of the neuron and we need to present as many images as possible. But the overall number of pairs that were presented during binocular rivalry was very small. The number of errors during binocular rivalry was typically higher than in the preceding paradigms (the subjects were not trained in the task previously). Just to give an idea of the amount of data that we have gathered, let us give a few numbers. Each rivalry session was 40 seconds. Of those, approximately 10 seconds corresponded to piecemeal rivalry. We also need to discard the boundary effects; we therefore only analyzed each session starting 2 seconds after stimulus presentation and up to 2 seconds before the end. With a dominance of 4 seconds, this yields approximately 2-3 perceptual dominance periods to analyze per picture. For a 25 minute experiment, adding 10 second between sessions and at least a 5% error rate, it is quite clear that the amount of data that is obtained is extremely low. Due to this, most of the data that will be presented throughout the rest of this chapter, corresponds to the flash suppression paradigm which is closely related to binocular rivalry and allows the electrophysiologist to get much more data.

6.5 Neuronal activity during flash suppression

Flash suppression offers the advantage that one can collect much more data in an electrophysiological experiment (Sheinberg and Logothetis, 1997). This is due to the fact that one can present more pairs of pictures, there are no intervening piecemeal rivalry periods and training is easier. Flash suppression is typically very strong basically for any pair of pictures as we discuss below (Wolfe, 1984, Sheinberg and Logothetis, 1997), see also Appendix 2 and Figure 6-7). In contrast, binocular rivalry can be quite weak for

some specific pairs without appropriately adjusting the contrast and colors of the pictures. Since time is one of the major limiting factors in our work with the patients, we opted to concentrate on collecting most of the data during flash suppression rather than binocular rivalry. An obvious question arises: are we investigating the same phenomenon? That is, to what extent does what we learn from flash suppression extrapolate to binocular rivalry? This is discussed in Appendix 2⁹⁹.

6.5.1 Behavioral responses during flash suppression

As noted above, the flash suppression phenomenon is very strong. The monocular stimulus is invariably suppressed by the contralateral flash. The average behavioral results from our subjects during flash suppression are illustrated in Figure 6-8. From a total of 9579 presentations, 906 (9.5%) were controls in which only a monocular stimulus was presented. Of those, subjects reported that there were two images in 7% of the trials. This is surprisingly high and is probably due to mistakes in pressing the buttons. Since they had to indicate a change in most of the trials, they typically made many mistakes in these controls. This is evident from the debriefing responses where they reported no change whatsoever in the monocular image in the control trials in 100% of the trials as expected. The remaining 90.5% of the presentations corresponded to flash suppression trials. In those, subjects reported in over 95% of the trials a suppression of the monocular stimulus both by pressing a button in a two-alternative forced choice and when giving a verbal report. These results are comparable to the large set of psychophysical data that we have collected during flash suppression using different types of reports and stimuli (see Appendix 2).

⁹⁹ I spent several months with two summer undergraduate students studying this question at the psychophysical level (see Appendix 2). An eminent scientist whose name I cannot name solved the problem in a much more elegant and efficient manner. He asked one of the world authorities in binocular rivalry and flash suppression. The response was definitive and conclusive. The world authority thought that binocular rivalry and flash suppression are basically the same phenomenon. Furthermore, he added, he thinks there is no one in this world that would claim that there is a difference between the two. For those who are not satisfied, psychophysical support for the similarity of the two phenomena is presented in Appendix 2.

6.5.2 Examples of neuronal responses during flash suppression

An example of the responses of a neuron during two trials of the flash suppression experiment is shown in Figure 6-9. This was a neuron that responded selectively to four different faces (but not all faces) including that of famous ex-Beatles singer Paul McCartney. When the photograph of Paul McCartney was shown monocularly and then a different stimulus was flashed (Figure 6-9a), the new stimulus, in this case the house, perceptually suppressed the image of the singer. The neuron responded during the monocular presentation and stopped firing when the percept of the singer disappeared. In another trial, a different stimulus was presented monocularly, in this particular case a grating, and the photograph of Paul McCartney was flashed to the contralateral eye suppressing the grating (Figure 6-9b). The neuron started firing only when the picture of Paul McCartney was seen.

The activity of a different neuron is illustrated in Figure 6-10. This unit was recorded from the left PHG and increased its firing rate upon presentation of pictures of spatial layouts and not to other stimuli. As discussed in Chapter 3, some neurons such as the one we are illustrating here, responded in a broader manner without being able to differentiate based on the spike count the different individual pictures within a category of stimuli (Kreiman *et al.*, 2000a). The neuron responded when the effective stimuli were presented monocularly and its activity went back to baseline when the effective stimulus was suppressed (Figure 6-10a). Monocular presentation of stimuli from other categories did not elicit a response (Figure 6-10c). When the effective stimuli were flashed after a different monocular stimulus was presented the neuron responded vigorously (Figure 6-10b). There was no change in the firing rate for any of the other groups of stimuli (Figure 6-10d). Although there was some variability among individual pictures, the neuron responded vigorously to the monocular presentation of most houses and natural scenes within this group (Figure 6-10d). An ANOVA analysis to address the question of whether there was a difference between individual stimuli within the spatial layouts group was not significant ($p > 0.1$). The neuron did not enhance its activity when these stimuli were suppressed by the flash of an ineffective stimulus (Figure 6-10e). Furthermore, there was a strong response during the flash of most individual houses and

natural scenes (Figure 6-10f) and there was no statistically significant difference between these distinct individual stimuli.

The neuron illustrated in Figure 6-11 was selective to a drawing of the character Curly from the TV series "The three stooges." The strong selectivity of this amygdala neuron with respect to approximately 50 other stimuli presented in the experiment was already shown and discussed in Chapter 3 (Figure 3-8). The unit responded transiently to the monocular presentation of Curly. Note that the main response was quite reliable with a few spikes within a latency slightly longer than 200 ms. There was no response when the picture was Curly was flashed but suppressed from perception by an ineffective stimulus. However, when the flash of the image of Curly was perceived and suppressed an ineffective stimulus (Figure 6-11, right), the neuron also responded with a transient but reliable and vigorous burst with a latency slightly shorter than 200 ms.

My final example of a selective unit in the medial temporal lobe (for this Chapter and for this thesis!) is shown in Figure 6-12. The selectivity of this neuron was discussed in Chapter 3, Figure 3-7. This amygdala unit also followed the percept, responding strongly when Clinton was shown monocularly and not when it was suppressed by another stimulus. For two of the figures, the neuron also responded when Bill Clinton was flashed, suppressing another stimulus. The photograph of Clinton's face was only presented twice during the flash period and therefore we cannot perform any reliable statistical analysis for it (see above for a longer discussion of the number of presentations). Given that the unit responded to very different images in which Clinton was present, was suppressed when Clinton was suppressed and responded vigorously when the image of Clinton was flashed, there is a very strong correlation between its activity and the perception of Clinton.

6.5.3 There is life also outside the temporal lobe

Most of our work has been concentrated on the medial temporal lobe. This is because that is where most of the electrodes are implanted (see Chapter 2)¹⁰⁰. But I would like to briefly mention an example of the activity of a neuron in the supplementary motor

¹⁰⁰ And also because we think the MTL is a fascinating area of the brain as we hope we have illustrated throughout this work!

area. The response of this neuron is shown in Figure 6-13. All stimuli were pooled together because we failed to observe any statistically significant difference for different categories of stimuli or different individual pictures. Note that the unit responded non-selectively but quite reliably after the onset of both the monocular stimulus and the flashed stimuli. We have already shown in Chapter 3 some examples of non-selective responses by neurons in the SMA. Given that the SMA is strongly associated with the motor output, it was tempting to speculate that the response was associated with the behavioral response (see Chapter 3 for discussion). However, here, we observe a transient response to both the monocular and the flashed stimuli. The latency in the first increase was 72 ms and in the second one it was 76 ms. The corresponding response durations were 305 ms and 321 ms. The subject only had to respond after the flashed stimuli so it is not clear that this response could be related to the preparation of a response. The unit responded irrespectively of the response (pressing one or the other button depending on whether the stimuli changed or not, see experimental procedures above); a two-tailed *t* test comparing the count of spikes in a 300 ms centered on the response peak for pressing one or the other button yielded $p > 0.2$. Halfway through the experiment we asked subjects to switch which hand they used for one or the other button. This did not affect the responses either ($p > 0.3$). It would seem as if the activity of the neuron were simply indicating a change in the visual world.

On a different note, it has been suggested that the activity of the anterior cingulate could be involved in the processing of errors (Paus, 2001). We therefore explored the activity of the few AC units (see Table 2) during the flash suppression test. None of these units showed visual selectivity as we have discussed in Chapter 3. The definition of error in flash suppression could be somewhat tricky. There are several types of errors that could occur. In the control trials in which there was no presentation of a different flashed picture, subjects could report seeing a second picture. As argued above, this is most likely due to the strong bias in the number of times they had to press one or the other button and reflects the automatic nature of the response or a motor mistake. I would be extremely skeptic to believe that subjects hallucinated and indeed saw a second picture. During the flash suppression trials, the phenomenon is so strong that it could be argued that the monocular stimulus should be suppressed in 100% of the trials. Then one could consider

those cases where the monocular stimulus was not suppressed as an error trial. It should be noted that the proportion of all these “error trials” is very low (see Figure 6-8). Finally, during several of these error trials, subjects made some movements and/or verbal exclamations to notify us of their mistake so this could easily be a confounding factor. It is therefore not surprising that we did not find any clear specific neuronal response during these error trials. When performing a direct statistical comparison of error trials versus non-error trials, there were two anterior cingulate neurons (out of the total of 27 neurons) that showed up as being statistically significant. There were also 5 neurons in the MTL (out of the total of almost 400 neurons, see Table 1) that yielded a significant value. Close visual inspection of the PSTHs from both the AC and MTL responses during these trials suggest that these were probably spurious activation or some artifact and not a real neuronal response.

6.5.4 Comparison of monocular responses and flash responses

After this brief interlude, we go back to the medial temporal lobe to directly compare the responses during the monocular presentation and the flash. We compared the latency of the neuronal response during the monocular presentation of the effective stimuli to that during the flash period when the effective stimulus became dominant. The distribution of the ratio of these two values (Figure 6-14a) is centered on a value of 1 and although there is considerable spread, there was no significant difference between the latencies in the two groups (two tailed t test, $p > 0.15$). Similarly, there was no significant difference between the duration of the neuronal responses (Figure 6-14b) in the two cases (two tailed t test, $p > 0.3$).

There was a very strong correlation between the changes in firing rate in response to the effective stimuli during the monocular presentation and those when the effective stimuli were flashed (Figure 6-14c). The r^2 value between the peak firing rates was 0.96 and the slope was 0.92. In spite of the fact that there is a completely different stimulus present in the retina during the flash period, the neuron parallels the perception and is completely oblivious to this ineffective stimulus and responds as in the monocular case when nothing was shown to the other eye.

In contrast, the correlation coefficient between the response to the flash of the effective stimulus when it was perceived and when it was suppressed was 0.08 and the two distributions were clearly different (two tailed t test, $p < 10^{-4}$). Note that in these two situations, the retinal input is basically¹⁰¹ the same and yet the neuronal activity is strikingly different. The neuron responds strongly if and only if the effective stimulus is perceived. In those cases, where the effective stimulus was presented monocularly, it is suppressed by the ineffective stimulus and the neuron does not change its firing rate.

6.5.5 Predicting the percept during flash suppression

It has been hypothesized that those neurons that represent the correlates of the contents of conscious perception may respond in an all-or-none manner. If this is the case, then one should be able to read out from this neuron (or this group of neurons) the on/off state and from that the contents of perception. We addressed this question quantitatively by performing an ROC analysis (see Chapter 3 for details). We computed the probability of misclassification p_e in predicting the presence of the effective stimulus based on the firing rate of the unit for those neurons that showed a statistically significant selective response. The results are summarized in Figure 6-15.

The results are divided into those neurons that were selective to a category of stimuli such as the one illustrated in Figure 6-10 and those that responded selectively only to one or more individual stimuli but not to all stimuli within a category such as the ones illustrated in Figure 6-11 and Figure 6-12. There was a strong correlation between the values of p_e during the monocular presentations and those during the flash period (Figure 6-15 **a,d**). Interestingly, the values of p_e during either period were clearly lower than those obtained during the flash period when the effective stimulus was suppressed from perception. When the preferred stimulus of the neuron was not perceived, the

¹⁰¹ Note that because of the design of our experiment the stimulus is not exactly the same. That is, an effective stimulus is paired randomly in each trial with a different ineffective stimulus (we do not know a priori which stimuli are effective and which ones are not in contrast to the work of Logothetis, see the discussion in the experimental procedures above). Therefore, in the two situations compared in the text the stimulus is always the same effective stimulus paired with different ineffective stimuli, in one case, the effective stimulus was shown monocularly and is therefore suppressed whereas in the other the ineffective stimulus was presented monocularly and therefore the effective stimulus dominates perception.

probability of misclassification was basically at chance levels (gray shaded area in Figure 6-15).

It would be interesting to know if there is a direct correlation between the activity of the neuron and the percept even in those trials in which subjects make a mistake. That is, in two trials with exactly the same stimuli both during the monocular presentation and during the flash period, but in which the behavior is different because the subject did not perceive the suppression in one of them, would the neuron also miss the suppression? However, the flash suppression phenomenon is too strong to accumulate a sufficient number of mistakes for each individual stimulus or even for a category of stimuli (see Figure 6-8). This type of analysis has been done in monkeys where the number of errors is sufficient to permit a statistical analysis (Celebrini and Newsome, 1994, Britten *et al.*, 1992).

6.5.6 Summary of responses during flash suppression

We computed an average of the neuronal responses to the effective stimuli during the different periods (Figure 6-16). Only those neurons that followed the percept were averaged in the plot. As discussed above for some specific examples, three aspects of the neuronal response of these neurons are evident. These units show a marked increase in their response to the monocular presentation of the stimulus; they do not respond beyond baseline during the flash period when the effective stimulus is suppressed from perception and finally they show a strong enhancement in their firing rate during the flash when the effective stimulus comes into perception. These results are similar for those neurons selective to a category of stimuli and those selective to individual stimuli (compare Figure 6-16a,c with Figure 6-16b,d). One interesting difference between the two is that the mean sdf_N for the ineffective stimuli is smaller for those units selective to a few individual stimuli. This is due to the observation that these responses are typically stronger, with a higher peak-to-baseline activity ratio¹⁰². There is a small increase in the sdf_N signal for the average of all other stimuli presented monocularly (blue trace in a and

¹⁰² As discussed in Chapter 3, this suggests that the neuron may not respond equally to all stimuli within the category even if the responses are statistically indistinguishable due to the large variance. The “category-specific” neurons show broader tuning curves and a smaller peak to baseline ratio.

b, particularly clear in a) during the flash presentation. That is because when the ineffective stimuli were presented monocularly, there were some trials in which the effective stimulus was flashed (but only a small percentage, see Figure 6-10, Figure 6-11, Figure 6-12 for example). Similarly, there is a small increase for the average of all other stimuli presented during the flash period during the monocular period (blue trace in parts **c** and **d**) for the very same reason. These changes are minuscule if existent at all in **b** and **d** because the number of presentations of the effective stimuli was typically much smaller (compare the number of presentations in Figure 6-11 and Figure 6-12 to those in Figure 6-10). If those trials are removed from the average in the computation of the average for ineffective stimuli, then the blue line becomes flat.

The total number of selective neurons and the ones that follow the percept in each region is shown in Figure 6-17 and Table 6-2. As we have shown in Chapter 3, most (in this case, all) of the visually selective neurons were in the MTL. Therefore, in what follows in this Section, I only discuss those units in the MTL. Of the total of 376 neurons in the MTL, 39 (10.3%) showed a selective response to a category of stimuli and 29 (7.7%) showed a selective response to individual stimuli. It should be reminded that there is no overlap between the group of neurons that showed a category-selective response (Cat) and those that showed a selective response to individual stimuli (Ind; see Section 6.3.3). In several cases, the unit was selective to a stimulus that was not presented a sufficient number of times during both the monocular presentation and the flash period. This constitutes the difference between the groups labeled "Cat" and "Ind" and those labeled "Cat*" and "Ind*." The overall distribution of the number of stimuli presented during the monocular period and during the flash period as well as the joint distribution of these two values is shown in Figure 6-7. Of the 31 neurons that were category-selective and where the selective stimuli had been presented a sufficient number of times during the flash period, 21 neurons (68%) were also selective during the flash period when the preferred stimulus was not perceptually suppressed. In all these neurons, the selectivity during the monocular presentation and the flash period was the same. Of the 17 neurons that responded selectively to individual stimuli and where the preferred stimuli had been presented during the flash period, 11 neurons (65%) were also selective during the flash when the preferred stimulus was not perceptually suppressed. In all these

neurons, the selectivity during the flash and the monocular presentation was the same. In other words, 68% and 65% of the visually selective units responded selectively during the monocular presentation and when the preferred stimulus was perceived during the flash but not when the preferred stimulus was suppressed by another stimulus.

Therefore, there were approximately 1/3 of the selective neurons that did not follow the percept. All these neurons failed to fulfill requirement (iii) from Section 6.3.3. That is, these units responded selectively during the monocular presentation but they did not respond selectively when the effective stimulus was flashed (regardless of whether it was perceived or not). It is possible that the lack of response of these neurons is due to the shorter presentation during the flash. This seems unlikely given that the latencies of neurons in the MTL seems to be much shorter than the 500 ms of the flash presentation. Another possibility is that the response of these neurons was somehow inhibited by the presence of another stimulus in the other eye. Interestingly, in lower visual areas, Logothetis' group has described the existence of some neurons that are selective but whose activity is anti-correlated with the percept and the activity of other neurons that are selective and respond in a constitutive fashion regardless of the perceptual alternations (Leopold and Logothetis, 1996, Logothetis and Schall, 1989). The activity of the neurons that do not follow the percept in our sample seems to be different from those reported by Logothetis. These neurons seem to respond to the monocular presentation and not during the flash of the preferred stimulus. It is not that these cells are anti-correlated with the percept or fire constitutively. Upon close inspection, it appears that the activity of these cells seems to be weaker and more poorly selective than the responses of those selective neurons that do follow the percept. This can be quantitatively observed in the p values that are in most cases between 0.01 and 0.05 where as the p values of most of the selective neurons that follow the percept is below 0.01. Thus, it is possible that the lack of response during the flash period is simply due to the weaker response that is sufficient to reach significance during the monocular presentation but not during the flash period.

Neurons seemed to respond to the flash of the preferred stimulus with a burst of spikes when it was perceived. Based on the interspike interval distribution and auto-correlogram it is possible to discriminate the existence of bursts (see Chapter 2). A certain classification of a burst of spikes, however, requires intracellular recorded data

that we do not have. This was discussed in further detail in Chapters 2 and 3 where we also indicated the criteria to consider a unit as bursting and the number of spikes per burst and burst duration. In $42\pm 32\%$ of the individual repetitions (ranging from 13 to 92%), there was a burst of spikes upon the appearance of the perceived stimulus during the flash when it was perceived. This was in contrast to the trials in which the preferred stimulus was suppressed, in which case we only observed bursts in $7\pm 4\%$ of the repetitions. During the monocular presentation, neurons responded with a burst of spikes in $34\pm 29\%$ of the cases and there was no significant difference with the proportion of bursts during the flash when the preferred stimulus was perceived. Bursts appeared with a latency of 156 ± 134 ms.

Given the low number of units, it is very hard to make any claims about whether there is a differential distribution of the number of neurons that follow the percept among the four different areas in the medial temporal lobe. We have argued in Chapter 4 that the possibility of establishing a clear distinction among the four areas in the proportion of selective units depends on the number of units and the strength of the regional differences. There, we presented the number of selective neurons required to yield significance in a χ^2 test as a function of the degree of non-uniformity of the distribution of selectivity. The same values apply to the data presented in the current Chapter.

6.5.7 Seizure activity

All the neuronal activity that we have recorded comes from epileptic patients. It is unclear, however, to what extent this could influence the results if at all. In previous work (Kreiman *et al.*, 2000b, Kreiman *et al.*, 2000a), most of our recordings were outside the seizure focus area. This question was also discussed in more detail in Chapters 2 and 3. I cannot yet provide the proportion of electrodes and selective units that were in the seizure focus because I have not yet been provided with this information from UCLA.

6.5.8 Inter-hemispheric switching

Miller and colleagues have recently put forward the suggestion that alternations during binocular rivalry could be due to a competition between rather than within

hemispheres (Miller *et al.*, 2000). According to this hypothesis, only one hemisphere would represent one percept while the contralateral hemisphere would represent the other percept and only one would be active while the corresponding percept predominates. In our case we have had simultaneous recordings in the right and left hemispheres. What kind of physiological data would support or contradict this hypothesis? If we had two units, one recorded from the left hemisphere and one from the right hemisphere, both selective to the same stimulus, the inter-hemispheric switching model would suggest that only one of them would be active while the preferred stimulus is predominant. If both units were active at the same time, however, this would pose difficulties on the model. Unfortunately, we still have not found any case with two units selective to the same stimulus one in each hemisphere.

6.6 Chasing the percept inside the brain

6.6.1 Comparison with data from other methods and models

Our results show that a large proportion (at least 60%) of the selective units modulate their activity following the percept of the subject. These observations parallel those observed in the macaque electrophysiological experiments of Logothetis' group. In humans, several investigators have explored the changes in blood flow in the brain during binocular rivalry (but not flash suppression). Tong *et al.* (Tong *et al.*, 1998) showed that areas selective for faces and places are strongly modulated by the percept. These regions were located in the fusiform gyrus and parahippocampal area, two areas high in the extrastriate visual cortex that, at least in monkeys, project to the MTL. Recently, Polonsky *et al.* (Polonsky *et al.*, 2000) and other unpublished work by Blake's group shows that there is a progressive increase in the correlation between the signal during binocular rivalry and that obtained in non-rivalrous presentation from V1 to higher visual areas. Interestingly, the work of Polonsky *et al.* and also another recent report (Tong and Engel, 2001) suggest that activity in V1 is modulated during rivalry alternations. We would very much like to record the activity of single neurons in our patients in V1;

however, we will have to wait until clinical considerations require the placement of electrodes in that area. Given the poor temporal resolution of fMRI, Tononi *et al* used magnetoencephalographic (MEG) recordings from 148 channels covering the entire head while subjects perceived binocular rivalry alternations of a two gratings shown at different frequencies (Tononi *et al.*, 1998). MEG offers excellent temporal resolution but the price to pay is a poor spatial resolution. They observed that the power of the MEG signal was modulated by the subject's percept in several posterior visual regions but also in anterior non-visual areas. The spatial resolution does not allow a precise localization of which regions were modulated. In a follow-up study, they showed that there are also substantial changes in the coherence of sensors in different brain regions (Srinivassan *et al.*, 1999). The synchronization of neuromagnetic responses occurred both between areas in separate hemispheres as well as within widely separated regions within the same hemisphere. At the single neuron level, Fries *et al* have suggested that the synchronization between neurons rather than changes in firing rate is correlated with perceptual alternations in rivalry in V1 (Fries *et al.*, 1997).

We are recording in an association area that receives direct input from the higher stages of visual processing in the cortical visual hierarchy, and therefore it is not clear that our data can speak directly to the Helmholtz versus Herring debate on whether attention or lower-level signals are causally responsible for the alternations during binocular rivalry. In other words, it is important to emphasize that the strong modulations that we have observed and those observed by Logothetis' group show a correlation between the activity and the percept and not necessarily a direct causal link. The same can be said about the functional imaging and MEG evidence mentioned briefly above. Establishing a causal link between neuronal activity and perception is indeed not an easy feat and will be one of the biggest challenges in the research of binocular rivalry. This is further discussed in the last chapter.

Some observations point to a role for the frontal cortex in the perceptual transitions during rivalry, although its exact function is still unclear. Prefrontal cortex seems to be activated according to functional imaging evidence during piecemeal rivalry or the transition switch (Lumer *et al.*, 1998). Patients with frontal lesions show a lack of flexibility in different tasks (Heilman and Valenstein, 1993) and a severe difficulty in

switching percepts in bistable percepts like the Necker cube (Ricci and Blundo, 1990, Meenan and Miller, 1994, Wilkins *et al.*, 1987). Recently, it has been shown that switching during binocular rivalry is also problematic in frontal patients (Pettigrew and Miller, 1998). We explored the putative role of frontal cortex during binocular rivalry by applying a stimulation technique that can be used in normal subjects called repetitive transcranial magnetic stimulation. This is discussed in detail in Appendix 3.

One interesting and intriguing observation made by Logothetis *et al* (Logothetis, 1998) is the existence of neurons that fired in “anti-correlation” with perception. That is, some neurons fired strongly when the ineffective stimuli were perceived, effectively signaling the disappearance of the effective stimulus from perception. These neurons were prevalent in intermediate visual areas (Logothetis and Schall, 1989, Leopold and Logothetis, 1996) but virtually absent in V1 and in the inferior temporal cortex and superior temporal sulcus (Sheinberg and Logothetis, 1997). We did not observe such anti-correlated responses in our data, supporting the idea that there is a stronger positive correlation between the activity of neurons in higher areas of the temporal lobe and perception.

All the properties of the neuronal responses that we have described match quite closely those reported in the macaque by Logothetis’ group. In the macaque, there is a monosynaptic connection between these areas; that is, the inferior temporal cortex projects directly to the entorhinal cortex, hippocampus and amygdala (Saleem and Tanaka, 1996, Cheng *et al.*, 1997, Felleman and Van Essen, 1991, Logothetis and Sheinberg, 1996) in addition to many other projections like frontal cortex. Ignoring momentarily the important fact that the two papers come from different species, it seems that the percept is then carried quite accurately from IT to the medial temporal lobe. The proportion of neurons that “followed” the percept that we report here is clearly higher than the ones observed for areas V4 and MT in monkeys (Sheinberg and Logothetis, 1997, Leopold and Logothetis, 1996, Logothetis, 1998, Logothetis and Schall, 1989)¹⁰³. On the other hand, it is smaller than the values reported by Sheinberg and Logothetis (Sheinberg and Logothetis, 1997). A summary comparing these data sets is shown in

¹⁰³ It should be noted, however, that in the lower visual areas, they only report values for binocular rivalry and not for flash suppression. In contrast, our data was mostly recorded during flash suppression.

Figure 6-18. It is quite likely that these differences are simply due to the different criteria used to determine whether units were selective or not¹⁰⁴. Alternatively, it is possible that, indeed, the proportion of neurons that “follow the percept” is actually lower in the medial temporal lobe than in the inferior temporal cortex. Maybe the correlation with perception increases from the retina until some area like the inferior temporal cortex and then decreases in successive anatomical targets. Clearly, at the other extreme of the nervous system, the behavioral side is correlated with the motor response rather than the sensory input.

One should not forget that we are talking about different species here and it is possible that the physiology and perceptual properties of the visual system could be very different between the two species. However, it must be emphasized that the behavioral responses of the monkeys during binocular rivalry seem to resemble those of humans so strongly (Myerson *et al.*, 1981, Logothetis, 1998, Sheinberg and Logothetis, 1997, Leopold and Logothetis, 1996, Logothetis and Schall, 1989) that it is hard to believe that this would be so. Of course, this is one of the major assumptions in all our work exploring the electrophysiology of the monkey brain. Although none of us would like to even entertain this possibility for more than one short paragraph, it is conceivable although unlikely that there are some major differences between the monkey brain and the human brain. After all, monkeys don't play chess.

Another point that should be mentioned is that monkeys are extensively trained in the binocular rivalry / flash suppression task before the recordings. Of course, the experiment cannot be done otherwise in monkeys. In our case, we are dealing with completely naïve participants. Binocular rivalry is known to be strongly affected by training (Leopold and Logothetis, 1999, Blake, In press). In the case of flash suppression, however, the phenomenon seems to be so strong that it is unlikely that there would be major changes due to training. Indeed, we have unpublished data from subjects that performed the flash suppression task quite extensively and we have not noted any clear changes (some of these data will be discussed in Appendix 2). It is also to be noted that in the lower visual areas, the proportion of neurons plotted in Figure 6-18 includes the “anti-

¹⁰⁴ In our case, we have used standard statistical criteria that we have defined and applied uniformly to all the data (see Section 6.3). Most of the units that were visually selective but did not follow the percept actually showed very weak responses. With 2 exceptions, they all showed p values between 0.01 and 0.05.

correlated” units described before whereas that is not the case for the IT/STS data or for our data.

A related interesting question concerns the responses of neurons in the frontal cortex during binocular rivalry. Unfortunately, we did not record from many neurons in the frontal cortex¹⁰⁵. The few neurons that we recorded from did not show clear selective changes in firing rate for different visual pictures. As mentioned above, frontal cortex may play a role in the transitions (Lumer *et al.*, 1998), see also Appendix 3).

All the data presented in this chapter as well as in Chapters 3 and 4 rests on the assumption that the neuron carries some information in the firing rate. The statistical analysis was done by counting spikes in rather large time windows. While this point is going to be discussed in more detail in the next chapter, I do not want to let the opportunity pass without saying that this is likely to be an oversimplification and that there is likely to be much more information conveyed by the neurons in the specific timing and interactions between spikes in different units.

One of the stimuli is completely suppressed from perception, that is, it becomes invisible. This does not mean that it could not affect perception or behavior in some way. Indeed, the school of Freud (Freud, 1966) extensively described what they called “sub-conscious” processing of information that can strongly affect behavior even in the complete absence of awareness. The idea that brain processes that don’t reach perception or awareness can still show a marked effect on behavior is widely accepted now. Where do these sub-conscious processes take place in the brain? Well, during flash suppression at least, it does not seem very clear that they could take place in the human medial temporal lobe. We did not find any clear evidence of processing of the unperceived stimulus during the suppressed period. The activity of the selective units while their preferred stimuli were perceptually suppressed was indistinguishable from baseline. The few neurons that were selective but did not follow the percept (see Table 6-1) responded only during the monocular presentation and not during the flash. Thus, there was no neuron that responded to the preferred stimulus while it was perceptually suppressed.

¹⁰⁵ Let me remind the reader that we cannot choose where to record from since this is determined by clinical criteria of seizure localization (see Chapter 2).

In one scenario, there could be strong activity in a given brain area that does not reach perception. Alternatively, it could be the same area that represents the conscious and sub-conscious information, but with different intensities and a threshold being responsible for the strong difference (Crick and Koch, 1998, Koch, *In Preparation*). While the selective stimulus is suppressed, the neuronal response is indistinguishable from baseline in the activity of the neurons that we have recorded (see Figure 6-9 and Figure 6-16). The suppressed stimulus can however influence behavior (Blake, *In press*). One of the strongest pieces of evidence for this comes from several studies of visual adaptation. Perceptual suppression of a stimulus does not have any effect on its ability to induce different types of aftereffects including the grating threshold elevation aftereffect, the spatial frequency shift aftereffect, the translational motion aftereffect and the tilt aftereffect. That is, albeit the information about the suppressed stimulus does not reach consciousness, it is not completely lost and it can directly affect behavior. This constitutes yet another example of the existence of neuronal processes that can influence our actions and thoughts without leading to awareness (Koch and Crick, 2001, Crick and Koch, 1998, Goodale and Milner, 1992). Therefore, there has to be some neural representation of the perceptually invisible stimulus. Our evidence suggests that it is unlikely that this unconscious representation occurs in the human MTL; we hypothesize that is perhaps a property of lower visual areas or prefrontal areas.

Activity in the medial temporal lobe has been strongly associated with memory processes (Squire and Zola-Morgan, 1991, Zola-Morgan and Squire, 1993, Heilman and Valenstein, 1993, Kandel *et al.*, 2000). Therefore, it is conceivable that the neuronal responses that we have observed are not the direct result nor cause of the perceptual alternations. Maybe these occur early in the visual system and the information is then rapidly conveyed to the medial temporal lobe to be stored. Alternatively, it could be that the switch activated in some area like the frontal cortex is sent to the medial temporal lobe as the perceptual transition is activated and only then is this information conveyed back to the inferior temporal cortex and other visual areas. We are, however, entering the realms of philosophical and theoretical speculations, and I would prefer to leave that for the last chapter.

6.6.2 Perceiving IT?

As described previously in this chapter, there is a strong increase in the percentage of neurons that follow the percept from lower visual areas to the inferior temporal (IT) cortex in the monkey brain. This has led some to the idea that the activity in IT may be a good indicator of the contents of aware perception. How unique is IT in its strong correlation with perception? In contrast to functional imaging where it is possible to examine several areas of the brain simultaneously, monkey electrophysiology (at least the way it is done nowadays) requires the laborious surgery, insertion of electrodes and search for neurons, one area at a time. Furthermore, for most monkey electrophysiology labs, it is actually one neuron at a time. Therefore, in the majority of cases we only know in each experiment about the responses in one area (until a new graduate student comes and repeats the experiments in a different brain area). Neurons in IT show strong projections not only to lower visual areas as a feedback signal but also to the medial temporal lobe structures like the hippocampus, entorhinal cortex and amygdala and frontal cortex structures. Our data suggests that the activity in the medial temporal lobe is also strongly correlated with perception. Assuming that this result also extrapolates to monkeys this suggests that the correlation with perception can occur in several areas of the visual system including IT, the MTL and perhaps even frontal cortex neurons. This brings us back to the central question of causality. While it is easier to search for correlates rather than causes, and searching for correlates seems to be an excellent temporary strategy, the most fundamental question that we will eventually have to pursue is the search for the neuronal *causal* correlates of consciousness (nccc) rather than the neuronal correlates of consciousness (ncc, see Chapter 7).

6.6.3 Are monkeys conscious during binocular rivalry?

There are several important conceptual differences between our work and the one performed in monkeys by Logothetis. To begin with, monkeys are highly overtrained in the binocular rivalry and flash suppression tasks. Binocular rivalry is not an easy task for monkeys to learn. Binocular rivalry clearly does not constitute part of their normal everyday lives in monkeys' natural environment but they can learn to satisfy the

investigators for appropriate orange juice rewards. It is not untypical for the animals to be trained several hours a day for several months. To what extent this would have an effect in the neuronal responses is unclear at this moment. In the recordings in earlier visual areas, monkeys are vastly overtrained with the exact same type of stimuli and parameters that are used during the actual testing. In the inferotemporal cortex studies, it seems that monkeys were trained with somewhat different stimuli (Koch, personal communication). Yet, it is known that training can affect even in normal humans the responses during rivalry (see for example Leopold and Logothetis, 1999). There is no way to do the experiments in monkeys that report their percepts without this arduous and lengthy training. In our tests, as we have emphasized before, subjects were not trained before the actual recording sessions. The existence of highly selective neurons whose activity is correlated with the percept suggests that this neuronal activity is indeed related to the visual sensation and is not due to the overtraining in monkeys.

A second related but important distinction must be made. In our case, we can be sure that the subjects were indeed aware of the perceptual changes whereas this is less clear in monkeys. Given that this is one of the crucial issues in the thesis, I think we should not neglect the subject. One of the typical consequences in humans of being trained in a particular repetitive task over and over again is automatization. That is, the first time we drive a car, we are aware of every single part, every detail, and every noise. After driving the car 1000 times, it is not atypical for us to think about something else and be completely unaware of the car itself. The same thing happens when we dance or when we play tennis. Training can lead to faster, stereotyped and repetitive responses that lack the conscious component. Indeed, in several cases, it can be argued that this is a good feature of the nervous system; it is not unusual for tennis players to claim that they play better if they do not think about the movement. It has been shown extensively that the strength of synapses can be dynamically modified depending on their usage. Indeed, massive reorganization in the brain can occur due to training. In blind subjects, for example, the activity in V1 can be associated with tactile discriminations (Cohen *et al.*, 1997, Hamilton and Pascual-Leone, 1998). As a movement becomes more practiced, it is represented more extensively in primary motor cortex (Karni *et al.*, 1995). The neuronal responses in the areas responsible for finger movement can be dynamically modified by

training monkeys in tactile discrimination tasks (Mushiake *et al.*, 1991). Since a large fraction of our exploration of the visual system at the neuronal level is done in monkeys, we are prone to believe that the monkeys are perfectly conscious during the visual tasks. This is more an expression of wish that a scientifically founded proof. Most investigators prefer to simply ignore this highly problematic issue and sweep it under the carpet.

It has been argued that the anatomy of the monkey visual system is very similar to the human visual system. This is likely to be so but the available data is scarce. The anatomy of the human brain has not been studied at the same level of detail as the monkey visual anatomy. Actually, there are still plenty of important questions that remain to be answered about the anatomy in monkeys as well. For most of the visual areas, we still do not know with certainty what the analogous areas in the human brain are. One would think that at least the retinae would be completely conserved. And still, recent data suggests that not even this may be true since it has been reported that there are spiking photoreceptors in humans that have not been found in any other species. There are fascinating overall similarities in several properties of the higher visual areas in monkeys and humans. The effects of lesions in monkeys seem to parallel in many cases quite closely some of the problems observed in neurological patients. But the spatial resolution of these studies is not enough to be certain that the two systems are completely analogous. After all, there are clear differences between humans and monkeys.

6.7 Figures legends

Figure 6-1: Necker cube: example of a bistable percept

This is an example of a figure that can be seen in two different conformations as observed by a Swiss crystallographer named Necker in 1832. While there is no change in the image, the cube can be seen in two different conformations with the faces enclosed by red lines facing ahead. In this case, the alternations between one and the other interpretations can be volitionally induced.

Figure 6-2: Binocular (and monocular) rivalry

Seen through red/blue glasses, the horizontal red grating and the vertical blue grating are projected onto the two retinae. Upon fixating on the white cross for several seconds, perception alternates between the two so that for a period of time the horizontal grating becomes completely dominant and then it is suppressed as the vertical grating comes into perception. Even without the red/blue glasses, a weaker version of the phenomenon (monocular rivalry) can be experienced with enough patience.

Figure 6-3: Behavior during rivalry and the gamma distribution

(a) Two different stimuli (in this case a face and a house) are presented to corresponding areas of the two retinae during binocular rivalry. In our experiments with the patients we used liquid crystal shutter glasses to present the stimuli to each eye. During binocular rivalry, perception alternates between the two images in a seemingly random fashion (b). The dashed lines indicate the periods of piecemeal rivalry where a mixture of the two images is seen whereas the continuous lines show the periods of exclusive dominance of one of one of the two stimuli (gray for the house and black for the face). The normalized distribution of dominance times for the two images is shown in part (c). The inset shows the corresponding distribution for the piecemeal rivalry periods. The continuous line

corresponds to the fit of a gamma function with parameters 2.5 and 3.2 respectively. The distribution of piecemeal rivalry follows a distribution close to an exponential (gamma of order 1). The correlation coefficient between the data and the gamma fit was 0.89 and 0.94 and the mean dominance periods were 1574 ± 893 ms and 1335 ± 534 ms respectively. Data from all sessions within one experiment were averaged to compute the distribution illustrated here.

Figure 6-4: Binocular rivalry: single trial example

Example of the neuronal activity of a unit located in the entorhinal cortex during one session of the binocular rivalry experiment. **(a)** Stimulus presented to each eye (top=left eye and bottom=right eye). During some periods, only one stimulus was presented monocularly in order to verify the subject's responses (there was a blank square of the same color as the background in the other eye). The vertical dashed lines indicate the changes in the stimulus presentation. **(b)** Neuronal response. Each tick corresponds to an action potential from the neuron. The continuous line shows the estimation of the spike density function obtained by convolving the spike train with a gaussian of a fixed width of 200 ms. **(c)** Behavioral responses of the subject. Black lines show the periods where the face was dominant whereas gray lines show the periods where the house was dominant. During piecemeal rivalry, subjects were instructed not to press any button.

Figure 6-5: Binocular rivalry: selective neuron example

Average activity of the unit illustrated in the previous figure for different stimuli within each group. Post-stimulus time histograms and rasters are aligned to the behavioral switch as indicated by the time at which the subject pressed a button so that the category indicated in the top became dominant. This time is denoted as $t=0$ and is indicated by a dashed vertical line. Before $t=0$, subjects perceived either rivalry or stimuli from a different category. The number of periods over which the data is averaged is indicated in parenthesis next for each group of stimuli. The dashed horizontal line shows the mean firing rate over the whole experiment. Bin size=200 ms.

Figure 6-6: Flash suppression paradigm

A stimulus (in this particular case a horizontal grating) is presented monocularly for a time $t_1=1000$ ms. Immediately after the monocular presentation, the same stimulus is flashed onto the same (ipsilateral) eye and a different stimulus (in this case a photograph of an eagle) is flashed to the other (contralateral) eye for a time period $t_3=500$ ms. In this case, subjects perceive first the horizontal grating during the monocular presentation and then the eagle during the flash. Note that there was no time interval between the monocular image and the flash. Subjects were asked to indicate whether the flash was the same picture as the first one or not by pressing a button. In 20% of the experiments, they were also requested to give a verbal description of what they had seen. In less than 10% of the trials, only the monocular image was flashed without a second image as a simple behavioral control.

Figure 6-7: Distribution of the number of presented stimuli

Distribution of the number of stimuli presented during the different phases of the flash suppression experiment. **(a)** Distribution during the monocular presentation. The arrow indicates the mean value. **(b)** Distribution during the flash period. The arrow indicates the mean value. **(c)** Joint distribution of the number of presentations for each stimulus during monocular and flash periods.

Figure 6-8: Behavior results during flash suppression

Average results of the behavioral responses required from the subjects during the flash suppression paradigm. Two types of responses were requested. The first one was a two-alternative forced choice in which they had to push a button to indicate whether the first monocular image was suppressed and a second image was observed or not. This was done in 100% of the trials. In 20% of the trials, subjects were also debriefed and asked to give a verbal description of what they had observed. The percentage indicates the

proportion of cases where the monocular image was suppressed during the flash suppression trials (labeled “FS”) and during the control trials (labeled “C”).

Figure 6-9: Flash suppression: single trial example

The response of a neuron in the right amygdala during flash suppression in two individual trials is illustrated here. In each plot, the top part shows the stimulus presentation, the middle part indicates the neuronal response and the bottom part shows the perception. In **(a)**, a photograph of Paul Mc Cartney was shown monocularly to the right eye during t_1 ms and a house was flashed during t_3 . The house was perceived and suppressed the image of Paul Mc Cartney. The neuron responded vigorously (with 13 spikes in this particular trial) while the subject saw the singer’s face and it stopped responding when the face was perceptually suppressed. The vertical dashed line denotes a change in the stimulus; note however, that there was no time interval between the monocular stimulus and the flash. In **(b)**, a horizontal grating was shown monocularly and the photograph of Paul Mc Cartney was flashed, suppressing the image of the grating. The neuron fired strongly (with 17 spikes in this particular trial) when the face of the singer was perceived.

Figure 6-10: Example 1, responses to categories

Example of the responses of a unit in the left parahippocampal gyrus during the flash suppression test. This unit responded selectively to spatial stimuli. **(a)** Rasters and PSTHs aligned to the times at which a spatial stimulus was presented monocularly and an ineffective stimulus was flashed. The first vertical dashed line indicates the time of presentation of the monocular stimulus and the second vertical dashed line indicates the time at which the flashed stimuli were presented. Flashed stimuli were drawn from all other categories presented in this experiment including emotional faces, photos of animals, photos of famous people and patterns. The dashed horizontal line indicates the mean firing rate of the unit over the whole experiment. The number of presentations is shown above the rasters. **(b)** Rasters and PSTHs aligned to the times at which a spatial layout stimulus was flashed and a different picture (any stimulus except a spatial layout)

was presented monocularly. The conventions are the same as in part **a**. **(c)** Rasters and PSTHs aligned to the monocular stimuli in each category. Note that the second PSTH is the same one shown enlarged in part **a**. Conventions are as in part **a**. Stimuli presented during the flash for each histogram were drawn from all other categories except for the monocular stimulus. Thus, the effective stimulus was flashed in the contralateral eye in several of the trials for all but the second PSTH and can cause a small increase in the firing rate after the flash in the PSTH. This is particularly evident in the last PSTH (abstract patterns) because there were slightly more presentations of the spatial stimuli during the flash. **(d)** Rasters and PSTHs aligned to the flashed stimuli in each category. Note that the second PSTH is the same as the one shown in part **b**. Conventions are as in part **a**. Stimuli presented during the monocular period in these PSTHs were drawn from all other categories except for the one indicated above the rasters. Thus, spatial stimuli were presented monocularly in some of the trials in all the PSTHs except for the second one and therefore can give rise to a small increase in the firing rate during the monocular presentation. **(e)** Rasters and PSTHs aligned to the monocular stimulus for each of the individual stimuli from the effective category for this unit. Although there is a considerable degree of variability from one picture to another, there was no significant difference in the analysis of variance comparing the across-picture variability to the variability upon repeated presentations of the same picture. **(f)** Rasters and PSTHs aligned to the flashed stimulus for each of the individual stimuli from the effective category (spatial scenes) of this unit. The 5th PSTH is missing because the number of presentations of this stimulus was <4.

Figure 6-11: Example 2, responses to individual stimuli

Responses of a neuron in the amygdala to the picture of Curly, a character from a famous US TV series called "The Three Stooges." The selectivity of this neuron with respect to all the other stimuli that were presented in the experiment was already illustrated in Chapter 3, Figure 3-8. The neuron increased its activity transiently and reliably when Curly was shown monocularly and did not respond when Curly was flashed but not perceived because of the suppression by an ineffective stimulus (left). There was also a

transient and reliable increase in the activity when the image of Curly was flashed and it was perceived suppressing an ineffective stimulus (right). The first vertical dashed line shows the time at which the monocular image was presented. The second vertical dashed line indicates the time at which the flashed stimuli were presented. The horizontal dashed line indicates the mean firing rate of the unit during the whole experiment.

Figure 6-12: Example 3, responses to individual stimuli

Responses of a neuron in the amygdala to three different pictures of Clinton. The responses of this neuron to the monocular presentation were illustrated in Chapter 3, Figure 3-7; there we showed the responses to all the other ineffective stimuli. The format and conventions of the figure are the same as in the previous figures. The neuron increased its firing rate when Clinton was shown monocularly and stopped responding when it was suppressed by an ineffective stimulus (left). There was also an increase in the firing rate for two of the pictures when the photograph was flashed and suppressed other stimuli (right). For the other photograph, there were unfortunately only two presentations.

Figure 6-13: Example in the SMA

Response of a neuron in the left supplementary motor area during the flash suppression paradigm. No clear differences were observed among different stimuli; therefore, all figures ($n=250$) presented in the experiment were averaged in this figure. As in the previous figures, the histogram is aligned to the presentation of the monocular figure at $t=0$. The time of onset of the monocular stimulus and the flash are marked by a vertical dashed line. The horizontal dashed line shows the mean firing rate of the unit.

Figure 6-14: Comparison of response properties during monocular presentation and flash period

(a) Distribution of the neuronal response latencies during the monocular presentation (top) and during the flash period (bottom). The response latency was measured from the time at which the monocular picture was presented (top) and from the time at which the

flashed stimuli were presented (bottom). Bin size = 50 ms. There was no significant difference between the latencies (two tailed t test $p > 0.1$). The arrow indicates the mean value **(b)** Distribution of the neuronal response duration during monocular presentation (top) and during the flash period (bottom). Bin size = 50 ms. There was no significant difference between the response durations (two tailed t test, $p > 0.2$). The arrow indicates the mean value **(c)** Distribution of the firing rates in spikes/sec during monocular presentation (top) and during the flash period (bottom). Bin size = 2 spikes/sec. There was no significant difference between the firing rates (two tailed t test, $p > 0.4$). The arrow indicates the mean value. All the distributions in this figure were computed by considering all the selective units (see Table 6-2).

Figure 6-15: ROC analysis

Probability of misclassification (p_e) in predicting the presence of the effective stimulus for each cell during the monocular or flash presentation period based on the firing rate (see text and Chapter 3 for details) **(a-c)** Data for all the category selective units that followed the percept ($perc_{cat}$ see Table 6-2). **(a)** Direct comparison of p_e during the monocular presentation and during the flash period when the effective stimulus is perceived. The dashed diagonal shows the $y=x$ line. The solid line corresponds to the linear fit to the data with $r^2=0.81$. We also computed the p_e value for the preferred stimuli when they were presented during the flash but they were not perceived due to suppression by another stimulus. The shaded rectangle shows the 95% confidence interval for p_e when the preferred stimulus was perceptually suppressed. The distribution of values of p_e is shown in **(b)** for the monocular presentation and in **(c)** for the flash presentation. Bin size = 0.05. The arrow indicates the mean value. The shaded rectangle shows the 95% confidence intervals for p_e when the effective stimulus was suppressed from perception. That is, an ideal observer would be almost at chance level when asked to discriminate the presence or absence of the preferred stimulus based on the firing rate while it was suppressed from the subject's perception **(d-f)** Data for those units that were selective to individual stimuli and followed the percept ($perc_{ind}$ see Table 6-2). Conventions are the same as in the corresponding figures **a** through **c**. Only the neurons that followed the

percept were included in this figure. For those neurons that were selective but did not follow the percept (see Table 6-2), there was no correlation between the p_e values during the monocular presentation (when it was typically below 0.4) and during the flash (when it was typically above 0.45).

Figure 6-16: Normalized SDF summary for flash suppression

Mean normalized spike density function responses to effective and ineffective stimuli. The spike density function sdf was obtained by convolving the spike train with a gaussian of 200 ms fixed width (see Chapter 3). For each unit, the sdf was normalized by dividing by its peak activity yielding sdf_N . Note that the baseline was not subtracted and therefore $0 \leq sdf_N \leq 1$. Although the maximum of the sdf_N to the effective stimulus for each unit was 1 by definition, the maximum of the average value over all units is less than 1 because the maximum does not occur at the same time for every unit. The average was computed separately for those units that were category selective (a, c) and those that responded to some individual stimuli but not to all within a category (b,d). The red trace in parts a and b shows the average activity for those trials where the effective stimulus was shown monocularly and an ineffective stimulus was flashed. The blue trace shows the average response to all the other stimuli. The shaded regions correspond to 95% confidence intervals. The vertical dashed lines indicate the time of monocular presentation and the time of the flash respectively. The red trace in parts c and d shows the average activity for those trials in which an ineffective stimulus was shown monocularly and the effective stimuli were flashed. The blue trace shows the average response to all other stimuli.

Figure 6-17: Summary of number and proportion of neurons that follow the percept

Summary of number of responsive units in each location and those that followed the percept. Units that were selective to categories or individual stimuli are plotted separately. Only those units for which the number of presentations during both the

monocular period and the flash period was more than 4 are shown in here. The labels in the x-axis are defined in Table 6-1.

Figure 6-18: Comparison with data from monkeys

Comparison of our results from those obtained in monkeys. Nomenclature, LGN: lateral geniculate nucleus; V1, V2, V4 and V5 visual areas 1, 2, 4 and 5; IT/STS inferior temporal cortex and superior temporal sulcus. The data were taken from the following published reports: LGN (Lehky and Maunsell, 1996), V1/V2 and V4 (Leopold and Logothetis, 1996), V5 (Logothetis and Schall, 1989), IT/STS (Sheinberg and Logothetis, 1997). Note that the value for the LGN is not missing, Lehky and Maunsell report that none of the LGN units were correlated with the percept.

6.8 Tables

Table 6-1: Number of units analyzed during binocular rivalry

	Amy		AC		Hip		EC		Par		Occ		Fr		PHG		SMA		Tot	
	L	R	L	R	L	R	L	R	L	R	L	R	L	R	L	R	L	R	L	R
<i>N</i>	0	8	3	0	4	12	1	40	0	0	0	30	0	0	1	7	9	4	18	101

Number of units analyzed during the binocular rivalry test in each region. Units from the left and right hemispheres are shown separately. From left to right, the locations are amygdala, anterior cingulate, hippocampus, entorhinal cortex, parietal lobe, occipital lobe, frontal cortex, parahippocampal gyrus, and supplementary motor area.

Table 6-2: Number of units and selectivity during flash suppression test

	Amy	Hip	EC	PHG	AC	Fr	SMA	Total
<i>N</i>	143	85	120	28	27	5	18	403
<i>Cat</i>	17 (12%)	11 (13%)	6 (5%)	5 (18%)	0	0	0	39
<i>Cat*</i>	14	9	4	4	0	0	0	31
<i>Perc_{cat}</i>	10 (71%)	7 (78%)	2 (50%)	2 (50%)	0	0	0	21 (68%)
<i>Ind</i>	15 (10%)	6 (7%)	3 (3%)	5 (18%)	0	0	0	29
<i>Ind*</i>	9	4	1	3	0	0	0	17
<i>Perc_{ind}</i>	6 (67%)	3 (75%)	0 (0%)	2 (67%)	0	0	0	11 (65%)

n is the total number of units in each location. Right and left hemispheres were merged to simplify the table. The percentages of units in the right hemisphere were: Amy 77%; Hipp 46%; EC 55%; PHG 36%; AC 15%; Fr 0%; SMA 56%. *Cat* indicates the number of units that were category selective during the monocular presentation. The percentage is computed with respect to the total number of units in that region. *Cat** indicates those units for which the selective stimuli were presented enough number of times during the

flash for a statistical analysis (≥ 4 for each individual stimulus; see the distribution of the number of presentations in each phase in Figure 6-7). “*Perc_{cat}*” indicates those units that were category selective and followed the percept (see text for details); the percentage is computed with respect to *Cat**. *Ind* denotes those units that showed selectivity for individual stimuli during the monocular presentation (see text for details); the percentages are computed with respect to the total number of units in that location. *Ind** denotes those units that were selective to individual stimuli during the monocular presentation and were presented enough number of times (≥ 4) during the flash period for statistical analysis. *Perc_{ind}* indicates those units that were selective to individual stimuli and followed the percept (see text for details); the percentages are computed with respect to *Ind**. The total number of units in the MTL (first 4 columns) was 376.

7 Summary and Reflections

7.1 Brief summary of the results

It is always an interesting exercise to attempt to summarize four years of one's work and almost 500 pages into a single sentence. Recording from individual neurons in the human brain we have observed that there is a high degree of invariance in the representation of perceptual information in the medial temporal lobe to dramatic changes in the visual input. This invariance manifests itself in the selective responses to different stimuli beyond physical similarities (Chapter 3), in the selective responses during visual imagery in the absence of visual input (Chapter 4) and in the selective responses that followed the percept in the presence of a different stimulus in the contralateral eye (Chapter 6). The MTL is at the top of the visual information processing hierarchy (Felleman and Van Essen, 1991, Kandel *et al.*, 2000) and receives direct projections from the higher stages of the visual pathways. In particular, the inferior temporal cortex in monkeys projects directly to the perirhinal and entorhinal cortices (Saleem and Tanaka, 1996) as well as to the amygdala (Cheng *et al.*, 1997, Suzuki, 1996). Whether the detailed anatomy of the visual system in monkeys and humans is conserved or not is still unclear but it seems likely that the selective visual units that we have observed receive input from higher stages of processing in the visual system in humans as well. The MTL receives input from other modalities in addition to the visual ones and it is therefore referred to as a multimodal association area (Kandel *et al.*, 2000). As such, the activity of individual neurons as briefly described in this Thesis, seems to be closely linked to the contents of our perception. The neuronal activity in the MTL does not seem to correlate with the several physical changes in the world to which perception is oblivious. In Chapter 1 we emphasized the remarkable constancy of our perceptual discriminations even in the

presence of dramatic changes in the information that reaches the retinae. A neuron that is selective to a face can respond when the face is shown but also when it is consciously seen during rivalry in the presence of a completely different picture projected onto the other eye; when the face is perceptually suppressed during a flash suppression experiment, the neuron's activity is indistinguishable from baseline (Chapter 6). Furthermore, the neuron can respond even in the complete absence of the stimulus in the physical world during volitional visual imagery (Chapter 4). Thus, our results show that there is a strong correlation between the contents of perception and the responses of individual units in the MTL. Does this imply that the activity in the MTL directly underlies the perceptual changes? This is a possible hypothesis but it is not a necessary implication of our results and it worth considering alternative possibilities. There is still a long way to go until we can clearly understand the direct causal relationship between neuronal activity and perception. But before we discuss the fundamental distinction between correlation and cause, let us reflect and inquire more into the properties of the responses of the MTL.

7.2 Invariance in perception and neuronal response

As we have argued before, there are several situations where the stimulus as measured at the level of the retina can change dramatically, yet our perception remains unaltered (see for example Figure 1-1). The reverse is also true; the retinal input can remain constant and the brain can be engaged in large perceptual alterations (for example during visual imagery (Chapter 4), dreams (Chapter 5) and binocular rivalry (Chapter 6)). Some of the simplest invariance properties correspond to changes in the size, luminance and position of a stimulus. If we look at a picture of Clinton, we can recognize it regardless of its size, the luminance in the room, or the position of Clinton within the picture. Large changes can be imposed on the picture such as altering its color, spatial frequency distortions, inversions and so on; these may show very small if any influence of perception. This was illustrated for the Van Gogh Sunflowers in Chapter 1. Neurons in lower visual areas are highly sensitive to these parameters whereas the response of

neurons in higher areas remains largely the same. For example, a slight shift of a stimulus outside of the small receptive field of a V1 neuron will dramatically alter its response. We did not explore the neuronal responses in the MTL to these physical changes because of time constraints; that is, we did not attempt to systematically vary the size, position, color, contrast and other properties of the pictures that were being presented. While all the stimuli were presented at the fovea and at a fixed size, I speculate that the responses of our neurons would not be altered by large changes in these “elementary” parameters such as position, size, color or contrast. Several investigators have shown this type of invariance in recordings in the monkey inferior temporal cortex (Tanaka, 1996, Gross, 1994, Desimone *et al.*, 1984, Logothetis and Sheinberg, 1996). Furthermore, inferior temporal cortex neurons show a response that is insensitive to the specific cue (such as contrast, texture or motion) used to delimit the selective shape (Sary *et al.*, 1993).

We have observed a different type of invariance in our study of the visual responses in the MTL. Some neurons showed responses that were selective to several pictures within a group of stimuli; these were referred to as category-selective neurons in Chapter 3. As we have argued there, it is unclear whether these neurons would respond equivalently to every possible stimulus within the selective category and not at all to other categories. For several units, we did not observe a statistically significant difference in the responses to distinct individual stimuli within the category. However, a unit selective for pictures of animals may very well respond much more strongly to a particular animal (e.g., elephant) that we did not present in the experiment. While our results do not imply that the general concept of a category is necessarily distinguished by individual neurons, it is interesting to speculate that the broad tuning that we have observed shows the possibility for the brain to extract some common information from physically distinct stimuli.

This is a level of abstraction and invariance much harder to achieve than the size invariance referred to earlier. For example, it is comparatively simple to write an algorithm that will recognize a single identical stimulus, say a specific house, shown in different colors, positions in the screen, sizes and contrasts (but always in the same orientation). Another matter would be to write an algorithm to discriminate all spatial layouts from all the other stimuli. The recognition of objects and other visual stimuli in a

manner that shows the same invariant properties that our perception does has been an important challenge for the computer science community. While there are several algorithms that have been proposed and volumes written on the matter, I think it is still fair to say that no digital camera and computer can match the performance of a 4 year old kid in the accurate identification of different objects under widely different conditions of light, size, color and so on. The abstraction to categories of stimuli that do not necessarily share physical properties will probably present a challenging problem to computer science. There are physically similar stimuli that must be discriminated into different categories and physically dissimilar ones that are to be lumped together into a common group.

Animals can be trained to distinguish certain categories of stimuli (see for example Vogels, 1999a) and some of these distinctions can be also observed at the neuronal level after extensive training (Freedman *et al.*, 2001, Vogels, 1999b). It is still unclear how these categorical responses arise in these neurons. Vogels' group has proposed that the distinction between categories can arise from the ensemble activity of units that are only poorly category-selective (Thomas *et al.*, 2001). According to this model, category-selectivity should be searched for in targets of units that are poorly category-selective and that show highly selective visual responses such as IT. Both the MTL and frontal cortex receive strong input from the inferotemporal neurons.

Fortunately, when we turn off a computer, the information (at least an important part of the information) does not disappear; we can turn on the computer and our files are still there. Similarly, we do not need constant visual input in order to create a percept. We can volitionally re-create an image during visual imagery and perceptual impressions occur during dreams and hallucinations in the absence of concomitant visual input. This constitutes arguably one of the strongest possible levels of invariance that the visual system is subject to: the existence of a percept in the absence of any information whatsoever in the input. At least during the process of imagery, we have observed that some neurons can modulate their activity in a selective manner depending on the impression the subject is holding in their minds. Furthermore, the selectivity of most of these units during imagery was the same as that during vision. But we should now pause and re-consider some important alternative interpretations of these results.

7.3 Perceptual processing, memory or language?

We have argued along this Thesis that the neuronal responses that we have observed upon presentation of visual stimuli, those during flash suppression, and those during visual imagery represent a correlate to the perceptual sensation the subject experiences. This is not necessarily the only possible interpretation of the experimental data. I will focus here on reflecting about two additional possible interpretations that seem to be among the most important ones.

7.3.1 Memory

The medial temporal lobe has been suggested to play a crucial role in several memory processes. Evidence for this claim comes from several different experiments in both humans and other animals (Buckner *et al.*, 2001, Squire and Zola-Morgan, 1991, Zola-Morgan and Squire, 1993, Kandel *et al.*, 2000, Kandel and Hawkins, 1992). Damage to the hippocampus in particular causes severe learning deficits in both rats and monkeys. Several experiments in monkeys have also shown that the perirhinal and entorhinal cortices are involved in memory-dependent recognition tasks. In humans, there have been some studies of patients with lesions restricted to the MTL or where a resection was done in this area for the treatment of temporal lobe epilepsy. One of the most famous cases is patient H.M. (Milner, 1972, Penfield and Milner, 1958) who has severe anterograde amnesia.

It is therefore interesting to ponder whether the neuronal responses that we have observed are related to some memory related process rather than the actual perceptual sensation. First, we should further define what we mean by “memory related process.” This may depend on the specific experiment. Sensory information needs to be acquired, and processed in some way so that it can be stored for later retrieval. We can therefore at least theoretically distinguish three processes: acquisition, storage and retrieval. It is important to emphasize that these stages are not necessarily independent. During the

visual presentation of pictures, the perceptual sensation may be correlated with the neuronal activity in an earlier visual area (say inferotemporal cortex for example) and then this information is subsequently submitted to the MTL for memory acquisition and/or storage. This model does not imply that the MTL should be a permanent source for the information. Several pieces of evidence suggest that the MTL (at least the hippocampus) could be involved in the transient storage of information and during the process of memory consolidation. Old memories do not seem to be affected by damage to the MTL and therefore it is currently believed that neocortical areas constitute a more plausible site for permanent storage. According to this hypothesis, the activity in the MTL could be occurring at a stage shortly after the perceptual sensation has occurred. This is in contrast to the alternative interpretation whereby the perceptual sensation could arise as a direct consequence of the activity in the MTL.

During visual imagery, the selective activation of a neuron may indicate the specific retrieval of information rather than the visual percept being held in the mind's eye. Following this model, it would be interesting to understand whether the MTL is involved in the retrieval of information from recent memories only or whether it is also activated during the reverie of old information. The observation that patients like H.M. can still recognize people and events from his infancy would suggest that the hippocampus may not be necessary for the reprocessing of information that was acquired long ago. A drawback in this interpretation, however, is the enormous plasticity of the brain that could accommodate for the usage of other brain areas for tasks that would normally require the hippocampus. This could be particularly important in several cases where the patients are studied several months or years after the lesion or resective procedure.

This memory-processing hypothesis could also account for the results described in Chapters 5 and 6. During flash suppression, the perceptual information (the results of the competition between the two stimuli) could be conveyed to the MTL for storage in the same way as the normal presentation of a picture does. We are not aware of all the information around us (this is beautifully demonstrated in several experiments like the striking "change blindness" phenomenon, see O'Regan *et al.*, 1999). There are several information filters (attention being one of the most prominent ones) throughout the

different visual processing stages in the human brain. It makes sense that the only bits of information that should be sent to the memory processing centers are the contents of our percept after the filtering stages. Thus, during the binocular rivalry or flash suppression experiments, no information about the suppressed stimulus seems to be present in the MTL.

As we have discussed in Chapter 5, the synchronous activation that we have observed during SWS could be related to the prominent role that has been suggested for sleep in memory consolidation processes. Therefore, all the data that we have presented so far could be interpreted as revealing part of the mechanism for storing visual information or as a direct process underlying perception. The distinction between these two possibilities brings us back to the fundamental issue of causality that I will postpone once again to a later section.

7.3.2 Language and concepts

One of the most astounding capabilities of our brains is the possibility of creating and understanding language. Several animal species possess some form of communication skill. These communication processes can be extremely complex, rich in information and share several characteristics with the human language. However, none compare to the human language in the power that has derived from the use of language. To some extent, the cultural repertoire of technological advances, tools and knowledge that has been accumulated and thanks to which a new human being does not need to re-invent the wheel from scratch, depends on language. This has been called “extelligence” (Cohen and Stewart, 2001) and has been a major factor in the progress of humanity. While, as the word implies, extelligence is in the culture, outside the brain, the capacity to generate and understand language and concepts is within the exclusive domain of the central nervous system.

Another possible interpretation that has been put forward by several colleagues is that the activity (or at least some of the neuronal activity) that we have observed is indeed not derived from the visual perceptual state of the subject but from language processing. For example, the neuron that responded to “animals” illustrated in Chapter 4 might well be selective activated by the "concept" of an animal rather to the visual presentation of an

animal. The fact that it was activated selectively during visual presentation and imagery would not necessarily imply thus a common mechanism for vision and imagery but the fact that the concept is part of the intersection between these two processes. An area involved in representing the general idea of animals would be recruited upon visual observation of a picture of an animal, during visual imagery of an animal and also during other processes that involve understanding or using the concept of the word animal. Maybe the neuron would respond equally strongly to the presentation of a word suggesting the concept “animal” such as “dog” or “animal” itself¹⁰⁶. Maybe it would respond to a picture of a zoo even in the absence of animals. Maybe it would respond also to the sound made by animals such as barking or meowing. Similar questions can be posed for several of the other selective neuronal responses that we have observed. But in order to fully define this alternative hypothesis we need to understand what a “concept” is.

To what extent can we dissociate the concept from the image? If we present the word “dog,” or a barking sound, or dog food, or dog smell, how can we be certain that this is not giving rise to a visual image? Blind people, particularly, congenitally blind people, of course show a rich variety of sensations and imaginations in the absence of any visual images. But the significant rearrangements in the brain organization make it difficult to derive any clear conclusions about the brains of non-blind people. For example, V1 may be activated by touch sensations in blind people; this illustrates the astounding plastic capability of the brain but it does not imply that V1 does not process visual information in normal sighted people. Several philosophers have actually proposed that it is not possible to generate concepts that are not linked to images (Finger, 2000). Locke argued that there can be no conceptual processing devoid of sensation (Locke, 1995). There may be abstract ideas such as “liberty” or “education” that can go beyond a

¹⁰⁶ Several people have suggested to me that we should do this experiment. The experiment would involve presenting words instead of images for those units that showed visually selective responses and study whether the neurons also respond in a selective way to the words. We have not done this experiment yet (but we probably will in the future). The main reason has been time but the other one has been the difficulty in preventing the subjects from forming a visual image while processing the word. If we ask subjects to look at the word “dog”, how can we be certain that they are not actually forming a visual image of a dog? If we force subjects to do some task like counting the number of letters or inverting the order of the letters, then the lack of a response may be related to this task. There are many other possibilities of course and I am not going to present a full discussion of these other possible experiments. This briefly illustrates the difficulty in distinguishing between an abstract concept and a visual image.

specific image¹⁰⁷. A related question is whether we can actually generate a mental image of something that we have not seen before and that is not a combination of parts observed before. We can all imagine a specimen with the head of a man, the body of a lion and the legs of a horse (this kind of creatures abound in mythological literature). But can we imagine an alien that does not look like the combination of anything we have seen on earth? Philosophical arguments for and against this notion abound. It is usually quite remarkable how aliens in movies usually still keep several earth-like features.

Language processing has been ascribed to areas in the brain well outside the MTL (Broca, 1861, Damasio and Damasio, 1992, Pinker, 2000, Pinker, 1997). No specific language deficit, either in the generation or understanding of language in any form seems to have been observed in patients with damage to the MTL and evidence from functional imaging studies in the human brain do not seem to suggest the involvement of the MTL in any language process. However, maybe this alternative hypothesis should be formulated in a more abstract way related to the encoding of concepts. The model would therefore suggest that information in the human brain could be stored at a level beyond the visual representation, in a way that associates multimodal characteristics of a stimulus. The selective neuronal responses that we have observed would thus represent a level of abstraction beyond vision and neurons could be activated by other cues. It will be fascinating to explore these other possibilities.

7.4 Neural codes

Throughout this work we have assumed that the number of spikes fired by a neuron in a relative long window of several hundred ms constitutes an important variable to understand the information the unit is representing and communicating to other neurons. There has been considerable debate particularly in the last decade about the types of neuronal codes that are used in the brain to represent information (Shadlen and

¹⁰⁷ This does not mean that the word “liberty” does not give rise to visual images in our brains such as the statue of liberty and so on. But maybe there is a more abstract representation of our ideas beyond these images. For the moment, I will leave the argument regarding the existence of abstract concepts or lack thereof to the philosophers.

Newsome, 1994, Shadlen and Newsome, 1998, Rieke *et al.*, 1997, Bialek *et al.*, 1991, Softky, 1995, Koch, 1999, Kreiman *et al.*, 2000c, Krahe *et al.*, submitted, deCharms and Zador, 2000, Konishi, 1991)¹⁰⁸. The timing of spikes could contribute significantly to the information conveyed by a neuron about a stimulus. Furthermore, populations of neurons could combine their responses to make a meaningful representation beyond the firing of individual neurons. The synchronous activity of ensembles of neurons is one of the multiple ways of encoding information by groups of neurons that has been suggested to play an important role in the nervous system (Singer and Gray, 1995, Engel and Singer, 2001, Singer, 1999, Koch and Laurent, 1999, Laurent, 1996b). We have explored some of these other mechanisms of encoding information beyond the firing rates but space constraints preclude from expanding here on this issue here.

7.5 Response duration

There is one important aspect of a large fraction of the electrophysiological responses that have been recorded so far (by us and by monkey electrophysiologists) that has been largely ignored. Most of the neurons both in lower as well as in higher visual areas seem to show very transient responses to the presentation of visual stimuli. The

¹⁰⁸ These are just a few of the many possible models neurons could use to encode information. The nervous system is probably much more complex than what we currently consider. The fascinating world of possible encoding processes is largely unexplored. And the details will matter to achieve a true understanding of how the brain works. This does not mean that the right variable to look at is the firing rate of a single neuron. I don't think we still know what the relevant variables are. I am just expressing a bias, a hunch, that it will not be the average activity of a million neurons. Let us say that we are interested in understanding the musical preferences of people in the United States. We could simply ask every person to shout the name of his favorite singer/ band/ performer and then measure the most voted ones in each state. Then we may end up learning that Californians like, say, Madonna whereas Massachusetts prefers Mozart. But it is clear that this does not capture the whole story (even after re-counting). There are several different cities within California (and neighborhoods within each city and so on) where the preferences may be different. It may even be more complicated. Maybe some people prefer to listen to some kind of music while they are on the car, another type when they are studying and yet another when they go dancing. And it may be that some people listen to some kind of music when they are alone but to a completely different one when they are in groups. Several of the techniques that we use nowadays to study the brain follow this example. In the same way that we cannot determine the three-dimensional structure of a protein with the naked eye, we will need accurate techniques with a high level of spatial and temporal resolution to ultimately understand the details about the brain. Without nuclear magnetic resonance or X-ray crystallography, we could discriminate a globular protein from a non-globular one, we can even have accurate assays for the function of the protein and determine when its activity has been deteriorated, but we cannot ultimately understand the fundamental link between structure and function.

duration of the response is typically much shorter than the lapse of time the stimulus remains on the monitor and on the retina. This is in stark contrast with our perception. If we present a picture for several seconds, it is possible that the neuronal response be as short as 100 ms (or even less in some cases). Note that here I am referring to the instantaneous change in the firing rate of one unit as the “neuronal response”; I will come back to this point below. While the investigator claims that the activity of the neuron may be representing a complex visual stimulus and thus be close to perception (see for example the previous Chapters), there seems to be a discrepancy between the neuronal duration and the perceptual one. In our data, for example, the mean response durations were close to 500 ms while the stimulus was on for 1000 ms.

One possible interpretation of this and many other observations is that the activity of a single neuron by itself cannot fully underlie perception. This is born from many other results as well. For example, while the probability of errors of individual neurons in the classification tasks from the ROC analysis were quite remarkable, they were clearly far worse than the overall behavior of a subject. In other words, a single neuron could make a mistake in determining whether its preferred stimulus, let’s say a particular face was present or not in one out of four or five cases approximately whereas the subject would only make one mistake in 50 trials and probably not because he failed to recognize the face. Maybe the duration of the response from the average firing rate of a group of neurons is better correlated with the perceptual duration. Indeed, when we computed the normalized average firing rate (see Chapters 4 and 5), we observed that the response duration was longer than the average duration of the individual responses. The probability of error in the classification tasks by averaging from a group of neurons is also significantly lower than the values for individual units. Note that this does not mean that perception should be distributed across millions of neurons. On the contrary, it suggests that a sparse representation could very well be more than enough¹⁰⁹.

¹⁰⁹ The word sparse is widely used in Neuroscience in a more literary and poetic than quantitative way. How many neurons “sparse” really means constitutes an important question that needs to be addressed in a quantitative way. While I do not have the data to seriously characterize the answer to this question here, I would like to venture some numbers. For a number of reasons that I do not have space to expand on here, I think the number of neurons will be remarkably small. I think less than 100 neurons will be enough to represent very complex visual stimuli. I would even dare say that a remarkable degree of processing can be achieved by less than 10 neurons. This does not mean that the brain actually uses only 10-100 neurons to represent any complex stimulus. This would not be robust to neuronal death or small injuries. The claim is

It could very well turn out to be that there is much more information in the ensemble activity of small groups of neurons than the simple average described here. Indeed, the synchronous discharge of neurons has been proposed to play a crucial role in the encoding of information in the brain (Singer and Gray, 1995, Singer, 1999, Engel and Singer, 2001, Laurent, 1996a, Stopfer *et al.*, 1997). Perhaps the synchronous discharge of two or a small group of neurons shows a duration that is closer to perception than the responses of individual neurons. Experiments where both multiple units and local field potentials have been simultaneously recorded during prolonged stimulus presentation times of several seconds show that only the latter shows lack of adaptation and continues to correlate with the stimulus (Logothetis *et al.*, 2001).

Maybe there is no actual discrepancy between the perceptual and neuronal response duration. In other words, perhaps the neuron simply signals a change in the environment and the constant perception is the default for the brain unless a neuron signals another change. A neuron selective to a particular face, for example, could simply fire shortly after the face comes into perception and then the percept is not changed until a neuron (most likely another one) signals the disappearance of the stimulus or another neuron (or group of neurons) signals the appearance of a new stimulus. According to this, there should be no correlation between the duration of neuronal firing and the permanence of a percept.

There are several other possibilities. What is important is that one should keep in mind that if our purpose is to be able to explain our perceptions from the neuronal activity, we should be able to account for every aspect of perception, not only its contents but also its duration, its robustness and invariance.

7.6 H.M., cause and correlations

We finally must come to grips with the central distinction between correlation and causality. The observation that the neuronal activity in a given brain region, the MTL in

that I venture the idea that with 10-100 nearby neurons it would be possible to very accurately represent a particular stimulus so that the probability of error from the ROC analysis would be as negligible as the behavior is. Ok, enough of speculation.

our case, is correlated with a given percept or behavior, does not necessarily imply that these neurons are directly involved in causing the specific sensation or movement. Some other alternative interpretations of the data were discussed earlier in this Chapter. Two important types of studies that can provide valuable data to address this issue are the direct stimulation of the brain and lesions or examination of neurological patients.

One of the most famous neurological patients is the case of patient H.M. who had extensive damage in the hippocampus and some of the adjacent brain areas (Milner, 1972, Kandel *et al.*, 2000). H.M. shows severe anterograde amnesia, yet his perceptual abilities seem to remain completely intact¹¹⁰. Based on this, it could be argued that we recorded from brain areas that are not needed for perception. There is no doubt that the medial temporal lobe plays a fundamental role in declarative memory (Squire and Zola-Morgan, 1991, Zola-Morgan and Squire, 1993) but we argue that we cannot assign the *exclusive* role of memory to the MTL. Furthermore, perceptual and memory processes may well be part of a continuum. There are strong direct connections between inferior temporal cortex and the MTL (Suzuki, 1996, Saleem and Tanaka, 1996, Cheng *et al.*, 1997). Furthermore, electrical stimulation in the human temporal lobe elicits visual percepts and memory recall and can interfere with visual perception (Penfield and Perot, 1963, Penfield and Jasper, 1954). However, it should be noted that most of the perceptual effects were caused by stimulation in more anterior (neocortical) areas. Furthermore, it should be mentioned that the perceptual effects could be caused by feedback to lower visual areas. While neuronal activation does not imply a role in perception, changes in neuronal activity in the MTL structures during vision have been shown by single neuron recordings in humans and monkeys and also by fMRI studies in humans.

While there is strong evidence for a most prominent role of the MTL in memory processes (Zola-Morgan and Squire, 1993, Squire and Zola-Morgan, 1991, Heilman and Valenstein, 1993), the rarity of patients, the poor localization capacity of neurological studies, the confounding effects of plasticity and uncontrolled nature of the lesions make it difficult to make a blunt statement about the problem in humans. In several

¹¹⁰ It should be noted that in several neurological studies the behavioral tests were gathered several years after the lesion or resection and there could be ameliorating effect due to plasticity and some other brain region taking over tasks that would normally be carried out by the MTL. Also, while most of the MTL was removed from H.M., some parts were spared and it should be kept in mind that this MRI study was done several decades after the initial resection (Corkin *et al.*, 1997).

neurological studies of patients with perceptual deficits, the precise localization of the lesion is quite poor (Heilman and Valenstein, 1993). A CT was used to localize the lesion in several studies (Ross, 1980, Warrington and Mc. Carthy, 1987, Warrington and Shallice, 1984, Warrington and Mc Carthy, 1983, Hart and Gordon, 1992, Sheridan and Humphreys, 1993). While in most cases the areas of damage were localized to the temporal lobe, no explicit detail regarding medial versus inferior regions is mentioned. In some patients with category-specific recognition deficits, damage in the MTL (but not always restricted to this region) is evident (Silveri and Gainotti, 1988, Gainotti and Silveri, 1996). Damage to the amygdala can lead to impairments in the recognition of facial expressions and emotions but not personal identity (Adolphs *et al.*, 1994). Damasio and colleagues described patients with amnesic associative agnosia that show impairments in perceptual recognition of faces and have bilateral damage *only* in the entorhinal cortex, hippocampal formation and amygdala (Tranel *et al.*, 1990, Damasio, 1990). A particular type of topographical perceptual process required for navigation (Whiteley and Warrington, 1978; Landis *et al.*, 1986; Habib M and Sirigu, 1987; McCarthy *et al.*, 1996; Aguirre *et al.*, 1998; Aguirre and D'Esposito, 1999) or encoding of scenes (Ross, 1980, Habib and Sirigu, 1987) is impaired in patients with damage in the MTL. Finally, functional imaging evidence suggests that, within the MTL structures, the parahippocampal place area might play a prominent role in topographical perceptual processes (Epstein *et al.*, 1999, Epstein and Kanwisher, 1998, Maguire *et al.*, 1998). Ablation studies have been performed in monkeys and rats in the MTL. The entorhinal and perirhinal cortices seem to be required for visual object recognition (Meunier *et al.*, 1996, Murray, 1996, Buckley and Gaffan, 1998, Buckley and Gaffan, 1997). The exact anatomical and functional correspondence between monkeys and humans is not clear. For example, there are patients with extensive damage in areas corresponding to the inferior temporal cortex and no perceptual deficits.

We have argued in a previous section that the correlates of the contents of our percepts could be in some brain structure in the MTL or shortly before the MTL in the visual hierarchy. Given the strong similarities between the properties of perception and those of the activity in the MTL, my bias is that it is unlikely that the two can be very far apart (in number of synapses, not necessarily in real distance) in the brain. If the

perceptual correlates occur several synapses away from the MTL, it would be mysterious and interesting to understand why or what function is served by all the processing of the information in between and how this processing affects or not the contents of awareness or its storage.

7.7 Neuronal causal correlates of consciousness

One of the main motivations for me to start inquiring into the neuronal encoding of the world of perception has been the influential work of my advisor Christof Koch together with Francis Crick (Crick, 1994, Crick and Koch, 1990, Crick and Koch, 1998). They hypothesize the existence of a group of neurons with specific properties whose activity represents the neuronal correlates of consciousness or *ncc*.

As discussed in the previous section, the next question that we have to address is a far more difficult one, the search for the neuronal causal correlates of consciousness (*nccc*). The search for the *ncc* is insufficiently constrained and can therefore lead to multiple solutions. While the search for the *nccc* may turn out to be extremely difficult, it constitutes a fun challenge with fascinating implications. *Audentes fortuna iuvat*.

1 Appendix 1: Spike sorting

1.1 Introduction

Extracellular recordings monitor the electrical potential outside the neurons as the name indicates. This implies that the activity of any neuron surrounding the electrode will contribute to the signal. Those neurons that are far away will only have a small weight in the overall signal. Upon the other hand, if the electrode happens to be very close to an individual neuron, the recording is basically a single unit analysis since the contribution of any other unit would be negligibly small. This is the case in several animal electrophysiological experiments where the goal is to move and position the electrode as close as possible to one neuron to obtain a very large signal-to-noise ratio (*snr*) in that neuron without any influence from surrounding units. This approach does not extrapolate well to multiple units¹¹¹ since it is time consuming and difficult to attempt to isolate several neurons by moving an electrode. The approach with multiple electrodes that we and many other groups use is typically quite different (see Chapter 2 and Nicolelis *et al.*, 1999, Nicolelis *et al.*, 1997): a set of multiple separate electrodes (or multiple tetrode electrodes) are placed in the area or areas of interest. In some cases in animal recordings, the fine position of individual electrodes can still be moved to improve the quality of the recordings (McNaughton *et al.*, 1983, Wilson and McNaughton, 1993). In other cases, the electrode positions may remain more or less fixed. Even if some electrodes may register a very poor signal or no neuronal signal at all, the usage of many electrodes increases the probability that some of those will yield data of interest (a quantitative report of the probability of obtaining a neuron in different areas from our data was presented in

¹¹¹ Heroic attempts have been made to record from two and in some cases even three neurons at the same time by moving the electrodes to isolate individual cells (see for example Krahe *et al.*, *submitted*). This is quite remarkable and it has yielded very important data but it is quite unlikely that it will extrapolate to the analysis of tens of neurons.

Chapter 2). This approach typically yields "noisier" data because on average the electrode will be farther apart from the neuron than in single neuron recordings. But it offers the possibility of recording from multiple neurons from separate electrodes. Furthermore, in many cases, multiple neurons can also be isolated within the same electrode or tetrode. This requires the ability to differentiate and isolate the multiple units from the different waveforms that are recorded. The researcher does not know the number of neurons that is close to the electrode or the shape of the waveforms from each neuron being recorded. Efficient algorithms are required to address isolate the different neurons and assign each timestamp to a particular neuron or cluster. Several techniques have been proposed through the years to address this problem. These range in level of sophistication and effectiveness from the simple dual window discriminator used in several single neuron experiments to fully automatic algorithms such as the ones proposed by Lewicki (Lewicki, 1998, Lewicki, 1994), Sahani (Sahani, 1999) or Pouzat (personal communication). It should be noted that some investigators prefer not to deal with the problem at all and simply lump all the data obtained from a single channel into what they call "multi-unit activity" (MUA). While this is clearly simpler from an algorithmic point of view and requires fewer assumptions about the data (because it is not unusual to make strong assumptions in the process of spike sorting, see below), the interpretation of the results is somewhat less clear. To begin with, one would not be able to speak of the selectivity or response properties of individual units. Second, it is quite possible that waveforms corresponding to noise are also included in the MUA. Third, it is possible that some of the response properties of MUA are different from those of single unit activity (SUA). For example, it is conceivable that a weak selectivity at the single unit level could be lost in the average obtained in the MUA and vice versa. Finally, for the study of correlations between the responses of two (or more) neurons, the interpretations derived from MUA are particularly complicated.

Since every single result obtained from extracellular recordings ultimately depends on the assignment of timestamps to spikes from different units, we performed a basic comparison of different spike sorting procedures. The data presented here should be considered as preliminary since there are several questions in the comparison and quantitative evaluation of a spike sorting algorithm that have not been considered in

detail yet. In this Appendix, I will start by giving a brief overview of different methodologies and algorithms that are available for spike sorting. In the process, I will describe our own semi-automatic spike sorting algorithm, "Spiker" and its advantages and disadvantages with respect to the other methods. Then I will present preliminary observations on the quantitative comparison of different spike sorting procedures and several variables and parameters of relevance to the experimenter. Finally, I will describe the results of our analysis comparing the visual selectivity of a rather large number of units after performing the spike sorting with two different methods.

1.2 Brief review of available methods

1.2.1 Dual window discrimination

Dual window discrimination constitutes one of the oldest and simplest methods to separate different spikes. Yet, it is still widely used in electrophysiological recordings. The reason for this, as will be discussed later, is that if the *snr* is very large (for example in a single unit recording where each unit is basically isolated from the contributions of all surrounding units by moving the electrode), this method is fast, it can be performed on-line and yields excellent results. A window can be set around the peak of the waveform and only accept those events where the peak falls within that window. For a given waveform \mathbf{x} , let $p_v(\mathbf{x})$ represent the voltage at the peak and let $p_t(\mathbf{x})$ represent the time at which the peak occurs; then a waveform is accepted into the cluster if:

$$T_v^l < p_v(\mathbf{x}) < T_v^u \quad \text{and} \quad T_t^l < p_t(\mathbf{x}) < T_t^u$$

where T_v^l , T_v^u , T_t^l and T_t^u represent the boundaries of the window. These are manually set by the investigator.

In a slightly more sophisticated version of this method, it is possible to establish also a window of acceptance around the trough of the waveform. This can be generalized to establish subsequently more windows for each cluster. Indeed, some spike sorting algorithms use 4-8 windows of acceptance to define a cluster. This leads to the *template matching* algorithm to be discussed in the next Section. We implemented the dual-

window discrimination method in MATLAB for off-line data analysis in order to compare its results to those from other algorithms (see below). In our implementation, the user was allowed to select many clusters by successively creating a window for the peak and trough and repeating the procedure for each individual cluster. If a waveform was already selected for a given cluster (i.e., it fulfilled the above requirements for a cluster) it was not assigned to a new cluster.

1.2.2 Template-matching

The template-matching algorithm is a very simple yet powerful extrapolation of the approach that we have just discussed. Each data point is assumed to arise from one of a fixed set of templates plus the addition of noise. No attempt is made to characterize the different noise sources. Each waveform is assigned to its closest template. It is comparably fast, it does not require extensive computation and it is very easy to implement in hardware. For these reasons, it is commonly used by many real-time sorting software/hardware configurations such as the Alpha-Omega multi-unit recording system. Let the templates be defined as t_1, \dots, t_N where N is the total number of templates. Then a particular waveform x is classified to cluster i if:

$$(x-t_i)^T(x-t_i) < (x-t_j)^T(x-t_j) \quad \text{for every } j \neq i$$

The templates are typically selected by the experimenter at the beginning of the recording session for each cluster. Note that the experimenter has to select the number of templates N in addition to the shape of each template. As in the dual-window discrimination method, this algorithm is not very robust to slight changes in the shapes of the waveforms along the experiment. Adaptive templates have been suggested to solve this problem.

1.2.3 Cluster cutting (Datawave)

One of the most widely used methods for spike sorting is a manual procedure usually called “cluster cutting” which involves the visual discrimination of clusters in the data shown in two-dimensional plots of selected variables. This procedure has been

implemented in Datawave (Datawave, Denver, CO)¹¹². The spike sorting for all the data presented in the current thesis was performed using this method. A number of parameters are selected for each waveform. These parameters include¹¹³: (1) peak voltage, (2) the time at which the peak occurs, (3) trough voltage and the (4) time of occurrence of the trough, (5) the waveform width and (6-8) the first three principal components of the data¹¹⁴ (Figure A1-1a-b). These variables are displayed in 2D plots. In the Datawave implementation, the user can choose up to 8 variables (therefore generating up to $\binom{8}{2}=28$ different 2D plots). The data typically form different clusters in these projections and the user has to select the boundaries for each cluster. An example of a selected subset of all the possible 2D subplots for one recording session is shown in Figure A1-2. In Datawave's old implementation (the one used for all the data in the current thesis work), the boundaries were constrained to have rectangular shapes. A more modern implementation allows for elliptical shapes¹¹⁵. The experimenter has to select both the number of clusters as well as the boundaries for each cluster. The results for the example illustrated in the previous figure are shown in Figure A1-3 where each cluster is shown in a different color. The corresponding mean waveforms in each cluster are illustrated in Figure A1-1c.

1.2.4 Automatic algorithms

All the algorithms and heuristic procedures described so far require the constant intervention of the investigator to select the windows of acceptance, the templates or the boundaries for the clusters. In view of the repetitiveness, time cost and subjectivity of the above procedures, several investigators have developed automatic algorithms to sort spikes. Among them, some important ones¹¹⁶ are the K-means algorithm implemented by

¹¹² See for example (Wilson and McNaughton, 1993, Kay and Laurent, 1999, Fried *et al.*, 1997).

¹¹³ These are the most typical set of 8 variables that we utilized. Other variables can also be used including the voltage and time of a valley before the peak as well as the voltage at specific time values. The dual window discriminator can therefore be formally considered a sub-variant of this manual sorting procedure.

¹¹⁴ These typically account for approximately 90% of the variance.

¹¹⁵ Note that, in general, an N-dimensional gaussian with a full covariance matrix has an ellipsoidal shape. If the non-diagonal terms of the covariance matrix are negligible, then the shape is spherical (Bishop, 1995).

¹¹⁶ This does not attempt to be an exhaustive list, for a recent review see (Lewicki, 1998).

RC electronics, the latent variable model approach by Maneesh Sahani (Sahani, 1999), the Bayesian approach of Mike Lewicki (Lewicki, 1994), the “bisection and merge” method by Fee *et al.* (Fee *et al.*, 1996), the method developed by Pouzat (personal communication) and the one recently proposed by Buzsaki’s group (Harris *et al.*, 2000). I am not going to discuss the algorithms for these methods here; a good review has been written recently by Lewicki (Lewicki, 1998).

One of the many advantages of these methods is that they are typically faster than the manual cluster cutting procedure and they do not require the supervision of a human eye. Processing speed may turn out to be quite crucial for some particular applications. For example, in the situations where the data needs to be processed as it is being acquired (for example, for neural prosthetics applications), it is crucial to have fast access to the neuronal activity. Automatic methods offer the exciting possibility of accessing the timestamps on-line. Furthermore, these methods eliminate the subjectivity of the manual procedures in which the investigator has to decide about the number of clusters as well as their boundaries. If the data are clearly separated in clusters, this can be quite simple and the degree of subjectivity is very small, but if the clusters overlap significantly, then the problem becomes quite complicated.

Unfortunately, at least in some cases, several waveforms are misclassified in automatic sorting procedures and then arbitrary decisions have to be made to supervise the results and possibly make post-hoc changes in the parameters of the models. To some extent, this again requires the supervision of a human observer and somewhat subjective decisions. Several of these methods were particularly designed with the idea of tetrode recordings in mind. While the same algorithms can be directly applied to data from a single electrode, the performance of these (and any other algorithm) drops quite significantly. This prompted us to develop a new approach that is intermediate between the fully manual or fully automatic processes.

1.2.5 Spiker

The manual procedures require long processing times for the user and require subjective judgment about the separation of clusters. Of course, the problem of spike sorting can be formulated as a particular question of pattern recognition and it is well

known in general that expert human subjects can become quite good at recognizing templates and patterns (in most cases much better than a large fraction of the available algorithms). Therefore, it is unclear that the goal of an automatic spike sorting method should be to surpass the performance of human experts. In many cases, the decision resulting from pattern recognition and the costs of a poor classification could be quite crucial. This is the case in several pattern recognition problems for clinical applications. It is therefore not surprising that there has been yet little penetration of automatic pattern recognition algorithms¹¹⁷. But in several cases, we do not require a perfect classification to obtain accurate results from electrophysiology¹¹⁸. That is, the electrophysiological results may be robust to some slight misclassification of spikes. On the other hand, a misclassification of 50% of the spikes could be very dangerous for the subsequent data analysis. This is addressed quantitatively in Section 1.3.5 below.

We therefore implemented¹¹⁹ a semi-automatic spike sorting algorithm that we named “*Spiker*.” This was developed together with a SURF student, Angela Yu¹²⁰. *Spiker* is a program to sort data into different categories. It was designed with the specific aim of sorting data recorded from extracellular single electrode recordings to extract activity from single neurons in a neurophysiology experiment. Instead of fine-tuning the parameters of the models for each data set, *Spiker* performs a classification of the data, shows some relevant statistics (such as the firing rates, percentage of short interspike intervals, power spectra and so on) and then allows the user to re-sort a specific cluster or merge clusters if this is necessary. *Spiker* comes with a user-friendly graphical interface (see Figure A1-4), which displays the clustering results and allows the user to manually modify the classification criteria if desired. The program allows the user to choose

¹¹⁷ One example of this is the scoring of sleep stages (described briefly in Chapter 5). Here the goal is to recognize specific patterns such as different stages of sleep in the scalp electroencephalographic recordings. In spite of the availability of some algorithms to do this, most MDs trust their eyes better.

¹¹⁸ In most cases (see however (Henze *et al.*, 2000) and (Wehr *et al.*, 1999)), we do not actually know how many neurons we have or what the actual timestamps are. That is, we do not know what a perfect classification would be.

¹¹⁹ As the reader can observe, several of the spike sorting algorithms have been developed at Caltech. Interestingly, there has been hardly a single electrophysiology lab at Caltech that has not developed its own spike sorting procedure (Mike Lewicki in Mark Konishi’s lab, Maneesh Sahani in Richard Andersen’s lab, Christophe Pouzat in Giles Laurent’s lab). While Christof Koch’s is not an electrophysiology lab, we still felt that there had to be a spike sorting algorithm from this lab.

¹²⁰ One of the algorithms for the first automatic step was provided by Christophe Pouzat. Maneesh Sahani and John Pezaris were very nice to provide their *spikesort* code for comparison purposes.

between an expectation-maximization algorithm and a vector quantization algorithm for the initial automatic classification step (Bishop, 1995). When the program reaches a solution, the user is given the option of merging or further sub-dividing the clusters. These user-input options were designed based on the empirical observation that the fully automatic clustering algorithms (Datawave autosort, Maneesh Sahani's *spikesort* algorithm) sometimes tend to overcluster and subdivide a cluster into two when they are apparently similar to our eyes. This may be due sometimes to small changes of the amplitude or shape of the waveform during an experiment or due to a poor underestimation of the noise (most of the spike sorting algorithms that have been implemented so far assume stationary conditions and are not adaptive). Also, this allows the user to decide to re-cluster when a particular cluster does not seem to be divided correctly. The interspike interval distribution, power spectrum and percentage of spikes within the refractory period are computed for each cluster (these are parameters that we have used in the criteria used to define a cluster as a unit, see Chapter 2). *Spiker* is freely available at (http://www.klab.caltech.edu/~gabriel/academia/spike_sorting.html) and easy to install and use. Further details about the algorithms and usage of *Spiker* can be obtained from this web page as well.

1.3 Comparison of spike sorting methods

Given the large array of possibilities for spike sorting it is legitimate to wonder which one provides more accurate results and how they compare to each other. To investigators analyzing neurophysiological data, it is also important to understand to what extent the conclusions from an experiment will or will not depend on the choice of spike sorting procedure and parameters. We performed a preliminary comparison of the results obtained by different sorting algorithms. Different procedures could yield different number of units for the same data set. Even with the same number of neurons, it is possible that the mean waveforms for each cluster and the assignment of timestamps to different clusters could be different. It should be noted again that for our data set and in most practical situations we do not know the “correct answer” to the problem; that is, we

do not know the underlying number of neurons or the mean waveform source from each unit responsible for generating each timestamp.

We compared the performance of four different spike sorting procedures that encompass a wide range of algorithms and strategies. These were the dual window discriminator, Maneesh Sahani's *spikesort* algorithm (Sahani, 1999), the manual cluster cutting procedure as implemented in Datawave and our *Spiker* algorithm (http://www.klab.caltech.edu/~gabriel/academia/spike_sorting.html). These methods are briefly described in Section 1.2. We applied these four methods to extracellular data recorded from four different microwires in our patients, with at least 30 minutes of recording in each data set. The types of electrodes as well as the recording methods are described in Chapter 2. Each data set contained at least 20000 waveforms (range 20031-47209). Dual window discrimination, *Spiker* and Datawave were performed by G.K. while *spikesort* was run by Jason Davis.

1.3.1 Which clusters should we compare?

The first question that arises is which clusters to compare from each algorithm. It is not trivial to rigorously decide how to compare the different methods. To begin with, there is no "absolute truth" about the number of neurons or their properties in the recordings. This is perhaps one of the reasons why, in spite of the existence of numerous methods for spike sorting, nobody (to the best of my knowledge) has yet attempted to perform a rigorous empirical comparison of different sorting algorithms and strategies¹²¹.

The problem is very simple, let's say that we obtain 4 clusters with method A and we obtain 3 clusters with method B; which clusters from method A should be compared with which clusters from method B? It is clear that we cannot compare every cluster in A with every cluster in B. One way to work around the problem of not knowing the "true" number of neurons is to generate and analyze synthetic data (see below). Instead, here we have opted for visually comparing the waveforms and deciding which clusters to

¹²¹ There are however very interesting comparisons of extracellular recordings and the results of sorting their data with simultaneous intracellular recordings (Wehr *et al.*, 1999, Henze *et al.*, 2000, Harris *et al.*, 2000).

compare. This could be easily implemented into an algorithm that chooses to compare the waveforms whose Euclidian distance is small (relative to the noise estimate).

1.3.2 Number of clusters

A few issues should be mentioned before directly comparing the number of clusters obtained from the different algorithms. During the manual clustering, the user (GK) had some experience in deciding (based on the waveforms) which clusters seemed to be of neuronal origin and which ones were not. Therefore, no time was spent in attempting to further classify the outlier waveforms and we only discuss here the comparison of those clusters that represented neurons. This necessarily yields a smaller number of clusters. The dual window discriminator has a limited resolution compared to the other methods in terms of the number of clusters that can be obtained. For example, the Datawave manual clustering procedure allows the user to set the window limits for 8 different variables (the cluster boundaries are set by the study of 28 different pairwise comparison graphs; see Section 1.2.3) whereas the dual window discriminator only allows for setting window limits in two different variables (the voltage values at two specific points corresponding to a peak and trough).

The number of clusters representing neurons was very similar with the four clustering algorithms for the data from the four microwires. One example of the results is shown in Figure A1-5. We obtained two clusters with units using the dual window discrimination method, 3 with the Datawave cluster-cutting procedure, 4 with Maneesh's *spikesort*, and 3 with *Spiker*. Additionally, we obtained one extra cluster in both *spikesort* and *Spiker*. This cluster was considered to be noise and not a unit based on the power spectra and the excessive (>2%) number of interspike intervals that were less than 1.5 ms (there should be very few for a real neuron due to the absolute refractory period). These criteria are further discussed in Chapter 2. Two clusters from *spikesort* could not be visually distinguished from each other and therefore, they were merged for the plot in Figure A1-5.

There was a strong agreement among the different procedures in the proportion of the total number of waveforms in each cluster (see Table 1). In particular, upon comparing *Spiker* and Datawave, the maximum difference between any two clusters was

3%. This corresponded to a difference in the classification of 601 spikes in an experiment of approximately 46 minutes duration. This corresponds to a difference of 0.2 spikes/sec and represented less than 10% of the total number of spikes that the neuron fired in this interval. There was a relatively close agreement in the overall number of sources obtained with the different methods. The only exception is the dual-window discrimination method that consistently yielded a lower number of units (this suggests that the results from the dual-window discriminator at low *snr* should be better considered to be MUA). While at high *snr* ratios one could expect most procedures to converge on the number of units, in more ambiguous situations with low *snrs*, the results could be different. While the *snr* used in this study were average values compared to our overall sample (see Chapter 2), it would be interesting to perform a more extensive characterization of the concordance or lack thereof among these methods with different recording qualities.

1.3.3 Within user repeatability

Given that several of the procedures require subjective decisions, it is important to address how much variability there is in the results of spike sorting after repeated application of the same procedures. Several of the automatic or semi-automatic sorting procedures utilize random starting conditions and therefore it is also important to assess how robust the results are to different starting conditions and other parameters. In the case of subjective methods, one can distinguish between the across-user variability and the within-user variability. Here we report only on the latter. For the data from the four microwires analyzed in the previous Section, we repeated the sorting procedure ten times using the *Spiker* algorithm with an interval of at least five days between repetitions. An example of the mean waveform obtained from one of the microwires in these ten successive runs of the sorting procedure is illustrated in Figure A1-6. There is quite a strong agreement between the ten separate runs; in all ten cases we obtained three clusters corresponding to units (there was some variation in the number of clusters assigned to noise sources but that is less relevant and is not shown here). The mean waveforms of the clusters are quite reproducible and there is no overlap among them. The five day interval was imposed in order to attempt to alleviate the confound factor of memory; That is, accurate and repeatable results could be obtained in successive runs just because the

experimenter remembers the number of clusters and their shape. Presumably, this would be eliminated after randomly switching the order of the four microwires and the intervening time period (during which the researcher also worked on several other waveforms). A related question that we have not investigated is the across-user repeatability. Let us say that the same data is given to ten independent researchers that use the same algorithm. How much variability would there be in the results?

1.3.4 Comparison with models

The lack of knowledge about the underlying solution to the problem can be avoided by using the algorithms on model data. It should be possible, for example, to generate a given number of sources (based for example on theoretical simulations such as the one by Gary Holt (Holt and Koch, 1999) or on empirical data) and then add some gaussian noise (see however Fee *et al.*, 1996) to the sources. The question would then be to assess which clustering algorithm allows us to recover the original designed neuronal sources. This approach has the advantage of allowing the investigator to systematically vary and thus quantitatively study several parameters that can be important for the spike sorting process. Among them is the sampling frequency, the noise level and in particular the signal-to-noise ratio and the number of units in the recording. This has the disadvantage of assuming a specific form for the noise and waveform shapes. In addition to this, the situation is artificial compared to the real world scenario where there are other sources of noise such as movement artifacts, coincident firing (although this could be easily added to the models) and other real life issues. This approach was pursued in collaboration with Jason Davis and the preliminary results were summarized in his Masters Thesis work (Davis, 2000).

1.3.5 Comparison of visual selectivity between *Spiker* and Datawave spikes

Ultimately, the main question that investigators are likely to be most concerned about is how the spike sorting algorithm could influence the results of the neurophysiological experiments. If a neuron is determined to be visually selective after

clustering the data with a given procedure, will it still be selective after applying a different sorting algorithm. While it would be interesting to be able to characterize with 100% accuracy the source of each recorded waveform, it is possible that a small margin of error would not significantly alter the neurophysiological experiments. On the other hand, a misclassification of a large fraction of the spikes could lead to random, irreproducible and arbitrary results. What constitutes an acceptable degree of error is thus an important question that requires exploration. Note that this is not exactly the same level of analysis as the previous one. Of course, if two algorithms yield exactly the same number of clusters and assign each timestamp to exactly the same cluster for every single waveform, then the results of any posterior analysis will be trivially exactly identical. It is possible, however, that there could be a difference in the classification of a number of waveforms. The question we are posing here is to what extent a differential classification of $x\%$ of the waveforms would turn out to significantly affect the results or not.

To address this question, we studied a subset of data recorded from 237 microwires in the MTL from the data set analyzed in Chapter 3. The data were clustered using the Datawave cluster-cutting procedure and then independently re-clustered using *Spiker*. The latter was performed several months after the initial clustering of the data with Datawave by the same investigator (myself). I did not remember the number of clusters obtained in each of the 237 microwires or their shape; therefore, it is very unlikely that memory could play a role in the results to be reported here.

We obtained a total of 10% more units with *Spiker*. The number of units per microwire was 1.9 and 2.1 in Datawave and *Spiker* respectively (see Table A1-2). For each unit, we computed the firing rates, coefficients of variation CV and CV₂ as well as the percentage of spikes within the refractory period. There was no significant difference in any of these variables between the two sorting procedures (two-tailed t test, $p > 0.1$). The proportion of selective neurons was almost identical with the two procedures (14 and 13.7% respectively). Not all the neurons that were selective in one method were selective with the other one, though. There was, however, an overlap of almost 80% in the group of selective units after sorting the data with one or the other procedure. Of the units that were selective according to one procedure but not with the other, more than 90% showed only a weak selectivity with a p value > 0.01 . The range of variability across the two

procedures is only slightly larger than the one obtained upon repeating the same sorting procedure by the same subject. Thus, these results suggest that the results that we have presented in Chapter 3 regarding the visual selectivity of the neurons are not due to the specific spike sorting procedure that was used or the choice of the number of clusters and their boundaries.

1.4 Summary and further questions

A detailed understanding of the mechanisms and codes that the nervous system uses to represent information is likely to require the study of data from more than one single unit. It seems that neuroelectrophysiologists are finally¹²² moving in the direction of multiple unit recordings. Intracellular recordings can provide very interesting and useful data but cannot be extrapolated to the analysis of hundreds or even tens of neurons. In order to be able to monitor the activity of several neurons, extracellular recordings seem to be essential. Furthermore, it is not clear at this point that in those cases where the goal is to record from several tens of neurons, it would be practical to attempt to move each individual electrode to isolate individual neurons. Therefore, extracellular recordings combined with efficient spike sorting algorithms become essential.

There is still a large number of questions that should be addressed in order to complete a detailed quantitative comparison of different spike sorting methods. In particular, two important questions (and several other comparatively minor ones) have not been addressed here. One is the question of bursting. Bursting may constitute a fundamental aspect of neuronal firing used to convey information in a more secure and efficient way (see for example Metzner *et al.*, 1998, Steriade, 2001, Lisman, 1997, Reinagel *et al.*, 1999, Koch, 1999). Bursting is one of the most difficult problems for spike sorting because it is hard to distinguish between two units that fire within a short

¹²² Investigators have started performing dual unit recordings several decades ago (see for example (Palm *et al.*, 1988, Perkel *et al.*, 1967). However, a large fraction of the electrophysiological studies performed so far with notable exceptions has only looked at the firing of individual neurons. Anyhow, more and more investigators are starting to implant several electrodes or even several tetrodes to have access to the activity of several tens of neurons.

time period and a single neuron that fires in bursts. Heuristic approaches have been taken in this direction by using the refractory period and the interspike intervals within a burst (Henze *et al.*, 2000), Pouzat, personal communication). The detailed comparisons of extracellular recordings with simultaneous intracellular recordings (Henze *et al.*, 2000, Harris *et al.*, 2000, Wehr *et al.*, 1999) seem to suggest that at least some of these algorithms can efficiently predict the occurrence of bursts but the problem still deserved further analysis and study.

The other somehow related problem is the distinction of two units when some of their spikes occur almost simultaneously due to synchronous firing. A systematic approach to address this problem has been taken by Sahani (Sahani, 1999) but not in several of the other algorithms. This is also quite important because synchronous firing has been suggested to play a fundamental role in the encoding and binding of visual information (see for example Singer, 1999, Singer and Gray, 1995, Stopfer *et al.*, 1997). At least partly because of this uncertainty in the classification of multiple neurons that can fire synchronously or in bursts, most investigators do not report about the correlated firing of multiple neurons recorded from the same microwire. However, this kind of data is potentially very interesting because it allows studying the coincident firing of adjacent neurons whereas two neurons recorded from separate electrodes are usually much farther apart.

It is important to know if and how the results of experiments depend on the usage of different algorithms and to what extent fully automatic algorithms yield solutions that are as good as or even better than the ones obtained by expert observers. The repeatability of the results upon re-clustering by the same investigators and the across-experimenter variability need to be studied in more detail. Finally, the crucial issue is to what extent the final results and neuroscience questions depend or not on this step.

Our preliminary results sketched in the current appendix suggest that there is a strong degree of robustness of the results to most of these different parameters. Different clustering algorithms (with the exception of dual window discrimination for noisy data) seem to yield quite similar numbers of clusters and waveforms (Figure A1-5) and the results seem to be reproducible at least for the same investigator (Figure A1-6). Finally,

the extensive comparison of the results in our studies of visual selectivity in a large sample with two different methods has yielded quite similar results.

1.5 Figure legends

Figure A1-1: Manual spike waveform separation (Datawave)

(a) Each waveform can be characterized by several parameters, including the peak voltage and time, valley voltage and time, width, height, slopes, etc. (b) The projection of the data onto the principal components of the waveforms can also be used to characterize the waveforms. Here we show the first three principal components (in red, green and blue in that order of eigenvalues) from real data in one particular experiment. In most experiments the first three principal components accounted for more than 90% of the variability. (c) End result of sources for all clusters separated from this data set.

Figure A1-2: Two-dimensional projection plots

Each pair of variables characterizing the waveforms among the ones shown in the previous figure is plotted in a two-dimensional graph. Each point represents a different waveform. The variables that are plotted in this example are the scores of the first three principal components (illustrated in Figure A1-1b), the peak voltage (*peak*), the time at which the minimum voltage is achieved (*min_t*) and the minimum voltage (*valley*). Clusters are evident in the data. In the manual spike sorting procedure, the user selects the boundaries for each cluster in each 2D plot. This is a subset of all the 2D plots used in each experiment to sort the waveforms; only 15 2D plots are shown here; up to 28 pairs of variables were used to cluster all the data discussed in this Thesis (it should be reminded that there are only 8 independent variables, though).

Figure A1-3: Two-dimensional plots after separating the clusters

The same two-dimensional plots illustrated in the previous figure are shown here after the manual selection of the cluster boundaries. Each of the clusters is shown in a different color. The mean waveform for each cluster was shown in Figure A1-1c.

Figure A1-4: Spiker menus

(a) Main graphic-user-interface of the *Spiker* semi-automatic spike sorting algorithm. The only parameters that are used are standard values like the sampling frequency, position of the peak and threshold. These can be filled in only once and do not need to be modified from cluster to cluster or from experiment to experiment. Details about *Spiker* can be found in http://www.klab.caltech.edu/~gabriel/academia/spike_sorting.html. (b) After the first automatic classification step, the results are displayed (not shown here) and the user has the option of revising the results of the clustering procedure by merging clusters, reiterating the algorithm to subdivide a specific cluster, eliminating the outliers in a cluster, eliminating a cluster and computing the interspike interval distribution and other parameters of the waveforms.

Figure A1-5: Example of waveforms obtained from different sorting procedures

Comparison of spike sorting results from four different algorithms (*spikesort*, black; dual window discrimination, red; Datawave cluster cutting, blue; *Spiker*, green) for one of the microwires. These spike sorting procedures are discussed in the text (see Section 1.2). Here we show the mean waveform for each cluster (the standard deviations are large enough to hide all the small differences among the different methods; given the large number of waveforms ($n=20031$ in this case), the standard errors are too small to be visible). Three units were observed with Datawave and *Spiker*. Using *spikesort*, we obtained four clusters, two of which were almost identical and were merged here as single cluster 1. Using the dual window discriminator, we only could discriminate two clusters.

Figure A1-6: Same-user repeatability

Repeatability of the spike sorting procedure for a single user. Here we show the mean waveforms obtained for three different clusters in the same microwire (in the same

experiment but a separate microwire than the one illustrated in the previous figure). The spike sorting algorithm used here was *Spiker*. The sorting procedure was repeated 10 times with an interval of at least five days in between by the same experimenter (G.K.). The three different clusters are shown in different colors.

Figure A1-7: Example of PSTHs obtained with two different spike sorting methods

Raster plots and post-stimulus time histograms (PSTH) for a selective unit after clustering the data with the Datawave cluster-cutting procedure (**a**) or the *Spiker* algorithm (**b**). These PSTHs show the response of a unit to visual presentation of stimuli from different categories. The unit showed a selective response to animals; the activity of this unit is discussed in more detail in Chapter 3.

1.6 Tables

Table A1-1: Proportion of spikes in each cluster

Cluster	Algorithm			
	Dual window discr.	Cluster cutting (Datawave)	<i>Spikesort</i> (Sahani)	<i>Spiker</i>
"1"	0.31	0.19	0.20	0.20
"2"	0.38	0.36	0.45	0.39
"3"	-	0.14	0.18	0.16
"rest"	0.31	0.31	0.17	0.25

Proportion of spikes in each cluster according to the four different algorithms (see the text for the description of each algorithm). The data correspond to the recordings from one microwire where 3 units were found according to the criteria specified in Chapter 2. There were 2 clusters using dual window discrimination, 3 clusters using Datawave, 4 clusters using *Spiker* (one corresponding to noise according to the power spectra and refractory period constraints) and 5 clusters in *spikesort* (one corresponding to noise according to the power spectra criterion and two being indistinguishable to our eye and lumped together here as cluster "1"). The waveforms for these clusters are shown in Figure A1-5. Here "rest" indicates all the waveforms that were not included in clusters "1", "2" or "3."

Table A1-2: Comparison of Datawave to Spiker

	Datawave cluster-cutting	<i>Spiker</i>
Microwires	237	
Units	450 (1.9)	496 (2.1)
firing rate (spikes/s)	3.6±5.7	3.3±5.1
Visually selective	63 (14%)	68 (13.7%)
Overlap	51	

Direct comparison of the results of spike sorting as well as visual selectivity of a set of data from 237 microwires between the Datawave cluster-cutting procedure and *Spiker*.

2 Appendix 2. Flash suppression: competition between eyes or patterns?

2.1 Introduction

If a stimulus is monocularly presented for a time period of approximately 1000 ms and a second stimulus is flashed to the contralateral eye while keeping the original stimulus in the original eye, the new stimulus suppresses the ipsilateral one almost completely. This phenomenon has been called flash suppression and was first described by Wolfe (Wolfe, 1984; see also Chapter 6). Flash suppression stemmed out of the explorations of a related phenomenon called binocular rivalry that refers to the perceptual alternations that occur when two dissimilar stimuli are presented to corresponding areas of the two retinae (Levelt, 1968, von Helmholtz, 1866).

While the stimulus presented to one of the eyes in binocular rivalry is perceptually suppressed, it is more difficult for information presented to that eye to be detected. For example, the detection threshold for stimuli presented to the suppressed eye during binocular rivalry is increased from 0.3 to 0.5 log units (several excellent reviews have been written regarding the properties of binocular rivalry; see for example Blake, 1989, Blake, In press, Logothetis, 1998, Levelt, 1968). The course of rivalry (the ongoing period of suppression) may remain unaffected by changes in the spatial configuration of either the dominant or suppressed stimulus. A stimulus that is presented to only one eye is a dominant stimulus; the dark field in the closed eye is suppressed (e.g., the world does not disappear nor does it appear darker upon closing one eye). Actually, thresholds are elevated in an eye that is seeing a blank field (Fox and Check, 1966). Following this evidence, it could be expected that if we show a stimulus on the left eye only, the right eye will be suppressed. Then, if we briefly flash two stimuli, one on each eye, we could expect the stimulus shown to the left eye to win the competition given the ongoing

suppression of the contralateral eye. This effect should disappear some time interval after showing the original monocular stimulus. However, the results are exactly the opposite; it is the flash presented to the contralateral eye that is perceived most clearly (Wolfe 1984).

This phenomenon is not a form of forward masking, fatigue or retinal adaptation. The parameters are just wrong for any of these possible mechanisms. A brief flash of a horizontal grating to the left eye simultaneously presented with a vertical grating to the right eye leads to a mixed percept in which the two gratings are fused or piecemeal fused. Thus, the previous period of monocular presentation is required for the flash suppression effect to take place. The monocular presentation length can be quite variable, though (see (Wolfe, 1984) and Section A2.3). The effect is not due to the occurrence of a short intervening blank period between the monocular presentation and the flash. It can occur in the absence of any intervening period; this suggests that the blank period is not reversing the dominance.

There are two extreme alternative explanations for the phenomenon. In one case, the ipsilateral eye or cortical neurons responding only to the ipsilateral input are adapted and therefore the contralateral eye or cortical neurons receiving contralateral input win the competition and suppress the ipsilateral stimulus. Alternatively, it could be that the competition actually takes place between neurons selective to the pattern or stimulus itself and thus reflects the neuronal interactions in regions that respond to what is being shown to each eye regardless of which eye the information comes through. These two possibilities are not necessarily mutually exclusive or comprehensive. One could imagine a scenario where there is a competition at different levels of the visual hierarchy and the final result of the competition depends on some weighted average of the preliminary conclusions from deliberations in the intermediate stages. Whether one explanation or the other better accounts for the data could also depend on the characteristics of the stimuli. The discussion between these two alternatives is reminiscent of the Helmholtz-Herring debate about the high-level versus low-level origin of the alternations that occur during binocular rivalry (see Chapter 6). Recent psychophysical data have suggested that binocular rivalry can be best accounted by a competition between patterns rather than between eyes (Logothetis *et al.*, 1996, Kovacs *et al.*, 1996, Andrews and Purves, 1997); see however (Lee and Blake, 1999.)

There is a physiological correlate to each of these two possibilities. The former suggests that the competition could occur between monocular neurons in early cortical visual areas (or even before cortex). The latter, on the contrary proposes a competition at a much higher level in the visual pathway, among neurons that are selective to the whole stimulus pattern that was presented. As we have discussed in Chapter 6, the physiological activity in monkeys during binocular rivalry suggests that a larger proportion of neurons is correlated with the percept in higher visual areas than in lower visual areas (Logothetis and Schall, 1989, Leopold and Logothetis, 1999, Sheinberg and Logothetis, 1997, Leopold and Logothetis, 1996, Logothetis, 1998). Functional imaging evidence, however, suggests that activity in early visual areas could also be correlated with the alternating percepts (Polonsky *et al.*, 2000, Tong and Engel, 2001).

We have discussed the physiological responses of neurons in the human MTL during binocular rivalry and flash suppression in Chapter 6. Here I would like to briefly describe some of the psychophysical investigations that we have carried out to attempt to better understand the phenomenon. We have performed a series of psychophysical tests where the flash suppression experiment was modified by swapping the stimuli between the eyes or presenting images composed of patches presented to both eyes. These tests were aimed at trying to quantify the relative contribution of eye versus pattern competition during flash suppression and thus constrain physiological models. A preliminary account of this work was presented previously (Kreiman and Koch, 1999). All the work described in this appendix was carried out with the enthusiastic help of a summer research student from Caltech, Nick Knouf, and a Polytechnic high school student, Shaun Lee. Michael Herzog provided valuable suggestions and insights.

2.2 Experimental methods

We performed two different types of experiments: flash suppression with patchy images and swapped flash suppression. Since the results of the experiment may depend on the actual characteristics of the stimuli (see for example Bonneh and Karni, 1999), for each type of experiment we used two distinct sets of stimuli: sinusoidal gratings and

complex images. Most of the experiments with complex images were done using a black and white photograph of a girl's face and a truck (see Figure A2-2). We also used different complex images in some other experiments with similar results to the ones reported here. The images subtended either 3 or 4 degrees of visual angle. The gratings were either horizontal or vertical; the spatial frequency ranged from 1 to 3.33 cycles per degree (cpd) in steps of 0.33 cpd and the contrast was varied between 20 and 100% in steps of 20%. The examples in the following figures show one particular configuration but all the trials were randomized and counterbalanced so that in each trial, the initial stimulation could go either to the right or left eye and it could be either a horizontal or vertical grating or any of the complex images. The monocular presentation occurred during a period of time that we will refer to as t_1 . The value of t_1 was varied between 500 and 2000 ms in steps of 500 ms unless otherwise stated. The flash was presented during a period that we will refer to as t_3 . The value of t_3 was 50, 100, 200 or 500 ms unless otherwise stated. A very long flash period leads to binocular rivalry alternations. In some cases, a blank period referred to as t_2 was interspersed between the monocular presentation and the flash. The value of t_2 was 0, 20, 50, 100 or 200 ms. If not stated its default value was 0 ms (indicating a continuity between the monocular presentation and the flash).

2.2.1 Flash suppression with patchy images

Most of the experiments on binocular rivalry have been performed by showing a whole image to one eye and a different whole image to the other eye. Diaz-Caneja showed that perceptual alternations during binocular rivalry could also be elicited if the stimuli were split and half was shown to one eye while the other half was shown to the other eye so that the two halves complemented to form a Gestalt figure (Diaz-Caneja, 1928). Binocular rivalry between stimuli that are composed of patches presented to the two eyes cannot easily be accounted by a complete suppression of one or the other eye during the alternations. Recently, Kovacs *et al.* confirmed and quantified these findings suggesting that the competition in binocular rivalry occurs between complex patterns and not independently within image patches (Kovacs *et al.*, 1996).

We extended this idea to the case of flash suppression. A scheme of the experiment is illustrated in Figure A2-1. A grating was uniformly split into n patches ($n=1,2$ or 3 where $n=1$ corresponds to the traditional whole image case). The patches were shown to the right and left eye without overlap so that a whole grating could be formed only by combining the information from the two eyes. A grating of a single orientation was shown during a period t_1 ($t_1=500, 1000, 1500$ or 2000 ms). There were four controls or catch trials to assess the accuracy of the perceptual reports from the subjects. In two of the controls, the same grating that had been shown monocularly was exclusively flashed during period t_3 ($t_3=50, 100, 200$ or 500 ms). These were referred to as “control 1” (horizontal grating) and “control 2” (vertical grating, see for example Figure A2-1a). In the other two controls, a different grating was shown during the flash and during t_1 . These were referred to as “control 3” (horizontal, then vertical) and “control 4” (vertical, then horizontal). During the two flash suppression conditions, the flash included both a horizontal and a vertical grating (Figure A2-1c-d). These were referred to as “FS 1” (horizontal grating first) or “FS 2” (vertical grating first).

If the flash suppression effect were due to a complete suppression of one eye during the flash period, then it should not work with patchy images shown to both eyes. It could be argued, however, that the suppression does occur completely between one eye and the other eye (or monocular neurons receiving input from one eye and the other) but within specific patches. We therefore also performed a second experiment that could not easily be explained by a local competition between patches in each eye.

2.2.2 Swapped flash suppression

Logothetis *et al.* presented evidence that the two images can be repeatedly swapped between the eyes at a relatively high frequency during binocular rivalry without affecting the perceptual alternations (Logothetis *et al.*, 1996). This observation has recently been contended by Blake’s group (Lee and Blake, 1999). We used a similar experimental strategy to evaluate the relative contribution of competition between eyes and patterns during flash suppression. The experimental design is illustrated in Figure A2-2. The novelty in this experiment is that during the swapped flash suppression conditions, the monocular stimulus was presented to the contralateral eye during the flash

and the new stimulus was flashed to the ipsilateral eye (Figure A2-2e-f). The same idea was used for the complex images (Figure A2-2h). This experiment also included several control or catch trials where only one grating was presented during the flash; these were similar to the ones described in the previous Section (see also Figure A2-2 and Figure A2-6).

Let us examine the predictions for this experiment from the two extreme models described in the Introduction. The eye competition model or pattern competition model yield the same results for the classic flash suppression experiment (where the monocular stimulus matches the picture shown to the ipsilateral eye). This is because the *new* stimulus is also shown to the *new* eye. Therefore both predict that the *new* stimulus in the *new* eye will win (although for different reasons). During the swapped flash suppression, however, the predictions from the two models are opposite. If the monocular stimulus shown to the right eye leads to some degree of adaptation in the right eye, during the flash period the ipsilateral eye will be suppressed, and therefore the subject should perceive whatever is shown to the contralateral eye and be blind to the information in the ipsilateral eye (Figure A2-3). Therefore, during the swapped flash suppression condition in which a horizontal grating is shown monocularly and then flashed to the contralateral eye (Figure A2-2e) while the vertical grating is flashed to the ipsilateral eye, the subject should perceive the horizontal grating during the flash (the old stimulus in the new eye is perceived). According to the competition between patterns model, there is adaptation or suppression of the old pattern regardless of which eye it was shown to. Therefore, in the above condition, the subject should perceive the new vertical grating in spite of the fact that it is shown to the old ipsilateral eye (the new stimulus in the old eye is perceived).

It could be argued that there is some impression of movement during the transition from the monocular image to the flashed stimuli. In other words, in the classic flash suppression experiment, there is basically no change and therefore no movement in one eye while there is a large sudden change from a blank image to a grating in the other eye. This change or movement could be responsible for a reversal of the dominance. During the swapped flash suppression, the argument is somewhat farfetched, but it is conceivable that the change from a horizontal to a vertical grating produces a smaller impression of change or movement than the change from a blank image to a grating. We

designed two new experiments to address this concern. In one, a short blank interval t_2 ($t_2=0, 20, 50, 100$ or 200 ms) was interspersed between the monocular image and the flash. Wolfe had already shown that a short blank interval does not affect the results of the classic flash suppression experiment. During this experiment, the images in both eyes change from a blank to a grating and therefore it cannot be claimed that the change or motion effect is responsible for the flash suppression phenomenon. In the second experiment to address this concern, images were flickered in the two eyes (this is similar to the technique used in Logothetis *et al.*, 1996) and Lee and Blake, 1999). This is illustrated in Figure A2-4. The flicker rate in most of the experiments was 50 Hz but we also tested flicker rates of 0, 10 and 18 Hz.

2.2.3 Behavioral responses

The results of the experiment could depend on the way the subject's perception is assessed. We therefore used several different methods to quantify what the subject had observed. Wolfe had requested subjects to respond by giving a rating from 1 to 10 where the extremes corresponded to a complete vertical or horizontal grating. Furthermore, he confirmed that the results were very similar to those obtained using a two-alternative forced choice (2-AFC) paradigm.

We asked subjects to indicate what they had seen during the flash. In one of the methods that we utilized to assess their responses we used a 2-AFC paradigm where subjects reported either image 1 or image 2 (where these corresponded to the horizontal and vertical gratings or either of the two complex images; see Section 2.2). If they saw neither or any degree of mixture between the two, they were still forced to guess or estimate whether their percept was closer to image 1 or image 2. In another set of experiments we also used a 3-AFC that included the possibility of a mixture (comparable to piecemeal rivalry). In yet another set of experiments, subjects were asked to rate the percept in a 5-point scale ranging from image 1 to image 2 in the two extremes. Before starting the experiments we showed the subjects the stimuli as well as different examples of mixtures.

All of these methods require some subjective evaluation from the subject. We therefore devised a more objective evaluation of the responses by requesting them to

perform an orientation discrimination task. Taking 0 degrees as vertical and 90 degrees as horizontal, the initial monocular stimulus was the same as in the previous experiments: for example, in one specific case, a vertical grating was presented monocularly while a blank field was shown to the contralateral eye. During the flash, however, the vertical grating was shown slightly rotated by a small angle θ ($0.5^\circ < |\theta| < 10^\circ$), while the horizontal grating was rotated by the same angle. These rotations occurred randomly either to the left or to the right for the vertical grating and either upwards or downwards for the horizontal one. In addition to asking the subjects to report their percept as described in the previous paragraph, they were requested to indicate the orientation of both the vertical and horizontal gratings. This gave us an objective measure of the extent to which subjects could detect orientation changes in the suppressed image. This allowed us to more accurately quantify the degree of suppression of the two pictures during the flash.

2.2.4 Subjects and stimulus presentation

A total of 15 subjects participated in the different experiments described in the current Appendix (but not every subject participated in every condition of every test). Three of these 15 were aware of the goals and methodology of the tests¹²³ while the remaining 12 were naïve. Stimuli were presented on a Power Macintosh 7100 attached to a Sony Trinitron 17" monitor. Images were separately presented to the right and left eyes through a custom built haploscope. Image presentation routines were written in MATLAB 5.2 (The MathWorks, Natick, MA) using the Psychophysics Toolbox (Brainard, 1997, Pelli, 1997). All subjects had normal or corrected to normal vision. No feedback was given to the subjects about their responses except in the first practice session. The data from the first practice session was not averaged in the data analysis.

An orientation discrimination threshold was determined for each subject prior to the experiments. This was done for the objective assessment of suppression described in the last Section. The requirement to determine a threshold for each individual stems from the observation that a blanket threshold may be too low for one subject or too high for

¹²³ These were Gabriel Kreiman, Nicholas Knouf and Shaun Lee.

another one, leading to spurious results in the flash suppression test. A threshold that is too low would lead to chance performance regardless of whether the subjects perceive one or the other grating as being suppressed. A threshold that is too high would lead to excellent orientation discrimination even in those cases where one of the gratings is largely suppressed. Three sets of 30 stimulus presentations were performed for both vertical horizontal gratings. Using the QUEST algorithm (Watson and Pelli, 1983), we calculated an estimate of the threshold for each subject in real-time. This threshold value was then used throughout the remainder of the orientation discrimination trials in the flash suppression experiments.

2.3 Results

There were a large number of parameters that we have tested as discussed in the previous Sections. I will only give a brief summary of the results here. Unless otherwise indicated, the reported values correspond to the following default parameters: $t_1=1000$ ms, $t_2=0$ ms, $t_3=100$ ms, flicker rate = 0 Hz, gratings of 2 cpd and 100% contrast, girl and truck as complex images, 3-AFC for the collection of responses.

2.3.1 Flash suppression with patchy images

The results for the flash suppression experiment with patchy images (illustrated in Figure A2-1) are shown in Figure A2-5. Here I am only showing the results of the 3-AFC response paradigm where the subjects had to indicate whether the flashed image was a horizontal grating, a vertical grating or a mixture (Figure A2-5a), or a face, a truck or a mixture (Figure A2-5b). The four controls were almost at the 100% performance level for the only picture that was presented during the flash. Subjects could perceive the flash suppression effect with the patchy images as indicated by the two flash suppression conditions. There was a significant number of trials where subjects reported perceiving a mixture, with an average of 20%, indicating that the suppression with the patchy images was not as strong as in the classical flash suppression trials. However, in almost no case

did subjects report perceiving the original image shown during t_1 exclusively during the flash. There was no clear difference between the trials with gratings and with complex images. The observation that the initial stimulus was perceptually suppressed supports the corresponding observations from binocular rivalry (Diaz-Caneja, 1928, Kovacs *et al.*, 1996) that a whole stimulus is perceptually suppressed regardless of the fact that this stimulus was being delivered in a split fashion. However, the significant increase in piecemeal rivalry also indicates that there could be a non-negligible contribution of inhibition of one or the other eye.

2.3.2 Swapped flash suppression

The results of the swapped flash suppression experiment (illustrated in Figure A2-2) are shown in Figure A2-6. The experiments using gratings are shown in Figure A2-6a while those with complex figures are shown in Figure A2-6b. The values illustrated in this figure correspond to an average of different parameters for the durations of t_1 , t_2 and t_3 . All the controls were basically at 100% for the only picture presented during the flash period. The results of the classic flash suppression trials (FS1 and FS2) were very strong, with an effective suppression of approximately 90% of the trials. The results of the swapped flash suppression conditions (Sw FS1 and Sw FS2) show that in most cases subjects saw the new pattern during the flash in spite of the fact that it was presented to the ipsilateral “old” eye. The results are much more variable and the proportion of responses lower than those in the classic flash suppression effect. The proportion of trials where subjects reported perceiving a mixture were significantly larger during the swapped trials compared to the classic flash suppression conditions. There were, however, almost no trials where subjects reported perceiving the monocular stimulus during the flash. The results of the swapped flash suppression conditions therefore match those of the “Complete pattern suppression” model (Figure A2-3b) more closely than those of the “Complete eye suppression” model (Figure A2-3a). However the degree of suppression seems to be somewhat weaker than in the classic flash suppression trials. The results were similar but slightly stronger for the swapped flash suppression experiment with complex images (Figure A2-6b).

These results were not affected by the introduction of a short blank interval between the monocular presentation and the flash. The results of the flash suppression experiment with flickering images (Figure A2-4) were somewhat less clear. However, the results were weaker not only for the swapped flash suppression condition but also for the classic flash suppression condition. This was only tested in very few naïve subjects ($n = 4$) and mostly at relatively high flickering rates. Most subjects found the flickering too annoying and it is unclear whether these highly variable results were due to some effect of flash suppression or some other effect that we did not thoroughly investigate (like changes in blinking or fixation, etc.) Several other parameters should be assessed during this condition in which images are flickered at relatively high frequencies in order to be able to draw a clear conclusion. However, as reported previously by Wolfe (Wolfe, 1984), the introduction of a short blank interval t_2 after the monocular image, did not affect the results. This observation suggests that flash suppression cannot readily be accounted by a motion impression or a larger "change" in the stimulation intensity in one of the eyes.

2.4 Discussion

The perceptual suppression of the monocular stimulus during classic flash suppression is very strong under a wide range of parameters. Large changes in the monocular presentation time, the flash time, the introduction of a short blank period, the nature of the images and other variables seem to have only little impact on the inhibition of the monocular stimulus. In spite of this, it has only been poorly investigated at the physiological or functional imaging level. Several of the experiments performed by Wolfe indicated that the phenomenon is not likely to be due to a form of light adaptation or forward masking (Wolfe, 1984).

The phenomenon of flash suppression also works when the image is split into patches and thus distributed between the two eyes (Figure A2-5). This result does not fit well with the model of complete suppression of the ipsilateral eye in the classic flash suppression paradigm. These results match the ones described during binocular rivalry

(Diaz-Caneja, 1928, Kovacs *et al.*, 1996). It could be argued, however, that there is a competition between eyes that occurs at the level of individual patches of the image and not in the whole field of view.

Our results suggest that the phenomenon cannot be explained by a complete suppression of the eye that received the monocular stimulation upon flashing any stimulus to the new eye. If this were the case, the results in the swapped flash suppression condition should be the opposite of those in the classic flash suppression condition (see Figure A2-3). However, the reports from our subjects matched much more closely the ones predicted by the “Complete pattern suppression” model. That is, even after swapping the images, subjects reported in most cases seeing the new stimulus during the flash. The suppression was somewhat weaker than in the classic flash suppression effect. This could indicate that there is a competition at different levels of the visual hierarchy and that there may be a small contribution of competition between eyes. Alternatively, there may be some other effect (such as a small after-effect of the monocular image) that could be influencing the increase in the mixed reports. It is interesting to emphasize that subjects do not report seeing the monocular stimulus during the flash period. These results match the ones observed by Logothetis *et al* during binocular rivalry (Logothetis *et al.*, 1996). However, during binocular rivalry, the observations seem to depend much more strongly on the stimulation parameters (Lee and Blake, 1999).

The perceptual experience during the flash depends on the stimulation pattern that occurred before. Two situations with identical stimulus during the flash can lead to exactly opposite perceptions depending on what the monocular stimulus was. In other words, there is dramatic change in the percept depending on the previous history of visual stimulation in spite of receiving the same retinal information. Thus, the phenomenon is relevant to our discussion of the dissociation between retinal information and perception. Our results in this chapter suggest that, while lower visual areas could play a role in the competition for awareness during this dissociation in the case of flash suppression, they cannot account for the whole perceptual suppression of one of the stimuli. Where would this competition take place then? The results of Sheinberg and Logothetis (Sheinberg and Logothetis, 1997) together with the electrophysiological data that we have presented in Chapter 6 suggest that the activity of most neurons in higher visual areas follows the

percept quite closely. These units are activated whenever the preferred stimulus appears into perception and they are not when the preferred stimulus is suppressed.

2.5 Figure legends

Figure A2-1: Flash suppression with patchy images

An image is uniformly subdivided into several square patches. The patches are then shown to the right and left eyes so that the whole pattern can be reconstructed after merging the information from the two eyes. The flash suppression experiment is then repeated with the patchy images. **(a)** Control situations where a vertical grating is shown during a time period t_1 and then the same vertical grating is flashed during a time period t_3 (this condition is referred to as “control 2” in Figure A2-5). **(b)** A horizontal grating is shown during t_1 and a vertical grating is flashed during t_3 (referred to as “control 3”). **(c)** A horizontal grating is followed by a flash of a horizontal and a vertical grating. Note that the patches that contained the horizontal grating are kept in the same eye. **(d)** A vertical grating is followed by a flash of a horizontal and a vertical grating.

Figure A2-2: Swapped flash suppression

Example of some of the conditions used in the swapped flash suppression experiment. **(a, b)** Control conditions where only one image is shown at a time. A horizontal grating was shown monocularly during a time t_1 . A blank period of duration t_2 ($t_2=0$ ms in several conditions; see text) was interspersed and then either a horizontal or a vertical grating was flashed during a time period t_3 . **(c, d)** Classic flash suppression experiment. A grating was shown monocularly and then a different grating was flashed to the contralateral eye while the original stimulus was flashed in the ipsilateral eye. **(e, f)** Swapped flash suppression conditions. A grating was shown monocularly and then a different grating was flashed to the ipsilateral eye while the original grating was flashed to the contralateral eye. **(g, h)** Flash suppression was also tested with more complex images instead of gratings. Here we only show two conditions for the complex images, a classic

flash suppression one (**g**) and one of the two possible swapped flash suppression conditions (**h**).

Figure A2-3: Predictions for swapped flash suppression

Predictions for the swapped flash suppression experiments from two extreme models. The predicted results are shown for the classic and swapped flash suppression tests (see Figure A2-2). The stimuli shown to each eye, here schematically indicated as image "1" or image "2" are shown at the top. The results indicate the proportion of cases in which subjects indicate they perceive image 1, image 2 or a mixture during the flash. The results of the classic flash suppression experiment are the same for the complete eye suppression model (**a**) or the complete pattern suppression model (**b**). However, during the swapped flash suppression experiment, the predictions are exactly opposite.

Figure A2-4: Swapped flash suppression with flickering

Swapped flash suppression experiment with flickering of the images. Each of the images was flickered on and off (for time periods p_1 and p_2 respectively; see text for details) instead of remaining always on as in the previous experiment. Images were flickered both during the monocular presentation period t_1 and during the flash period t_3 . (**a**) Classic flash suppression experiment where the monocular horizontal grating was shown to the same ipsilateral eye during the flash. (**b**) Swapped flash suppression experiment where the monocular horizontal grating was shown to the contralateral eye during the flash.

Figure A2-5: Patchy images, results

Results of the patchy flash suppression experiment (Figure A2-1). (**a**) Experiment using horizontal and vertical gratings. The proportion of responses where subjects reported seeing the horizontal, vertical or mixed grating during the flash period is shown on the y axis for each condition. There were four control conditions: C1 (horizontal grating during t_1 , horizontal grating during flash), C2 (vertical grating during t_1 , vertical grating during

flash), C3 (horizontal grating during t_1 , vertical grating during flash) and C4 (vertical grating during t_1 and horizontal grating during flash). There were two flash suppression conditions: FS1 (horizontal grating during t_1) and FS2 (vertical grating during t_1); both a horizontal and a vertical grating were shown during the flash for FS1 and FS2. Error bars show 95% percentiles. **(b)** Experiment with complex images (girl face and truck). The conventions and conditions are the same as in part **(a)**.

Figure A2-6: Swapped flash suppression, results

Results of the swapped flash suppression experiment (Figure A2-2). **(a)** Experiment with gratings. **(b)** Experiment with complex images (girl face and truck).

3 Appendix 3. The frontal lobe and binocular rivalry – Repetitive trans-cranial magnetic stimulation studies

3.1 Introduction

A major distinction should be drawn in our understanding of any phenomenon between correlation and cause. All the work that we have described in the previous chapters shows a correlation between neuronal activity and the occurrence of specific percepts in the subject's brains. In other words, we observed concomitant changes in neuronal activity when the subject's visual awareness changed due to a presentation of a specific visual stimulus (Chapter 3), the imagination of specific stimuli (Chapter 4), the different stages of sleep (Chapter 5), or the perceptual changes elicited during flash suppression (Chapter 6). As we have discussed in Chapter 7, these neuronal changes, however fascinating, do not necessarily imply a causal link between the two phenomena. Trying to address the difficult question of *cause* constitutes one of the major challenges in Neuroscience. Most of the work involving single or multiple neuron electrophysiology, psychophysics, functional imaging and scalp electro or magneto encephalographic recordings only yield data about correlations.

Repetitive transcranial stimulation (rTMS) offers the possibility of altering the activity in the brain in a temporary and non-invasive manner¹²⁴ (Pascual-Leone *et al.*, 1998, Walsh and Cowey, 1998). Therefore, rTMS has been used to produce rapidly

¹²⁴ By non-invasive here I mean that it is not necessary to do any kind of surgery, insert electrodes or drugs or any other equipment into the subject. This does not mean that the stimulation does not cause immediate changes in the neuronal activity in the brain (otherwise, there would be no perceptual or motor effects).

reversible local changes to study the functional importance of a given brain area¹²⁵. rTMS can potentially provide an ideal complementary tool to electrophysiological and functional imaging experiments to assess the causality relationship between neural activity and perception and/or behavior (Pascual-Leone *et al.*, 1998, Walsh and Cowey, 1998, Wassermann and Grafman, 1999).

Here we have used this technique to explore the function of the human frontal cortex during binocular rivalry. When different stimuli are projected onto corresponding areas of the two retinae, perception alternates between the two images, a phenomenon called binocular rivalry (see Chapter 6). The visual input remains constant (except for small eye movements) but the conscious percept changes from one image to a mixture (piecemeal rivalry) and then to the other image, making this an interesting phenomenon for the study of the neuronal correlates of visual awareness (Crick, 1994, Crick and Koch, 1998). A more complete description of the phenomenon was given in Chapter 6; there are also several excellent reviews on the topic (Blake, 1989, Blake, In press, Leopold and Logothetis, 1999, Logothetis, 1998, Levelt, 1968). Several models have been proposed for the origin of the alternations in binocular rivalry; these date back to the opposing alternatives high and low level views presented by Helmholtz and Herring respectively (von Helmholtz, 1866, Alais *et al.*, 2000). One of the most prevalent ideas is that alternation arises as a result of direct competition between two groups of neurons that represent one or the other percept. Whether this competition occurs in early visual areas or in later visual areas is still a matter of controversy (Crick and Koch, 1995, Logothetis, 1998, Leopold and Logothetis, 1999, Sheinberg and Logothetis, 1997, Tong *et al.*, 1998, Tong and Engel, 2001, Lee and Blake, 1999, Blake, 1989, Blake, In press, Polonsky *et al.*, 2000). Another explanation is that another area participates in biasing the percept in one way or the other facilitating, enhancing or inhibiting the perceptual switches. This proposal does not exclude the possible existence of a direct competition between two networks that represent the two percepts. Some evidence in favor of this is given by the observations that rivalry alternations can be influenced to some degree by some “high-level” cognitive processes such as selective visual attention, the “meaning” of the stimuli

¹²⁵ An important issue to consider in this as in any other technique that could affect the activity inside the brain is whether there are any collateral consequences or short-term or long-term effects due to the stimulation. This is discussed in Section 3.2.

and even mood¹²⁶. There are also recent data from a very interesting set of investigations by Lumer *et al* (Lumer *et al.*, 1998, Lumer and Rees, 1999) where they have observed in a functional MRI study that activity in right pre-frontal cortex¹²⁷ is correlated with perceptual transitions during binocular rivalry. This suggests the possibility that neurons selective for one of the percepts (for example in inferotemporal cortex or in the MTL) send the information to pre-frontal cortex where the involuntary decisions regarding a perceptual switch take place. Alternatively, the frontal cortex may play an important role in the modulation of the competition between sensory networks in the higher areas of the temporal lobe.

The present study aimed to address the putative role of the frontal cortex during binocular rivalry by altering the activity of neurons in pre-frontal cortex in human subjects performing a binocular rivalry experiment using rTMS. Our preliminary results suggest that rTMS in pre-frontal cortex increases the percentage of time that subjects spend in piecemeal rivalry. All the data in the current appendix was collected in collaboration with Julian Keenan and Alvaro Pascual-Leone of the Beth Israel Hospital at Harvard Medical School.

3.2 Experimental procedures

3.2.1 Safety of rTMS

An important issue to consider in this as in any other technique that could affect the activity inside the brain is whether there are any collateral consequences and short-term or long-term effects due to the stimulation. The fact that TMS can induce an electrical signal in the brain does not necessarily mean that it is bad for the brain or that it can harm it in any way. It could be claimed that talking to someone can also cause a

¹²⁶ I have once heard that if you present a picture of the woman you love in one eye and ANY other stimulus in the world on the other eye, the picture of the loved one never disappears from perception (Jeremy Wolfe, personal communication). Because of the potential implications of this and perhaps for fear that this may or may not be true, I have never tried to assess the accuracy of this statement.

¹²⁷ It should be noted that this is not the only area in the brain whose activity was correlated with perceptual switches (see (Lumer and Rees, 1999, Lumer *et al.*, 1998) for details).

neuronal signal in the listener and we have actually shown that we can elicit very specific changes in neuronal activity by presenting visual stimuli in Chapter 3. But the question is rather different in this case since TMS is not the normal way of eliciting neuronal activity in the brain and we still do not understand in detail the precise patterns of activation and inactivation caused by stimulation. In the case of rTMS, the possible negative short-term effects have been shown to depend on the parameters of the stimulation. Large currents at high frequencies can lead to seizure attacks in otherwise normal subjects. Therefore, a set of rigorous guidelines has been established that determine which parameters and combinations thereof are safe and which are not in the usage of rTMS (Chen *et al.*, 1997). All the work described in the current Appendix was done in conformity with these guidelines and after the informed consent of the subjects. Below these thresholds, I am not aware of any adverse short-term effects of rTMS (Pascual-Leone *et al.*, 2000, Chen *et al.*, 1997, Ruohonen, 1998, Walsh and Cowey, 1998, Wassermann and Grafman, 1999). The long-term effects are more difficult to pinpoint. The transcranial magnetic stimulation techniques as they are applied today are quite recent for a very long-term study¹²⁸.

3.2.2 Stimulus presentation and subject responses

We conducted two sets of experiments (involving five and three subjects respectively). Subjects were people associated to the Beth Israel Hospital, either as researchers, MDs or staff and had normal or corrected-to-normal vision. Subjects were naïve to the task of binocular rivalry. All except one subject (J.K.) were also naïve to the purposes of the experiment. Some of them (two out of eight) also experienced rTMS for the first time.

Stimuli were presented in a Macintosh Computer using the MATLAB Psychophysics Toolbox (Brainard, 1997, Pelli, 1997). Different stimuli were presented to each eye by using a custom built Helmholtz mirror haploscope (Figure A3-1). This enables us to present separate stimuli to each eye with no cross-leakage whatsoever of one stimulus onto the wrong eye. The mirrors were adjusted for each subject in each

¹²⁸ It could be argued also that we still do not know either whether there are any long-term effects of repeatedly being exposed to magnetic fields in functional imaging studies.

repetition to ensure that there is no double vision. Stimuli consisted of sinusoidal gratings of horizontal and vertical orientation with a spatial frequency of three cycles per degree. Subjects were requested to report what they saw by pressing two buttons in the computer keyboard. Keyboard responses were recorded at 75 Hz in the same computer that presented the figures. Subjects were asked to press and hold a specific key whenever they perceived exclusively the horizontal grating and to press and hold another key whenever they perceived a vertical grating. They were instructed to release both keys if they saw patches of vertical and horizontal images (piecemeal rivalry). The association between which hand the subject used to report his percept in each case and the stimulation conditions were randomly assigned and counterbalanced throughout the different sessions in each subject. The order of stimuli and stimulation were counterbalanced across subjects (Figure A3-2).

3.2.3 Motor threshold discrimination

The actual current induced inside the brain by magnetic stimulation depends on a number of subject specific factors as well as equipment specific ones. Among the subject related variables are the thickness of the dura and skull as well as the subject's arousal state. The specific type, shape and orientation of the coil also influence the degree and spread of the induced current (Ruohonen, 1998, Meyer *et al.*, 1991). It is empirically observed that the strength of the magnetic field required to elicit a movement in the hand of different subjects varies quite significantly. A constant TMS intensity across subjects may actually be correlated with a very different induced currents and thus potentially cause very different neural effects. Therefore, the TMS intensities are measured relative to the subject's motor threshold. We therefore determined the motor threshold (MT) for each subject. This was done by measuring the minimal field strength of single pulses applied in the motor cortex required to elicit movement in the hand or fingers. The results of this motor threshold estimation are described below. Given that the intensity of stimulation during the experiment is proportional to the MT, the accuracy in determining the MT could be very important in attempting to make the induced currents as homogeneous as possible across subjects.

3.2.4 Magnetic stimulation

We used a circular coil of 8 cm diameter (Ruohonen, 1998, Kosslyn *et al.*, 1999). The circular coil activates a larger cortical area than other coil shapes. It also allows to the spread of stimulation to reach into deeper areas of the brain (Roth *et al.*, 1991). The drawback is the lower spatial resolution and therefore the potential possibility of obtaining less specific results. For this study, we chose to use this type of coil because of the larger area and intensity of stimulation. The maximum magnetic field strength of the equipment is 2 Tesla. We used 100% MT intensity as the magnetic field strength for our experiments. In order to compare results across subjects it is important to know whether they are receiving the same stimulation strength. The method that we have used of relating the intensity to the motor threshold assumes that the proportionality constant between magnetic field and induced current in the brain is the same in the motor cortex and the area of interest. According to this assumption, by using the same intensity ratio between the area of interest and the motor cortex in different subjects we get an approximately similar induced current. It should be noted that there is no specific test for determining a threshold directly in the area of interest given our current knowledge.

3.2.5 Experimental design

Subjects performed four sessions of binocular rivalry. Each session consisted of 10 repetitions of 40 seconds each. Between the first and the second session and between the third and the fourth session, subjects received rTMS during 10 minutes at 1 Hz. The order of stimulation was counterbalanced across the different subjects. A scheme of the experiment is shown in Figure A3-2.

3.2.6 Anatomical localization

The anatomical localization of the stimulation site was different in the first and second series of experiments. During the first experiment (5 subjects), we used the average Tailarach coordinates of the area of maximum activation in the study by Lumer (Lumer and Rees, 1999). We obtained structural MRI information (1.5 Tesla) from the five subjects that participated in this experiment. The MRI data was converted to

Tailarach coordinates and the distance (projected to the skull) from specific fiducials was measured for each subject. We used the nose, ear lobes and teeth fiducials.

During the second experiment (3 subjects), we did not have the structural MRI data for any of the subjects. We therefore utilized a different technique to determine the loci of stimulation. Primary visual cortex was also stimulated in the second experiment (but not in the first). Localization of occipital cortex followed the procedure described by Kosslyn *et al.* (Kosslyn *et al.*, 1999). For the right and left pre-frontal cortex, we used relative distances to approximate the stimulation location. Using the data from the subjects that participated in the first experiment, we computed distances normalized with respect to the head circumference measured at the middle ear and right above the nose (C_{head}). Then we computed for each of the fiducial points, distances to the target stimulation areas normalized by c_{head} .

$${}_j p_i = {}_j x_i / C_{head}$$

where ${}_j x_i$ represents the distance in mm for subject j from fiducial point i to the locus of interest taken from the MRI data (i =ear, teeth, nose, nasion, glabella, porion). We averaged these normalized distances across subjects:

$$p_i = \langle {}_j p_i \rangle$$

where $\langle \rangle$ correspond to the mean over the 5 subjects. For the subjects that participated in experiment 2, we measured C_{head} and then we estimated the coordinates of the stimulation loci by computing

$${}_j x_i = p_i * C_{head}$$

We directly compared the results of this extrapolation to the ones obtained from the MRI structural data for the subjects in the first experiment. The values for one of the subjects are shown in Table A3-1. This probably turns out to be a small underestimation of the true error since we are using the same data for calibration and testing and there could be some overfitting. In all the subjects from experiment 1 the error was less than 6% from any fiducial point except for the porion. The standard deviation of p_i from this point was very large. We only considered the other fiducial points for localization. The localization error between the different reference points was < 1 cm.

3.2.7 Data Analysis

For each binocular rivalry session, we obtained 10 repetitions of 40 sec each (see Figure A3-2). We computed the intervals of dominance of each type of grating (horizontal and vertical) as well as the intervals of piecemeal rivalry¹²⁹. We computed the distribution of intervals of horizontal, vertical and piecemeal rivalry dominance and we fitted a gamma distribution to the each of the distributions (Levelt, 1968, Myerson *et al.*, 1981, Logothetis, 1998). Gamma distributions with orders between 2 and 4 were obtained for the horizontal and vertical periods (Figure A3-3). The gamma order for the distribution of piecemeal periods was close to 1. For each experiment we computed the following variables: m_h = mean period for horizontal dominance (ms), m_v = mean period for vertical dominance (ms), m_p = mean period for piecemeal rivalry (ms), s_h = standard deviation of the period for horizontal dominance (ms), s_v = standard deviation of the period for vertical dominance (ms), s_p = standard deviation of the period for piecemeal rivalry (ms), n_h = number of periods of horizontal dominance, n_v = number of periods of vertical dominance, n_p = number of periods of piecemeal rivalry, g_{h1} = parameter 1 of the gamma distribution of the periods of horizontal dominance, g_{h2} = parameter 2 of the gamma distribution of the periods of horizontal dominance, g_{v1} = parameter 1 of the gamma distribution of the periods of vertical dominance, g_{v2} = parameter 2 of the gamma distribution of the periods of vertical dominance, g_{p1} = parameter 1 of the gamma distribution of the periods of piecemeal rivalry, g_{p2} = parameter 2 of the gamma distribution of the periods of piecemeal rivalry, p_h = proportion of total time of horizontal dominance, p_v = proportion of total time of vertical dominance, p_p = proportion of total time of piecemeal dominance ($p_i = n_i * m_i / (n_h * m_h + n_v * m_v + n_p * m_p)$ where $i = h, v$ or p).

¹²⁹ We discarded the first 3 seconds of data in each repetition (this was because as the experiment starts the subject may take longer to respond or see a higher proportion of mixtures). We also discarded the first repetition within each session. Again, this is because the responses would be different in this case, particularly in the first session that was the first encounter of the subject with binocular rivalry.

3.3 Results

3.3.1 Normalized distribution of dominance periods

Since the subjective criteria for piecemeal rivalry and exclusive dominance may be different across subjects, it is typical in binocular rivalry to obtain inter-individual variations in the mean dominance periods. In order to compare data across subjects we normalized the dominance periods by the mean period duration (Levelt, 1968, Blake, 1989, Logothetis *et al.*, 1996, Lumer *et al.*, 1998).

The first qualitative observation is that subjects could still observe normal alternations between the images during binocular rivalry after rTMS (Figure A3-3). The distribution of dominance periods during binocular rivalry is typically well fit by a gamma distribution of low order (Levelt, 1968, Logothetis *et al.*, 1996). The parameters of the gamma distribution for the normalized distributions did not change significantly after TMS in either the left or right pre-frontal cortex. Thus, the normalized distribution of dominance periods is not significantly affected by frontal transcranial magnetic stimulation.

3.3.2 Piecemeal rivalry duration

While normalizing by the mean period seems to be necessary to pull data across subjects because of inter-individual variability, this procedure misses an important aspect of the information obtained in the experiment. In particular, it may turn out that the effect of TMS is to alter the mean dominance period of the gratings or the mean period of piecemeal rivalry.

We therefore computed the mean dominance time and the percentage of time spent in each percept for each subject under the different stimulation conditions. Our preliminary results suggest that there is an increase in the time that subjects observed piecemeal rivalry after right pre-frontal TMS (green bars in Figure A3-4). In Figure A3-4 we indicate the change in the number of periods, mean dominance period and total time spent in each percept after right and left pre-frontal TMS. Specifically, let t_{h-pl} be the percentage of time that a subject sees the horizontal grating before left pre-frontal TMS

and t_{h_al} the corresponding time after left TMS. Since the absolute value of t_{h_pl} varies from subject to subject we computed the normalized change, t_{h_pl}/t_{h_al} and this is what is plotted in Figure A3-4. A similar procedure is used for the number of periods and the mean dominance period¹³⁰.

The number of perceptual periods did not seem to be altered by TMS. Therefore, the increase in piecemeal rivalry for the mean dominance period is proportional to that in the total percentage of time in each percept¹³¹.

3.3.3 Comparison with a simple binocular rivalry model

In order to determine whether the results could be due to random fluctuations in the amount of time that subjects spent in each percept, we compared the values obtained with a basic null hypothesis. It is well established that the distribution of dominance times is well fitted by a gamma distribution (Figure A3-3 and Levelt, 1968, Logothetis *et al.*, 1996). As a null hypothesis, let's assume that there is a single gamma distribution for each subject regardless of whether TMS was applied or not and let's ask whether we can obtain the same results.

In order to compare the results we will use the following second order rivalry model (see scheme in Figure A3-5). The model starts randomly in the H, V or P states. The amount of time in a given state is randomly taken from a gamma distribution with parameters g_{1s} and g_{2s} where $s = h, v$ or p . After that, there is a transition to another state. The probability of switching from one state to another depends on the history of the process. That is, the probability of switching from, say, piecemeal to horizontal is larger if the previous state was vertical than if it was horizontal¹³². All the parameters were

¹³⁰ We also computed the ratio of the two pre-TMS conditions (sessions 1 and 3). This is not an accurate baseline condition since, because we counterbalanced the order of right and left TMS, in three subjects, this baseline includes a period of right TMS. Also, the time between the two rivalry periods that we compare in this baseline was longer than the one we compare in the other situations. Furthermore, subjects had to switch the key they were using to report horizontal and vertical between the two stimulation conditions. These factors explain the apparent increase in piecemeal rivalry of approximately 20% in this condition.

¹³¹ We asked all the subjects to report their subjective impressions about the test before and after TMS. Two subjects (FM, MH) indicated that "they were getting more confused towards the end." Subject JK reported that he "could not do the task" after right TMS. Subjects reported that it was complicated (at the beginning) to adapt to switching the key they had to press between condition 2 and 3.

¹³² We only consider in this model the history of one state previous to the current one. The probability of a given transition conditional on the previous two states is basically the same as the probability conditional

separately fitted for each subject with the behavioral data from all the different stimulation conditions described above. We used the average parameters over all conditions for each subject.

We randomly separated the “conditions” for the model into right and left TMS and we ran the model 100 times for each of the five subjects. We analyzed the data from the model in the same way as the experimental data. Only 1 out of 100 runs had a deviation beyond 20% (either positive or negative) in the time spent in piecemeal rivalry. This corresponded to an increase of 22% in the percentage of total time that the model perceived vertical between two conditions.

3.3.4 Experiment 2

The preliminary conclusions above were not supported by the results in our 2nd experiment (Nature would be too easy otherwise)¹³³. However, it is important to note that there were several important differences between the two experiments.

As we have discussed above, we did not have MRI localization data for the subjects in experiment 2 and we used a different procedure to locate the stimulation site (see Section 3.2.6). The error using this method in the first series subjects was < 1 cm (but this could be due to overfitting and therefore the error could be larger in those subjects who participated in experiment w, see Section 3.2.6). The location from the different fiducial points that was reached using this method did not match perfectly among the different fiducials, adding an error of 1-2 cm. It must be noted that, given the resolution of the circular coil that was used (diameter 8 cm) it is not clear that these localization errors would translate into large differential brain activation. But it is conceivable that stimulation in the subjects in the second experiment took place away from or at least not with the same intensity as in the first set of subjects.

Subject SC (naïve to TMS) seemed to be very nervous before the TMS procedure. Maybe, this yielded a lower Motor Threshold than usual and therefore she might have been stimulated with a very low intensity.

on only the previous state according to the behavioral data. This suggests that the process or the system only has memory for one previous state.

¹³³ "A man who has only one watch knows what time it is. A man who has two is never sure," anonymous proverb.

After the right TMS period in subject EB, one of the mirrors used in the rivalry setup was loose and we had to spend approximately 4 minutes to re-adjust it. This implies that the first few minutes after TMS (presumably the most important ones because of the decay of the effect of TMS with time) were lost in this case.

3.4 Discussion and very preliminary conclusions

We obviously need to do more experiments before we can draw any clear conclusion about the putative role of frontal cortex during binocular rivalry. Unfortunately, I have not been able to run these experiments again. Not that I think they were not interesting or worth pursuing but the pace of research has been quite fast and Boston is somewhat far from Caltech (even farther away than UCLA in spite of the traffic). I would like to briefly discuss however, the preliminary conclusions and potential implications of these results; everything that I ponder on below should be taken as preliminary and tentative given that we would need a larger n in order to confirm our findings.

In addition to TMS, another way of causally linking the neuronal activity in a brain area with electrophysiological or imaging results is the study of neurological patients or lesions in animals. However, the resolution of lesions may not be accurate enough in many cases to perform an accurate direct comparison to neurophysiological findings (and there might be several other confounding factors too). Our preliminary results, together with the imaging of Lumer *et al.* (Lumer *et al.*, 1998) suggest that people with lesions in frontal cortex may have difficulties in switching their percept. We would therefore expect an increase in the mean dominance of the stimuli or in piecemeal rivalry.

Indeed, it is interesting to note that a patients with damage to the frontal lobe seem to show difficulties in decision taking process or switching perception and attention (Meenan and Miller, 1994, Smith and Jonides, 1999, Wilkins *et al.*, 1987). Ambiguous figures such as the Necker cube or the vase-face illusion show a similar gamma distribution of dominance intervals to the ones observed during binocular rivalry (Borselino *et al.*, 1972). Interestingly, frontal patients show a clear difficulty in switching

their percept in these other types of bistable stimuli (Ricci and Blundo, 1990, Meenan and Miller, 1994, Wilkins *et al.*, 1987). Recent work in patients with bipolar disorder has shown that the rate of perceptual alternation during binocular rivalry is slow compared with normal controls (Pettigrew and Miller, 1998).

Frontal cortex has been generally associated with high-level cognitive processes and the ability to make volitional conscious decisions (Wilkins *et al.*, 1987, Miller, 2000). It would be very interesting to study the neuronal activity in frontal cortex during binocular rivalry. In our studies (see Chapter 6), we unfortunately had very little probes implanted in frontal lobe targets and most of those were in the supplementary motor areas and anterior cingulate and not in pre-frontal cortex. In monkeys, to our knowledge, the role of frontal cortex during binocular rivalry has not been explored yet. The preliminary evidence presented here suggests that it will be very interesting to study the activity of ensembles of pre-frontal neurons during perceptual transitions and the possible correlations in the activity of frontal cortex and the temporal lobe.

*Quod scripsi, scripsi*¹³⁴.

¹³⁴ What I have written, I have written.

3.5 Figure legends and tables

Figure A3-1: Helmholtz/Wheatstone mirror haploscope and TMS round coil

(a) Mirror setup used to deliver different stimuli to the two eyes during the binocular rivalry experiments (von Helmholtz, 1866, Wheatstone, 1838.) Stimuli were presented on the two sides of a computer monitor and were reflected on the mirrors to reach each eye. This optic setup to show stimuli to the two eyes has no cross-leakage to the other eye. (b) Schematic view of the TMS procedure. A changing magnetic field will induce, according to Faraday's law, an induced current in the brain. The magnitude of the induced current depends on several parameters such as the shape of the coil, the stimulation parameters and the relative position with respect to the brain structures (Ruohonen, 1998).

Figure A3-2: Scheme of the experiments

Scheme of the order of the rivalry tests and rTMS. BR = binocular rivalry, L/R = left/right. The order of which hemisphere was stimulated first was counterbalanced across subjects.

Figure A3-3: Average normalized distribution of dominance periods

Average normalized distribution of dominance periods during experiment 1. Data from 5 subjects during horizontal and vertical dominance periods are lumped together (there was no statistically significant difference between the dominance period for horizontal and vertical gratings). Top, left: before left TMS. The data corresponds to 664 intervals, the order of the gamma distribution was $n = 2.33$ (95% confidence interval = [2.14, 2.53]), $\lambda=0.42$, (95% confidence interval= [0.39, 0.47]), the correlation coefficient between the data and the fit was $r^2 = 0.97$. The bin size in all the distributions shown in this figure was 0.1. n, λ are the two parameters of a gamma distribution¹³⁵. Top, right: after left TMS.

¹³⁵ $f(x) = e^{-x/\lambda} x^{n-1} / \lambda^n \Gamma(n)$ where $\Gamma(n)$ is the gamma function: $\Gamma(n) = \int_0^\infty x^{n-1} e^{-x} dx$, $n > 0$

776 intervals. $n = 2.72$, CI = [2.52, 2.93]. $\lambda = 0.37$, CI= [0.34, 0.40]. $r^2 = 0.96$. Bottom, left: before right TMS. 781 intervals. $n = 2.30$, CI = [2.12, 2.48]. $\lambda = 0.43$, CI= [0.39, 0.48]. $r^2 = 0.97$. Bottom, right: after right TMS. 745 intervals. $n = 2.15$, CI = [1.95, 2.35]. $\lambda = 0.47$, CI= [0.42, 0.52]. $r^2 = 0.97$.

Figure A3-4: Average changes in dominance duration with TMS

For each subject we computed the change in the number of dominance periods (left), the mean period duration (middle) and the percentage of total time in a given percept (right). Results show the average values across subjects (mean \pm S.E.M.) for each TMS treatment (R, right TMS; L, left TMS; B, baseline). The baseline condition corresponds to a comparison of the binocular rivalry test at the beginning of the whole experiment for each subject and before the second application of TMS.

Figure A3-5: Schematic of second order rivalry model

A given percept (Horizontal, Vertical or Piecemeal) is maintained for a random amount of time taken from a gamma distribution. The parameters of the gamma distribution are the average fitted values from the data assuming a single distribution regardless of the TMS treatment. The transition probabilities take into account the history of the process (see text for details).

3.6 Tables

Table A3-1: Localization procedures

MRI (mm)	jx_i (mm)	Diff. D (mm)	Diff. (%)	From
103.2	101.52	-1.68	-1.6	Ear
148.3	146.98	-1.32	-0.9	Teeth
147.3	142.69	-4.61	-3.1	Nose
99	104.78	5.78	5.8	Nasion
99.5	102.67	3.17	3.2	Glabella
49.9	64.47	14.57	29.2	Porion

Comparison of the localization procedures used in the experiment 1. Coordinates are shown for one of the subjects in experiment 1 with the extrapolation method described in the text. Column 1 indicates the MRI coordinates (as used in experiment 1) from each of the fiducial markers shown in the last column. Column 2 indicates the coordinates as computed from the extrapolation (see text for details). Column 3 indicates the difference in mm and column 4 the relative difference.

References

Abeles, M. (1982). Quantification, smoothing, and confidence limits for single-units histograms. *Journal of Neuroscience Methods*, **5**:317-325.

Abeles, M. (1991). *Corticonics*, Cambridge University Press, Cambridge.

Adolphs, R., Tranel, D., Damasio, H. and Damasio, A. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the amygdala. *Nature*, **372**:669-672.

Aersten, A., Gerstein, G., Habib, M. and Palm, G. (1989). Dynamics of neuronal firing correlation - Modulation of effective connectivity. *Journal of Neurophysiology*, **61**:900-917.

Aguirre, G. K. and D'Esposito, M. (1997). Environmental knowledge is subserved by separable dorsal/ventral neural areas. *Journal of Neuroscience*, **17**:2512-2518.

Aguirre, G. K., Detre, J. A., Alsop, D. C. and D'Esposito, M. (1996). The parahippocampus subserves topographical learning in man. *Cerebral Cortex*, **6**:823-829.

Alais, D., O'Shea, R., Mesana-Alais, C. and Wilson, I. (2000). On binocular alternation. *Perception*, **29**:1437-1445.

Albright, T., Jessell, T., Kandel, E. and Posner, M. (2000). Neural Science: A Century of Progress and the Mysteries that Remain. *Cell*, **100**:S1-S55.

Allison, T., Ginter, H., McCarthy, G., Nobre, A. C., Puce, A., Luby, M. and Spencer, D. D. (1994). Face recognition in human extrastriate cortex. *Journal of Neurophysiology*, **71**:821-825.

Allison, T., Puce, A., Spencer, D. and McCarthy, G. (1999). Electrophysiological studies of human face perception. I: Potentials generated in occipitotemporal cortex by face and non-face stimuli. *Cerebral Cortex*, **9**:415-430.

Allman, J. and McGuinness, E. (1988). Visual cortex in primates. *Comparative Primate Biology*, **4**:279-326.

Allman, J., Miezin, F. and McGuinness, E. (1985). Stimulus specific responses from beyond the classical receptive field: neurophysiological mechanisms for local-global comparisons in visual neurons. *Annual Review of Neuroscience*, **8**:407-430.

Alvarez, P. and Squire, L. R. (1994). Memory consolidation and the medial temporal lobe: a simple network model. *Proceedings of the National Academy of Sciences U.S.A.*, **91**:7041-7045.

Andrews, T. J. and Purves, D. (1997). Similarities in normal and binocularly rivalrous viewing. *Proceedings of the National Academy of Sciences U.S.A.*, **94**:9905-9908.

Aserinsky, E. and Kleitman, N. (1953). Regularly occurring periods of eye motility and concomitant phenomena during sleep. *Science*, **118**:273-274.

Bair, W. and Koch, C. (1996). Temporal precision of spike trains in extrastriate cortex of the behaving macaque monkey. *Neural Computation*, **8**:1185-1202.

Barlow, H. (1972). Single units and sensation: a neuron doctrine for perception. *Perception*, **1**:371-394.

Bartolomeo, P., Bachoud-Levi, A.-C., de Gelder, B., Denes, G., Dalla Barba, G., Brugieres, P. and Degos, J.-D. (1998). Multiple-domain dissociation between impaired visual perception and preserved mental imagery in a patient with bilateral extrastriate lesions. *Neuropsychologia*, **36**:239-249.

Bastian, J. and Nguyenkim, J. (2001). Dendritic Modulation of Butst-Like Firing in Sensory Neurons. *Journal of Neurophysiology*, **85**:10-22.

Behrmann, M., Winocur, G. and Moscovitch, M. (1992). Dissociation between mental imagery and object recognition in a brain-damaged patient. *Nature*, **359**:636-637.

Berger, H. (1930). Ueber das Elektroenkephalogramm des Menschen. *Journal of Psychology and Neurology*, **40**:160-179.

Berkeley, G. (1709). *An essay towards a new theory of vision*, Christopher D. Green, [electronic source].

Berry, M. J. and Meister, M. (1998). Refractoriness and neural precision. *Journal of Neuroscience*, **18**:2200-2211.

Berry, M. J., Warland, D. K. and Meister, M. (1997). The structure and precision of retinal spike trains. *Proceedings of the National Academy of Sciences U.S.A.*, **94**:5411-5416.

Bialek, W., Steveninck, R. and Warland, D. (1991). Reading a neural code. *Science*, **252**:1854-1857.

Binder, J., Frost, J., Hammeke, T., Bellgowan, P., Rao, S. and Cox, R. (1999). Conceptual processing during the conscious resting state: A functional MRI study. *Journal of Cognitive Neuroscience*, **11**:80-93.

Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*, Clarendon Press, Oxford.

Bisiach, E. and Luzzatti, C. (1978). Unilateral neglect of representational space. *Cortex*, **14**:129-133.

Blake, R. (1989). A neural theory of binocular rivalry. *Psychological review*, **96**:145-167.

Blake, R. (*In press*). A primer on binocular rivalry, including current controversies. *Brain and Mind*.

Blake, R. and Cormack, R. (1979a). On utricular discrimination. *Perception and Psychophysics*, **26**:53-68.

Blake, R. and Cormack, R. (1979b). Psychophysical evidence for a monocular visual cortex in stereoblind humans. *Science*, **203**:274-275.

Bliss, T. and Collingridge, G. (1993). A synaptic model of memory: long-term potentiation in the hippocampus. *Nature*, **361**:31-39.

Bliss, T. and Lomo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *Journal of Physiology*, **232**:331-356.

Bogen, J. (1994). Descartes' fundamental mistake: Introspective singularity. *Behavioral and Brain Sciences*, **17**:175-176.

Bogen, J. E. (1995). On the neurophysiology of consciousness: I. An overview. *Consciousness and Cognition*, **4**:52-62.

Bonneh, Y., Sagi, D. and Karni, A. (2001). A transition between eye and object rivalry determined by stimulus coherence. *Vision Research*, **41**:981-989.

Borselino, A., De Marco, A., Allzetta, A., Rinesi, S. and Bartolini, B. (1972). Reversal time distribution in the perception of visual ambiguous stimuli. *Kybernetik*, **10**:139-144.

Bradley, D. C., Chang, G. C. and Andersen, R. A. (1998). Encoding of 3D structure from motion by primate area MT neurons. *Nature*, **392**:714-717.

Brainard, D. (1997). The Psychophysics Toolbox. *Spatial Vision*, **10**:433-436.

Braun, A., Balkin, R., Wesensten, N., Carson, R., Varga, M., Baldwin, P., Selbie, S., Belenky, G. and Herscovitch, P. (1997). Regional cerebral blood flow throughout the sleep-wake cycle: An H₂-15-O PET study. *Brain*, **120**:1173-1197.

Braun, A. R. and Herscovitch, P. (1998). Dissociated pattern of activity in visual cortices and their projections during human rapid eye movement sleep. *Science*, **279**:91-95.

Britten, K. H., Shadlen, M. N., Newsome, W. T. and Movshon, J. A. (1992). The analysis of visual motion: a comparison of neuronal and psychophysical performance. *Journal of Neuroscience*, **12**:4745-4765.

Broca, M. (1861). Remarques sur le siege de la faculte du langage articule, suivies d'une observation d'aphemie (perte de la parole). *Bulletin de la Societe Anatomique*, **6**:330-357.

Brodmann, K. (1909). *Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues*, Barth, Leipzig.

Brody, C. (1997). Analysis and modeling of spike train correlations in the lateral geniculate nucleus. Thesis, Computation and Neural Systems, Caltech, Pasadena.

Brody, C. (1999). Correlations without synchrony. *Neural Computation*, **11**:1537-1551.

Brothers, L. and Ring, B. (1993). Mesial temporal neurons in the macaque monkey with responses selective for aspects of social stimuli. *Behavioral Brain Research*, **57**:53-61.

Brown, M. and Aggleton, J. (2001). Recognition memory: What are the roles of the perirhinal cortex and hippocampus? *Nature Reviews Neuroscience*, **2**:51-61.

Buckley, M. and Gaffan, D. (1997). Impairment of visual object-discrimination learning after perirhinal cortex ablation. *Behavioral Neuroscience*, **111**:467-475.

Buckley, M. J. and Gaffan, D. (1998). Perirhinal cortex ablation impairs object identification. *Journal of Neuroscience*, **18**:2268-2275.

Buckner, R. L., Raichle, M. E., Miezin, F. M. and Petersen, S. E. (1996). Functional anatomical studies of memory retrieval for auditory words and visual pictures. *Journal of Neuroscience*, **16**:6219-6235.

Buckner, R. and Wheeler, M. (2001). The cognitive neuroscience of remembering. *Nature Neuroscience Reviews*, **2**:624-634.

Buffalo, E., Reber, P. and Squire, L. (1998). The human perirhinal cortex and recognition memory. *Hippocampus*, **8**:330-339.

Butler, N. a. H., W. (1996). *Comparative Vertebrate Neuroanatomy*, Wiley-Liss Inc., New York.

Buzsaki, G. (1989). Two-stage model of memory trace formation: a role for "noisy" brain states. *Neuroscience*, **31**:551-570.

Calvin, W., Ojemann, G. and Ward, A. (1973). Human cortical neurons in epileptogenic foci: comparison of inter-ictal firing patterns to those of "epileptic" neurons in animals. *Electroencephalography and Clinical Neurophysiology*, **34**:337-351.

Campbell, F. W. and Howell, E. R. (1972). Monocular alternation: A method for the investigation of pattern vision. *Journal of Physiology, London*, **225**:19-21.

Celebrini, S. and Newsome, W. (1994). Neuronal and psychophysical sensitivity to motion signals in extrastriate area MST of the macaque monkey. *Journal of Neuroscience*, **14**:4109-4124.

Chamberlain, J. (2001). A biologist looks at the study of consciousness. *BioEssays*, **23**:297-298.

Chapin, J. K., Moxon, K. A., Markowitz, R. S. and Nicolelis, M. A. L. (1999). Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex. *Nature Neuroscience*, **2**:664-670.

Charlot, V., Tzourio, N., Zilbovicius, M., Mazoyer, B. and Denis, M. (1992). Different mental imagery abilities result in different regional cerebral blood flow activation patterns during cognitive tasks. *Neuropsychologia*, **30**:565-580.

Chen, R., Gerloff, C., Classen, J., Wassermann, E. M., Hallett, M. and Cohen, L. G. (1997). Safety of different inter-train intervals for repetitive transcranial magnetic

stimulation and recommendations for safe ranges of stimulation parameters. *Electroencephalography and Clinical Neurophysiology*, **105**:415-421.

Chen, W., Kato, T., Zhu, X.-H., Ogawa, S., Tank, D. and Ugurbil, K. (1998). Human primary visual cortex and lateral geniculate nucleus activation during visual imagery. *Neuroreport*, **9**:3669-3674.

Cheng, K., Saleem, K. S. and Tanaka, K. (1997). Organization of Corticostriatal and Corticoamygdalar Projections Arising from the Anterior Inferotemporal Area TE of the Macaque Monkey: A *Phaseolus vulgaris* Leucoagglutinin Study. *Journal of Neuroscience*, **17**:7902-7925.

Cohen, J. and Stewart, I. (2001). Where are the dolphins? *Nature*, **409**:1119-1122.

Cohen, L. G., Celnik, P. and Hallett, M. (1997). Functional relevance of cross-modal plasticity in blind humans. *Nature*, **389**:180-183.

Corkin, S., Amaral, D., Gonzalez, R., Johnson, K. and Hyman, B. (1997). HM's medial temporal lobe lesion: Findings from magnetic resonance imaging. *Journal of Neuroscience*, **17**:3964-3979.

Coslett, H. B. (1997). Neglect in vision and visual imagery: a double dissociation. *Brain*, **120**:1163-1171.

Cotterill, R. (1998). *Enchanted looms. Conscious networks in brains and computers.*, Cambridge University Press, New York.

Crick, F. (1994). *The astonishing hypothesis*, Simon & Schuster, New York.

Crick, F. and Koch, C. (1990). Towards a neurobiological theory of consciousness. *Seminars in the Neurosciences.*, **2**:263-275.

Crick, F. and Koch, C. (1995). Are we aware of neural activity in primary visual cortex ? *Nature*, **375**:121-123.

Crick, F. and Koch, C. (1998). Consciousness and neuroscience. *Cerebral Cortex*, **8**:97-107.

Crick, F. and Mitchison, G. (1983). The function of dream sleep. *Nature*, **304**:111-114.

Csicsvari, J., Hirase, H., Czurko, A., Mamiya, A. and Buzsaki, G. (1999). Oscillatory coupling of hippocampal pyramidal cells and interneurons in the behaving rat. *Journal of Neuroscience*, **19**:274-287.

Csicsvari, J., Hirase, H., Mamiya, A. and Buzsaki, G. (2000). Ensemble patterns of hippocampal CA3-CA1 neurons during sharp wave-associated population events. *Neuron*, **28**:585-594.

Damasio, A. (1990). Category-related recognition defects as a clue to the neural substrates of knowledge. *Trends in Neurosciences*, **13**:95-98.

Damasio, A. and Damasio, H. (1992). Brain and language. *Scientific American*, **267**:89-95.

Dan, Y., Alonso, J. M., Usrey, W. M. and Reid, R. C. (1998). Coding of visual information by precisely correlated spikes in the lateral geniculate nucleus. *Nature Neuroscience*, **1**:501-507.

Dave, A., Yu, A. and Margoliash, D. (1998). Behavioral state modulation of auditory activity in a vocal motor system. *Science*, **282**:2250-2254.

Davis, J. (2000). Comparison of Three Spike-Sorting Algorithms with Simulated Data. Thesis, Computation and Neural Systems, Caltech, Pasadena.

Davis, K. D., Hutchison, W., Lozano, A., Tasker, R. and Dostrovsky, J. (2000). Human Anterior Cingulate Cortex Neurons Modulated by Attention-Demanding Tasks. *Journal of Neurophysiology*, **83**:3575-3577.

deCharms, R. and Zador, A. (2000). Neural representation and the cortical code. *Annual Review of Neuroscience*, **23**:613-+.

Dement, W. and Kleitman, N. (1957a). Cyclic variations in EEG during sleep and their relations to eye movements, body motility and dreaming. *EEG Clinical Neurophysiology*, **9**:673-690.

Dement, W. and Kleitman, N. (1957b). The relation of eye movements during sleep to dream activity: An objective method for the study of dreaming. *Journal of Experimental Psychology*, **53**:339-346.

Descartes, R. (1994). [1637] *Discurso del metodo [Discours de la methode pour bien conduire la raison et chercher la verite dans les sciences]*, RBA Editores, Barcelona.

Desimone, R. (1991). Face-selective cells in the temporal cortex of monkeys. *Journal of Cognitive Neuroscience*, **3**:1-8.

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

D'Esposito, M., Detre, J. A., Aguirre, G. K., Stallcup, M., Alsop, D. C., Tippet, L. J. and Farah, M. J. (1997). A fMRI study of mental image generation. *Neuropsychologia*, **35**:725-730.

Diaz-Caneja, E. (1928). Sur l'alternance binoculaire [On binocular alternation]. *Annales d'Oculistique*, **165**:721-731.

DiCarlo, J. and Maunsell, H. (2000). Form representation in monkey inferotemporal cortex is virtually unaltered by free viewing. *Nature Neuroscienc*, **3**:814-821.

Dierckx, P. (1993). *Curve and surface fitting with splines*, Clarendon, Oxford.

Ditchburn, R. and Protchard, R. (1960). Binocular vision with two stabilized retinal images. *Quarterly Journal Experimental Psychology*, **12**:26-32.

Dixon, M., Bub, D. and Arguin, M. (1998). Semantic and visual determinants of face recognition in a prosopagnosic patient. *Journal of Cognitive Neuroscience*, **10**:362-376.

Duda and Hart (1973). *Pattern classification and scene analysis*, Wiley-Interscience.

Duncan, J., Seitz, R., Kolodny, J., Bor, D., Herzog, H., Ahmed, A., Newell, F. and Emslie, H. (2000). A Neural Basis for General Intelligence. *Science*, 457-460.

Dutour, E. (1760). Discussion d'une question d'optique [Discussion on a question of optics]. *l'Academie des Sciences. Memoires de Mathematique et de physique presentes par Divers Savants*, **3**:514-530.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*, Chapman & Hall/CRC, London.

Eichenbaum, H. (1996). Is the rodent hippocampus just for 'place'? *Current Opinion in Neurobiology*, **6**:187-195.

Eichenbaum, H. (1997). How doe the brain organize memories? *Science*, **277**:330-332.

Ekman, P. (1976). *Pictures of facial affect*, Consulting Psychologists Press, Palo Alto, CA.

Engel, A. and Singer, W. (2001). Temporal binding and the neural correlates of sensory awareness. *Trends in Cognitive Sciences*, **5**:16-25.

Engel, J. (1996). Surgery for seizures. *New England Journal of Medicine*, **334**:647-652.

Epilepsy Foundation (1999). Epilepsy: a report to the nation. *Epilepsy USA*, **Nov. 1999**.

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

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

Fahle, M. (1982). Binocular Rivalry: Suppression depends on orientation and spatial frequency. *Vision Research*, **22**:787-800.

Farah, M., Soso, M. and Dasheiff, R. (1992). Visual angle of the mind's eye before and after unilateral occipital lobectomy. *Journal of Experimental Psychology: Human Perception and Performance*, **18**:241-246.

Farah, M. J. (1988). Is visual imagery really visual? Overlooked evidence from neuropsychology. *Psychological Reviews*, **95**:307-317.

Fee, M. (2000). Active stabilization of electrodes for intracellular recording in awake behaving animals. *Neuron*, **27**:461-468.

Fee, M., Mitra, P. and Kleinfeld, D. (1996). Automatic sorting of multiple unit neuronal signals in the presence of anisotropic and non-Gaussian variability. *Journal of Neuroscience Methods*, **69**:175-188.

Felleman, D. J. and Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, **1**:1-47.

ffytche, D., Howard, R., Brammer, M., David, A., Woodruff, P. and Williams, S. (1998). The anatomy of conscious vision: an fMRI study of visual hallucinations. *Nature Neuroscience*, **1**:738-742.

Finger, S. (2000). *Minds behind the brain. A history of the pioneers and their discoveries*, Oxford University Press, New York.

Finke, R. (1985). Theories Relating Mental Imagery to Perception. *Psychological Bulletin*, **98**:236-259.

Fitzpatrick, D. C., Batra, R., Stanford, T. R. and Kuwada, S. (1997). A neuronal population code for sound localization. *Nature*, **388**:871-874.

Fletcher, P. C., Frith, C. D., Baker, S. C., Shallice, T., Frackowiak, R. S. J. and Dolan, R. J. (1995). The mind's eye - Precuneus activation in memory-related imagery. *Neuroimage*, **2**:195-200.

Forde, E. and Humphreys, G. (1999). Category-specific recognition impairments: a review of important case studies and influential theories. *Aphasiology*, **13**:169-193.

Freedman, D., Riesenhuber, M., Poggio, T. and Miller, E. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science*, **291**:312-316.

Freud, S. (1955). [1900] *The interpretation of dreams*, Oxford University Press, Oxford.

Freud, S. (1966). [1895] *Aus den Anfängen der Psychoanalyse*. [electronic source]

Fried, I., Kreiman, G. and Koch, C. (2000). Imagery neurons in the human brain: A proposed substrate for visual experiential phenomena in epilepsy. In *American Epilepsy Society, Annual Meeting*, Vol. 41 *Epilepsia*, Los Angeles.

Fried, I., MacDonald, K. A. and Wilson, C. (1997). Single neuron activity in human hippocampus and amygdala during recognition of faces and objects. *Neuron*, **18**:753-765.

Fried, I., Mateer, C., Ojemann, G., Wohms, R. and Fedio, P. (1982). Organization of visuospatial functions in human cortex. *Brain*, **105**:349-371.

Fried, I., Wilson, C. L., Maidment, N. T., Engel, J., Behnke, E., Fields, T. A., MacDonald, K. A., Morrow, J. M. and Ackerson, L. (1999). Cerebral microdialysis combined with single-neuron and electroencephalographic recording in neurosurgical patients. *Journal of Neurosurgery*, **91**:697-705.

Fries, P., Roelfsema, P. R., Engel, A. K., Konig, P. and Singer, W. (1997). Synchronization of oscillatory responses in visual cortex correlates with perception in interocular rivalry. *Proceedings of the National Academy of Sciences U.S.A.*, **94**:12699-12704.

Frith, C. and Dolan, R. J. (1997). Brain mechanisms associated with top-down processes in perception. *Philosophical Transactions of the Royal Society of London*, **352**:1221-1230.

Fujita, I., Tanaka, K., Ito, M. and Cheng, K. (1992). Columns for visual features of objects in monkey inferotemporal cortex. *Nature*, **360**:343-346.

Gabbiani, F. and Koch, C. (1998). In *Methods in Neuronal Modeling* (Eds, Koch, C. and Segev, I.). MIT Press, Cambridge.

Gabbiani, F., Metzner, W., Wessel, R. and Koch, C. (1996). From stimulus encoding to feature extraction in weakly electric fish. *Nature*, **384**:564-567.

Gaffan, D. (1994). Dissociated effects of perirhinal cortex ablation, fornix transection and amygdalotomy - evidence for multiple memory systems in the primate temporal lobe. *Experimental Brain Research*, **99**:411-422.

Gainotti, G. and Silveri, M. (1996). Cognitive and anatomical locus of lesion in a patient with a category-specific semantic impairment for living beings. *Cognitive Neuropsychology*, **13**:357-&.

Gais, S., Plihal, W., Wagner, U. and Born, J. (2000). Early sleep triggers memory for early visual discrimination skills. *Nature Neuroscience*, **3**:1335-1339.

Gauthier, I., Behrmann, M. and Tarr, M. (1999). Can face recognition really be dissociated from object recognition? *Journal of Cognitive Neuroscience*, **11**:349-370.

Gauthier, I. and Logothetis, N. (2000). Is face recognition not so unique after all? *Cognitive Neuropsychology*, **17**:125-142.

Gauthier, I. and Nelson, C. (2001). The development of face expertise. *Current Opinion in Neurobiology*, **11**:219-224.

Gegenfurtner, K., Kiper, D. and Levitt, J. (1997). Functional properties of neurons in macaque area V3. *Journal of Neurophysiology*, **77**:1906-1923.

Ghaem, O., Mellet, E., Crivello, F., Tzourio, N., Mazoyer, B., Berthoz, A. and Denis, M. (1997). Mental navigation along memorized routes activates the hippocampus, precuneus, and insula. *Neuroreport*, **8**:739-744.

Ghose, G. and Tso, D. (1997). Form processing modules in primate area V4. *Journal of Neurophysiology*, **77**:2191-2196.

Goebel, R., Khorram-Sefat, D., Muckli, L., Hacker, H. and Singer, W. (1998). The constructive nature of vision: direct evidence from functional magnetic resonance imaging studies of apparent motion and motion imagery. *European Journal of Neuroscience*, **10**:1563-1573.

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

Gorno-Tempini, M., Ciolotti, L. and Price, C. (2000). Category differences in brain activation studies: where do they come from? *Proceedings of the Royal Society of London Series B*, **267**:1253-1258.

Graziano, M., Hu, X. and Gross, C. (1997). Coding the locations of objects in the dark. *Science*, **277**:239-241.

Green, D. and Swets, J. (1966). *Signal detection theory and psychophysics*, Wiley, New York.

Gross, C. (2000). Coding for visual categories in the human brain. *Nature Neuroscience*, **3**:855-856.

Gross, C. G. (1994). How inferior temporal cortex became a visual area. *Cerebral cortex*, **5**:455-469.

Halgren, E., Walter, R. D., Cherlow, D. G. and Crandall, P. H. (1978). Mental phenomena evoked by electrical stimulation of the human hippocampal formation and amygdala. *Brain*, **101**:83-117.

Hamilton, R. and Pascual-Leone, A. (1998). Cortical plasticity associated with Braille learning. *Trends in Cognitive Sciences*, **2**:168-174.

Harris, K. D., Henze, D. A., Csicsvari, J., Hirase, H. and Buzsaki, G. (2000). Accuracy of Tetrode Spike Separation as Determined by Simultaneous Intracellular and Extracellular Measurements. *Journal of Neurophysiology*, **84**:401-414.

Hart, J. and Gordon, B. (1992). Neural subsystems for object knowledge. *Nature*, **359**:60-64.

Hatsopoulos, N., Gabbiani, F. and Laurent, G. (1995). Elementary computation of object approach by a wide-field visual neuron. *Science*, **270**:1000-1003.

Haxby, J., Hoffman, E. and Gobbini, M. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, **4**:223-233.

Hedge, J. and Van Essen, D. (2000). Selectivity for Complex Shapes in Primate Visual Area V2. *Journal of Neuroscience*, **20:RC61**:1-6.

Heilman, K. M. and Valenstein, E. (1993). *Clinical Neuropsychology*, Oxford University Press, New York.

Heit, G., Smith, M. E. and Halgren, E. (1988). Neural encoding of individual words and faces by the human hippocampus and amygdala. *Nature*, **333**:773-775.

Henze, D. A., Borhegyi, Z., Csicsvari, J., Mamiya, A., Harris, K. D. and Buzsaki, G. (2000). Intracellular Features Predicted by Extracellular Recordings in the Hippocampus In Vivo. *Journal of Neurophysiology*, **84**:390-400.

Hobson, J. (1995). *Sleep*, Scientific American Library, New York.

Hobson, J. A. (1988). *The dreaming brain*, Basic Books, Inc. Publishers, New York.

Hobson, J. A., Pace-Schott, E. F., Stickgold, R. and Kahn, D. (1998). To dream or not to dream? Relevant data from new neuroimaging and electrophysiological studies. *Current Opinion in Neurobiology*, **8**:239-244.

Hofle, N., Paus, T., Reutens, D., Fiset, P., Gotman, J., Evans, A. C. and Jones, B. E. (1997). Regional cerebral blood flow changes as a function of delta and spindle activity during slow wave sleep in humans. *Journal of Neuroscience*, **17**:4800-4808.

Holmes, M. D., Ojemann, G. A. and Lettich, E. (1996). Neuronal activity in human right lateral temporal cortex related to visuospatial memory and perception. *Brain Research*, **711**:44-49.

Holt, G. R. and Koch, C. (1999). Electrical interactions via the extracellular potential near cell bodies. *Journal of Computational Neuroscience*, **6**:169-184.

Holt, G. R., Softky, W. R., Koch, C. and Douglas, R. J. (1996). Comparison of discharge variability in vitro and in vivo in cat visual cortex neurons. *Journal of Neurophysiology*, **75**:1806-1814.

Honderich, T. (1995). *The Oxford companion to philosophy*, Oxford University Press, Oxford.

Hopfield, J. J. (1995). Pattern recognition computation using action potential timing for stimulus representation. *Nature*, **376**:33-36.

Hopfield, J. J., Feinstein, D. I. and Palmer, R. G. (1983). 'Unlearning' has a stabilizing effect in collective memories. *Nature*, **304**:158-159.

Horne, J. (1988). *Why we sleep: The functions of sleep in humans and other mammals*, Oxford University Press, Oxford.

Howard III, M., Volkov, I., Granner, M., Damasio, H., Ollendieck, M. and Bakken, H. (1996). A hybrid clinical-research depth electrode for acute and chronic in vivo microelectrode recording of human brain neurons. *Journal of Neurosurgery*, **84**:129-132.

Howard, R., ffytche, D., Barnes, J., McKeefry, D., Ha, Y., Woodruff, P., Bullmore, E., Simmons, A., Williams, S., David, A. and Brammer, M. (1998). The functional anatomy of imagining and perceiving colour. *Neuroreport*, **9**:1019-1023.

Hubel, D. and Wiesel, T. (1959). Receptive fields of single neurons in the cat's striate cortex. *Journal of Physiology (London)*, **148**:574-591.

Humphreys, G. (1996). Object recognition: The man who mistook his dog for a cat. *Current Biology*, **6**:821-824.

Humphreys, G. and Riddoch, M. (1993). Object agnosias. *Bailliere's Clinical Neurology*, **2**:339-359.

Hutchison, W., Davis, K., Lozano, A., Tasker, R. and Dostrovsky, J. (1999). Pain-related neurons in the human cingulate cortex. *Nature Neuroscience*, **2**:403-405.

Insausti, R., Herrero, M. and Witter, M. (1997). Entorhinal cortex of the rat: cytoarchitectonic subdivisions and the origin and distribution of cortical efferents. *Hippocampus*, **7**:146-183.

Insausti, R., Tunon, T., Sobreviela, T., Insausti, A. and Gonzalo, L. (1995). The human entorhinal cortex - A cytoarchitectonic analysis. *Journal of Comparative Neurology*, **355**:171-198.

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

Ishai, A. and Sagi, D. (1995). Common mechanisms of visual imagery and perception. *Science*, **268**:1772-1774.

Ishai, A., Ungerleider, L., Martin, A. and Haxby, J. (2000). The representation of objects in the human occipital and temporal cortex. *Journal of Cognitive Neuroscience*, **12**:35-51.

James, W. (1890). *The principles of psychology*, Dover, New York.

Johnston, D. and Wu, S. M. (1995). *Foundations of cellular neurophysiology*, The MIT Press, Cambridge.

Kahana, M., Sekuler, R., Caplan, J., Kirschen, M. and Madsen, J. R. (1999). Human theta oscillations exhibit task dependence during virtual maze navigation. *Nature*, **399**:781-784.

Kamitani, Y. and Shimojo, S. (1999). Manifestation of scotomas created by transcranial magnetic stimulation of human visual cortex. *Nature Neuroscience*, **2**:767-771.

Kandel, E. and Hawkins, R. (1992). The biological basis of learning and individuality. *Scientific American*, **267**:79-86.

Kandel, E. and Squire, L.R. (2000). Neuroscience: breaking down scientific barriers to the study of brain and mind. *Science*, **290**:1113-1131.

Kandel, E., Schwartz, J. and Jessell, T. (2000). *Principles of Neural Science*, McGraw-Hill, New York.

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

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

Kanwisher, N. and Moscovitch, M. (2000). The cognitive neuroscience of face processing: An introduction. *Cognitive Neuropsychology*, **17**:1-11.

Karni, A., Meyer, G., Jezzard, P., Adams, M., Turner, R. and Ungerleider, L. (1995). Functional MRI evidence for adult motor cortex plasticity during motor skill learning. *Nature*, **377**:155-158.

Karni, A., Tanne, D., Rubenstein, B. S., Askenasy, J. J. M. and Sagi, D. (1994). Dependence on REM sleep of overnight improvement of a perceptual skill. *Science*, **265**:679-681.

Kawasaki, H., Adolphs, R., Kaufman, O., Damasio, H., Bakken, H., Howard III, M. and Hori, T. (2000). Responses to emotional visual stimuli recorded in human prefrontal cortex. *Cognitive Neuroscience Annual Meeting*, San Francisco.

Kawasaki, H., Adolphs, R., Kaufman, O., Damasio, H., Damasio, A., Granner, M., Bakken, H., Hori, T. and Howard III, M. (2001). Single-neuron responses to emotional visual stimuli recorded in human ventral prefrontal cortex. *Nature Neuroscience*, **4**:15-16.

Kay, L. M. and Laurent, G. (1999). Odor- and context-dependent modulation of mitral cell activity in behaving rats. *Nature*, **2**:1003-1009.

Keeping, E. S. (1995). *Introduction to Statistical Inference*, Dover, New York.

Keysers, C., Xiao, D. K., Foldiak, P. and Perret, D. I. (2001). The speed of sight. *Journal of Cognitive Neuroscience*, **13**:90-101.

Kluver, H. and Bucy, P. C. (1939). Preliminary analysis of the functions of the temporal lobes in monkeys. *Archives of Neurology and Psychiatry*, **42**:979-1000.

Koch, C. (1999). *Biophysics of Computation*, Oxford University Press, New York.

Koch, C. (In Preparation). *Towards a neuronal theory of consciousness*.

Koch, C. and Crick, F. (2001). The zombie within. *Nature*, **411**:893.

Koch, C. and Laurent, G. (1999). Complexity and the Nervous System. *Science*, **284**:96-98.

Konishi, M. (1991). Deciphering the Brain's Codes. *Neural Computation*, **3**:1-18.

Konishi, M. (1993). Listening with two years. *Scientific American*, **268**:66-72.

Konishi, M. (1995). In *The cognitive neurosciences* (Ed, Gazzaniga, M.) , pp. 269-277.

Kosslyn, S., Alpert, N., Thompson, W., Maljkovic, V., Weise, S., Chabris, C., Hamilton, S., Rauch, S. and Buonanno, F. (1993). Visual mental imagery activates topographically organized visual cortex: PET investigations. *Journal of Cognitive Neuroscience*, **5**:263-287.

Kosslyn, S. and Ochsner, K. (1994). In search of occipital activation during visual mental imagery. *Trends in Neuroscience*, **17**:290-292.

Kosslyn, S., Thompson, W., Kim, I., Rauch, S. and Alpert, N. (1996). Individual differences in cerebral blood flow in area 17 predict the time to evaluate visualized letters. *Journal of Cognitive Neuroscience*, **8**:78-82.

Kosslyn, S. M. (1994). *Image and brain*, MIT Press, Cambridge.

Kosslyn, S. M., Pascual-Leone, A., Felician, O., Camposano, S., Keenan, J. P., Thompson, W. L., Ganis, G., Sukel, K. E. and Alpert, N. M. (1999). The Role of Area 17 in Visual Imagery: Convergent Evidence from PET and rTMS. *Science*, **284**:167-170.

Kosslyn, S. M., Thompson, W. L. and Alpert, N. M. (1997). Neural Systems Shared by Visual Imagery and Visual Perception: A PET Study. *Neuroimage*, **6**:320-334.

Kosslyn, S. M., Thompson, W. L., Kim, I. J. and Alpert, N. M. (1995). Topographical representations of mental images in primary visual cortex. *Nature*, **378**:496-498.

Kosslyn, S., Ganis, G. and Thompson, W. (2001). Neural foundations of imagery. *Nature Neuroscience Reviews*, **2**:635-642.

Kovacs, I., Papatomas, T. V., Yang, M. and Feher, A. (1996). When the brain changes its mind: Interocular grouping during binocular rivalry. *Proceedings of the National Academy of Sciences*, **93**:15508-15511.

Krahe, R., Kreiman, G., Gabbiani, F., Koch, C. and Metzner, W. (*submitted*). Stimulus encoding and feature extraction by multiple pyramidal cells in the hindbrain of weakly electric fish. .

Kreiman, G. (2001). Moveo ergo sum. *BioEssays*, **23**:662.

Kreiman, G. and Koch, C. (1999). Flash Suppression: Competition Between Eyes or Patterns? In *Investigative Ophthalmology and Visual Sciences*, Vol. 40 Fort Lauderdale, **S421**.

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

Kreiman, G., Koch, C. and Fried, I. (2000b). Imagery neurons in the human brain. *Nature*, **408**:357-361.

Kreiman, G., Krahe, R., Metzner, W., Koch, C. and Gabbiani, F. (2000c). Robustness and Variability of Neuronal Coding by Amplitude Sensitive Afferents in the Weakly Electric Fish *Eigenmannia*. *Journal of Neurophysiology*, **84**:189-204.

Kreiman, G., Staba, R., Wilson, C. and Fried, I. (*submitted*). Synchrony between single neurons in the human brain during sleep.

Kryger, M., Roth, T. and Dement, W. C. (1994). *Principles and Practice of Sleep Medicine.*, W.B. Saunders Company, Philadelphia.

LaBerge, S. (1985). *Lucid Dreaming*, Ballantine Books, New York.

LaBerge, S. (1990). *Exploring the world of lucid dreaming*, Ballantine books, New York.

LaBerge, S. P., Dement, W. C., Nagel, L. E. and Zarcone, V. P. (1981). Lucid dreaming verified by volitional communication during rem sleep. *Perceptual and motor skills*, **52**:727-732.

Lack, L. (1978). *Selective attention and the control of binocular rivalry*, The Hague, The Netherlands.

Laurent, G. (1996a). Dynamical representation of odors by oscillating and evolving neural assemblies. *Trends in Neuroscience*, **19**:489-496.

Laurent, G. (1996b). Odor images and tunes. *Neuron*, **16**:473-476.

LeDoux, J. (2000). Emotion circuits in the brain. *Annual Review of Neuroscience*, **23**:155-184.

Lee, S. and Blake, R. (1999). Rival ideas about binocular rivalry. *Vision Research*, **39**:1447-1454.

Lehky, S. R. and Maunsell, J. H. R. (1996). No binocular rivalry in the LGN of alert monkeys. *Vision Research*, **36**:1225-1234.

Lehmann, E. L. (1975). *Nonparametrics: statistical methods based on ranks*, McGraw-Hill, New York.

Leonard, C. M., Rolls, E. T., Wilson, F. A. W. and Baylis, G. C. (1985). Neurons in the amygdala of the monkey with responses selective for faces. *Behavioural Brain Research*, **15**:159-176.

Leopold, D. A. and Logothetis, N. K. (1996). Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. *Nature*, **379**:549-553.

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

Levelt, W. J. M. (1968). *On Binocular Rivalry*, Mouton & Co., Paris.

Levy, R., Hutchison, W., Lozano, A. and Dostrovsky, J. (2000). High-frequency synchronization of neuronal activity in the subthalamic nucleus of parkinsonian patients with limb tremor. *Journal of Neuroscience*, **20**:7766-7775.

Lewicki, M. (1994). Bayesian modeling and classification of neural signals. *Neural Computation*, **6**:1005-1030.

Lewicki, M. S. (1998). A review of methods of spike sorting: the detection and classification of neural action potentials. *Network: Computation and Neural Systems*, **9**:R53-R78.

Li, L., Miller, E. and Desimone, R. (1993). The representation of stimulus familiarity in anterior inferior temporal cortex. *Journal of Neurophysiology*, **69**:1918-1929.

Libet, B. (1982). Brain stimulation in the study of neuronal functions for conscious sensory experiences. *Human Neurobiology*, **1**:235-242.

Lipton, L. and Meyer, L. (1984). A flicker-free field-sequential stereoscopic video system. *SMPTE Journal*, **Nov**:1047.

Lisman, J. (1997). Bursts as a unit of neural information: making unreliable synapses reliable. *Trends in Neuroscience*, **20**:38-43.

Livingstone, M. (1996). Differences between stereopsis, interocular correlation and binocularity. *Vision Research*, **38**:1127-1140.

Livingstone, M. and Hubel, D. (1988). Segregation of form, color, movement and depth: anatomy, physiology and perception. *Science*, **240**:740-749.

Locke, J. (1995). [1689] *An essay concerning human understanding*, ILT Digital Classics, New York.

Logothetis, N., Pauls, J., Augath, M., Trinath, T. and Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, **412**:150-157.

Logothetis, N. K. (1998). Single units and conscious vision. *Philosophical Transactions of the Royal Society of London Series B*, **353**:1801-1818.

Logothetis, N. K., Leopold, D. A. and Sheinberg, D. L. (1996). What is rivalling during binocular rivalry? *Nature*, **380**:621-624.

Logothetis, N. K. and Pauls, J. (1995). Psychophysical and physiological evidence for viewer-centered object representations in the primate. *Cerebral Cortex*, **3**:270-288.

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

Logothetis, N. K. and Schall, J. D. (1989). Neuronal correlates of subjective visual perception. *Science*, **245**:761-763.

Logothetis, N. K. and Sheinberg, D. L. (1996). Visual Object Recognition. *Annual Review of Neuroscience*, **19**:577-621.

Louie, K. and Wilson, M. (2001). Temporally structured replay of awake hippocampal ensemble activity during rapid eye movement sleep. *Neuron*, **29**:145-156.

Lumer, E. D., Friston, K. J. and Rees, G. (1998). Neural Correlates of perceptual rivalry in the human brain. *Science*, **280**:1930-1934.

Lumer, E. D. and Rees, G. (1999). Covariation of activity in visual and prefrontal cortex associated with subjective visual perception. *Proceedings of the National Academy of Sciences U.S.A.*, **96**:1669-1673.

MacLeod, K. and Laurent, G. (1996). Distinct mechanisms for synchronization and temporal patterning of odor-encoding neural assemblies. *Science*, **274**:976-979.

Robert MacNish, *The philosophy of Sleep*, 1834.

Madison, D., Malenka, R. and Nicoll, R. (1991). Mechanisms Underlying Long-Term Potentiation of Synaptic Transmission. *Annual Review of Neuroscience*, **14**:379-397.

Magarinos-Ascone, C., Figueiras-Mendez, R., Riva-Meana, C. and Cordoba-Fernandez, A. (2000). Subthalamic neuron activity related to tremor and movement in Parkinson's disease. *European Journal of Neuroscience*, **12**:2597-2607.

Magee, J. (1997). A synaptically controlled associative signal for hebbian plasticity in hippocampus. *Science*, **257**:209-213.

Maguire, E. A., Burgess, N., Donnett, J. G., Frackowiak, R. S., Frith, C. D. and O'Keefe, J. (1998). Knowing where and getting there: a human navigation network. *Science*, **280**:921-924.

Mainen, Z. F. and Sejnowski, T. J. (1995). Reliability of spike timing in neocortical neurons. *Science*, **268**:1503-1506.

Maquet, P. and Franck, G. (1996). Functional neuroanatomy of human rapid eye movement sleep and dreaming. *Nature*, **383**:163-166.

Maquet, P., Degueldre, C., Delfiore, G., Aerts, G., Peters, J. M., Luxen, A. and Franck, G. (1997). Functional neuroanatomy of human slow wave sleep. *Journal of Neuroscience*, **17**:2807-2812.

Maquet, P., Laureys, S., Peigneux, P., Fuchs, S., Petiau, C., Phillips, C., Aerts, J., Del Fiore, F., C, D., Meulemans, T., Luxen, A., G, F., Van der Linden, M., Smith, C. and Cleeremans, A. (2000). Experience-dependent changes in cerebral activation during human REM sleep. *Nature Neuroscience*, **3**:831-836.

Markram, H., Lubke, J., Frotscher, M. and Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science*, **275**:213-215.

Martin, A., Wiggs, C. L., Ungerleider, L. G. and Haxby, J. V. (1996). Neural correlates of category-specific knowledge. *Nature*, **379**:649-52.

Martin, A. and Chao, L. (2001). Semantic memory and the brain: structure and processes. *Current Opinion in Neurobiology*, **11**:194-201.

Martinez-Conde, S., Macknik, S. L. and Hubel, D. H. (2000). Microsaccadic eye movements and firing of single cells in the striate cortex of macaque monkeys. *Nature Neuroscience*, **3**:251-258.

Mayford, M. and Kandel, E. (1999). Genetic approaches to memory storage. *Trends in Genetics*, **15**:463-470.

McCarthy, G., Puce, A., Belger, A. and Allison, T. (1999). Electrophysiological studies of human face perception. II: Response properties of face-specific potentials generated in occipitotemporal cortex. *Cerebral Cortex*, **9**:431-444.

McCarthy, R. and Warrington, E. (1994). Disorders of semantic memory. *Philosophical Transactions of the Royal Society of London Series B*, **346**:89-96.

McHugh, T. J., Tonegawa, S. and Wilson, M. A. (1996). Impaired hippocampal representation of space in CA1-specific NMDAR1 knockout mice. *Cell*, **87**:133-1349.

McNamara, J. O. (1999). Emerging insights into the genesis of epilepsy. *Nature*, **399**:A15-A22.

McNaughton, B. (1998). The neurophysiology of reminiscence. *Neurobiology of Learning and Memory*, **70**:252-267.

McNaughton, B., O'Keefe, J. and Barnes, C. (1983). The stereotetrode - a new technique for simultaneous isolation of several single units in the central nervous system from multiple unit records. *Journal of Neuroscience Methods*, **8**:391-397.

McNaughton, B. L. and Nadel, L. (1990). In *Neuroscience and Connectionist Theory* (Ed, Seifert, W.). Academic Press, New York, pp. 1-64.

Meenan, J. P. and Miller, L. A. (1994). Perceptual flexibility after frontal or temporal lobectomy. *Neuropsychologia*, **32**:1145-1149.

Mellet, E., Tzourio, N., Crivello, F., Joliot, M., Denis, M. and Mazoyer, B. (1996). Functional anatomy of spatial mental imagery generated from verbal instructions. *Journal of Neuroscience*, **16**:6504-6512.

Mellet, E., Tzourio, N., Denis, M. and Mazoyer, B. (1998). Cortical anatomy of mental imagery of concrete nouns based on their dictionary definition. *Neuroreport*, **9**:803-808.

Metzner, W., Koch, C., Wessel, R. and Gabbiani, F. (1998). Feature extraction by burst-like spike patterns in multiple sensory maps. *Journal of Neuroscience*, **18**:2283-2300.

Meunier, M., Bachevaer, J., Mishkin, M. and Murray, E. A. (1993). Effects on visual recognition of combined and separate ablations of the entorhinal and perirhinal cortex in rhesus monkeys. *Journal of Neuroscience*, **13**:5418-5432.

Meunier, M., Hadfield, W., Bachevalier, J. and Murray, E. (1996). Effects of rhinal cortex lesions combined with hippocampectomy on visual recognition memory in rhesus monkeys. *Journal of Neurophysiology*, **75**:1190-1205.

Meyer, B., Britton, T., Klotten, H., Steinmetz, H. and Benecke, R. (1991). Coil placement in magnetic brain-stimulation related to skull and brain anatomy. *Electroencephalography and Clinical Neurophysiology*, **81**:38-46.

Milders, M. and Perrett, D. (1993). Recent developments in the neuropsychology and physiology of face processing. *Bailliere's Clinical Neurology*, **2**:361-384.

Miller, E. (2000). The prefrontal cortex and cognitive control. *Nature Reviews Neuroscience*, **1**:59-65.

Miller, S., Liu, G., Ngo, T., Hooper, G., Riek, S., Carson, R. and Pettigrew, J. (2000). Interhemispheric switching mediates perceptual rivalry. *Current Biology*, **10**:383-392.

Milner, B. (1972). Disorders of learning and memory after temporal lobe lesions in man. *Clinical Neurosurgery*, **19**:421-446.

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

Mishkin, M. (1982). A memory system in the monkey. *Philosophical Transaction of the Royal Society of London Series B*, **298**:85.

Miyashita, Y. (1988). Neuronal correlate of visual associative long-term memory in the primate temporal cortex. *Nature*, **335**:817-820.

Miyashita, Y. (1993). Inferior Temporal Cortex: Where visual perception meets memory. *Annual Review of Neuroscience*, **16**:245-263.

Miyashita, Y. (1995). How the brain creates imagery: projection to primary visual cortex. *Science*, **268**:1719-1720.

Miyashita, Y. and Chang, H. S. (1988). Neuronal correlate of pictorial short-term memory in the primate temporal cortex. *Nature*, **331**:68-71.

Miyashita, Y., Rolls, E. T., Cahusac, P. M. B., Niki, H. and Feigenbaum, J. D. (1989). Activity of hippocampal formation neurons in the monkey related to a conditional spatial response task. *Journal of Neurophysiology*, **61**:669-678.

Morris, R., Garrud, P., Rawlins, J. and O'Keefe, J. (1982). Place navigation impaired in rats with hippocampal lesions. *Nature*, **297**:681-683.

Mountcastle, V. (1959). Neural mechanisms subserving cutaneous sensibility, with special reference to the role of afferent inhibition in sensory perception and discrimination. *Bulletin Johns Hopkins Hospital*, **105**:201-232.

Mountcastle, V. (1995). The parietal system and some higher brain functions. *Cerebral cortex*, **5**:377-390.

Movshon, J. A. and Newsome, W. T. (1992). Neural Foundations of Visual Motion Perception. *Current Directions in Psychological Science*, **1**:35-39.

Muller, R. (1996). A quarter of a century of place cells. *Neuron*, **17**:813-822.

Murata, A., Fadiga, L., Fogassi, L., Gallese, V., Raos, V. and Rizzolatti, G. (1997). Object representation in the ventral premotor cortex (area F5). of the monkey. *Journal of Neurophysiology*, **78**:

Murray, E. (1996). What have ablation studies told us about the neural substrates of stimulus memory? *Seminars in the Neurosciences*, **8**:13-22.

Murray, E., Baxter, M. and Gaffan, D. (1998). Monkeys with rhinal cortex damage or neurotoxic hippocampal lesions are impaired on spatial scene learning and object reversals. *Behavioral Neuroscience*, **112**:1291-1303.

Murray, E. and Mishkin, M. (1998). Object recognition and location memory in monkeys with excitotoxic lesions of the amygdala and hippocampus. *Journal of Neuroscience*, **18**:6568-6582.

Mushiake, H., Inase, M. and Tanji, J. (1991). Neuronal activity in the primate premotor, supplementary and precentral motor cortex during visually guided and internally determined sequential movements. *Journal of Neurophysiology*, **66**:705-718.

Myerson, J., Miezin, F. and Allman, J. (1981). Binocular rivalry in macaque monkeys and humans: a comparative study in perception. *Behavioral Analysis Letters*, **1**:149-159.

Nachson, I. (1995). On the modularity of face recognition: the riddle of domain specificity. *Journal of Clinical and Experimental Neuropsychology*, **17**:256-275.

Nadasdy, Z., Hirase, H., Czurko, A., Csicsvari, J. and Buzsaki, G. (1999). Replay and time compression of recurring spike sequences in the hippocampus. *Journal of Neuroscience*, **19**:9497-9507.

Nakamura, K., Mikami, A. and Kuota, K. (1992). Activity of single neurons in the monkey amygdala during performance of a visual-discrimination task. *Journal of Neurophysiology*, **67**:1447-1463.

Nakayama, K., He, Z. and Shimojo, S. (1995). In *Visual cognition*, Vol. 2 (Eds, Kosslyn, S. and Osherson, D.). The MIT press, Cambridge.

Newsome, W., Britten, K. and Movshon, J. (1989). Neuronal correlates of a perceptual decision. *Nature*, **341**:52-54.

Nicolelis, M., Ghazanfar, A., Faggin, B., Votaw, S. and Oliveira, L. (1997). Reconstructing the engram: Simultaneous, multisite, many single neuron recordings. *Neuron*, **18**:529-537.

Nicolelis, M. A. L., Stambaugh, C. R., Brisben, A. and Laubach, M. (1999). In *Methods for Neural Ensemble Recordings* (Ed, Nicolelis, M. A. L.). CRC Press.

Nishijo, H., Ono, T. and Nishino, H. (1988). Single neuron responses in amygdala of alert monkeys during complex sensory stimulation with affective significance. *Journal of Neuroscience*, **8**:3570-3583.

Nolte, J. (1998). *The human brain: an introduction to its functional anatomy*, Mosby, New York.

O'Craven, K. and Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *Journal of Cognitive Neuroscience*, **12**:1013-1023.

O'Craven, K. M., Rosen, B. R., Kwong, K. K., Treisman, A. and Savoy, R. L. (1997). Voluntary attention modulates fMRI activity in human MT-MST. *Neuron*, **18**:591-598.

Ojemann, G. and Mateer, C. (1979). Human language cortex: localization of memory, syntax, and sequential motor-phoneme identification systems. *Science*, **205**:1401-1403.

Ojemann, G. A. (1997). Treatment of temporal lobe epilepsy. *Annual Review of Medicine*, **48**:317-328.

Ojemann, G. A. and J., S.-M. (1999). Activity of neurons in human temporal cortex during identification and memory for names and words. *Journal of Neuroscience*, **19**:5674-5682.

Ojemann, G. A., Ojemann, S. G. and Fried, I. (1998). Lessons from the human brain: neuronal activity related to cognition. *The Neuroscientist*, **4**:285-300.

Ojemann, J. G., Ojemann, G. A. and Lettich, E. (1992). Neuronal activity related to faces and matching in human right nondominant temporal cortex. *Brain*, **115**:1-13.

O'Keefe, J. and Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, **34**:171-175.

Ono, H. and Barbeito, R. (1985). Utrocular discrimination is not sufficient for utrocular identification. *Vision Research*, **25**:289-299.

O'Regan, J., Rensink, R. and Clark, J. (1999). Change-blindness as a result of 'mudsplashes'. *Nature*, **398**:34.

Ouyang, Y., Kreiman, G., Kantor, D. B., Rosenstein, A. J., Schuman, E. M. and Kennedy, M. B. (1997). Tetanic stimulation produces a similar increase in both nonphosphorylated and autophosphorylated CAMKII in dendrites of hippocampal CA1 neurons. In *Annual Meeting of the Society for Neuroscience*, Vol. 23 New Orleans, LA, pp. 708.

Ouyang, Y., Rosenstein, A., Kreiman, G., Schuman, E. M. and Kennedy, M. B. (1999). Tetanic stimulation leads to increased accumulation of Ca²⁺ calmodulin-dependent protein kinase II via dendritic protein synthesis in hippocampal neurons. *Journal of Neuroscience*, **19**:7823-7833.

Palm, G., Aersten, A. and Gerstein, G. (1988). On the significance of correlations among neuronal spike trains. *Biological Cybernetics*, **59**:1-11.

Pascual-Leone, A., Tormos, J. M., Keenan, J., Tarazona, F., Canete, C. and Catala, M. D. (1998). Study and Modulation of Human Cortical Excitability With Transcranial Magnetic Stimulation. *Journal of Clinical Neurophysiology*, **15**:333-343.

Pascual-Leone, A., Walsh, V. and Rothwell, J. (2000). Transcranial magnetic stimulation in cognitive neuroscience - virtual lesion, chronometry, and functional connectivity. *Current Opinion in Neurobiology*, **10**:232-237.

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

Paus, T. (2001). Primate anterior cingulate cortex: where motor control drive and cognition interface. *Nature Reviews Neuroscience*, **2**:417-424.

Pavrides, C. and Winson, J. (1989). Influences of hippocampal place cell firing in the awake state on the activity of these cells during subsequent sleep episodes. *Journal of Neuroscience*, **9**:2907-2918.

Pelli, D. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, **10**:437-442.

Penfield, W. (1937). The cerebral cortex in man. I. The cerebral cortex and consciousness. *Archives of Neurology and Psychiatry*, **40**:417-442.

Penfield, W. and Jasper, H. (1954). *Epilepsy and the functional anatomy of the human brain*, Little, Brown and Company, Boston.

Penfield, W. and Milner, B. (1958). Memory deficit produced by bilateral lesions in the hippocampal zone. *Archives in Neurology and Psychology*, **79**:475-479.

Penfield, W. and Perot, P. (1963). The brain's record of auditory and visual experience. A final summary and discussion. *Brain*, **86**:595-696.

Perkel, D., Gerstein, G. and Moore, G. (1967). Neuronal spike trains and stochastic point processes II: Simultaneous spike trains. *Biophysical Journal*, **7**:419-440.

Perky, C. W. (1910). An experimental study of imagination. *American Journal of Psychology*, **21**:422-452.

Perrett, D., Rolls, E. and Caan, W. (1982). Visual neurones responsive to faces in the monkey temporal cortex. *Experimental Brain Research*, **47**:329-342.

Peterhans, E. and von der Heydt, R. (1991). Subjective contours - bridging the gap between psychophysics and physiology. *Trends in Neuroscience*, **14**:112-119.

Pettigrew, J. and Miller, S. (1998). A 'sticky' interhemispheric switch in bipolar disorder? *Proceedings of the Royal Society of London Series B*, **265**:2141-2148.

Pikkarainen, M., Ronkko, S., Savander, V., Insausti, R. and Pitkanen, A. (1999). Projections from the lateral, basal and accessory basal nuclei of the amygdala to the hippocampal formation in rat. *Journal of Comparative Neurology*, **403**:229-260.

Pinker, S. (1997). Words and rules in the human brain. *Nature*, **387**:547-548.

Pinker, S. (2000). *The language instinct: How the mind creates language*, HarperTrade, New York.

Plato (1999). [360 BCE] *Timaeus*, [classics.mit.edu].

Poe, G., Nitz, D., McNaughton, B. and Barnes, C. (2000). Experience-dependent phase-reversal of hippocampal neuron firing during REM sleep. *Brain Research*, **855**:176-180.

Poggio, T. (1990). A theory of how the brain might work. *Cold Spring Harbor Symposia on Quantitative Biology*, **55**:899-910.

Polonsky, A., Blake, R., Braun, J. and Heeger, D. (2000). Neuronal activity in human primary visual cortex correlates with perception during binocular rivalry. *Nature Neuroscience*, **3**:1153-1159.

Porac, C. and Stanley, C. (1986). Sighting dominance and utrocular discrimination. *Perception and Psychophysics*, **39**:449-451.

Porro, C., Francescato, M., Cettolo, V., Diamond, M., Baraldi, P., Zuiani, C., Bazzocchi, M. and di Prampero, P. (1996). Primary motor and sensory cortex activation during motor performance and motor imagery: A functional magnetic resonance imaging study. *Journal of Neuroscience*, **16**:7688-7698.

Potter, M. and Levy, E. (1969). Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology*, **81**:10-15.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. (1996). *Numerical Recipes in C*, Cambridge University Press, Cambridge.

Puce, A., Allison, T. and McCarthy, G. (1999). Electrophysiological studies of human face perception. III: Effects of top-down processing on face-specific potentials. *Cerebral Cortex*, **9**:445-458.

Rainer, G., Rao, S. and Miller, E. (1999). Prospective coding for objects in primate prefrontal cortex. *Journal of Neuroscience*, **19**:5493-5505.

Ramon y Cajal, S. (1968). *The structure of Ammon's horn*, Charles C. Thomas, Illinois.

Rao, S., Williams, G. and Goldman-Rakic, P. (1999). Isodirectional Tuning of Adjacent Interneurons and Pyramidal Cells During Working Memory: Evidence for Microcolumnar Organization in PFC. *Journal of Neurophysiology*, **81**:1903-1916.

Rayport, M. and Waller, H. (1967). Technique and Results of Micro-Electrode Recordign in Human Epileptogenic Foci. *Electroencephalography and Clinical Neurophysiology*, **25**:143-151.

Rechtschaffen, A., Gilliland, M. A., Bergmann, B. M. and Winter, J. B. (1983). Physiological correlates of prolonged sleep deprivation effects in rats. *Science*, **221**:182-184.

Rechtschaffen, A. and Kales, A. (1968). *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects.*, NIH Publication 204, Washington, D.C.

Reinagel, P., Godwin, D., Sherman, S. and Koch, C. (1999). Encoding of visual information by LGN bursts. *Journal of Neurophysiology*, **81**:2558-2569.

Ricci, C. and Blundo, C. (1990). Perception of ambiguous figures after focal brain lesions. *Neuropsychologia*, **28**:1163-1173.

Riches, I., Wilson, F. and Brown, M. (1991). The effects of visual stimulation and memory on neurons of the hippocampal formation and the neighboring parahippocampal gyrus and inferior temporal cortex of the primate. *Journal of Neuroscience*, **11**:1763-1779.

Richmond, B. J., Optican, L. M. and Spitzer, H. (1990). Temporal encoding of two-dimensional patterns by single units in primate primary visual cortex. I. Stimulus-response relations. *Journal of Neurophysiology*, **64**:351-369.

Rieke, F., Warland, D., van Steveninck, R. and Bialek, W. (1997). *Spikes*, The MIT Press, Cambridge, Massachusetts.

Ringo, J., Sobotka, S., Diltz, M. and Bunce, C. (1994). Eye Movements Modulate Activity in Hippocampal, Parahippocampal and Inferotemporal Neurons. *Journal of Neurophysiology*, **71**:1285-1288.

Rizzolatti, G., Fadiga, L., Fogassi, V. and Gallese, V. (1997). The space around us. *Science*, **277**:190-191.

Roland, P. and Gulyas, B. (1995). Visual Memory, Visual Imagery and Visual Recognition of Large Field Patterns in the Human Brain: Functional Anatomy by Positron Emission Tomography. *Cerebral Cortex*, **1**:79-93.

Roland, P. E. and Gulyas, B. (1994). Visual imagery and visual representation. *Trends in Neuroscience*, **17**:281-287.

Rolls, E., Cahusac, P., Feigenbaum, K. and Miyashita, Y. (1993). Responses of single neurons in the hippocampus of the macaque related to recognition memory. *Experimental Brain Research*, **93**:299-306.

Rolls, E., Judge, S. and Sanghera, M. (1977). Activity of neurones in the inferotemporal cortex of the alert monkey. *Brain Research*, **130**:229-238.

Rolls, E., Perrett, D. and Wilson, F. (1982). Neuronal responses related to visual recognition. *Brain*, **105**:611-646.

Rolls, E., Sanghera, M. and Roper-Hall, A. (1979). The latency of activation of neurones in the lateral hypothalamus and substantia innominata during feeding in the monkey. *Brain Research*, **164**:121-135.

Rolls, E. T. (1984). Neurons in the cortex of the temporal lobe and in the amygdala of the monkey with responses selective for faces. *Human Neurobiology*, **3**:209-222.

Rolls, E. T., Miyashita, Y., Cahusac, P. M. B., Kesner, R. P., Niki, H., Feigenbaum, J. D. and Bach, L. (1989). Hippocampal neurons in the monkey with activity related to the place in which a stimulus is shown. *Journal of Neuroscience*, **9**:1835-1845.

Ross, E. D. (1980). Sensory-Specific and Fractional Disorders of Recent Memory in Man. *Archives of Neurology*, **37**:193-200.

Rossi, G., Walter, R. and Crandall, P. (1968). Generalized spike and wave discharges and nonspecific thalamic nuclei. *Archives of Neurology*, **19**:174-183.

Roth, B., Saypol, J., Hallett, M. and Cohen, L. (1991). A theoretical calculation of the electric-field induced in the cortex during magnetic stimulation. *Electroencephalography and Clinical Neurophysiology*, **81**:47-56.

Ruohonen, J. (1998). Transcranial Magnetic Stimulation: Modelling and Techniques. Thesis, Department of Engineering Physics and Mathematics, Helsinki University of Technology, Espoo, Finland.

Sacks, O. (1998). *The man who mistook his wife for a hat*, Touchstone Books, New York.

Sahani, M. (1999). Latent variable models for neural data analysis. Thesis, Computation and Neural Systems, Caltech, Pasadena.

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

Sakai, K. and Miyashita, Y. (1994a). The memory storehouse. *Trends In Neuroscience*, **17**:513-514.

Sakai, K. and Miyashita, Y. (1994b). Visual imagery: an interaction between memory retrieval and focal attention. *Trends in Neuroscience*, **17**:287-289.

Saleem, K. S. and Tanaka, K. (1996). Divergent projections from the anterior inferotemporal area TE to the perirhinal and entorhinal cortices in the macaque monkey. *Journal of Neuroscience*, **16**:4757-4775.

Salzman, C., Britten, K. and Newsome, W. (1990). Cortical microstimulation influences perceptual judgments of motion idirection. *Nature*, **346**:174-177.

Salzman, C. D., Murasugi, C. M., Britten, K. H. and Newsome, W. T. (1992). Microstimulation in visual area MT: Effects on direction discrimination performance. *Journal of Neuroscience*, **12**:2331-2355.

Sary, G., Vogels, R. and Orban, G. A. (1993). Cue-invariant shape selectivity of macaque inferior temporal neurons. *Science*, **260**:995-997.

Scheffe, H. (1953). *The Analysis of Variance*, Wiley, New York.

Schmolesky, M., Wang, Y., Hanes, D., Thompson, K., Leutgeb, S., Schall, J. and Leventhal, A. (1998). Signal timing across the macaque visual system. *Journal of Neurophysiology*, **79**:3272-3278.

Schwartz, E., Desimone, R., Albright, T. and Gross, C. (1983). Shape-recognition and inferior temporal neurons. *Proceedings of the National Academy of Sciences U.S.A.*, **80**:5776-5778.

Scoville, W. B. and Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery and Psychiatry*, **20**:11-21.

Sejnowski, T. and Destexhe, A. (2000). Why do we sleep? *Brain Research*, **886**:208-223.

Shadlen, M. N. and Newsome, W. T. (1994). Noise, neural codes and cortical organization. *Current Opinion in Neurobiology*, **4**:569-579.

Shadlen, M. N. and Newsome, W. T. (1998). The variable discharge of cortical neurons: implications for connectivity, computation and information coding. *Journal of Neuroscience*, **15**:3870-3896.

Sheinberg, D. and Logothetis, N. (2000). The other cells in inferotemporal cortex. In *Society for Neuroscience Annual Meeting*, New Orleans, pp. 106.5.

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

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, **237**:1317-1323.

Sheridan, J. and Humphreys, G. (1993). A verbal-semantic category-specific recognition impairment. *Cognitive Neuropsychology*, **10**:143-184.

Shimojo, S. and Nakayama, K. (1990). Real world occlusion constraints and binocular rivalry. *Vision Research*, **30**:69-80.

Siapas, A. and Wilson, M. (1998). Coordinated interactions between hippocampal ripples and cortical spindles during slow-wave sleep. *Neuron*, **21**:1123-1128.

Silbersweig, D., Stern, E., Frith, C., Cahill, C., Holmes, A., Grootenck, S., Seaward, J., McKenna, P., Chua, S., Schnorr, L., Jones, T. and Frackowiak, R. (1995). A functional neuroanatomy of hallucinations in schizophrenia. *Nature*, **378**:176-179.

Silveri, M. and Gainotti, G. (1988). Interaction between vision and language in category-specific semantic impairment. *Cognitive Neuropsychology*, **5**:677-709.

Singer, W. (1999). Neuronal synchrony: A versatile code for the definition of relations? *Neuron*, **24**:49-65.

Singer, W. and Gray, C. (1995). Visual feature integration and the temporal correlation hypothesis. *Annual Review of Neuroscience*, **18**:555-586.

Skaggs, W. E. and McNaughton, B. L. (1996). Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience. *Science*, **271**:1870-1873.

Smith, E. E. and Jonides, J. (1999). Storage and Executive Processes in the Frontal Lobes. *Science*, **283**:1657-1661.

Sobotka, S., Nowicka, A. and Ringo, J. (1997). Activity Linked to Externally Cued Saccades in Single Units Recorded From Hippocampal, Parahippocampal and Inferotemporal Areas of Macaques. *Journal of Neurophysiology*, **78**:2156-2163.

Softky, W. R. (1995). Simple codes versus efficient codes. *Current Opinion in Neurobiology*, **5**:239-247.

Solms, M. (1996). *The neuropsychology of dreams*, Lawrence Erlbaum Association, New Jersey.

Sperry, R. (1982). Some effects of disconnecting the cerebral hemispheres. *Science*, **217**:1223-1226.

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

Srinivassan, R., Russell, D., Edelman, G. and Tononi, G. (1999). Increased synchronization of neuromagnetic responses during conscious perception. *Journal of Neuroscience*, **19**:5435-5448.

Steinmetz, P., Roy, A., Fitzgerald, P., Hsiao, S., Johnson, K. and Niebur, E. (2000). Attention modulates synchronized neuronal firing in primate somatosensory cortex. *Nature*, **404**:187-190.

Steriade, M. (2001). To burst, or rather, not to burst. *Nature Neuroscience*, **4**:671.

Steriade, M., McCormick, D. A. and Sejnowski, T. J. (1993). Thalamocortical oscillations in the sleeping and aroused brain. *Science*, **262**:679-685.

Sternberg, R. and Kaufman, J. (1998). Human abilities. *Annual Review of Psychology*, **49**:479-502.

Stickgold, R., LaTanya, J. and Hobson, J. (2000). Visual discrimination learning requires sleep after training. *Nature Neuroscience*, **3**:1237-1238.

Stickgold, R., Scott, L., Rittenhouse, C. and Hobson, J. (1999). Sleep-induced changes in associative memory. *Journal of Cognitive Neuroscience*, **11**:182-193.

Stopfer, M., Bhagavan, S., Smith, B. H. and Laurent, G. (1997). Impaired odour discrimination on desynchronization of odour-encoding neural assemblies. *Nature*, **390**:70-74.

Sutherland, G. and McNaughton, B. (2000). Memory trace reactivation in hippocampal and neocortical neuronal ensembles. *Current Opinion in Neurobiology*, **10**:180-186.

Suzuki, W., Miller, E. and Desimone, R. (1997). Object and place memory in the macaque entorhinal cortex. *Journal of Neurophysiology*, **78**:1062-1081.

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

Tamura, H. and Tanaka, K. (2001). Visual response properties of cells in the ventral and dorsal parts of the macaque inferotemporal cortex. *Cerebral Cortex*, **11**:384-399.

Tanaka, K. (1993). Neuronal mechanism of object recognition. *Science*, **262**:685-688.

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience*, **19**:109-139.

Templeton, W. B. and Green, G. A. (1968). Chance results in utrocular discrimination. *Quarterly Journal of Experimental Psychology*, **20**:200-203.

Thomas, E., van Hulle, M. and Vogels, R. (2001). Encoding of categories by noncategory-specific neurosn in the inferior temporal cortex. *Journal of Cognitive Neuroscience*, **13**:190-200.

Thorpe, S., Fize, D. and Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, **381**:520-522.

Tomita, H., Ohbayashi, M., Nakahara, K., Hasegawa, I. and Miyashita, Y. (1999). Top-down signal from prefrontal cortex in executive control of memory retrieval. *Nature*, **401**:699-703.

Tong, F. and Engel, S. (2001). Interocular rivalry revealed in the human cortical blind-spot representation. *Nature*, **411**:195-199.

Tong, F., Nakayama, K., Moscovitch, M., Weinrib, O. and Kanwisher, N. (2000). Response properties of the human fusiform face area. *Cognitive Neuropsychology*, **17**:257-279.

Tong, F., Nakayama, K., Vaughan, J. T. and Kanwisher, N. (1998). Binocular Rivalry and Visual Awareness in Human Extrastriate Cortex. *Neuron*, **21**:753-759.

Tononi, G. and Edelman, G. (1998). Neuroscience - Consciousness and complexity. *Science*, **282**:1846-1851.

Tononi, G., Srinivasan, R., Russell, D. and Edelman, G. (1998). Investigating neural correlates of conscious perception by frequency-tagged neuromagnetic responses. *Proceedings of the National Academy of Sciences U.S.A.*, **95**:3198-3203.

Tranel, D., Damasio, H. and Damasio, A. (1990). Face agnosia and the neural substrates of memory. *Annual Review of Neuroscience*, **13**:89-109.

Treisman, A. M. and Kanwisher, N. G. (1998). Perceiving visually presented objects: recognition, awareness, and modularity. *Current Opinion in Neurobiology*, **8**:218-226.

Treves, A. and Rolls, E. T. (1994). Computational analysis of the role of the hippocampus in memory. *Hippocampus*, **4**:374-391.

van Steveninck, R., Lewen, G. D., Strong, S. P., Koberle, R. and Bialek, W. (1997). Reproducibility and variability in neural spike trains. *Science*, **275**:1805-1808.

vanRullen, R. and Thorpe, S. (In preparation). Surfing a Spike Wave down the Ventral Stream. *Cerebral Cortex (submitted)*, .

Vogels, R. (1999a). Categorization of complex visual images by rhesus monkeys: Part 1: behavioral study. *European Journal of Neuroscience*, **11**:1223-1238.

Vogels, R. (1999b). Categorization of complex visual images by rhesus monkeys: Part 2: single-cell study. *European Journal of Neuroscience*, **11**:1239-1255.

von der Heydt, R., Peterhans, E. and Baumgartner, G. (1984). Illusory contours and cortical neuron responses. *Science*, **224**:1260-1262.

von Helmholtz, H. (1866). *Physiological Optics*, Dover Publications, New York.

Walker, P. (1978). Binocular rivalry: central or peripheral selective processes? *Psychology Bulletin*, **85**:376-389.

Walsh, V. and Cowey, A. (1998). Magnetic Stimulation Studies of Visual Cognition. *Trends in Cognitive Sciences*, **2**:103-110.

Wandell, B. A. (1995). *Foundations of vision*, Sinauer Associates Inc., Sunderland.

Ward, A. and Schmidt, R. (1961). Some properties of single epileptic neurons. *Archives of Neurology*, **5**:90-95.

Ward, A. and Thomas, L. (1955). The electrical activity of single units in the cerebral cortex of man. *EEG Clinical and Neurophysiology*, **7**:135-136.

Warland, D. K., Reinagel, P. and Meister, M. (1997). Decoding visual information from a population of retinal ganglion cells. *Journal of Neurophysiology*, **78**:2336-2350.

Warrington, E. and Mc Carthy, R. (1983). Category specific access dysphasia. *Brain*, **106**:859-878.

Warrington, E. and Mc. Carthy, R. (1987). Categories of knowledge - Further fractionations and an attempted integration. *Brain*, **110**:1273-1296.

Warrington, E. and McCarthy, R. (1994). Multiple meaning systems in the brain - A case for visual semantics. *Neuropsychologia*, **32**:1465-&.

Warrington, E. and Shallice, T. (1984). Category specific semantic impairments. *Brain*, **107**:829-854.

Wassermann, E. M. and Grafman, J. (1999). Combining transcranial magnetic stimulation and neuroimaging to map the brain. *Trends in Cognitive Science*, .

Watson, A. and Pelli, D. (1983). Quest - A bayesian adaptive psychometric method. *Perception and Psychophysics*, **33**:113-120.

Wehr, M., Pezaris, J. and Sahani, M. (1999). Simultaneous paired intracellular and tetrode recordings for evaluating the performance of spike sorting algorithms. *Neurocomputing*, **26-7**:1061-1068.

Wellmer, J., Widman, G., von Oertzen, J., Van Rost, D., Schaller, C., Urback, H., Schramm, J. and Elger, C. (2000). Anatomically precise and reproducible localization of grid electrodes in invasive presurgical epilepsy diagnosis - A method of routine application. In *American Epilepsy Society*, Vol. 41 S7 (Ed, Epilepsia). Epilepsia, Los Angeles.

Wessel, R., Koch, C. and Gabbiani, F. (1996). Coding of time-varying electric field amplitude modulations in a wave-type electric fish. *Journal of Neurophysiology*, **75**:2280-2293.

Wheatstone, C. (1838). On some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosophical Transactions of the Royal Society of London*, **128**:371-394.

Wheeler, M., Petersen, S. and Buckner, R. (2000). Memory's echo: Vivid remembering reactivates sensory-specific cortex. *Proceedings of the National Academy of Sciences*, **97**:11125-11129.

Wilkins, A. J., Shallice, T. and McCarthy, R. (1987). Frontal lesions and sustained attention. *Neuropsychologia*, **25**:359-365.

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

Wilson, M. A. and McNaughton, B. L. (1994). Reactivation of Hippocampal Ensemble Memories During Sleep. *Science*, **265**:676-679.

Wolfe, J. (1984). Reversing ocular dominance and suppression in a single flash. *Vision Research*, **24**:471-478.

Yakovlev, V., Fusi, S., Berman, E. and Zohary, E. (1998). Inter-trial neuronal activity in inferior temporal cortex: a putative vehicle to generate long-term visual associations. *Nature Neuroscience*, **1**:310-317.

Yamane, S., Kaji, S. and Kawano, K. (1988). What facial features activate face neurons in the inferotemporal cortex of the monkey? *Experimental brain research*, **73**:209-214.

Young, A., Humphreys, G., Riddoch, M., Hellowell, D. and de Haan, E. (1994). Recognition impairments and face imagery. *Neuropsychologia*, **32**:693-702.

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

Zeki, S. (1983). Color coding in the cerebral cortex - The reaction of cells in monkey visual cortex to wavelengths and colors. *Neuroscience*, **9**:741-765.

Zeki, S. (1992). The visual image in mind and brain. *Scientific American*, **267**:69-76.

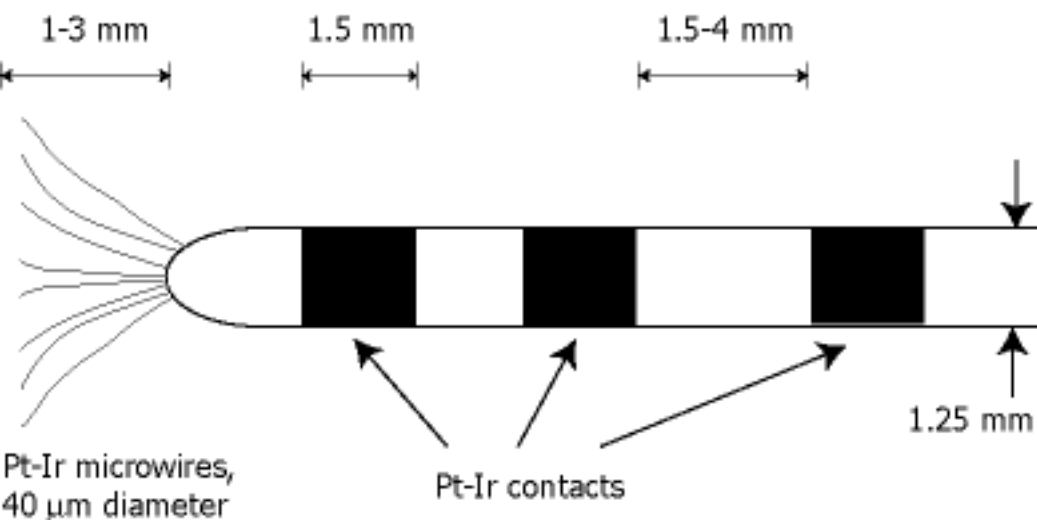
Zhang, L., Tao, H., Holt, C., Harris, W. and Poo, M. (1998). A critical window for cooperation and competition among developing retinotectal synapses. *Nature*, **395**:37-44.

Zirlinger, M., Kreiman, G. and Anderson, D. (2001). Amygdala-enriched genes identified by microarray technology are restricted to specific amygdaloid sub-nuclei. *Proceedings of the National Academy of Sciences U.S.A.*, **98**:5270-5275.

Zola-Morgan, S. and Squire, L. R. (1993). Neuroanatomy of memory. *Annual Review of Neuroscience*, **16**:547-563.

Gabriel Kreiman
Thesis - Chapter 1
Figure 1

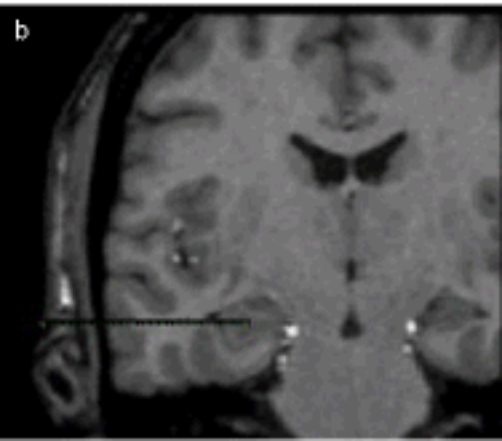
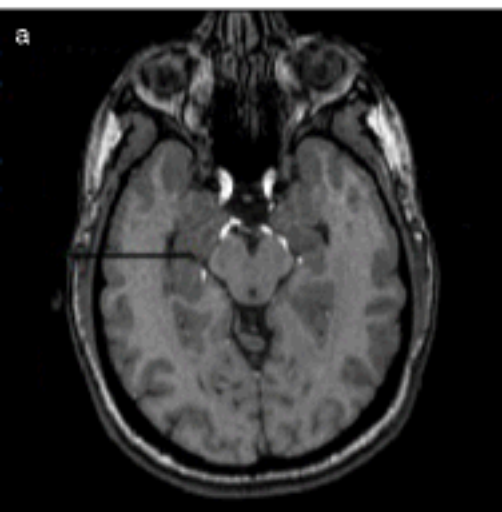


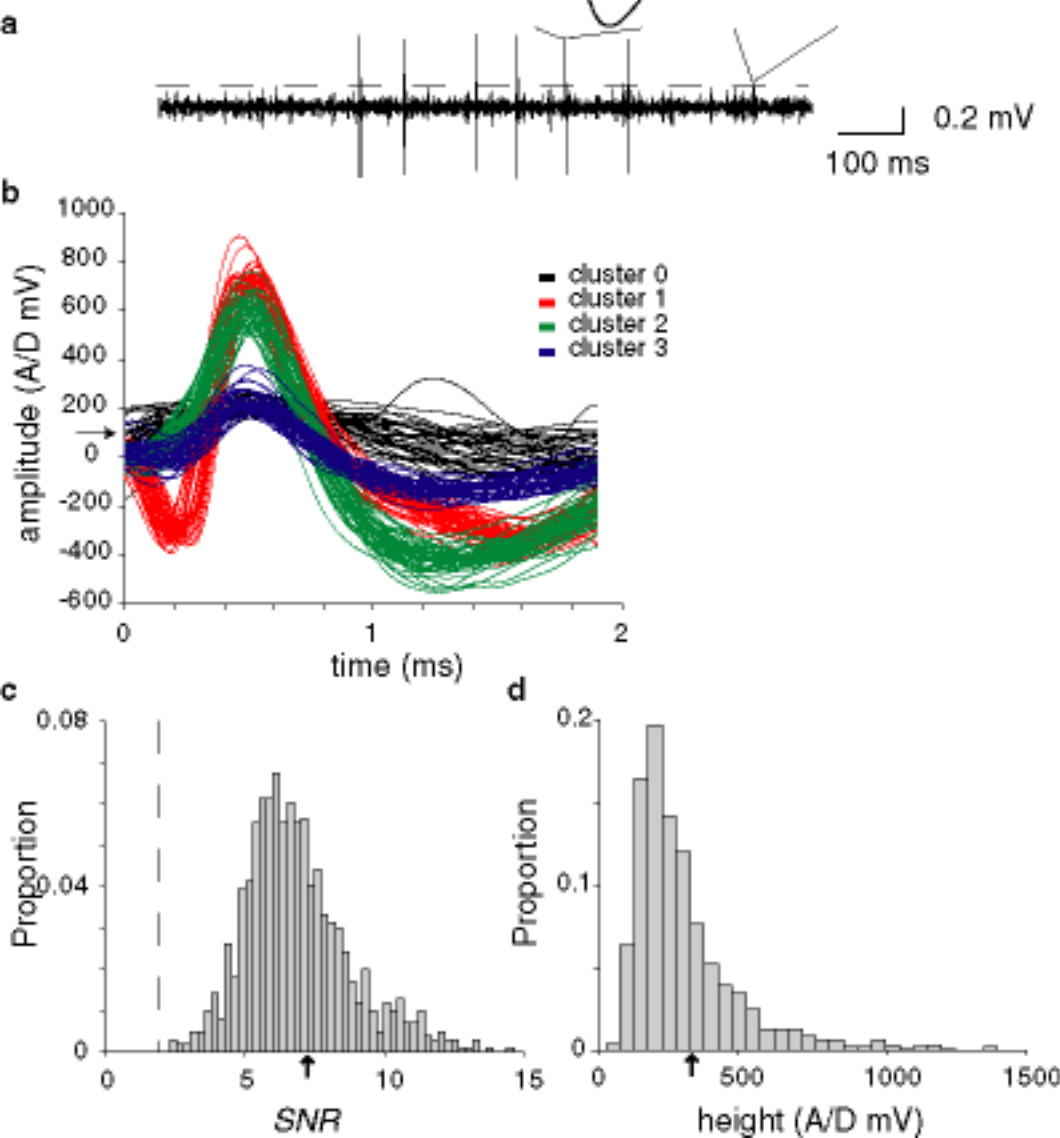


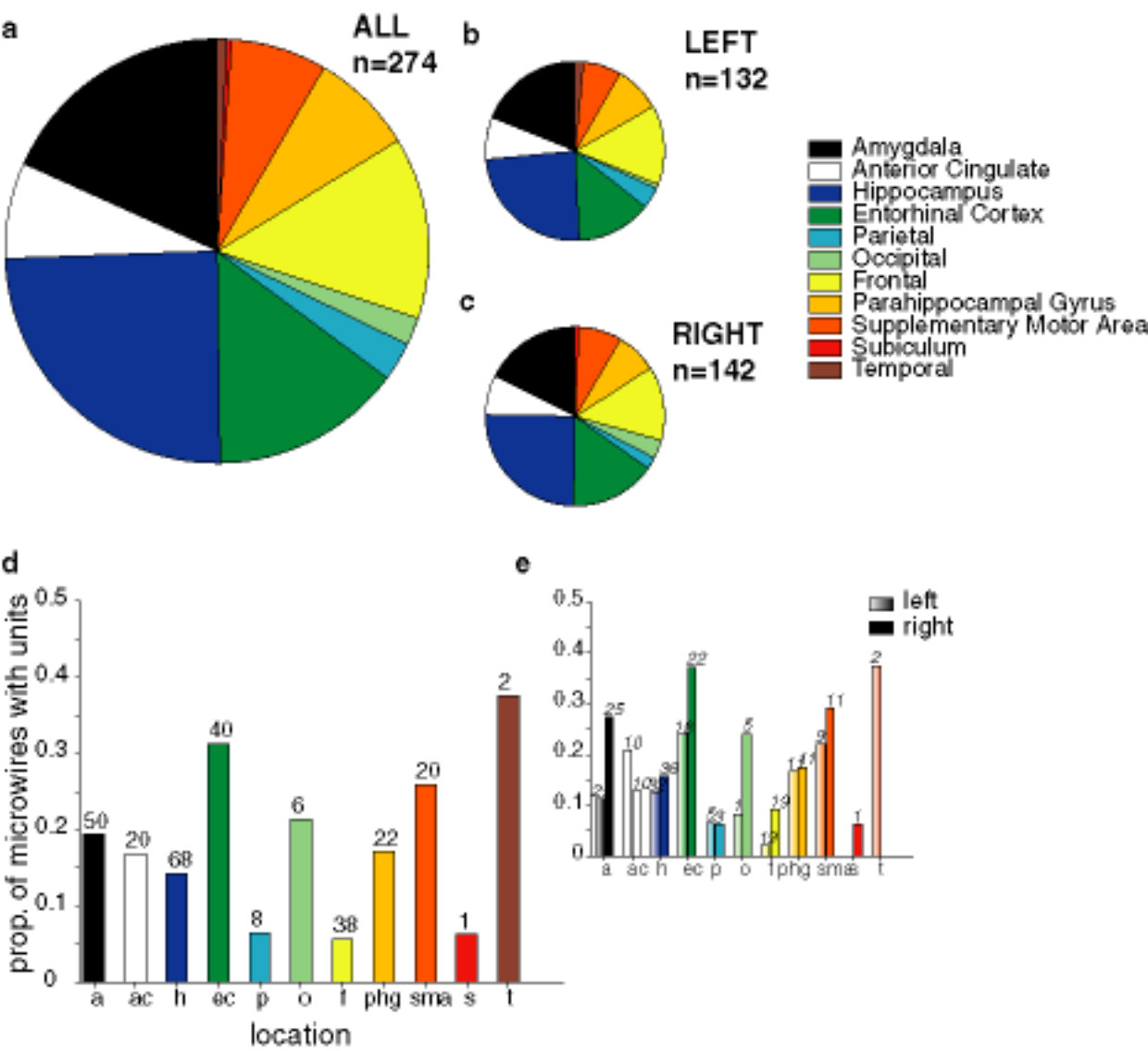
Scale 10:1
(except for microwire length)

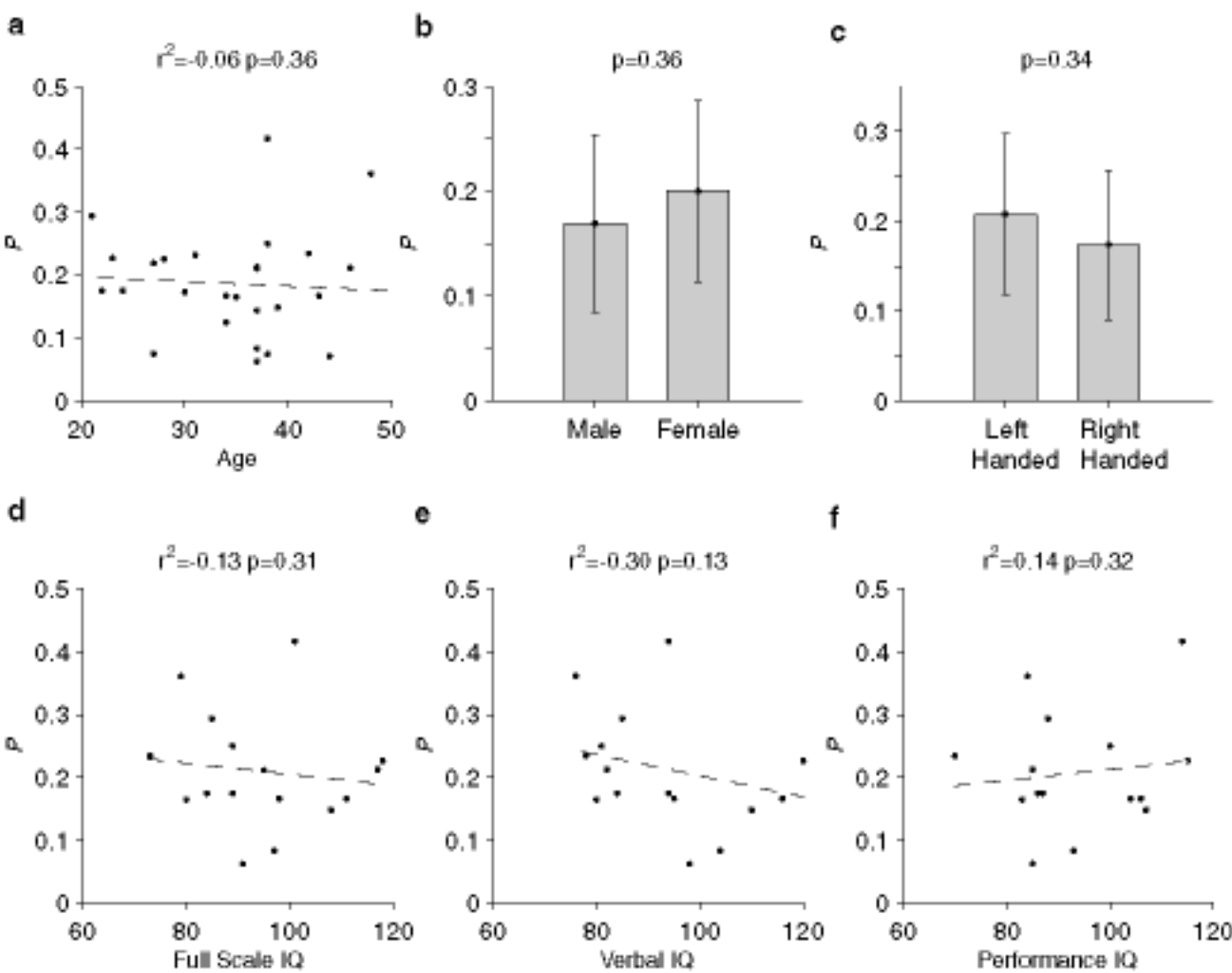
Gabriel Kreiman
Thesis - Chapter 2
Figure 2

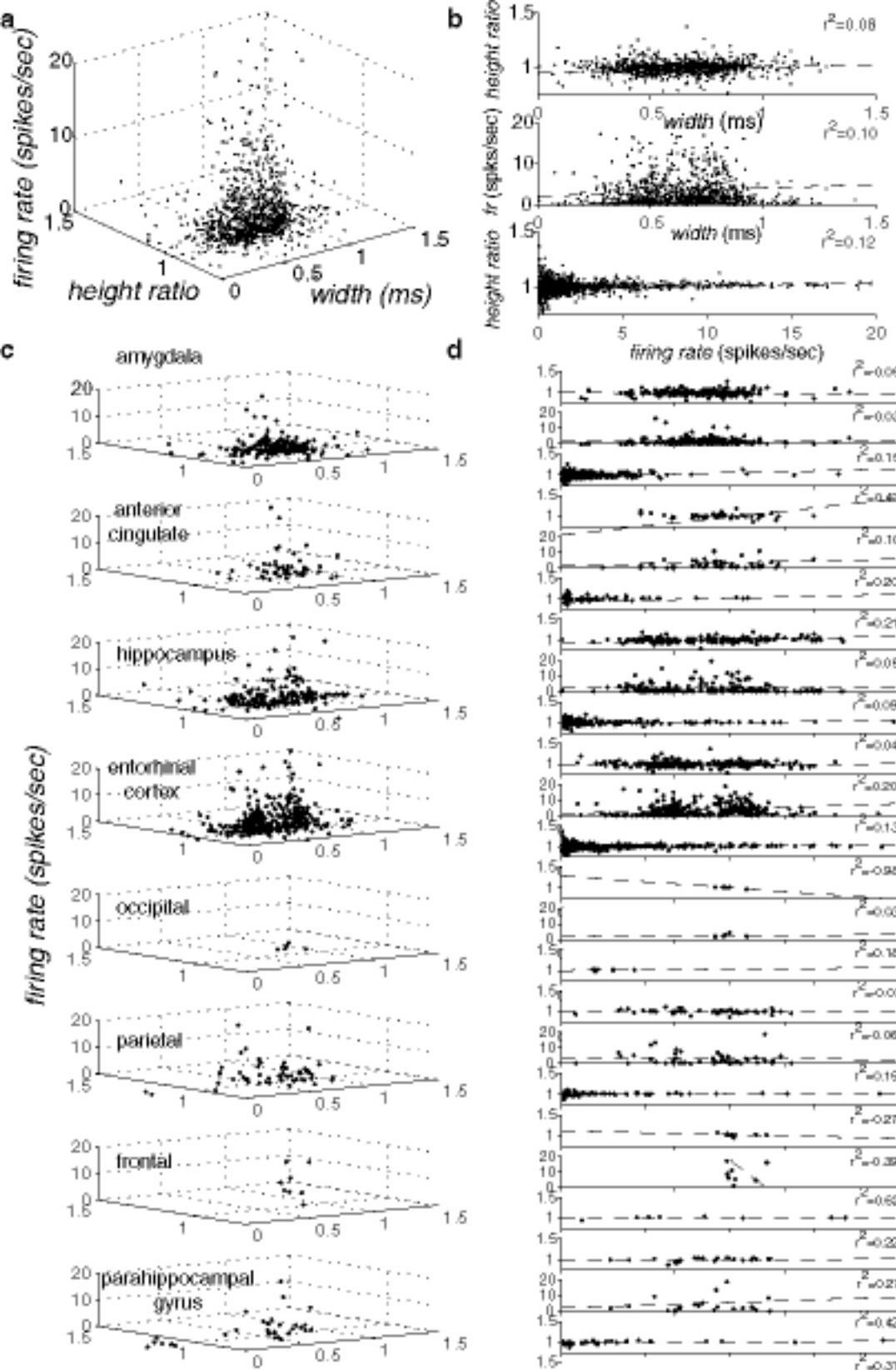


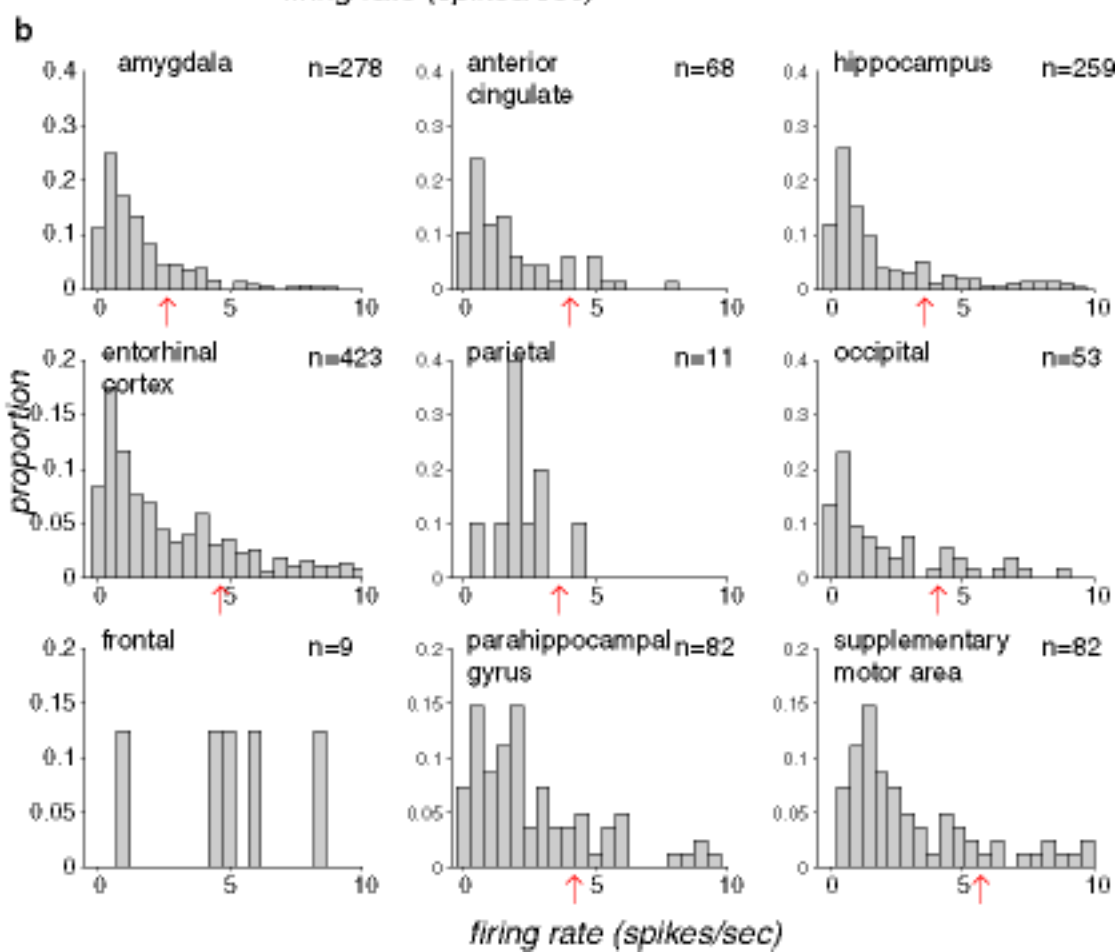
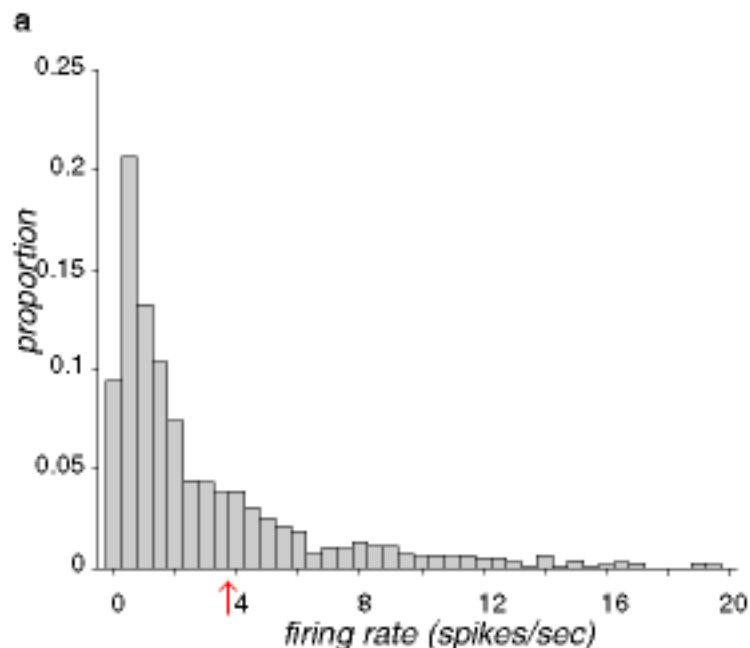


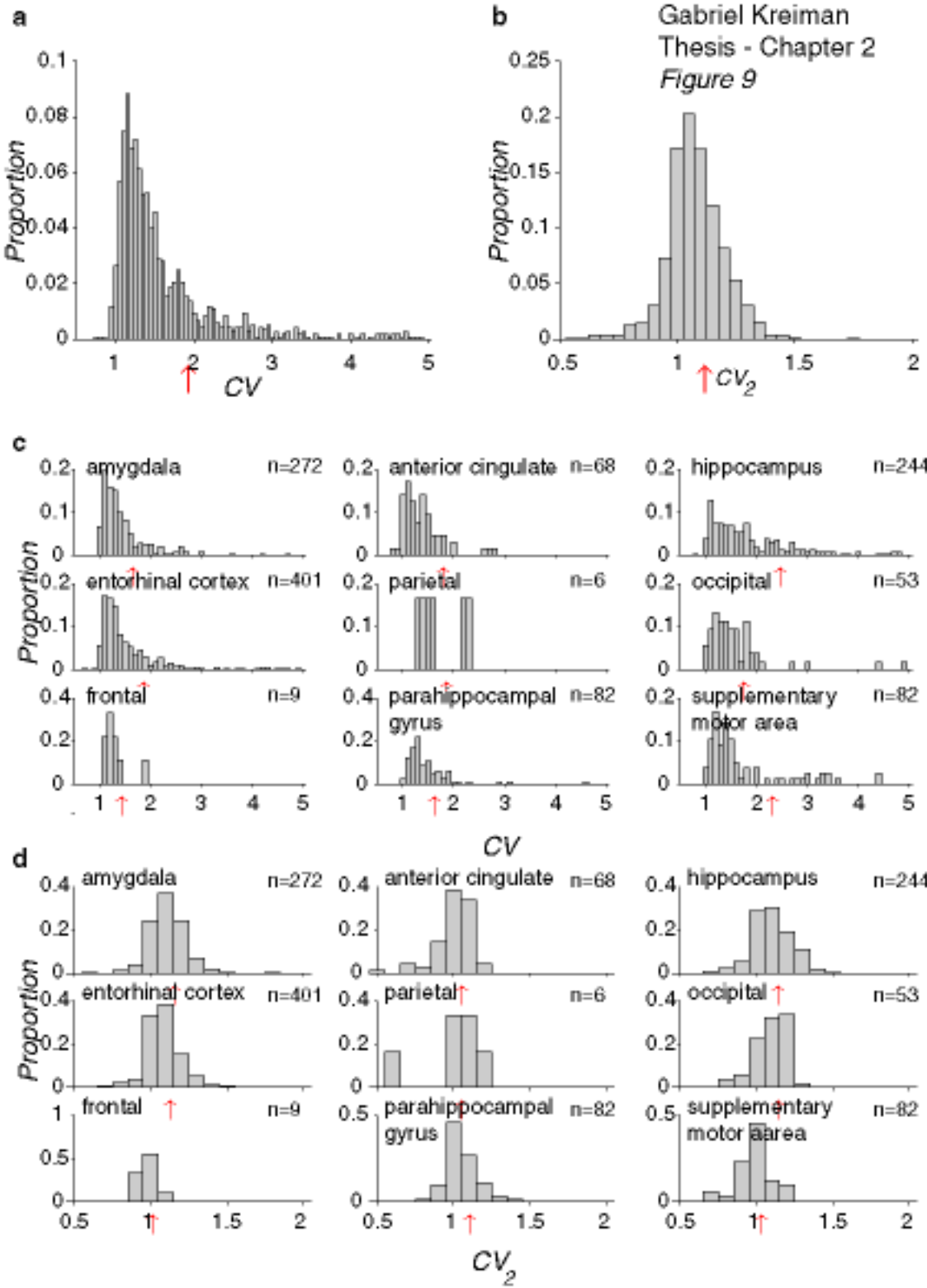


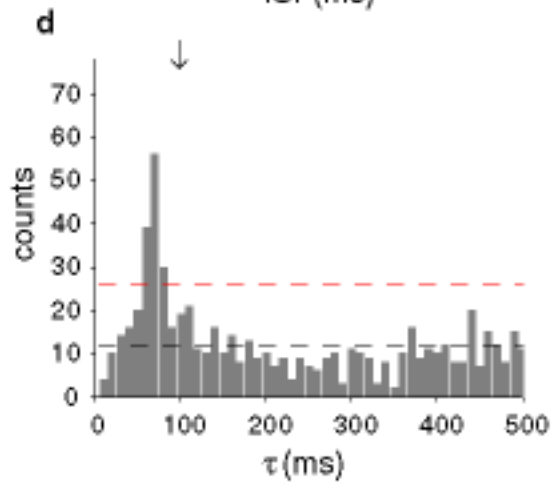
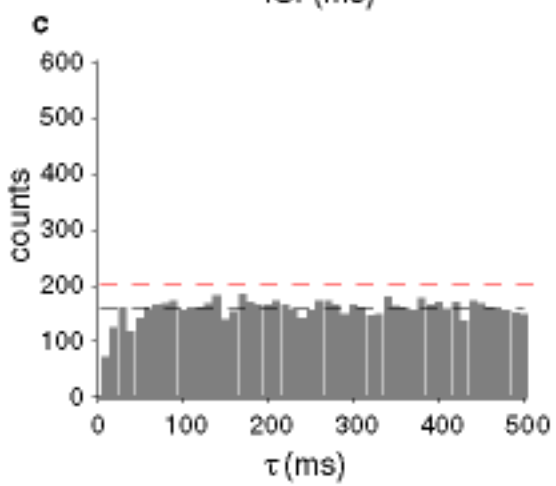
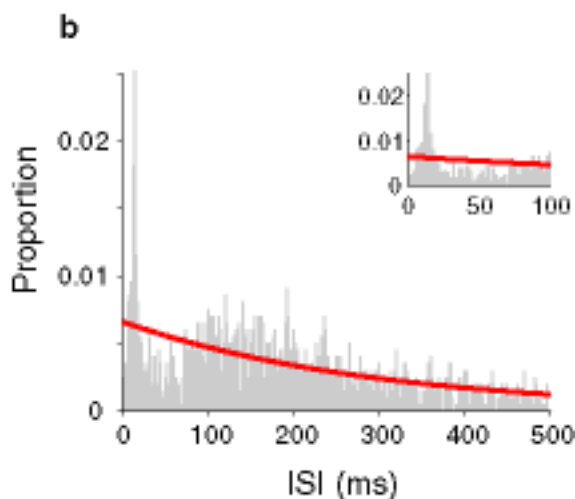
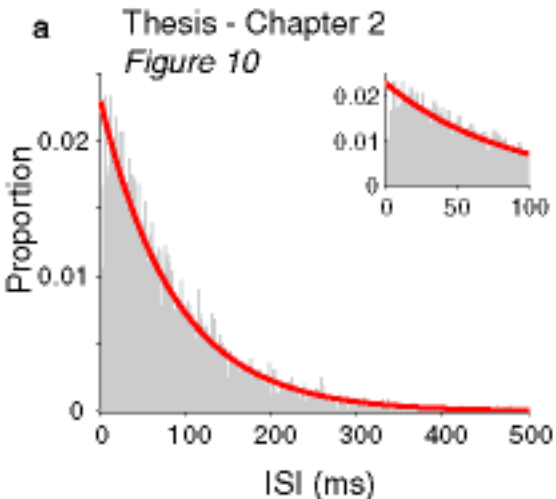


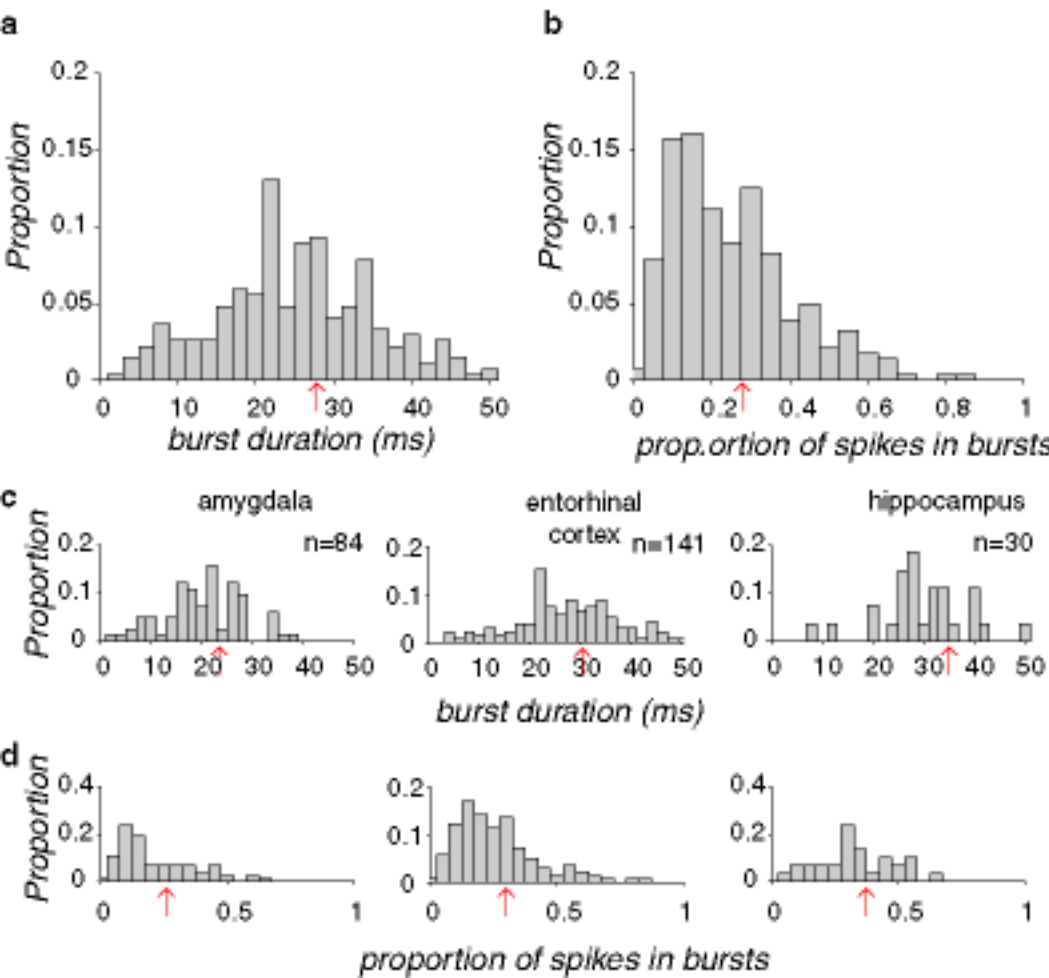


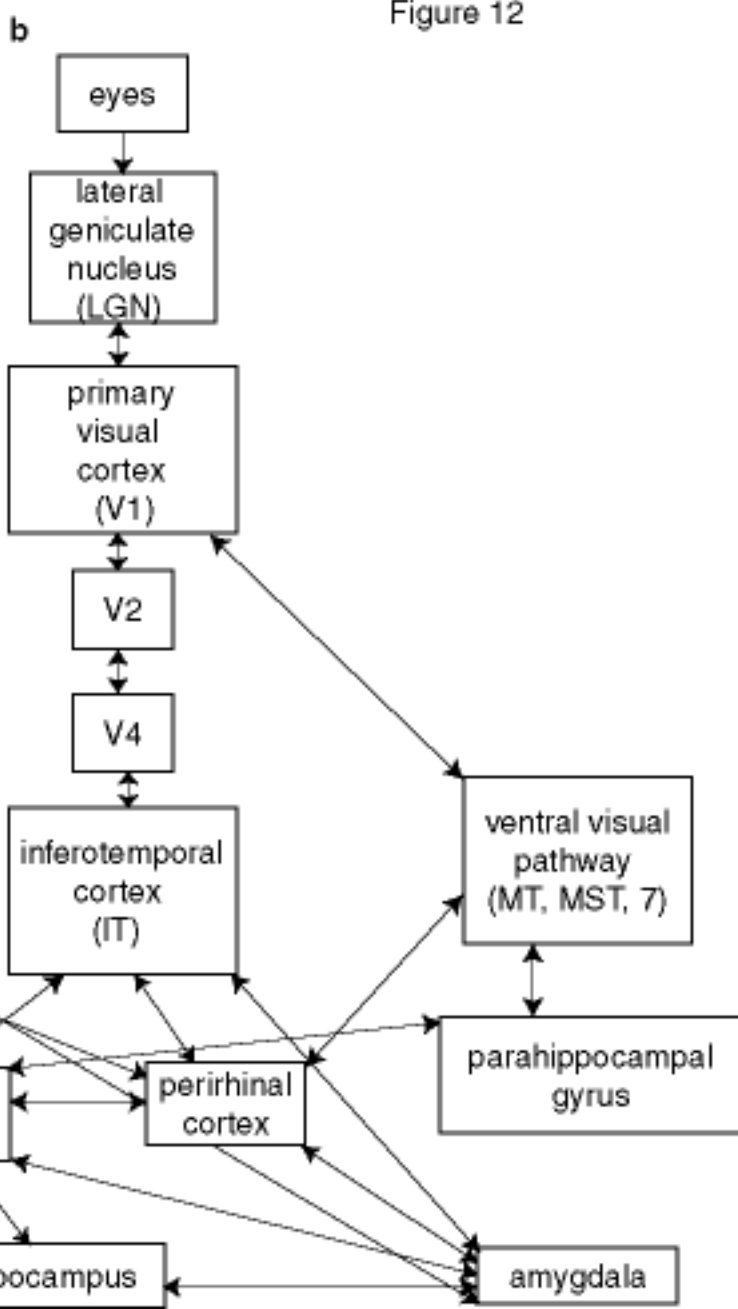






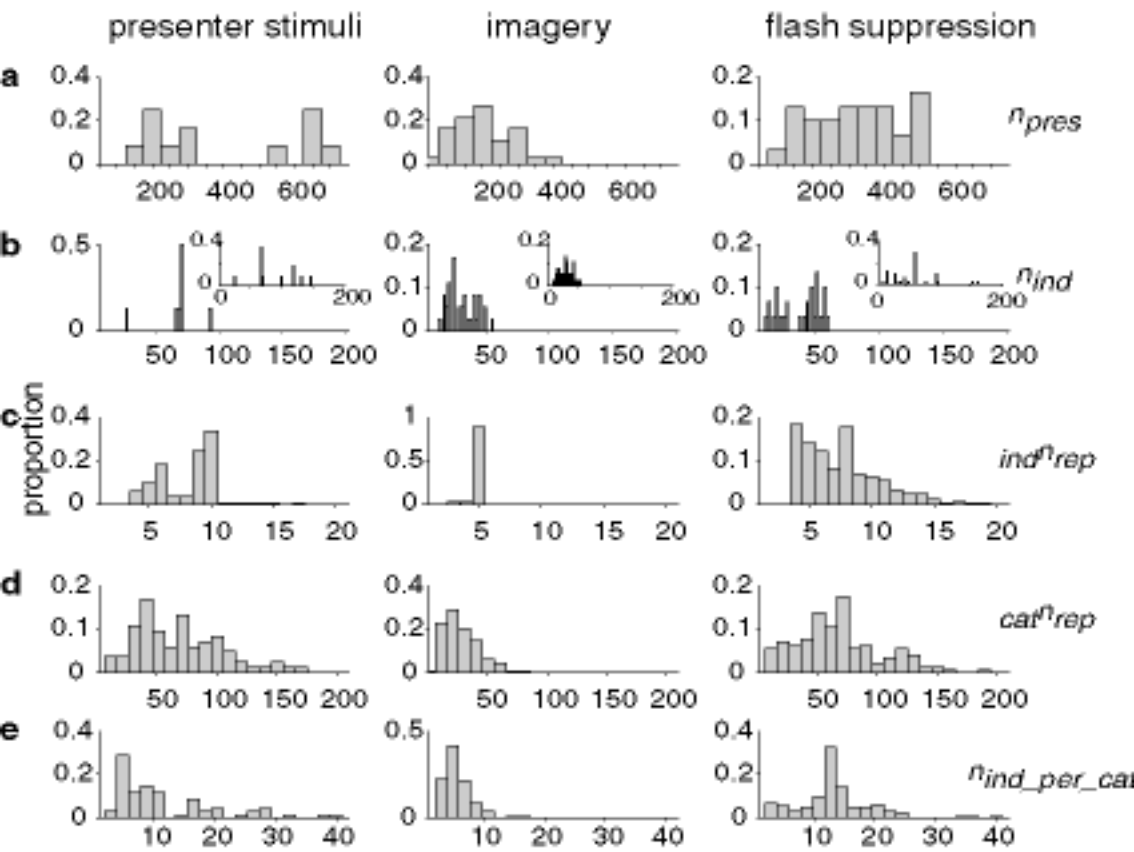




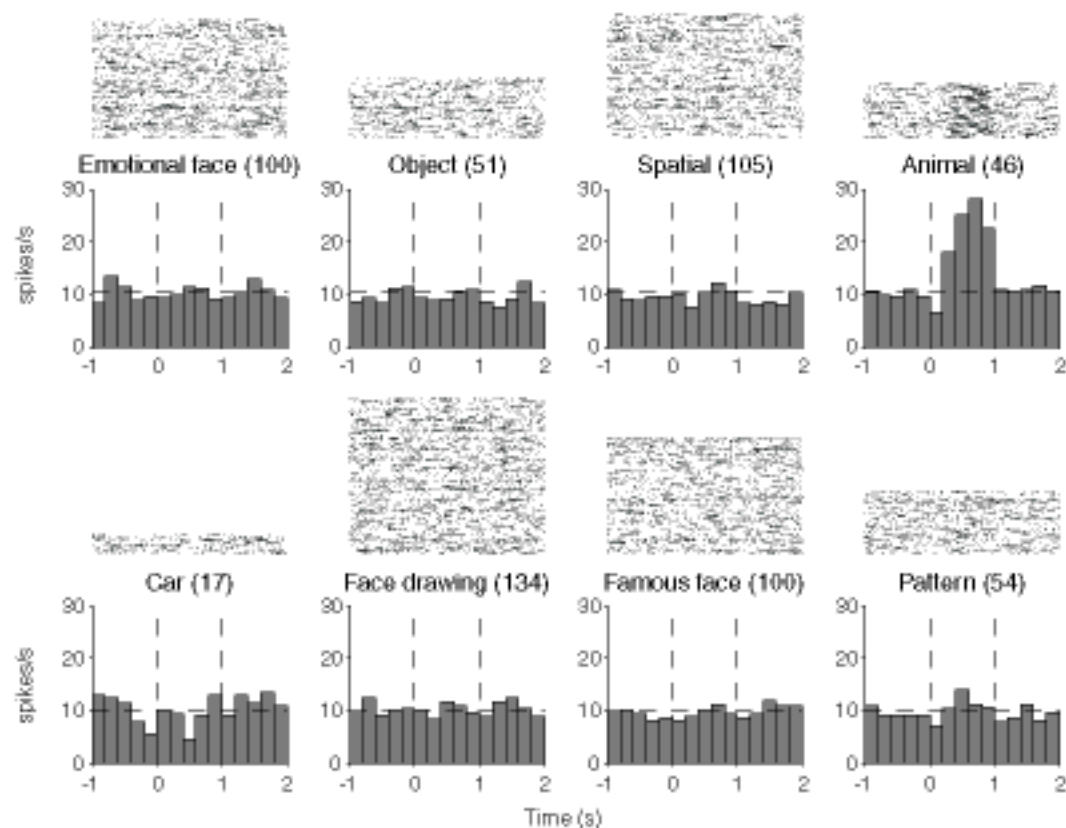


Gabriel Kreiman
Thesis - Chapter 3
Figure 1

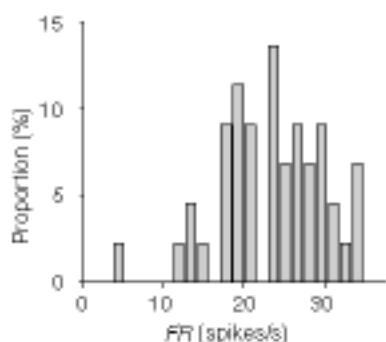




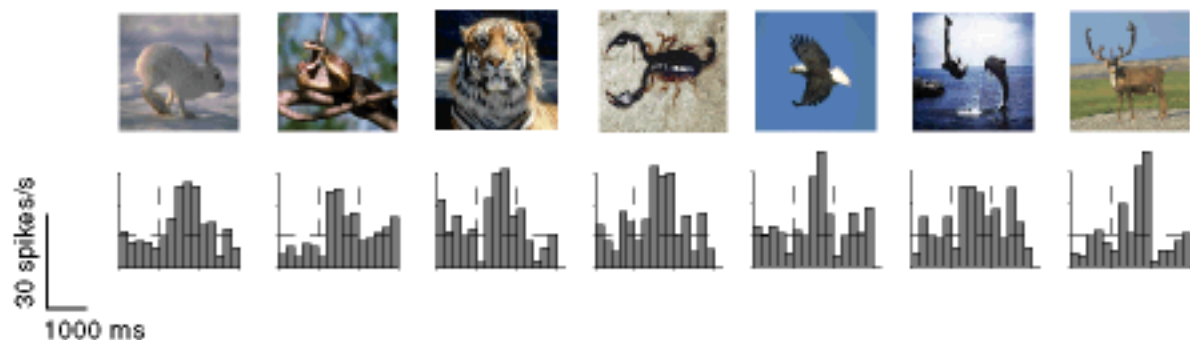
a



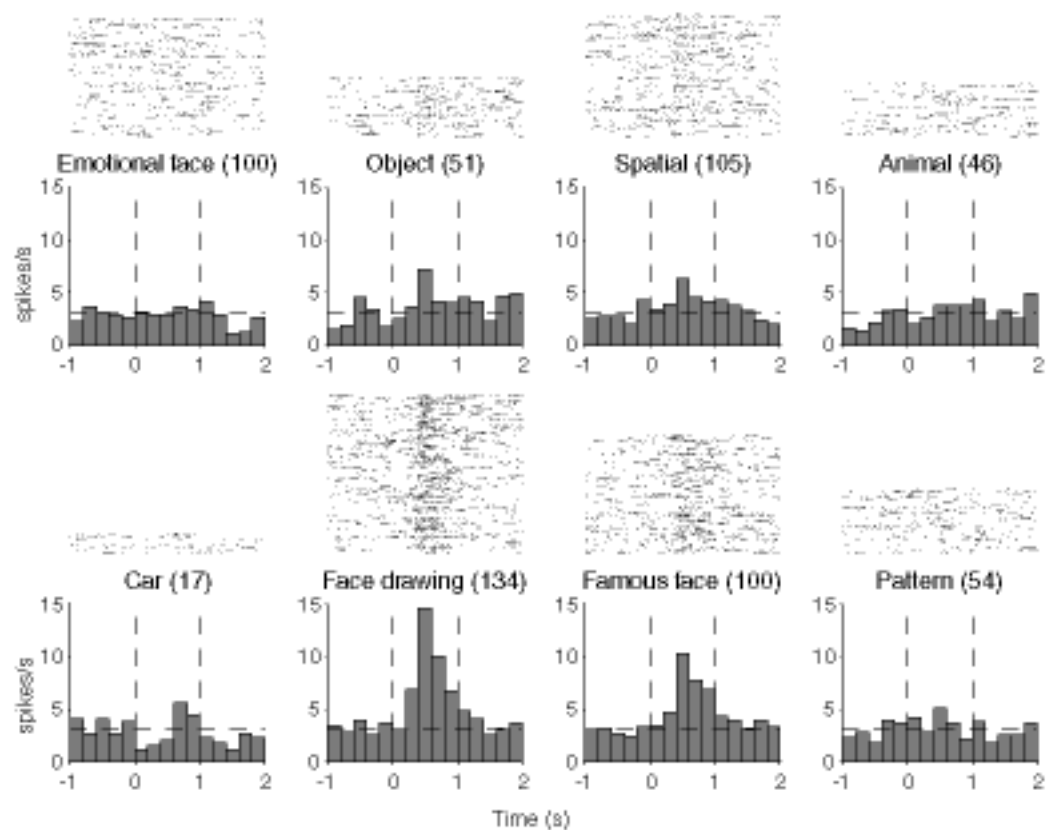
b



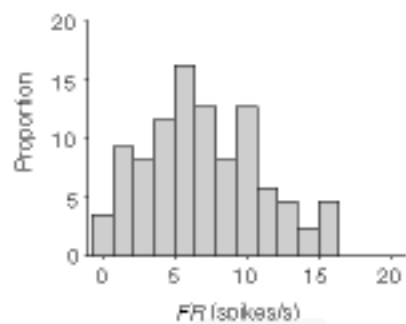
c



a



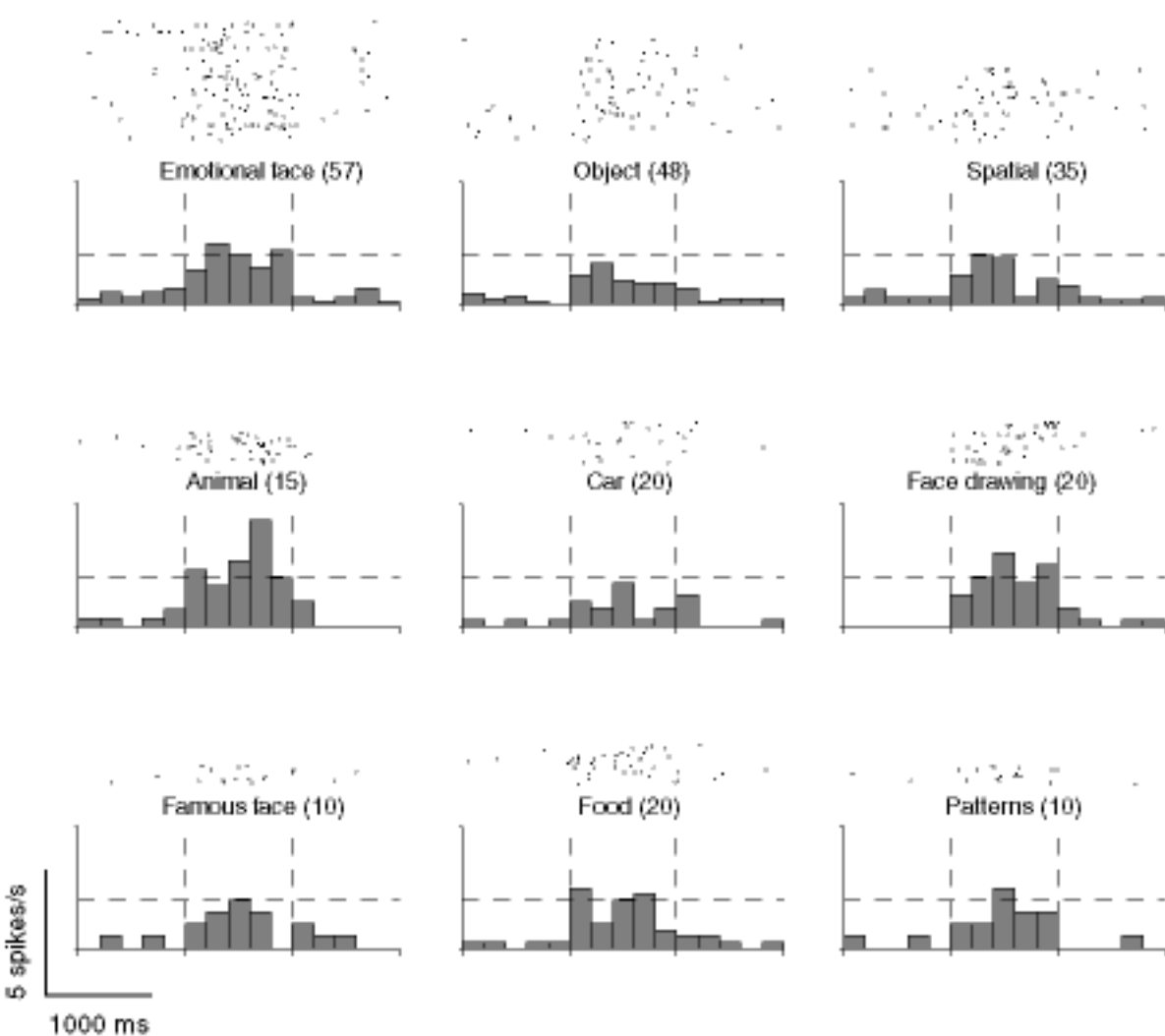
b



c



Gabriel Kreiman
Thesis - Chapter 3
Figure 5



Gabriel Kreiman
Thesis - Chapter 3
Figure 6

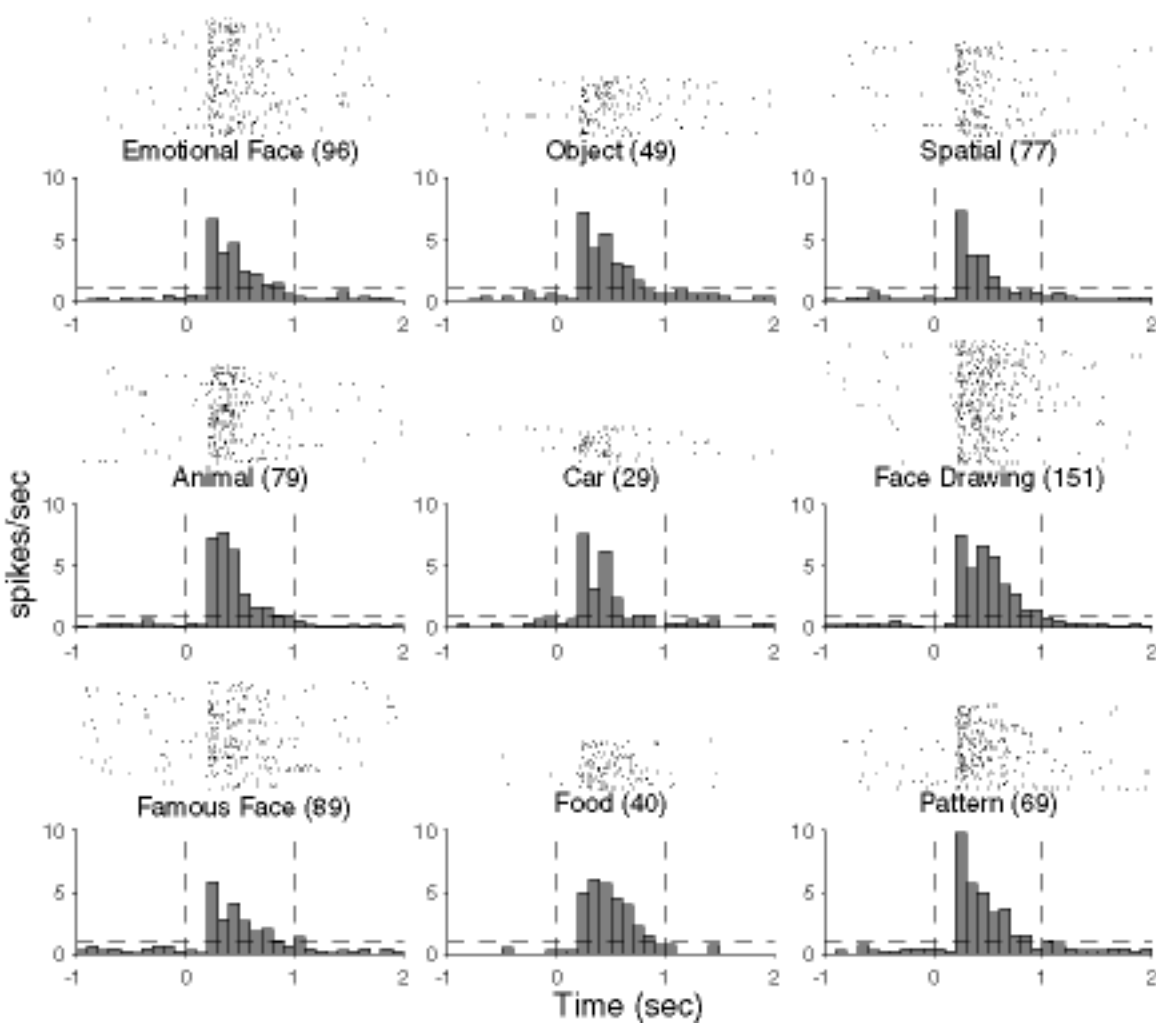
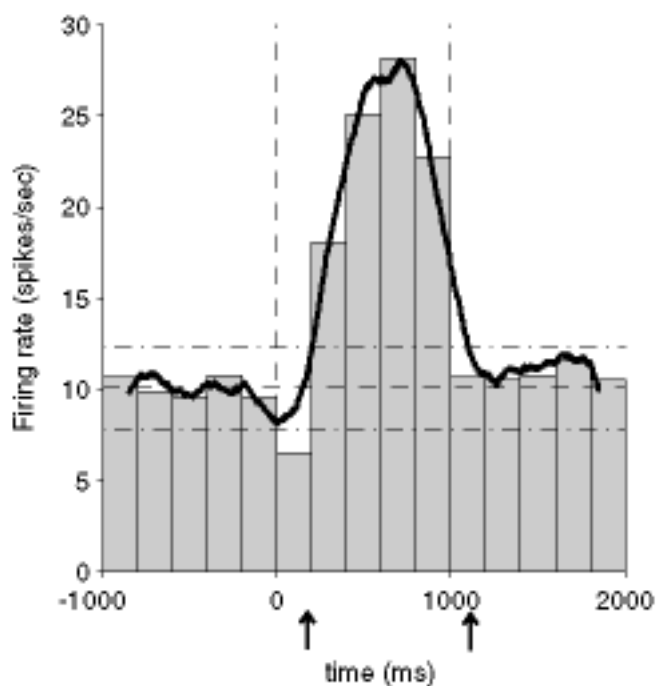


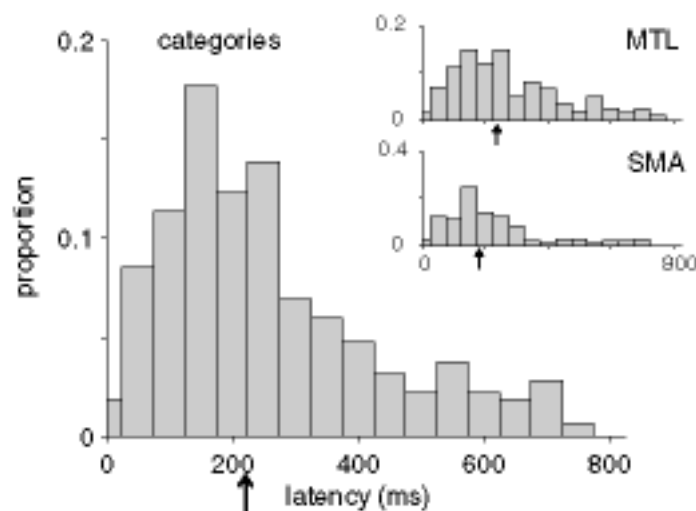
Figure 7



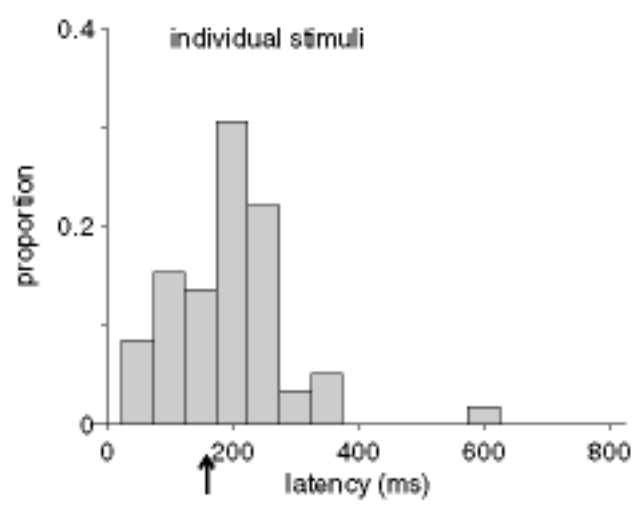
a



b

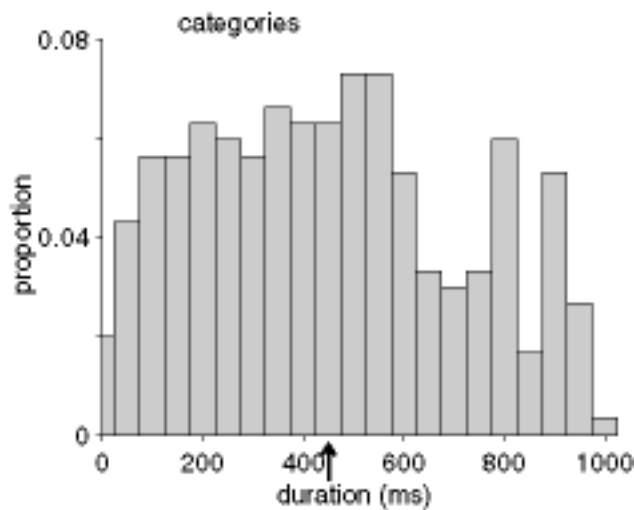


c

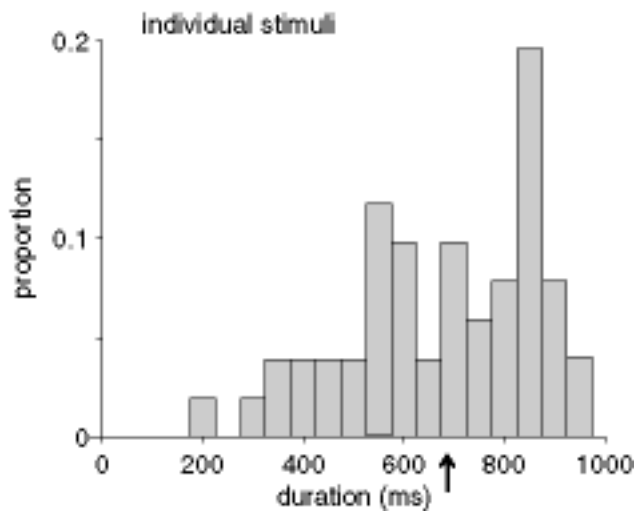


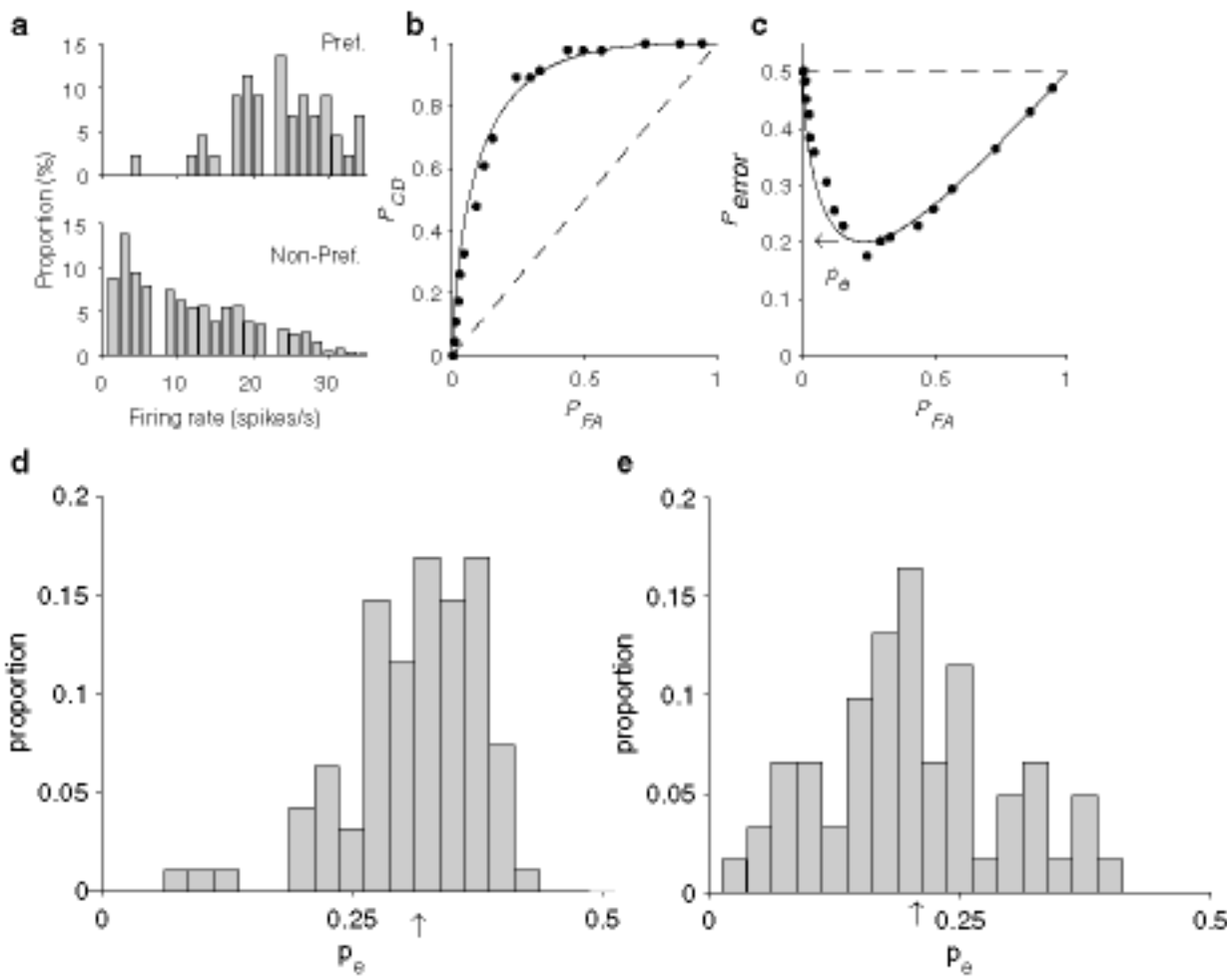
Gabriel Kreiman
Thesis - Chapter 3
Figure 10

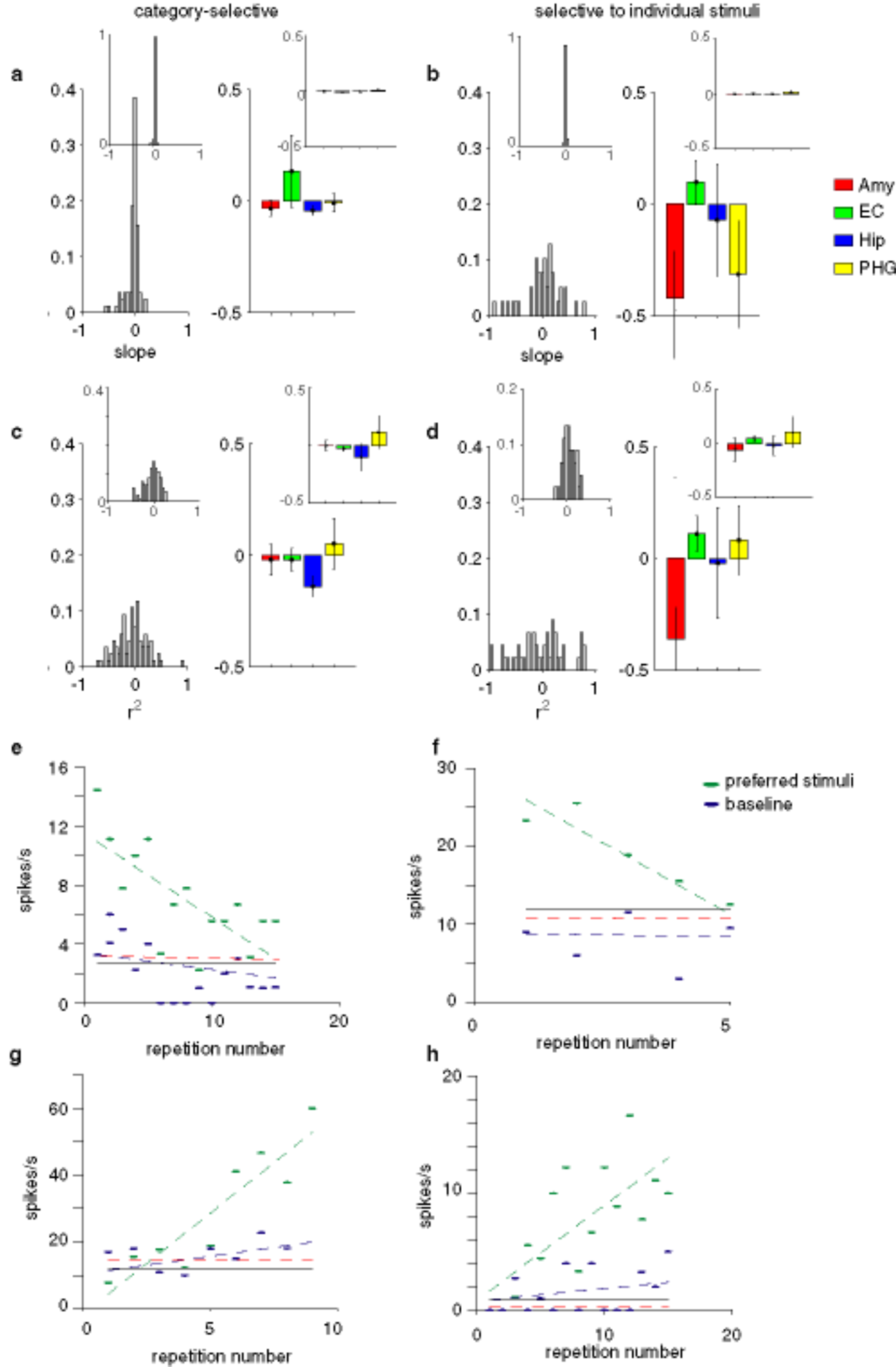
a

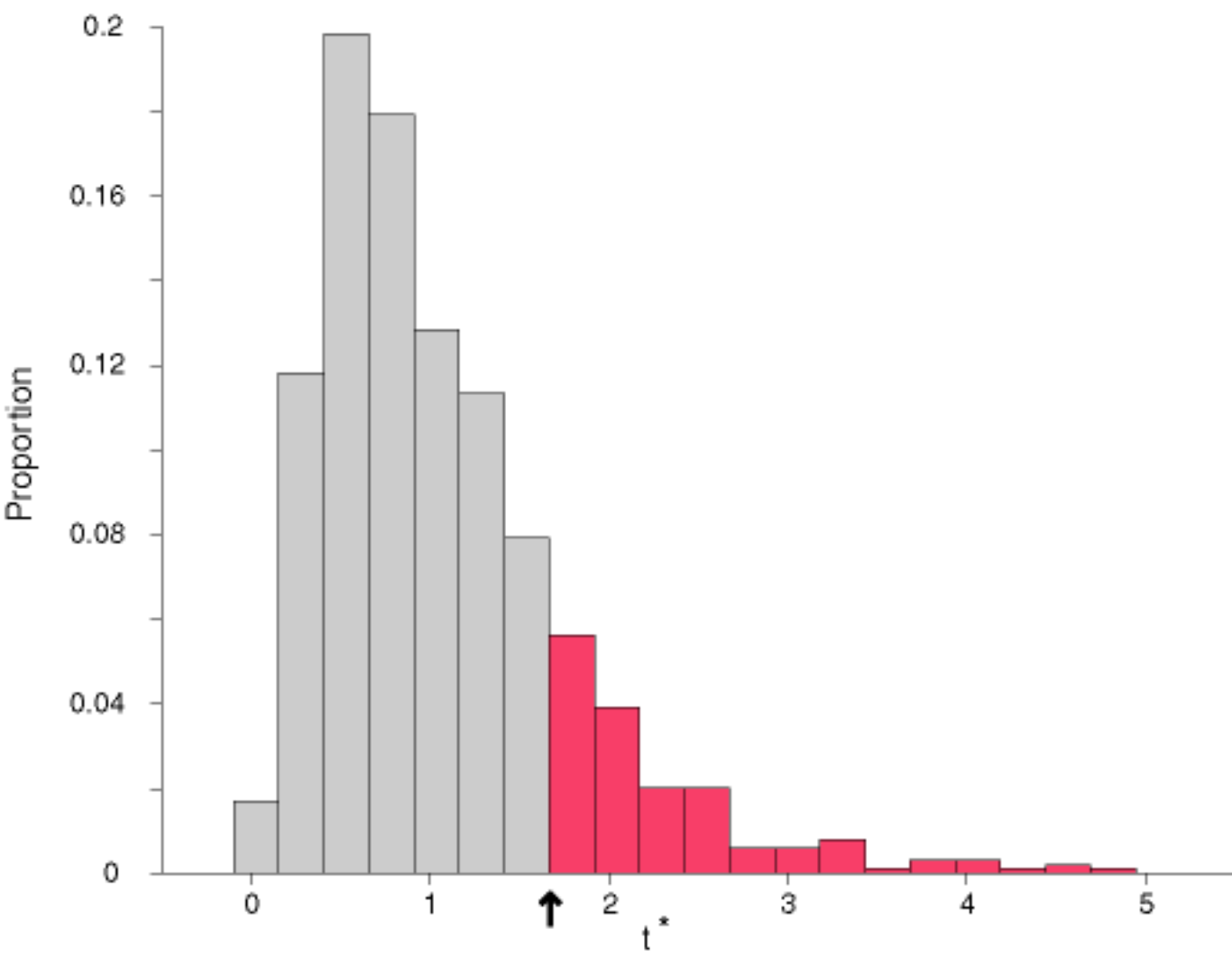


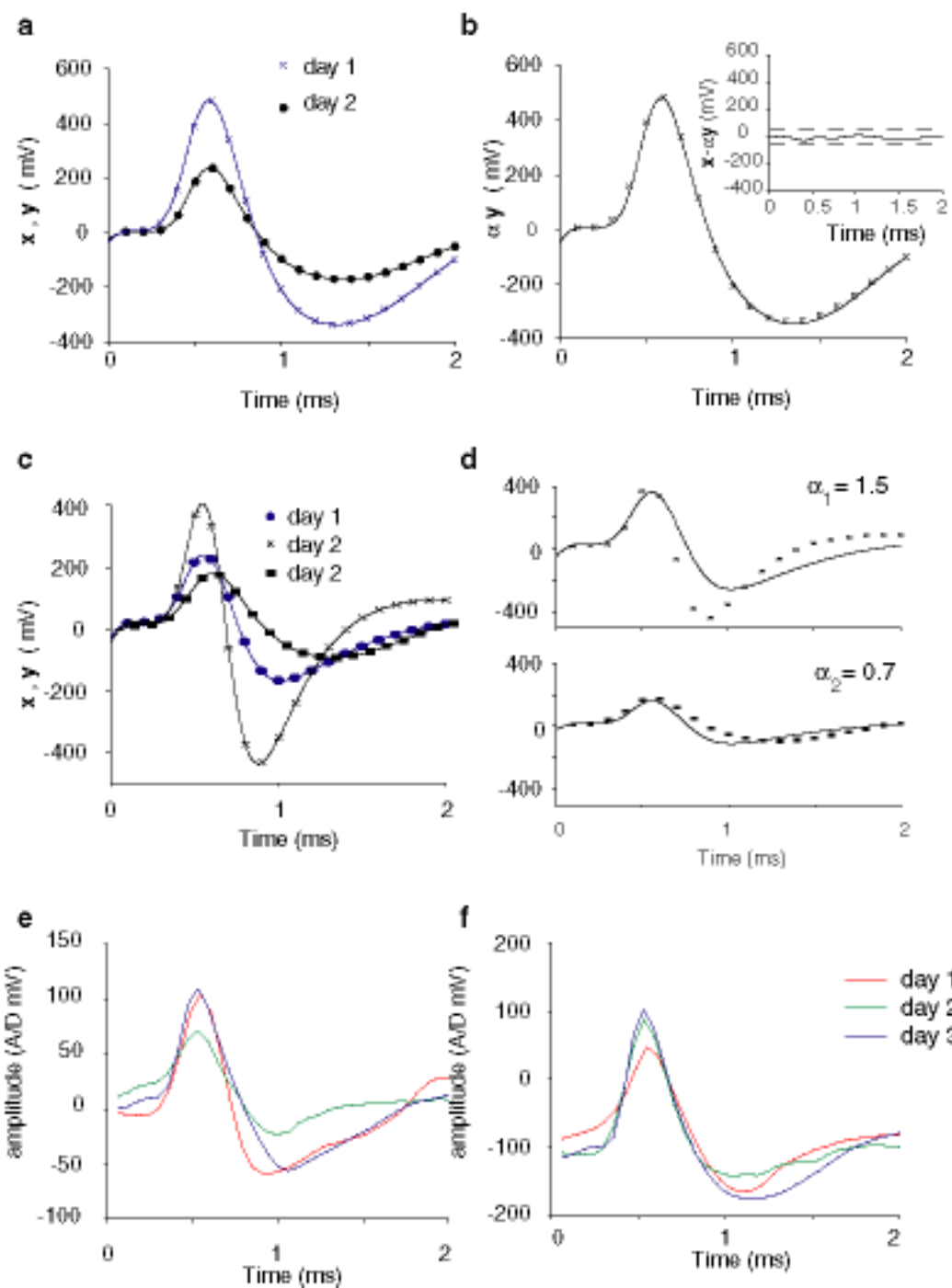
b

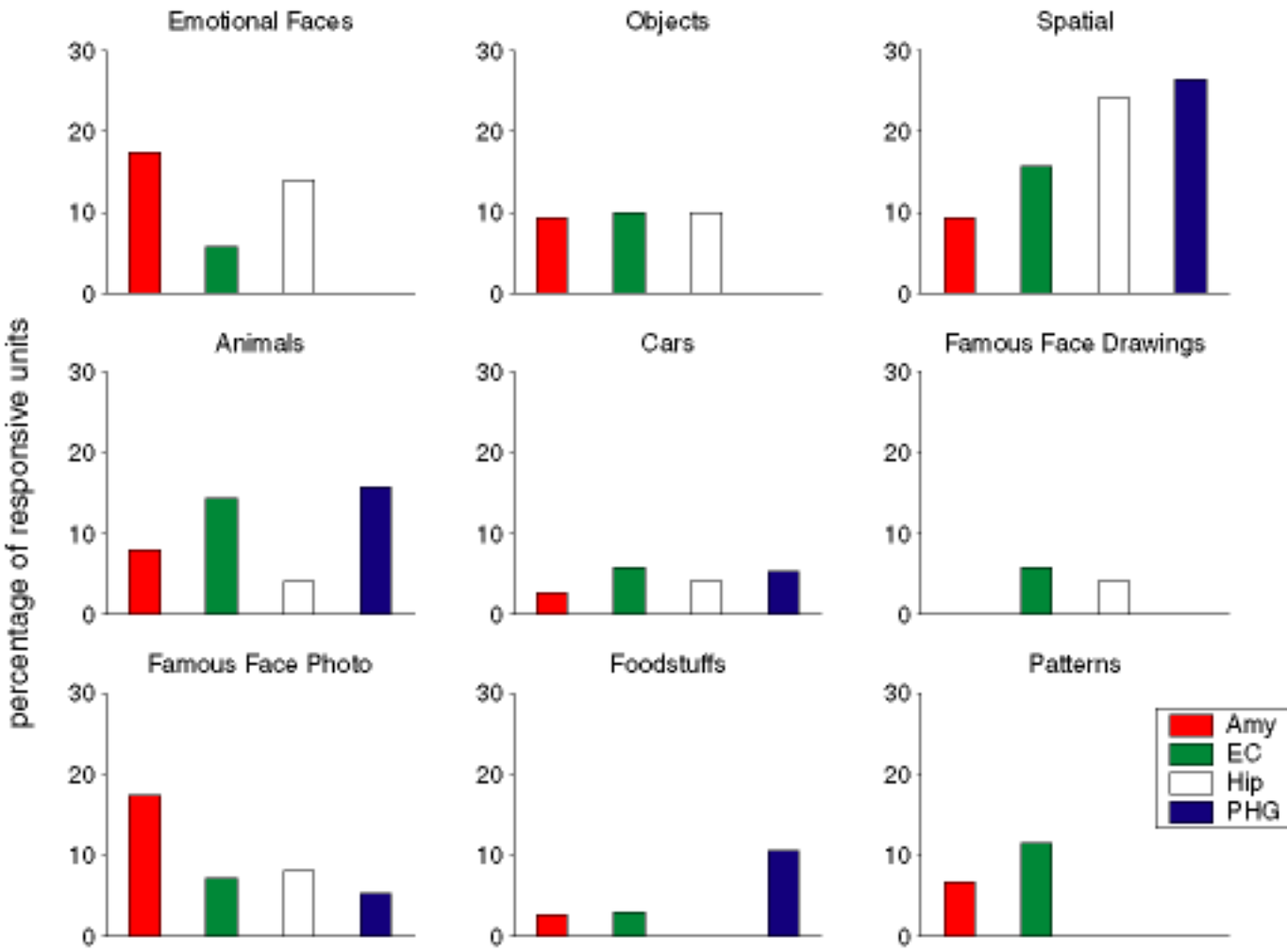




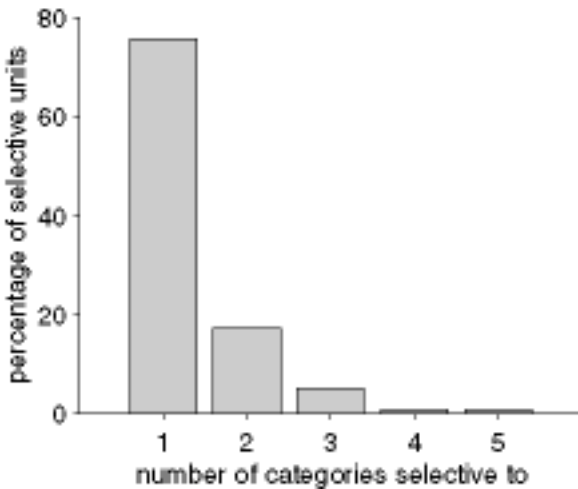


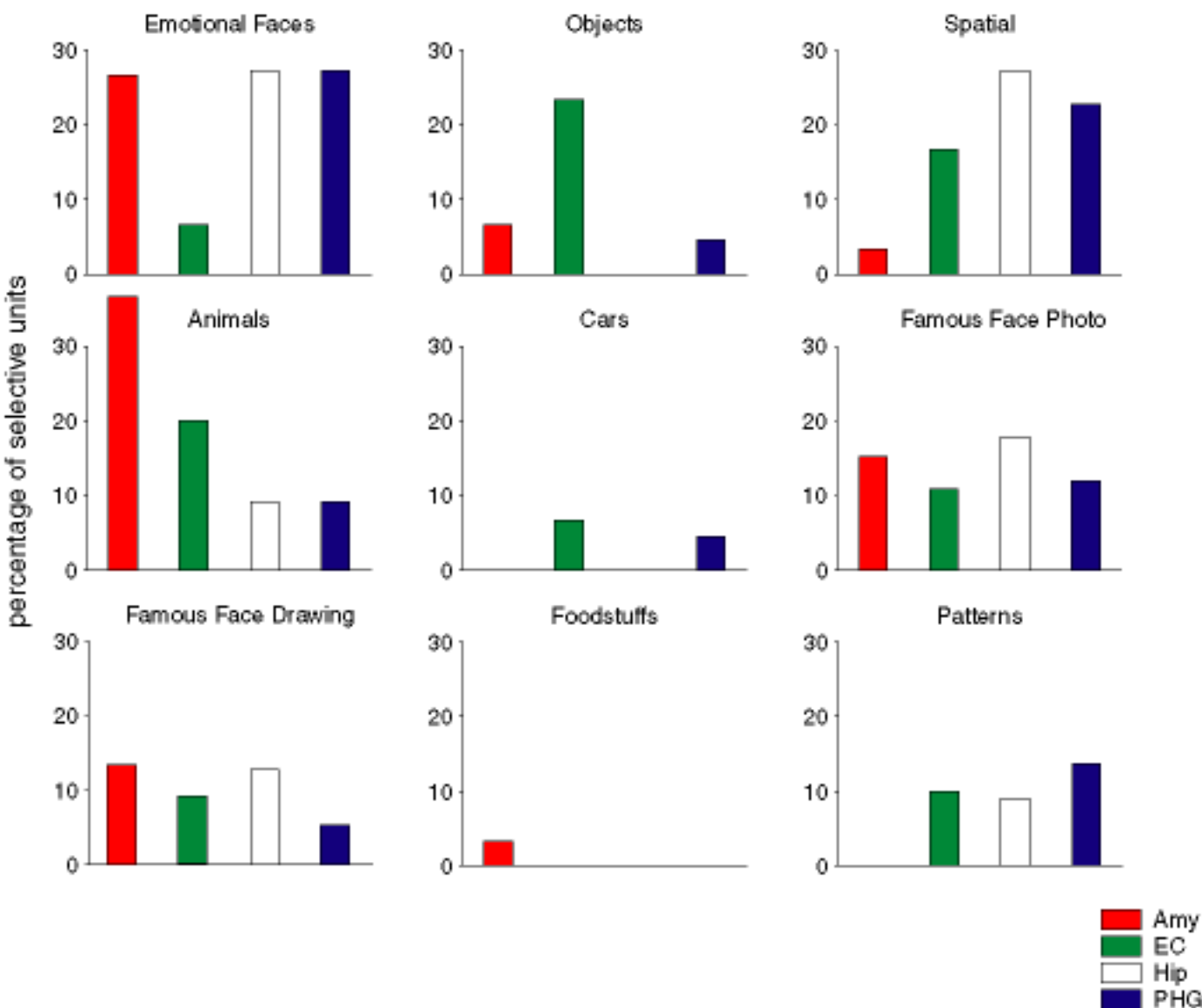




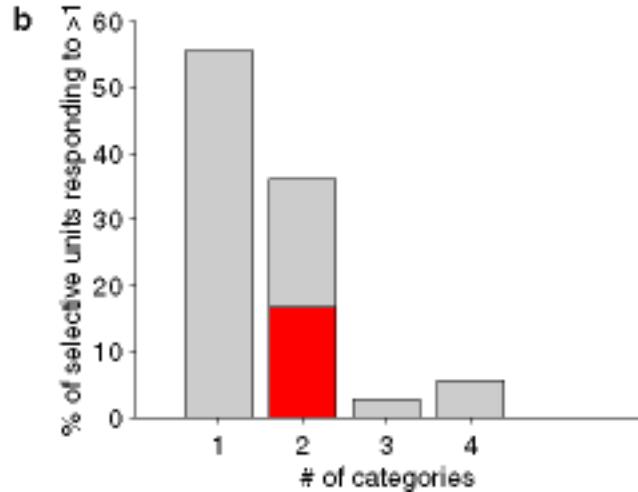
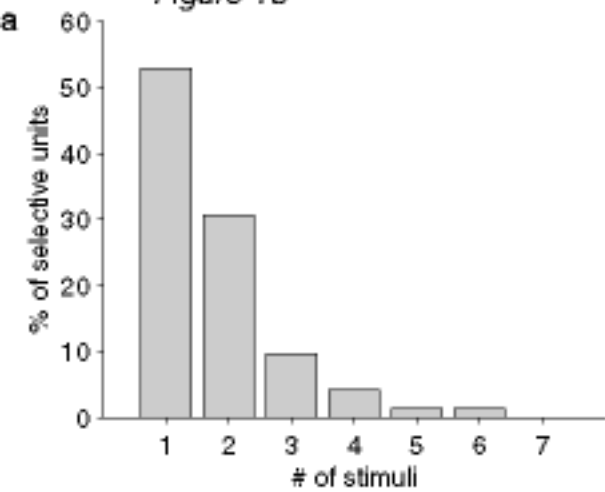


Gabriel Kreiman
Thesis - Chapter 3
Figure 16

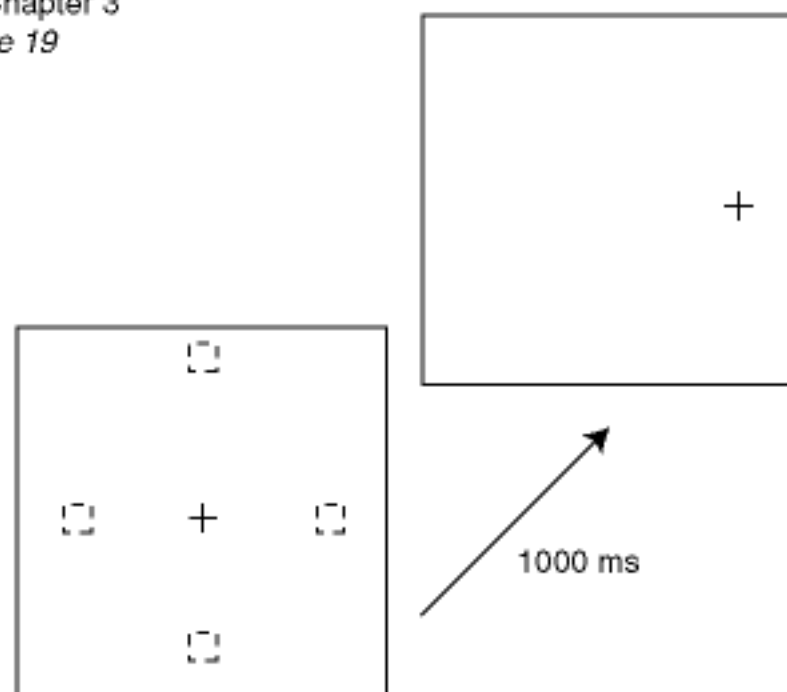




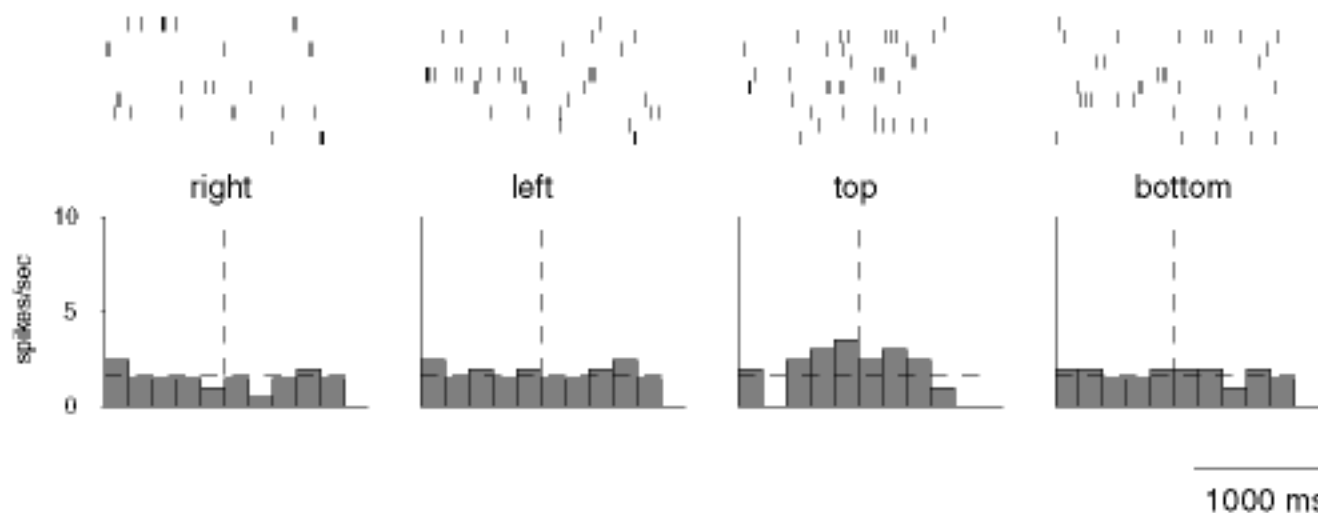
Gabriel Kreiman
Thesis - Chapter 3
Figure 18

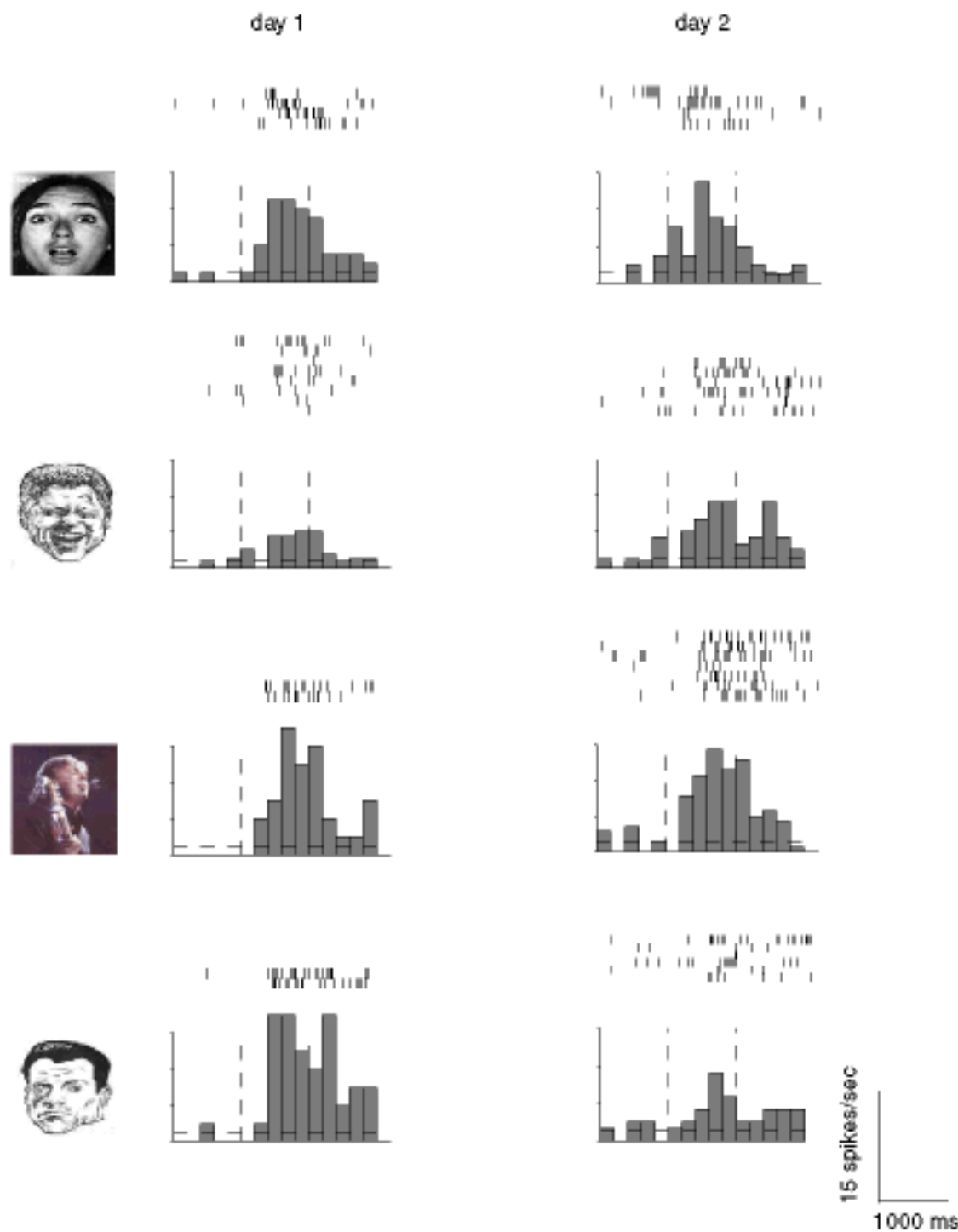


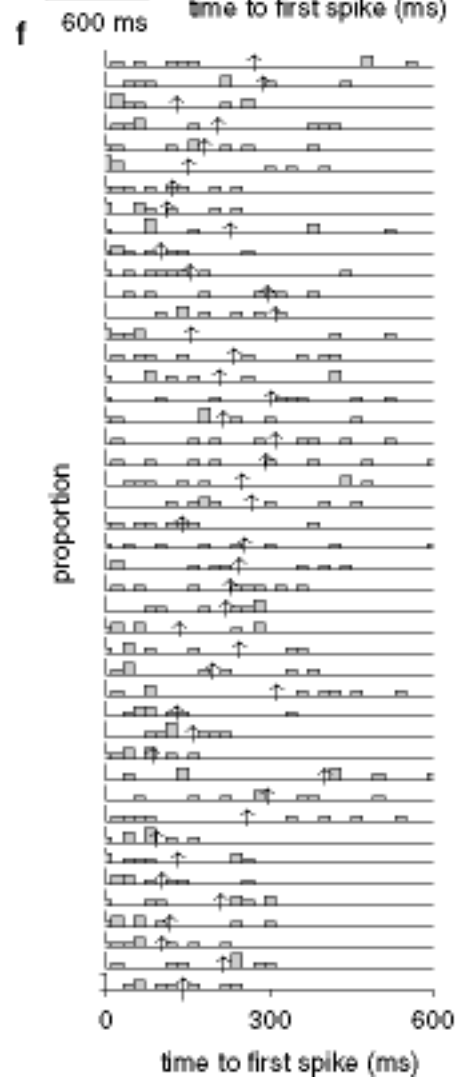
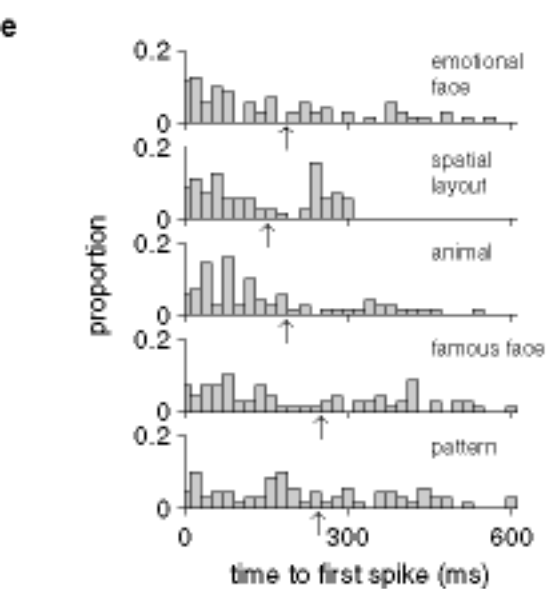
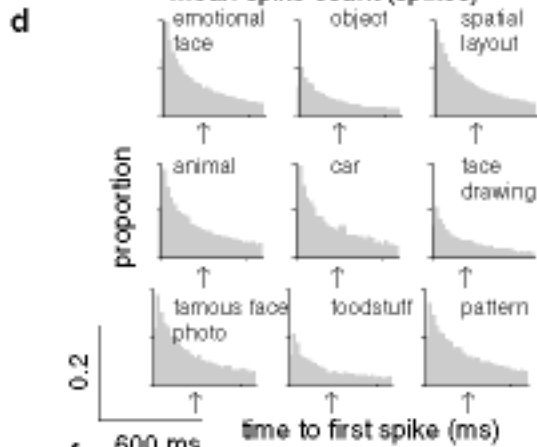
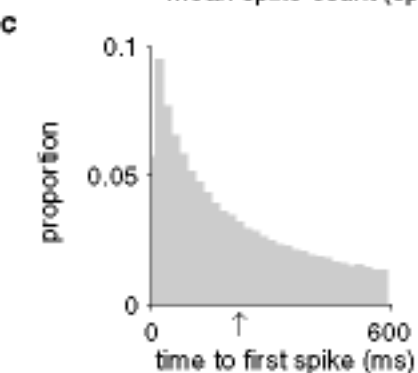
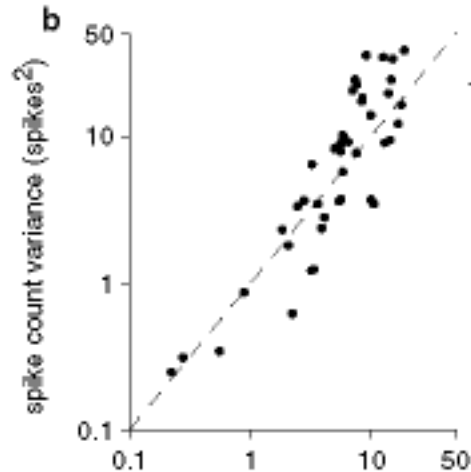
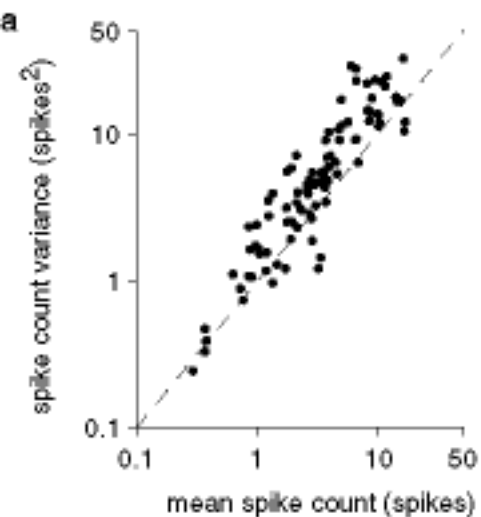
a



b







Gabriel Kreiman
Thesis - Chapter 3
Figure 22

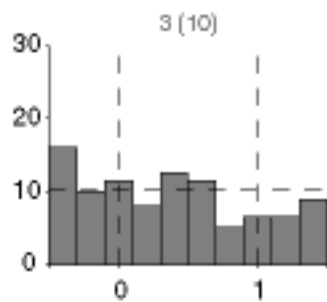
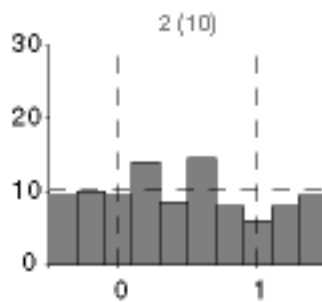
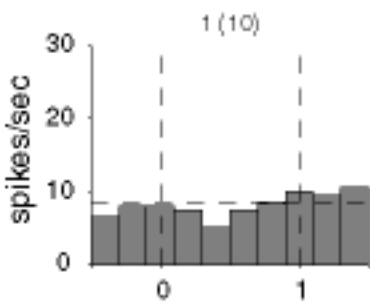
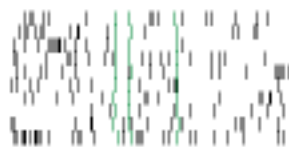
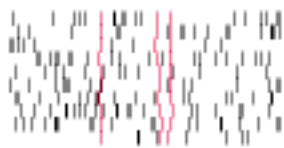
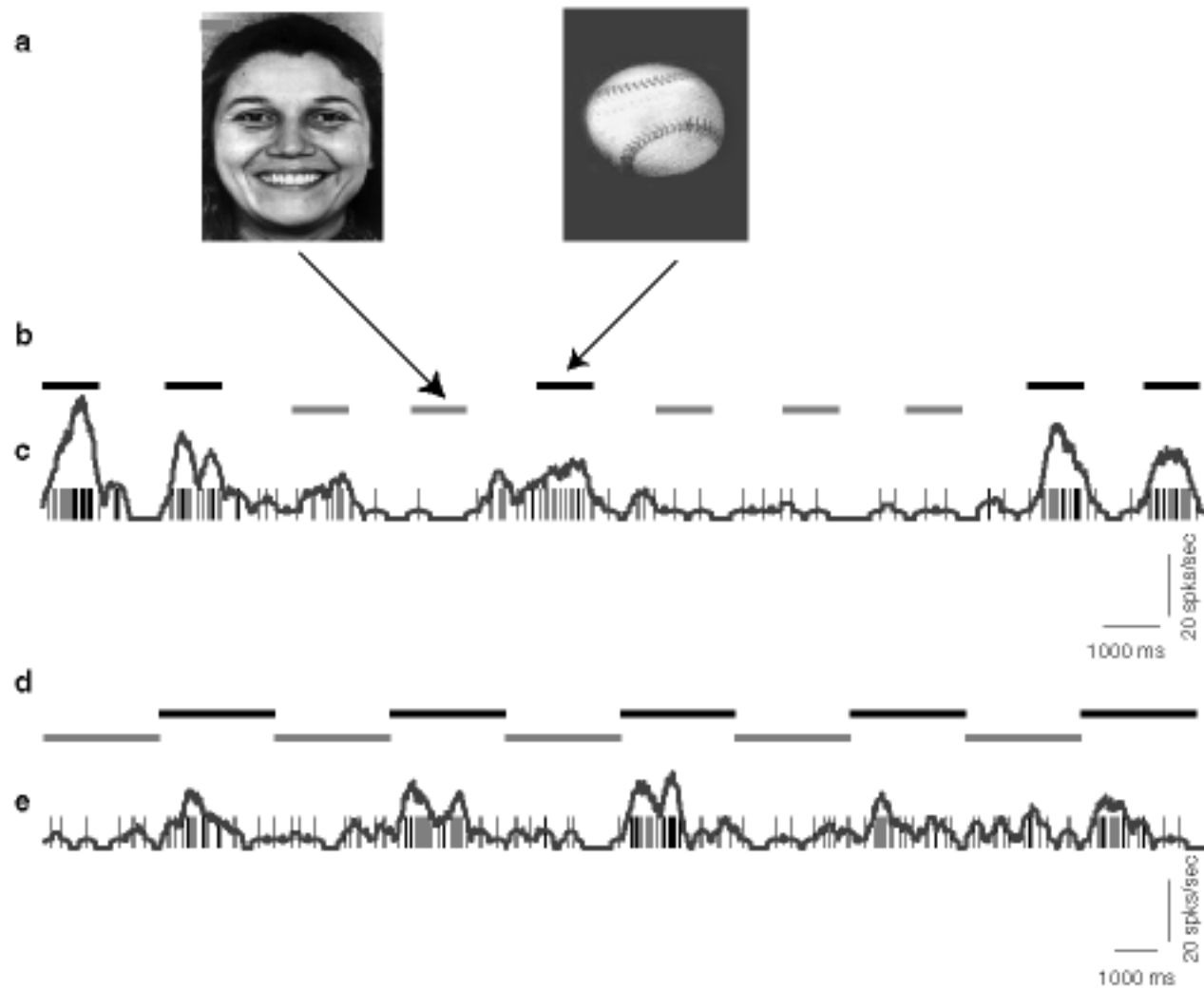
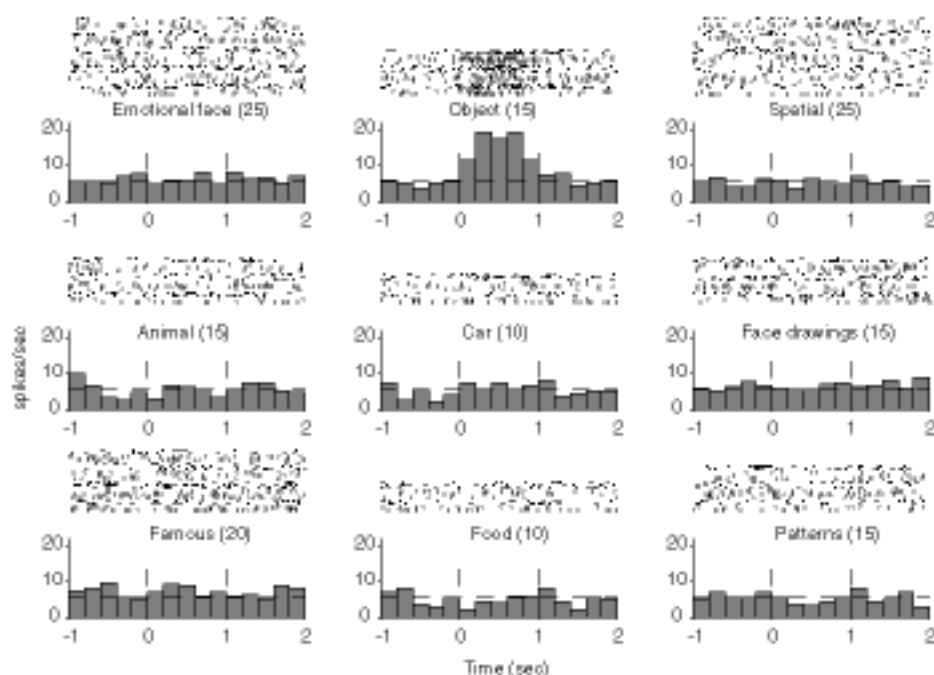


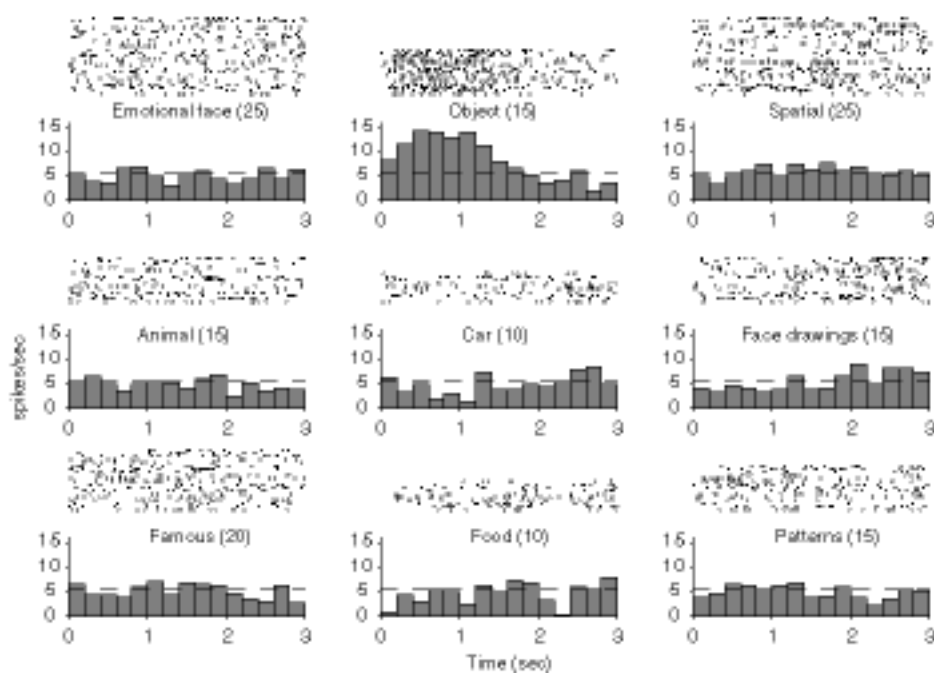
Figure 1



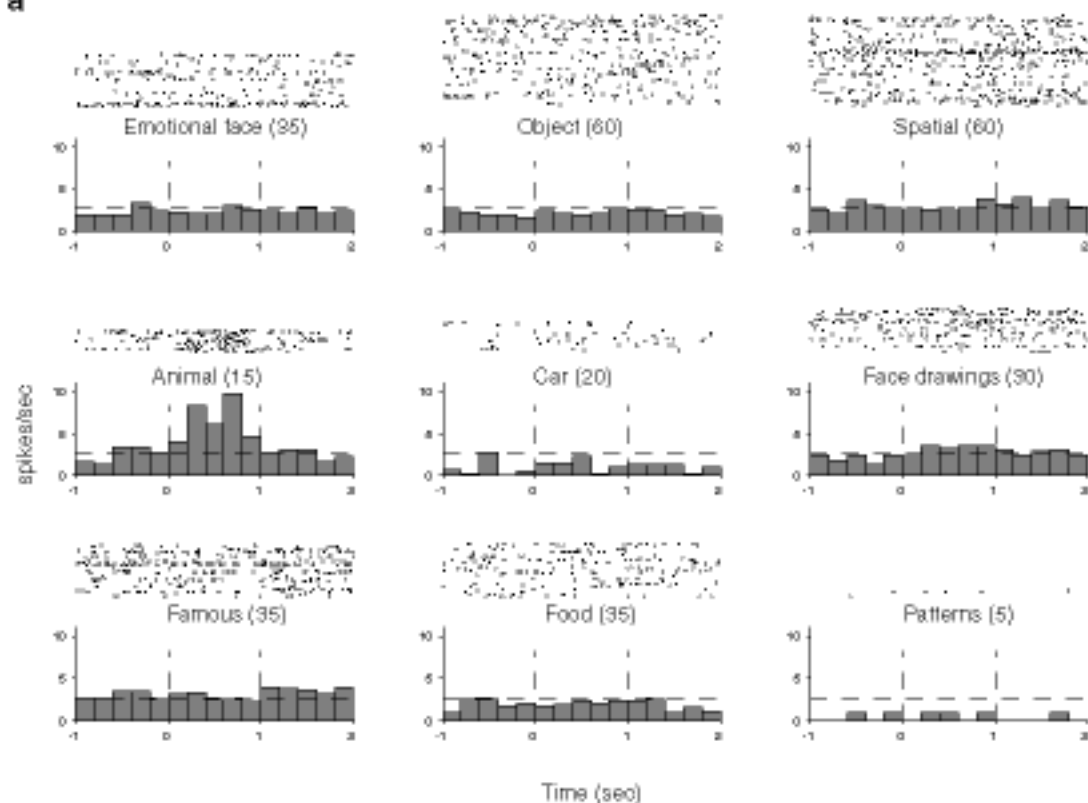
a



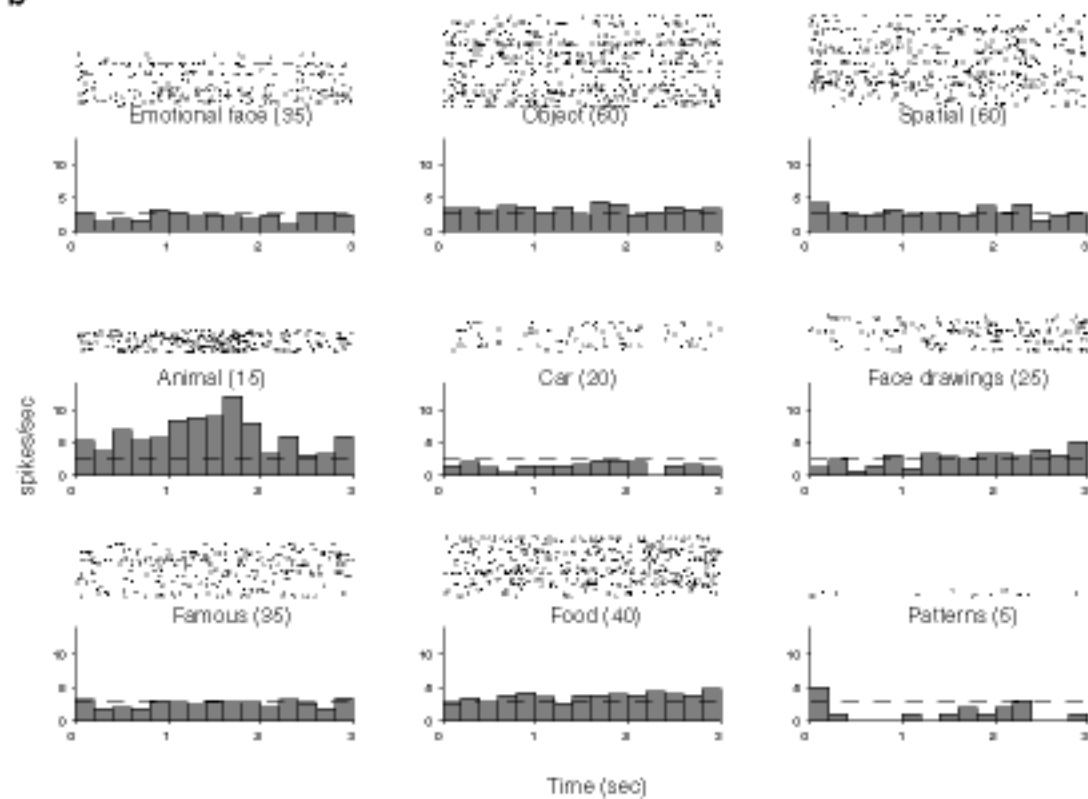
b



a

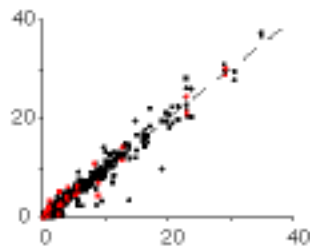
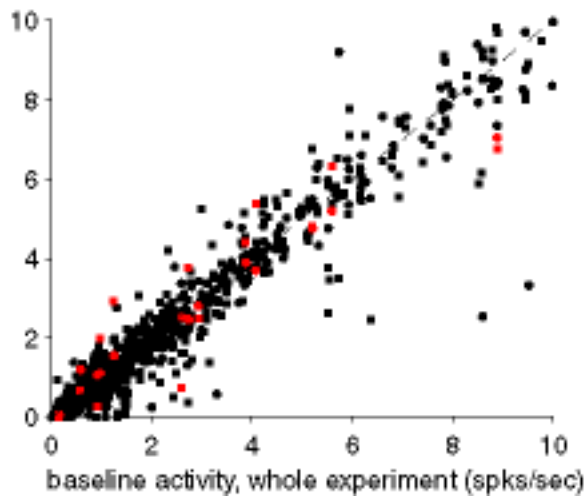


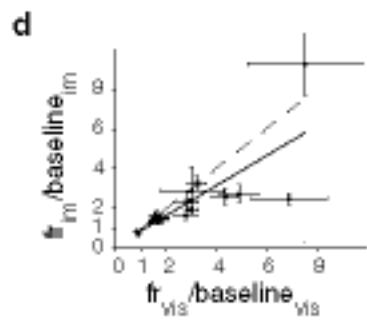
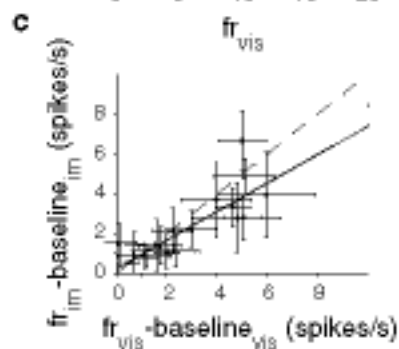
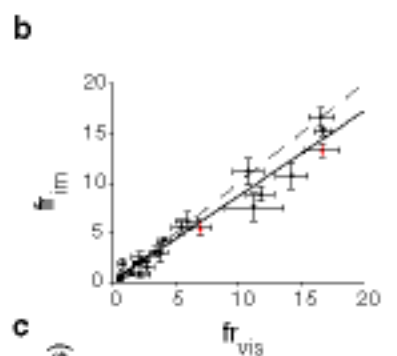
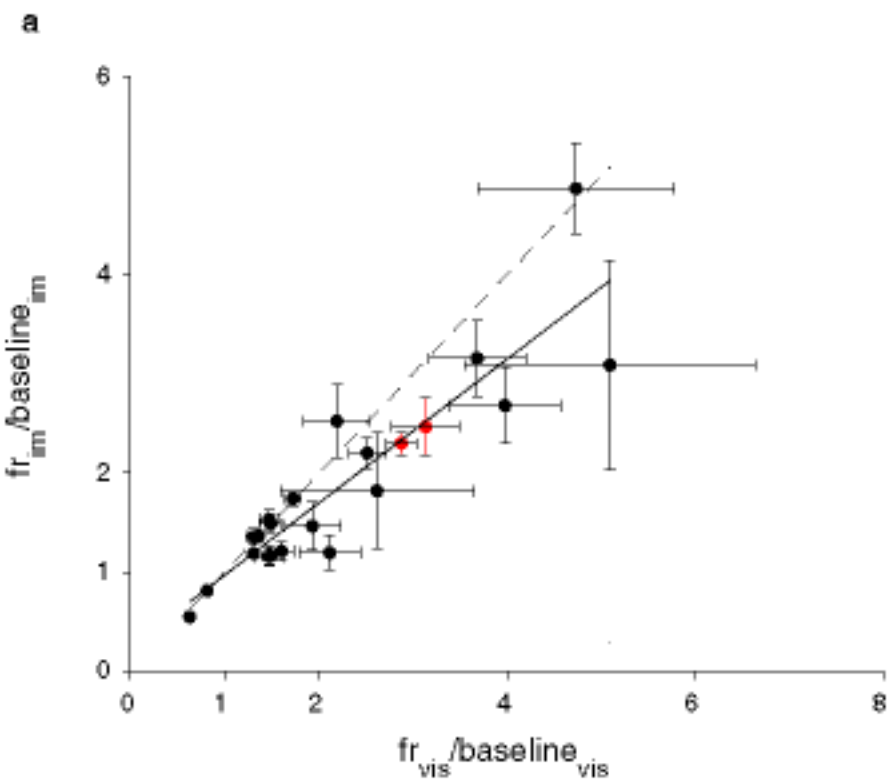
b



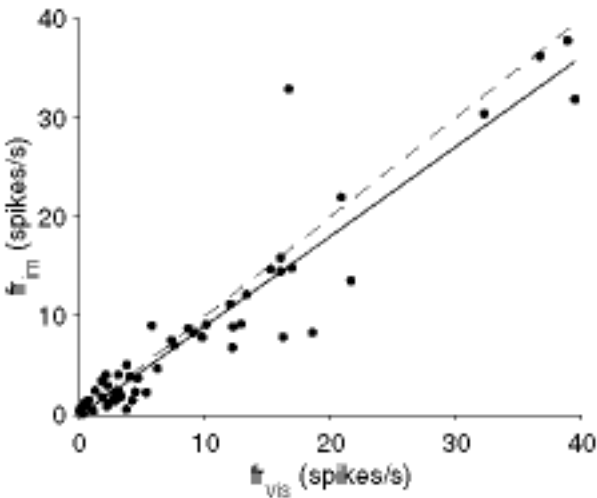
Gabriel Kreiman
Thesis - Chapter 4
Figure 4

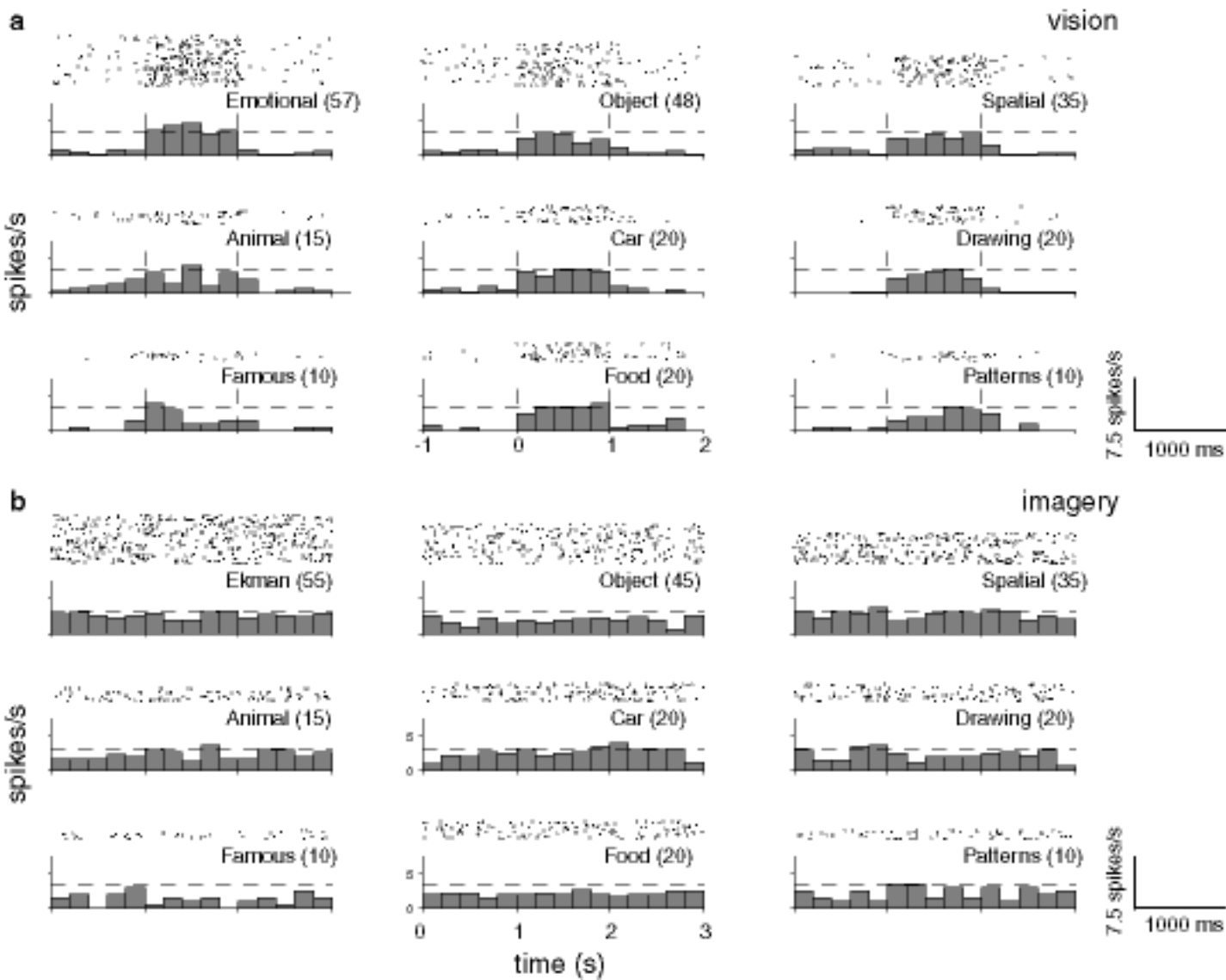
baseline activity, vision or imagery (spks/sec)



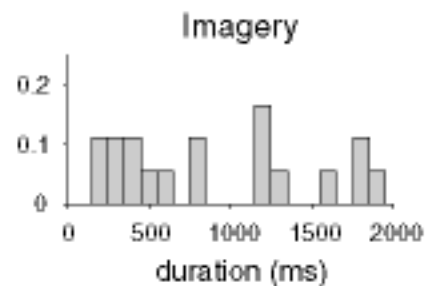
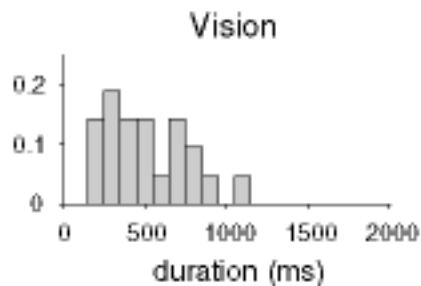
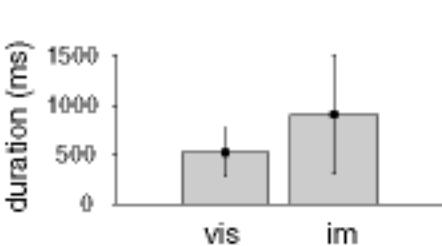


Gabriel Kreiman
Thesis - Chapter 4
Figure 6

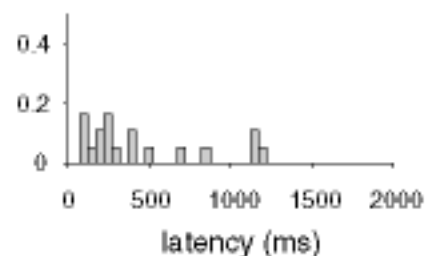
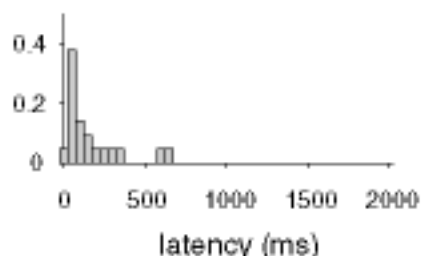
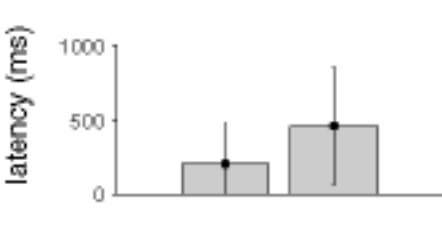




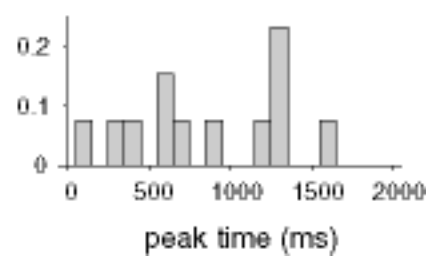
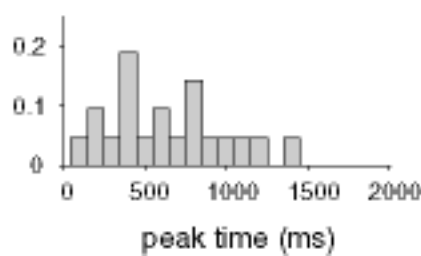
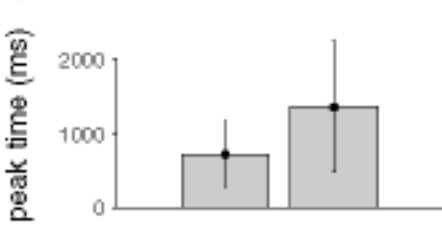
a

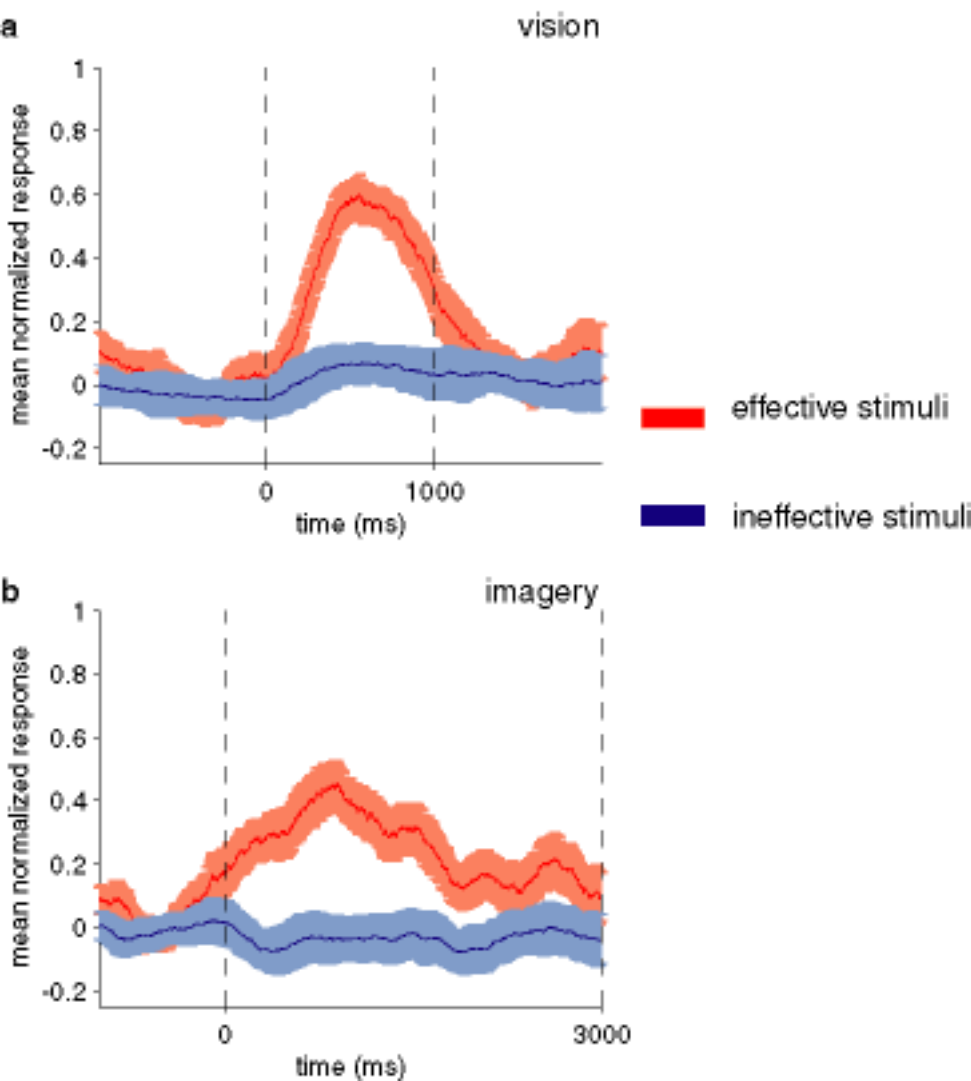


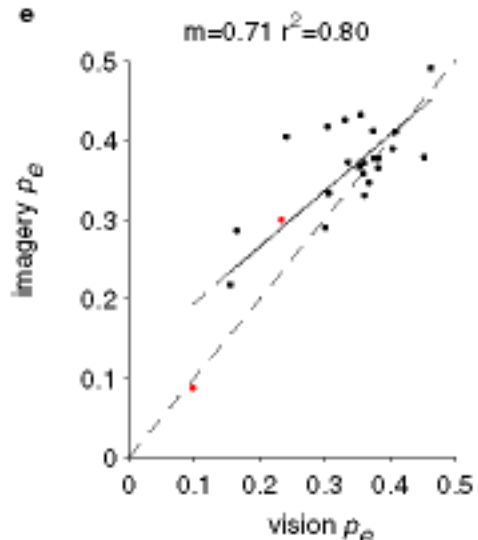
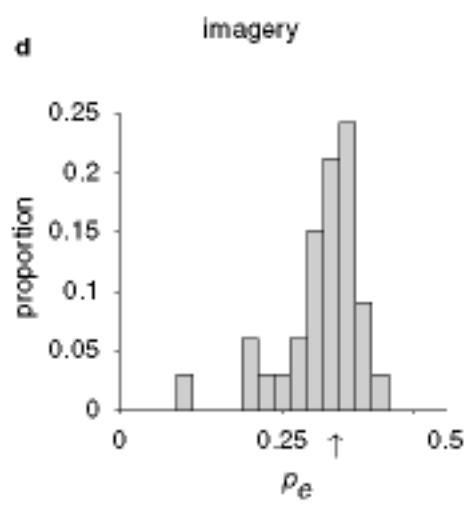
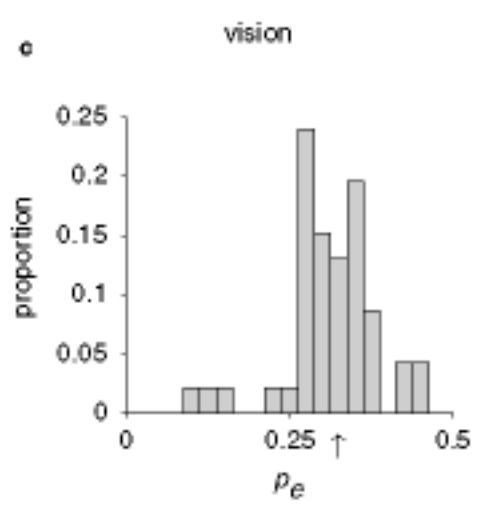
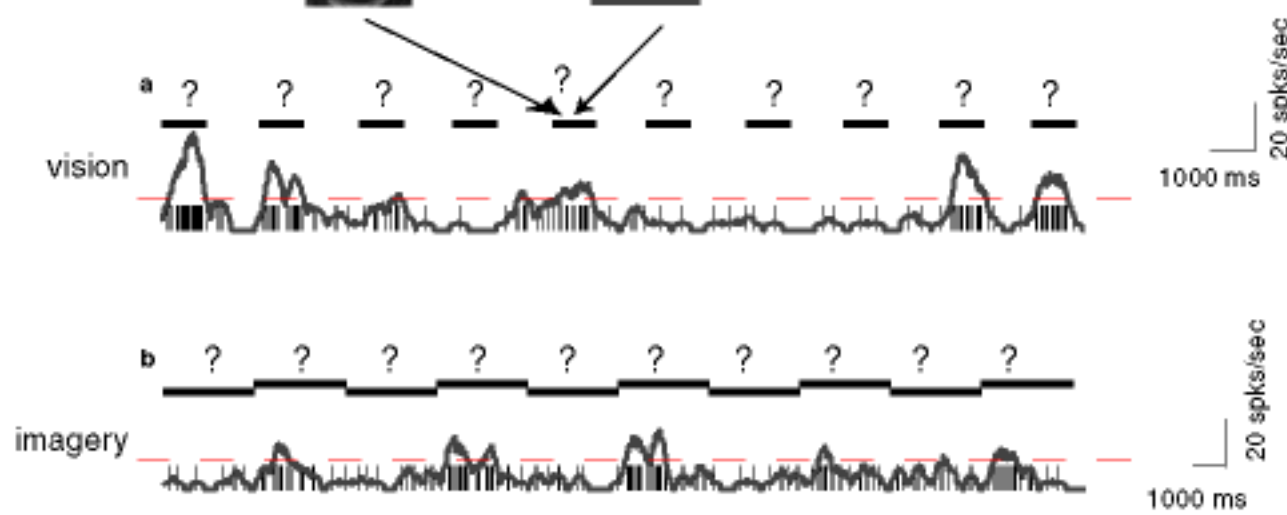
b



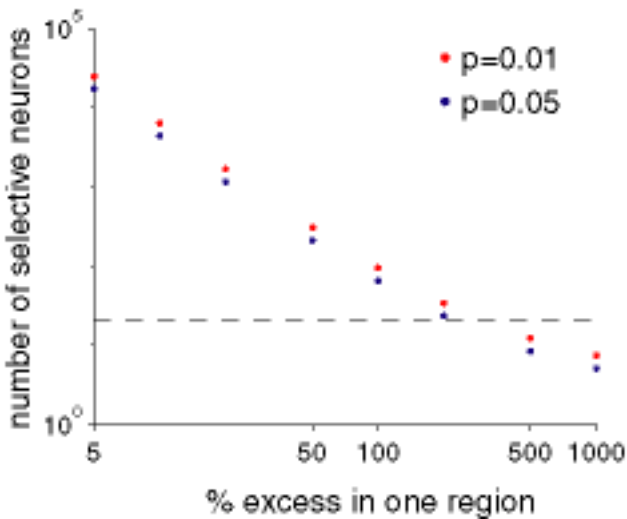
c



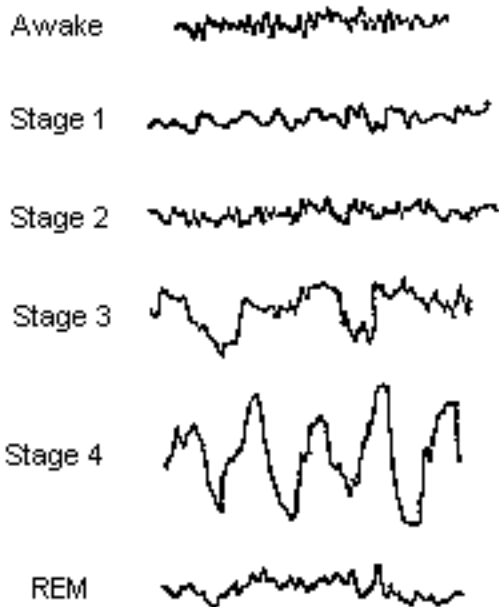


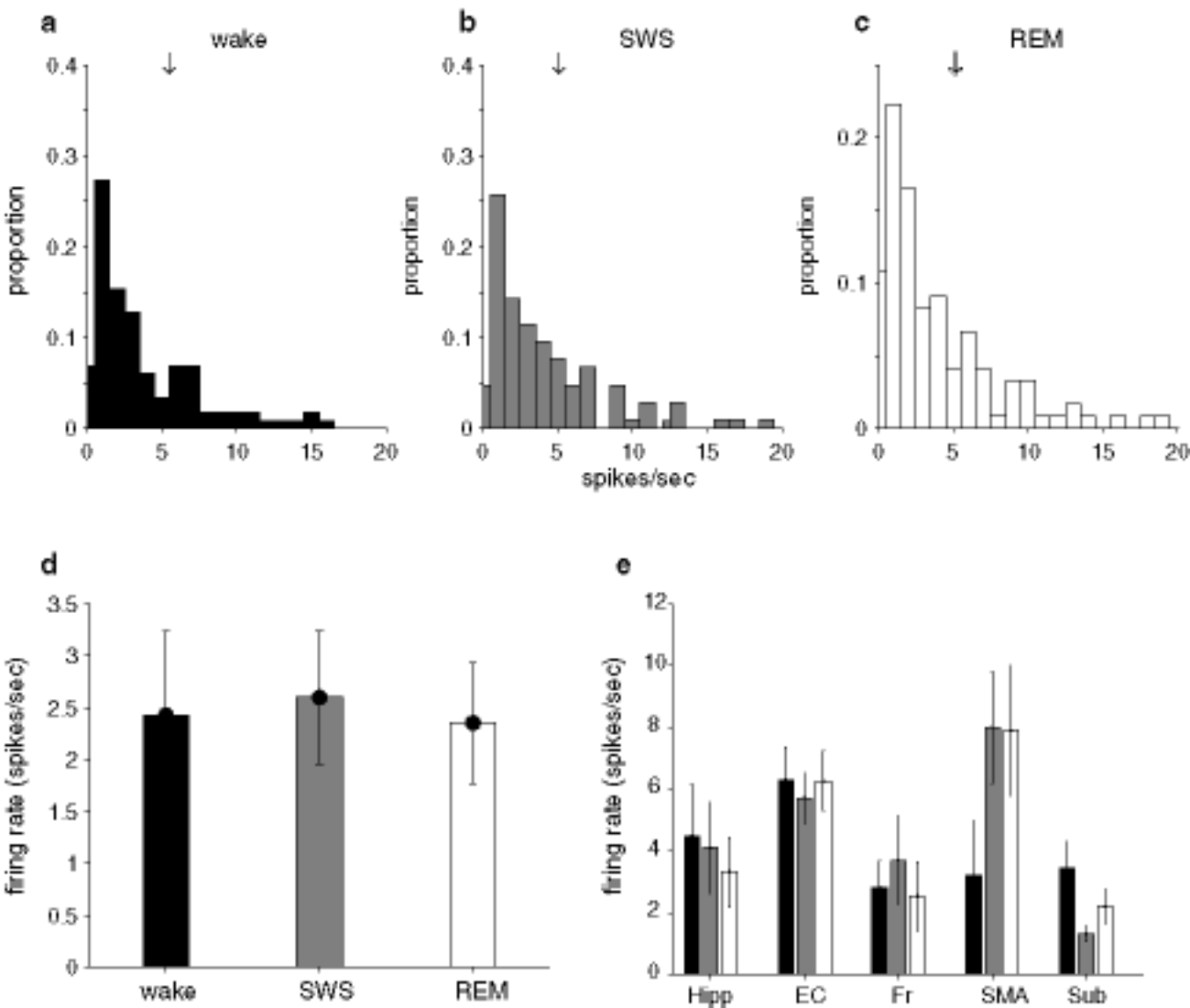


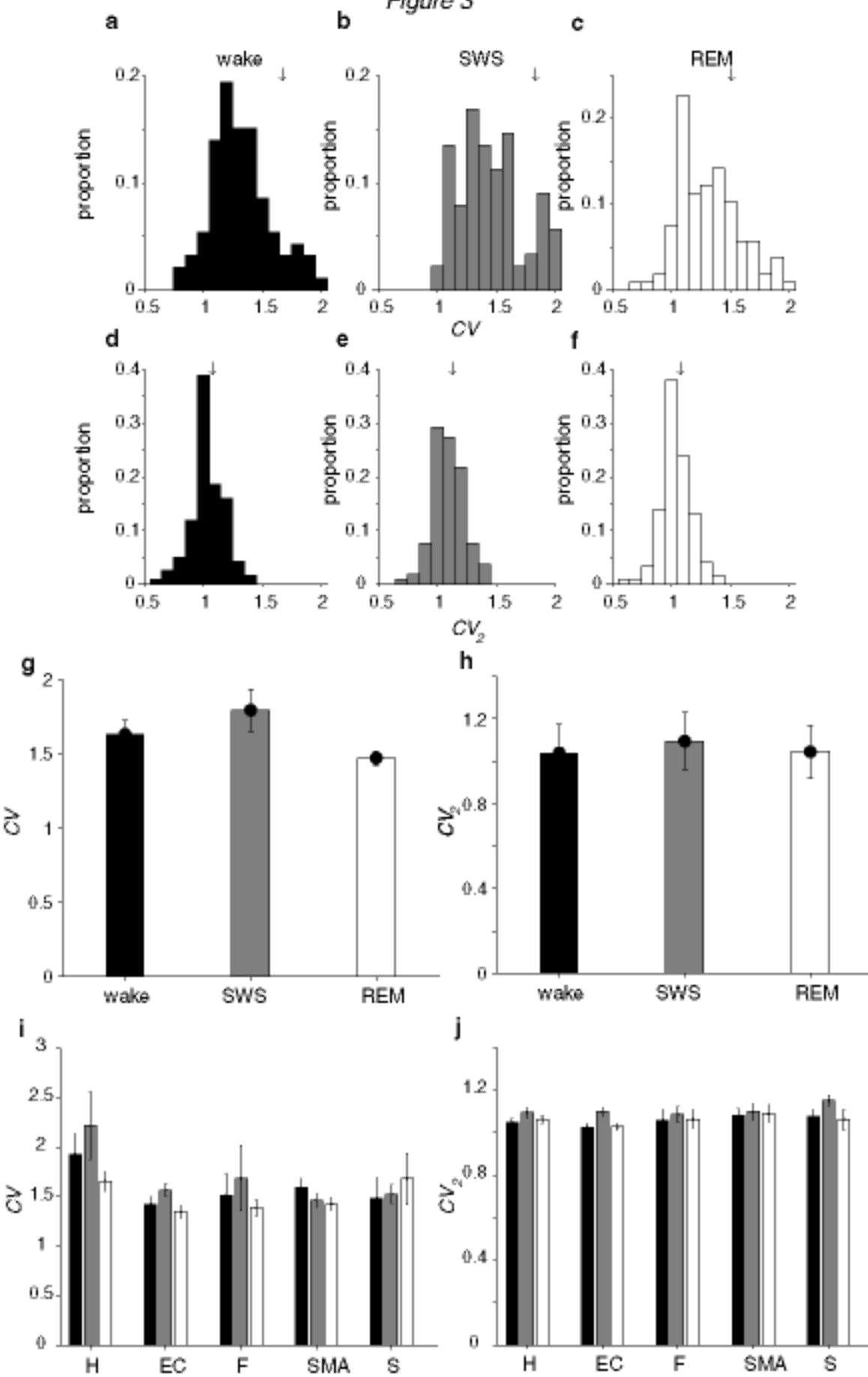
Gabriel Kreiman
Thesis - Chapter 4
Figure 11

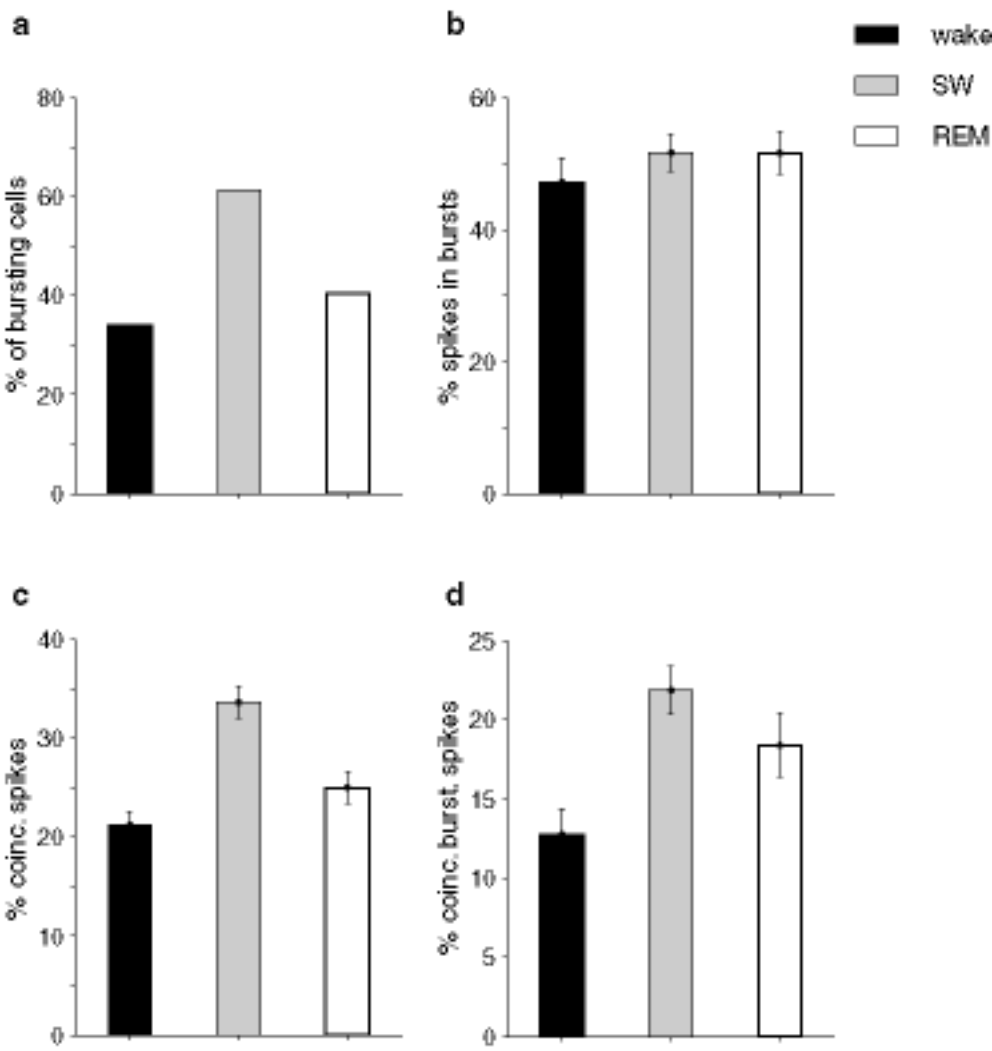


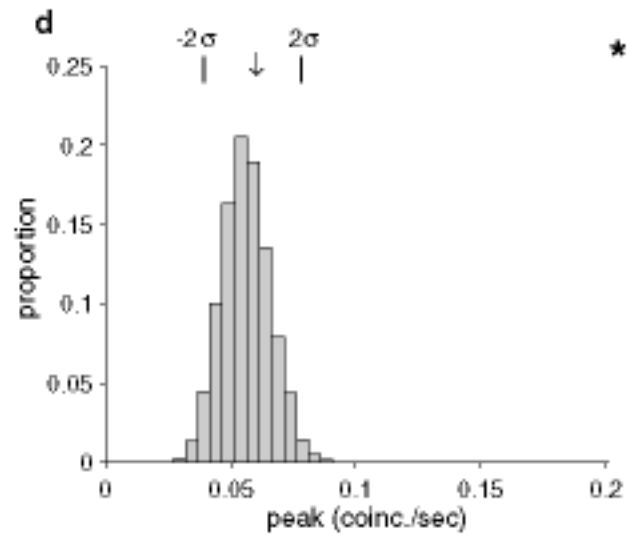
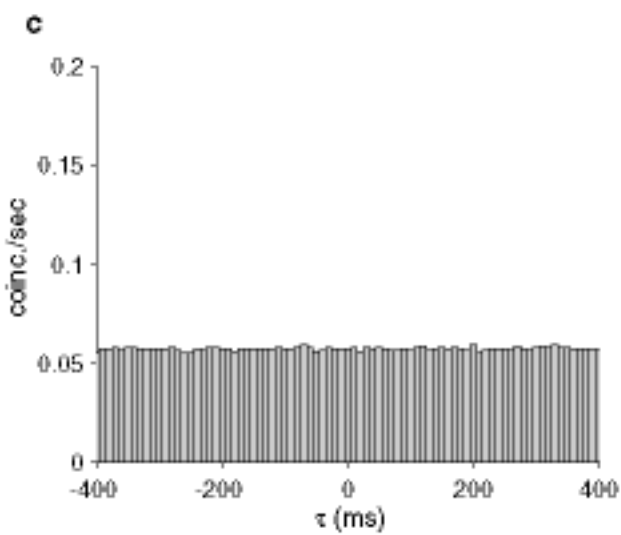
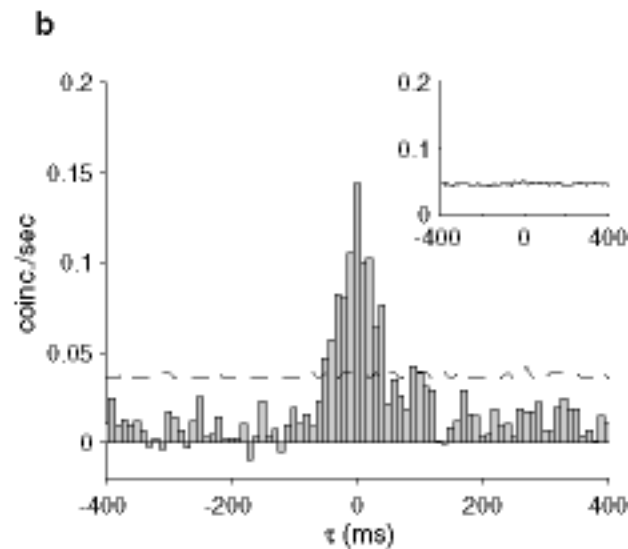
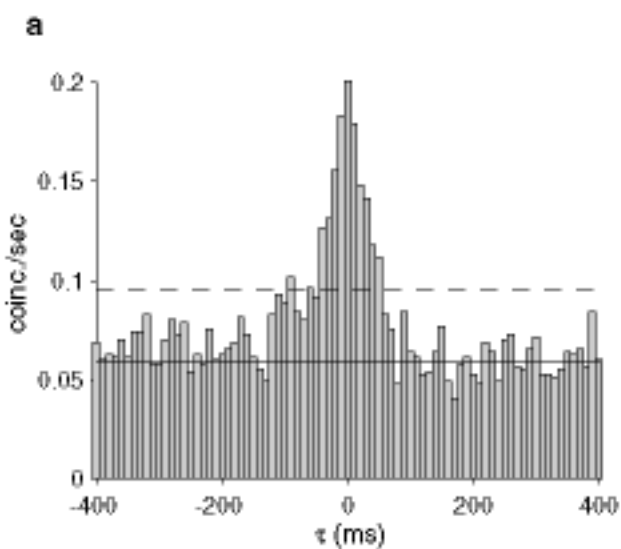
Gabriel Kreiman
Chapter 5
Figure 1

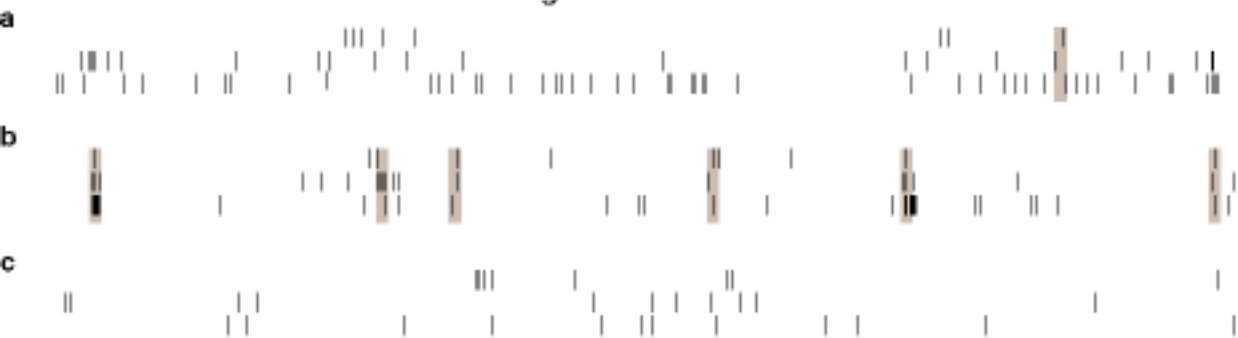






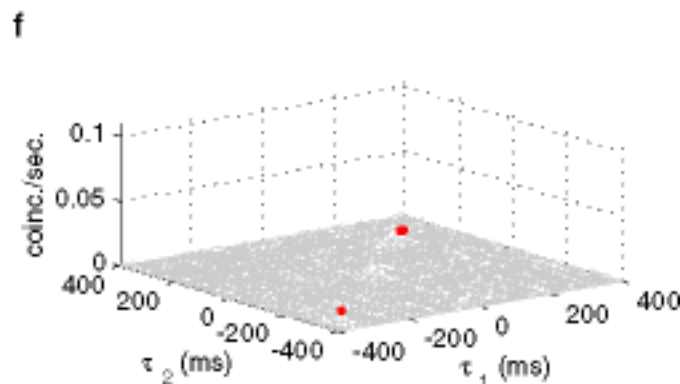
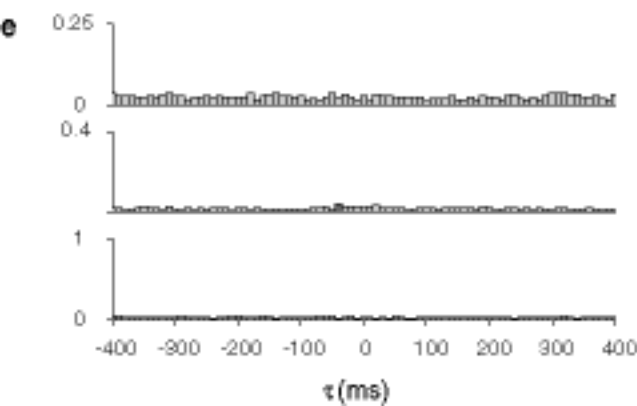
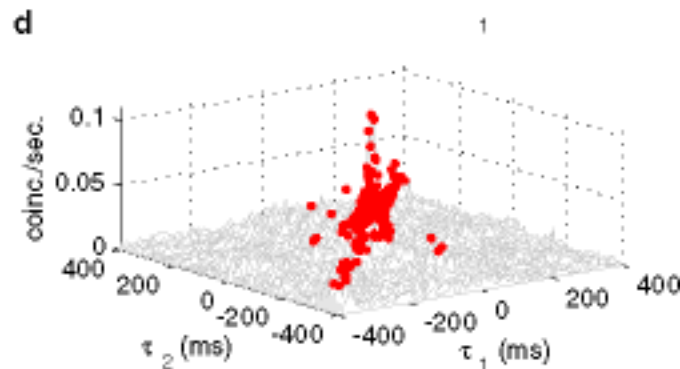
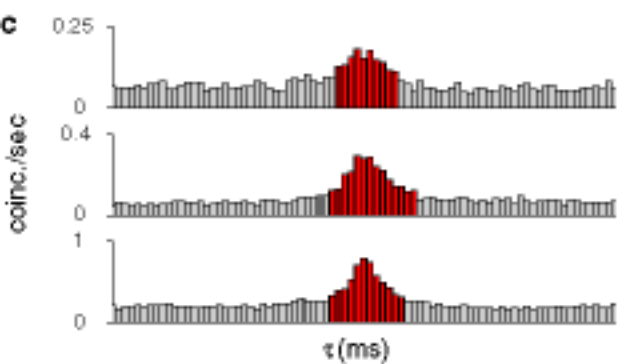
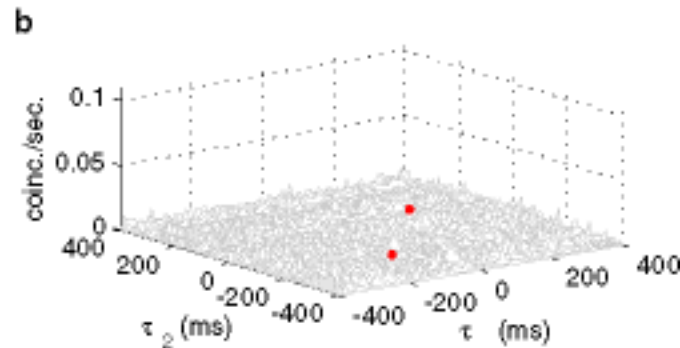
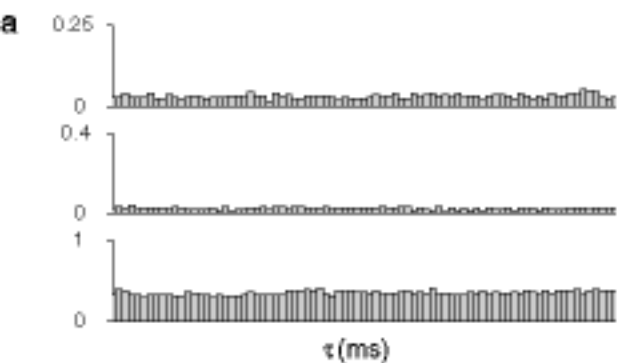




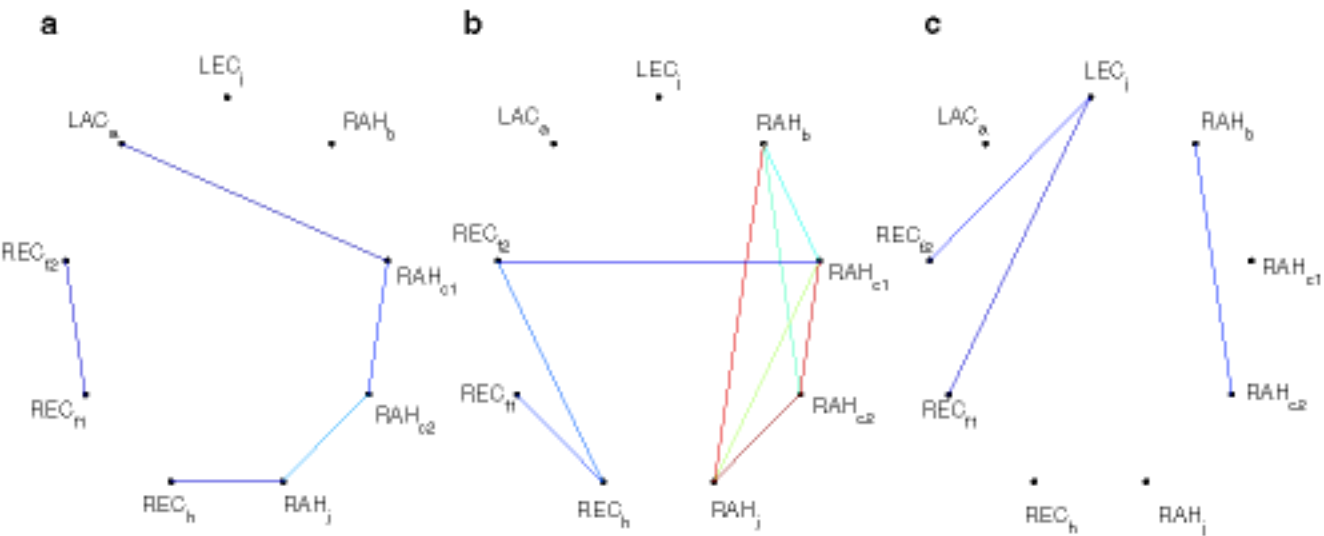


1000 ms

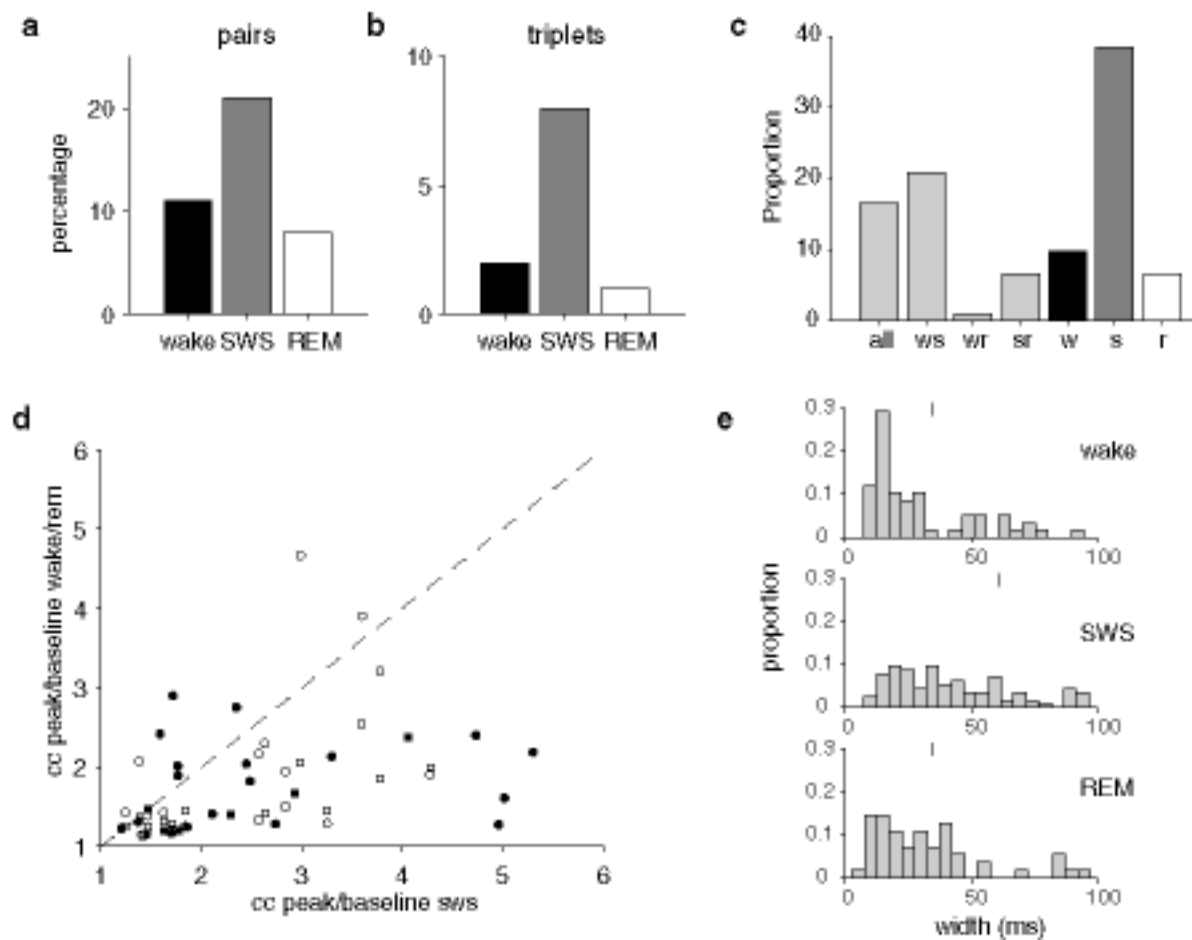
Gabriel Kreiman
Chapter 5
Figure 7

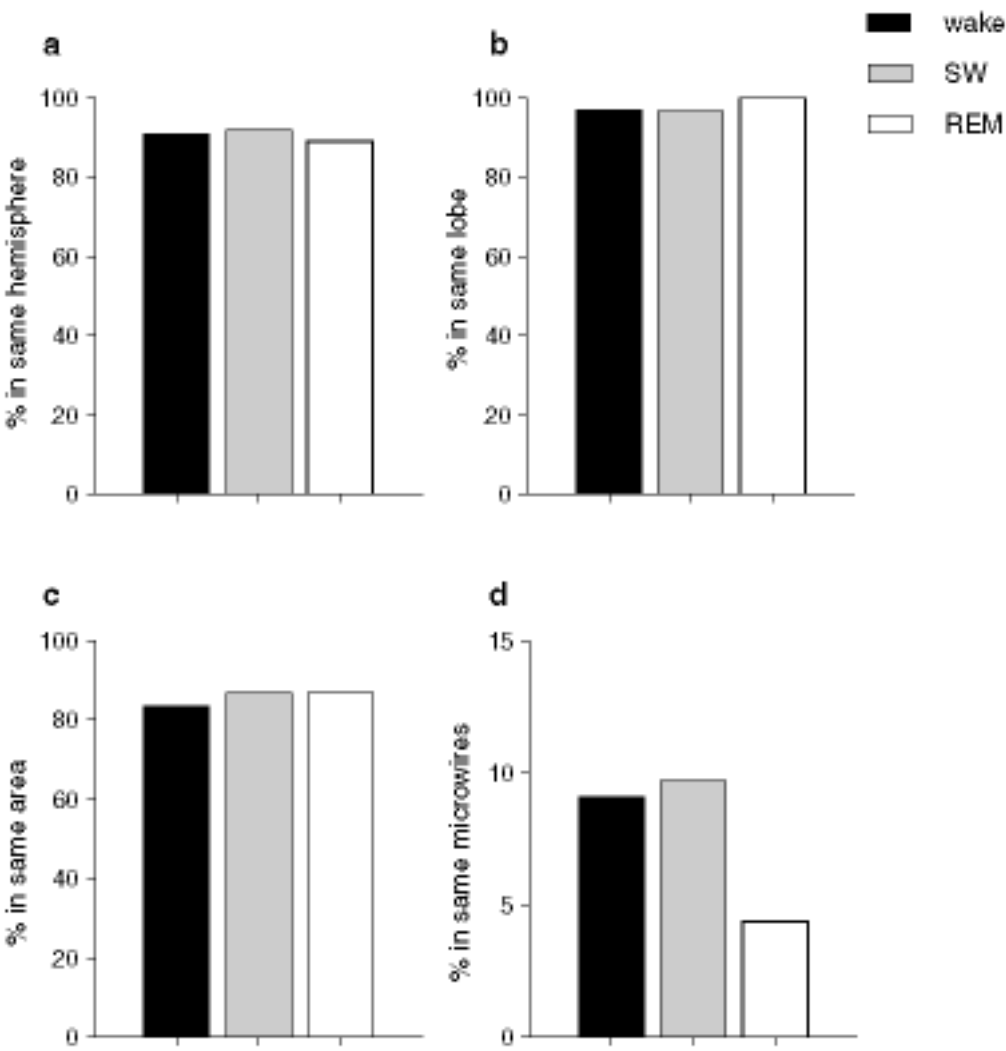


Gabriel Kreiman
 Chapter 5
 Figure 8

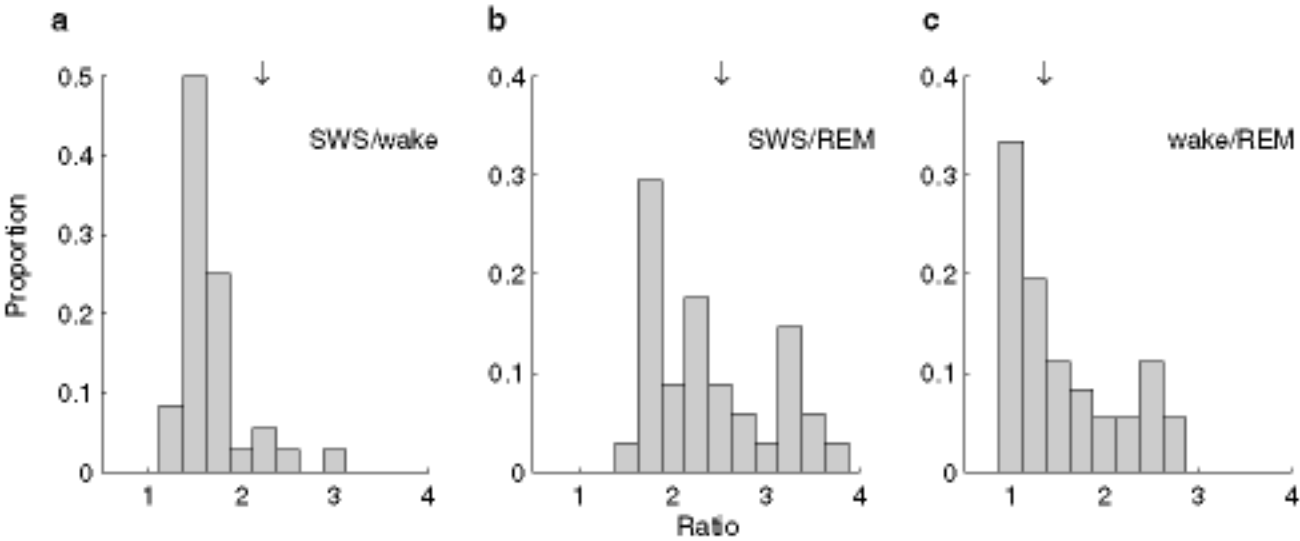


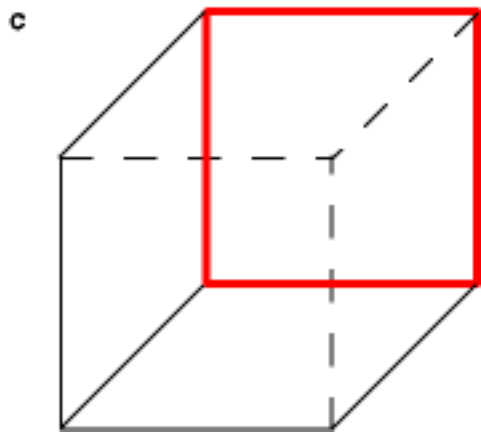
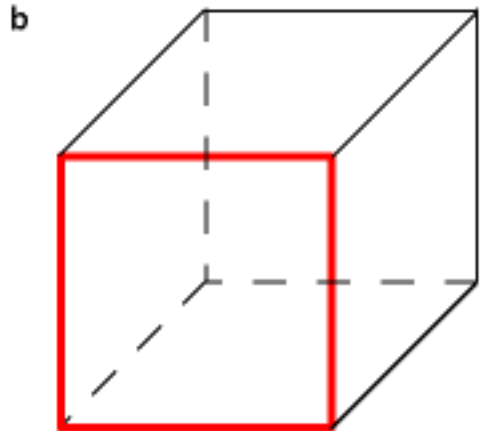
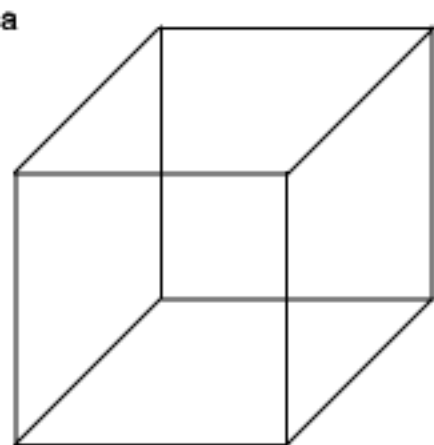
Gabriel Kreiman
Chapter 5
Figure 9



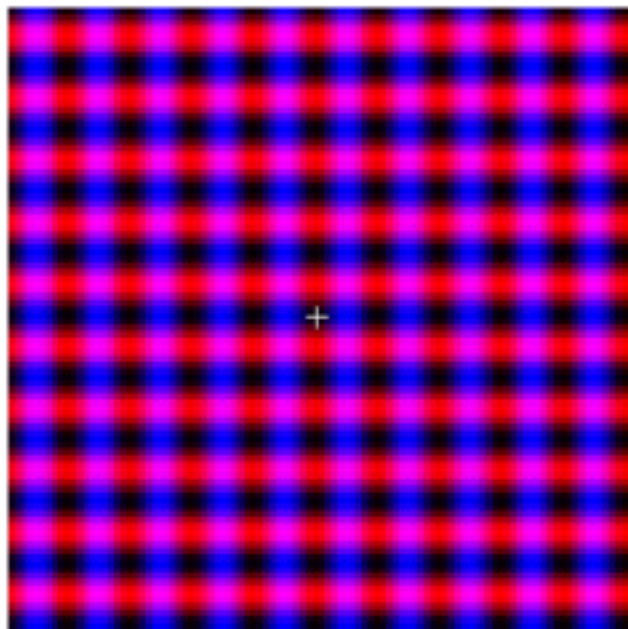


Gabriel Kreiman
Chapter 5
Figure 11

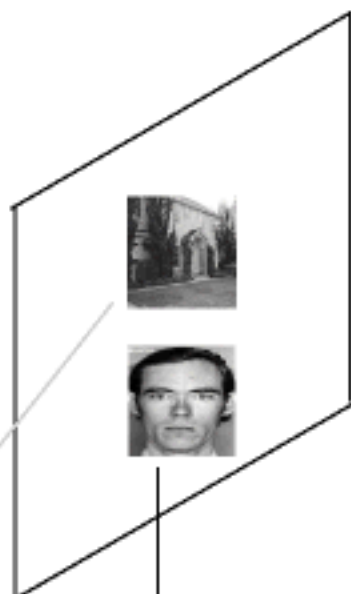




Gabriel Kreiman
Thesis - Chapter 6
Figure 2



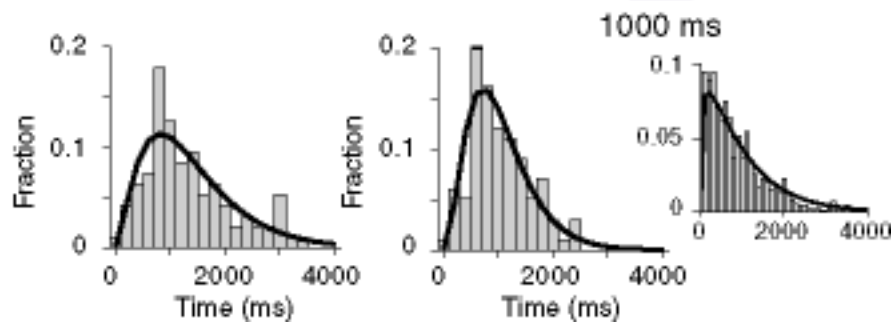
a

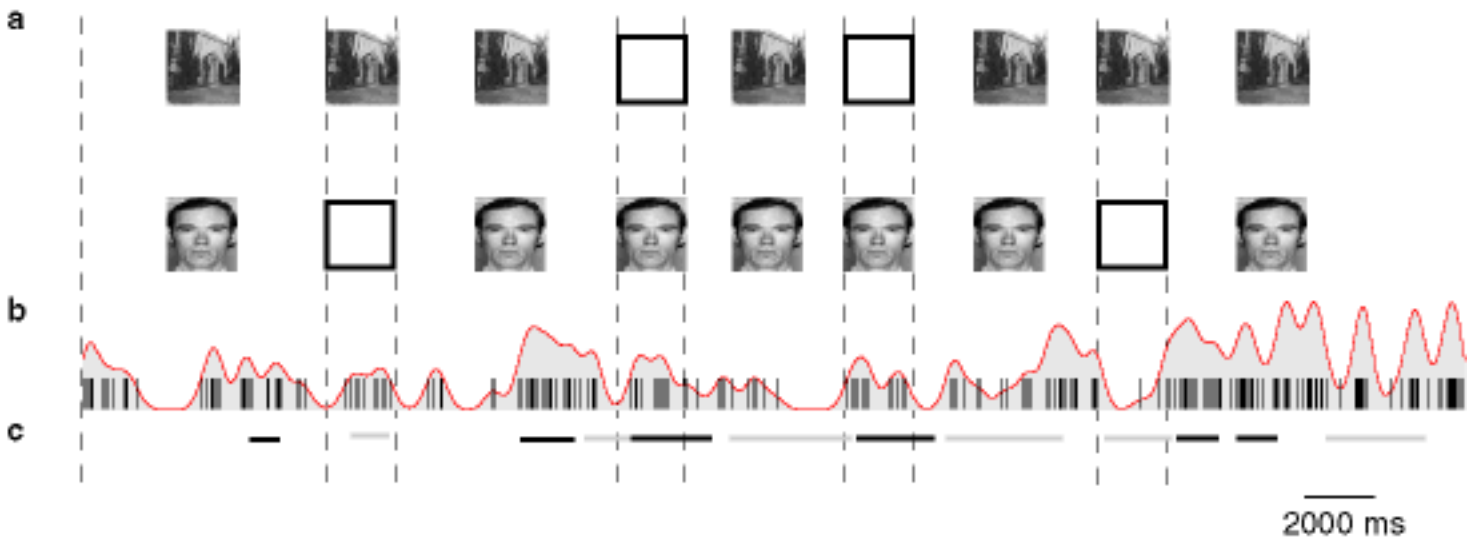


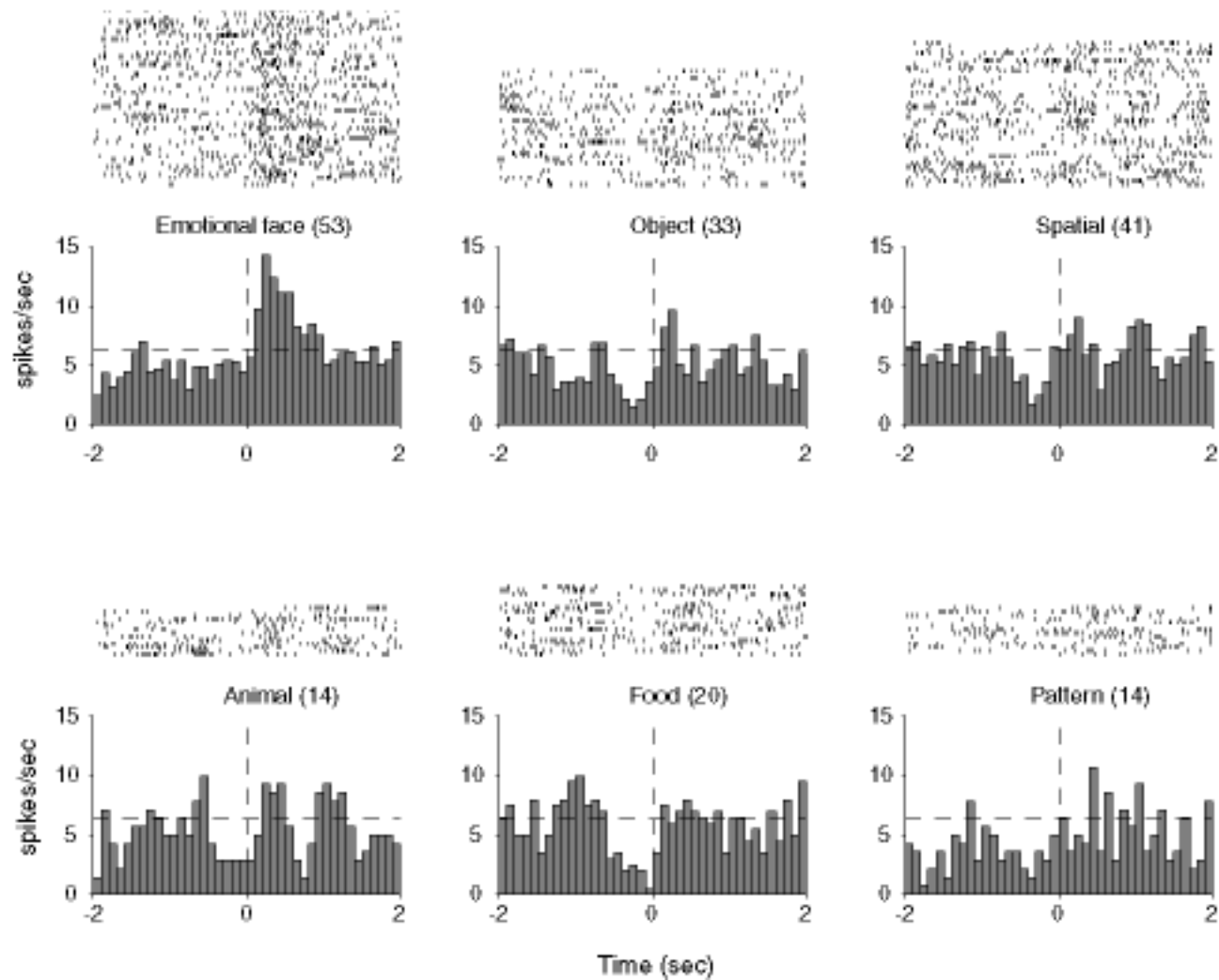
b



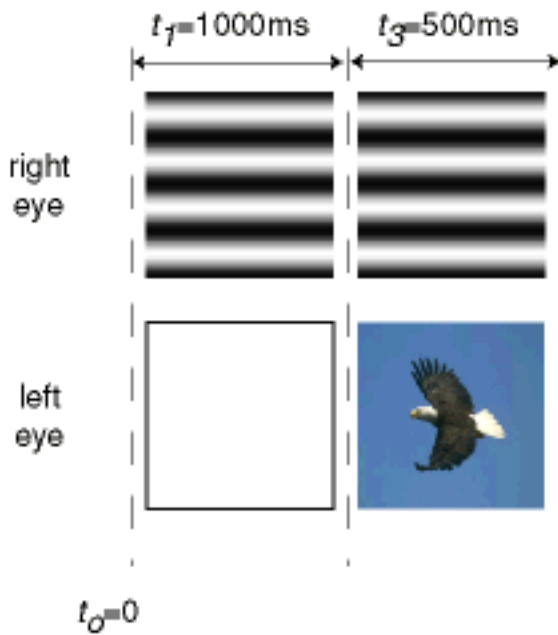
c



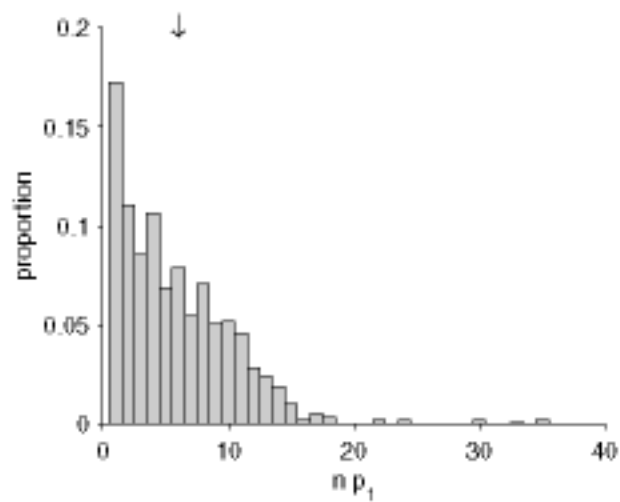




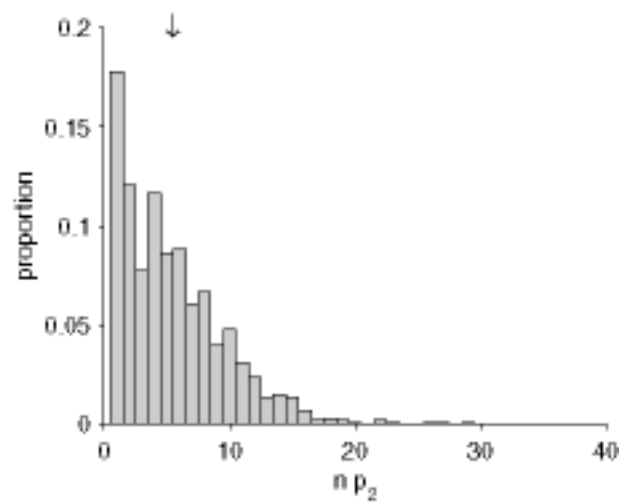
Gabriel Kreiman
Thesis - Chapter 6
Figure 6



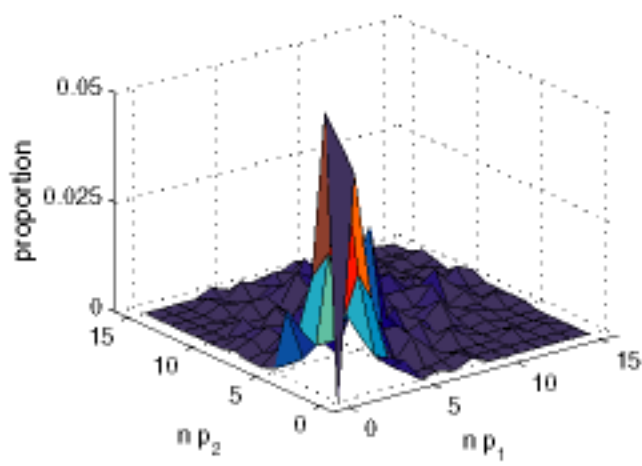
a

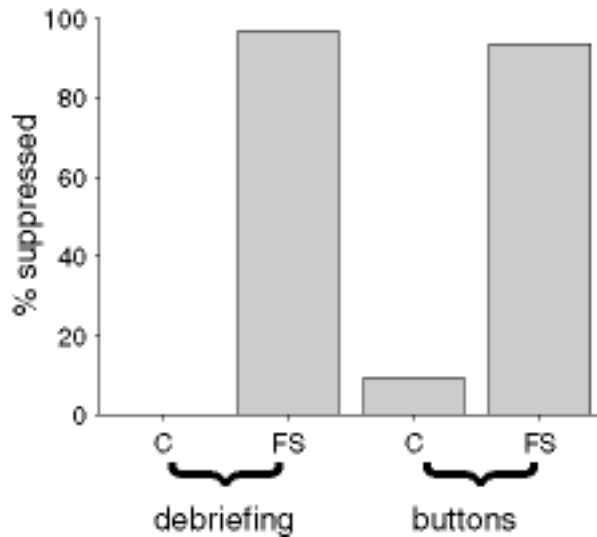


b



c





a

time →

right
eye



left
eye



stimulus

neuronal
response



perception



b

time →

right
eye



left
eye



stimulus

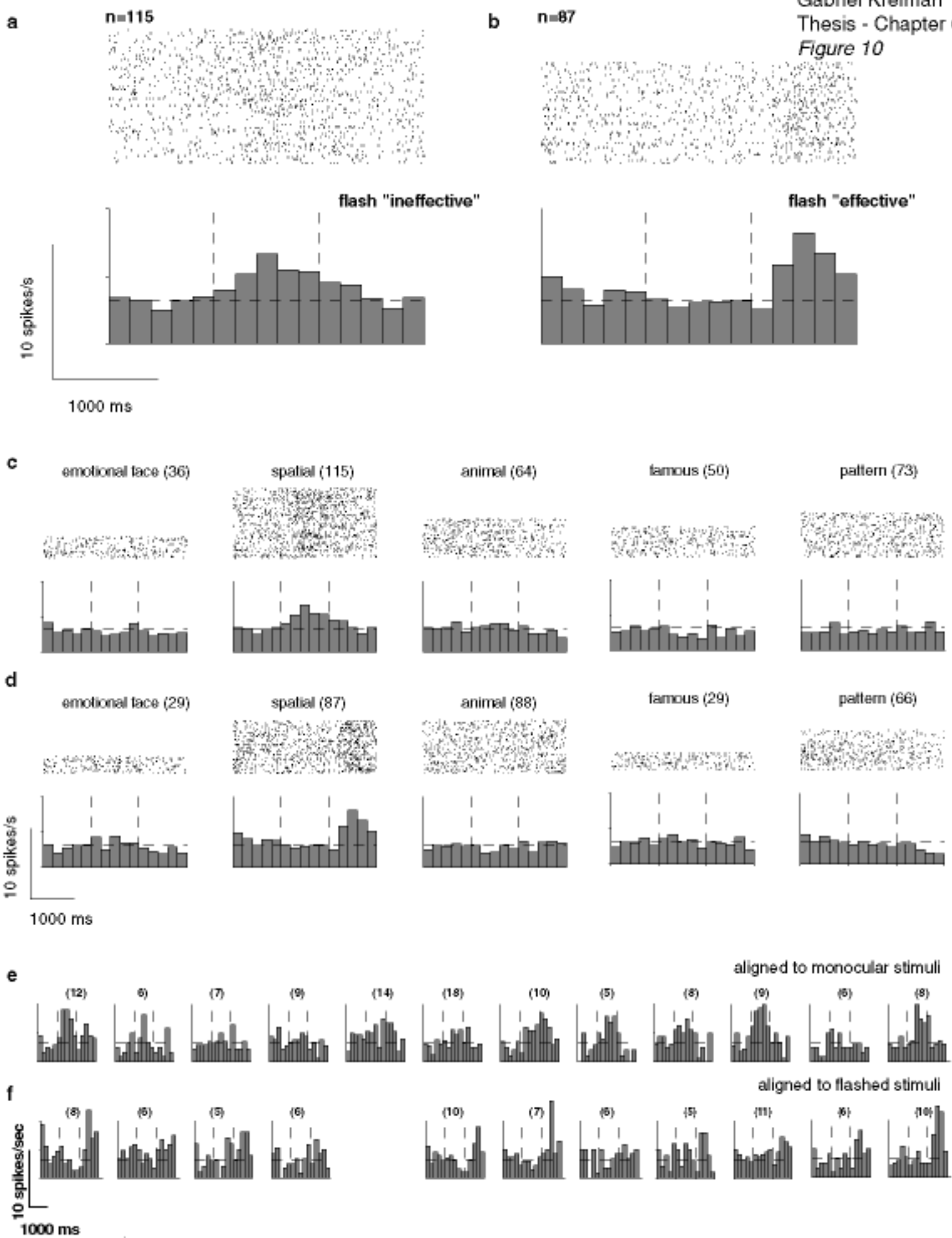
neuronal
response



perception



1000 ms



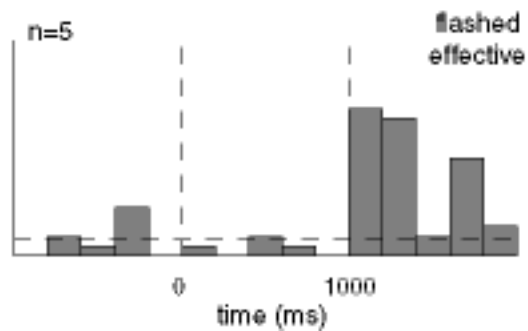
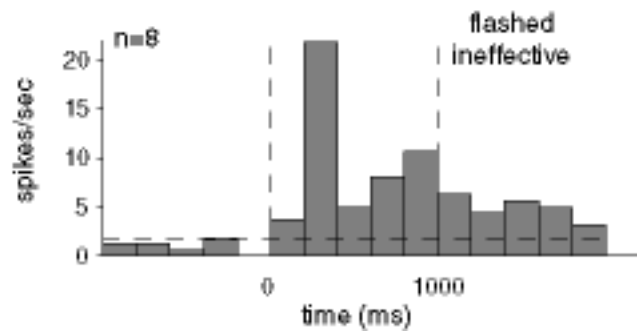
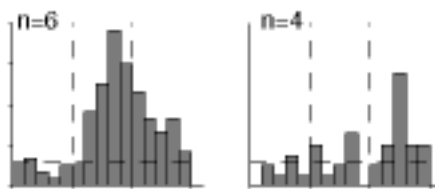
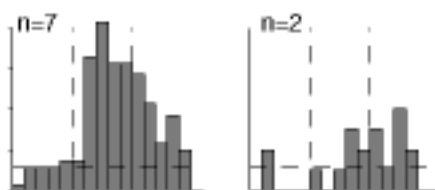
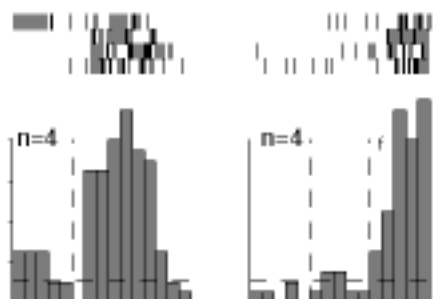


Figure 12

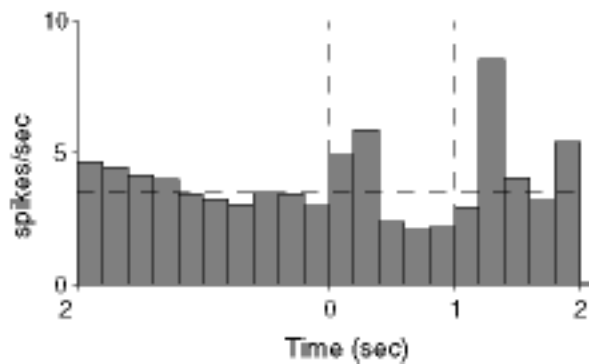


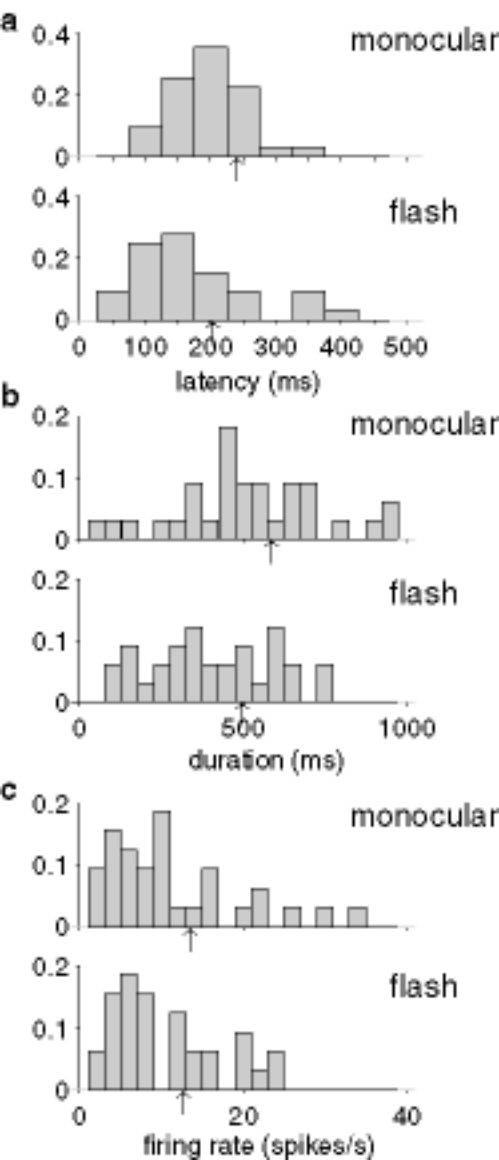
20 spikes/s
1000 ms

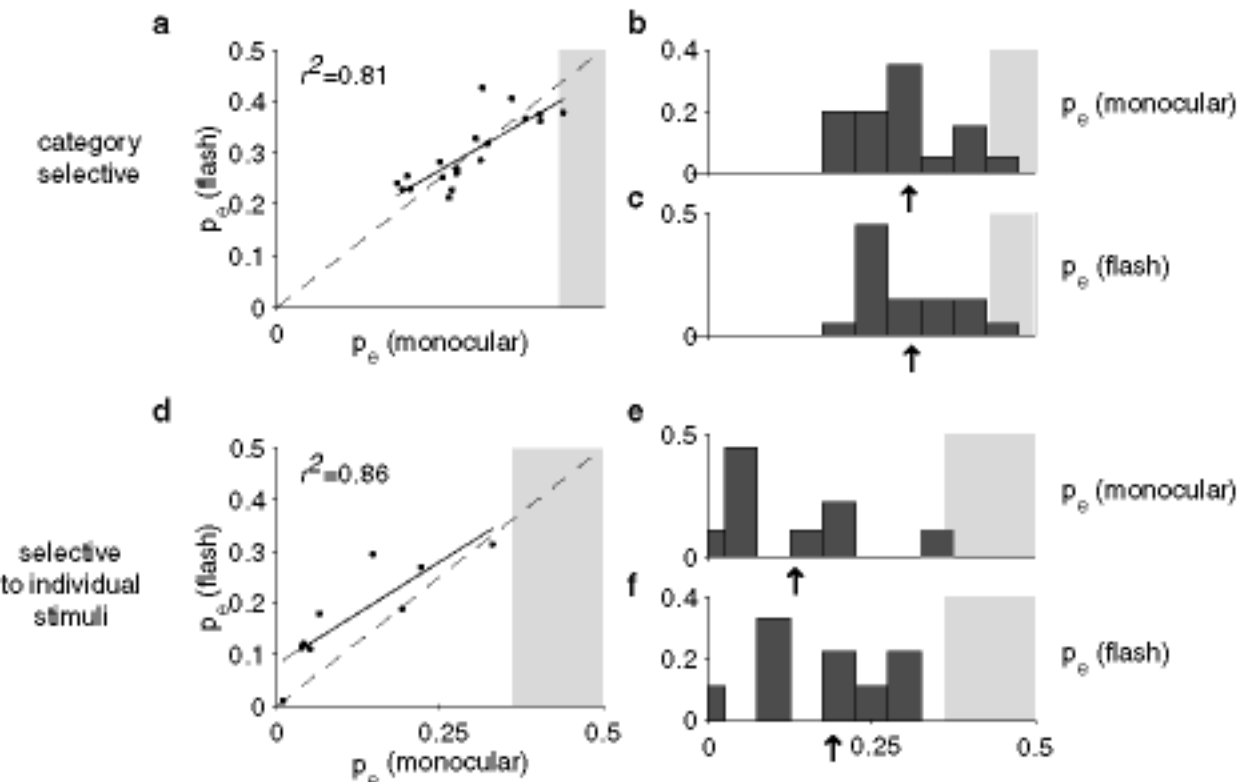
flashed
ineffective

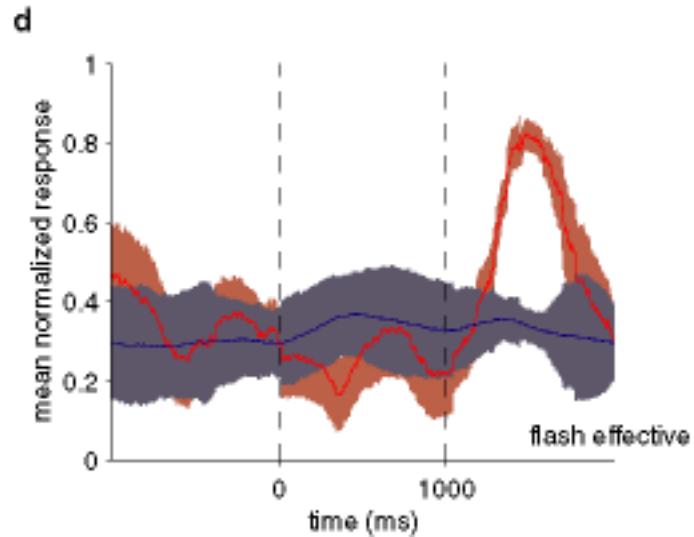
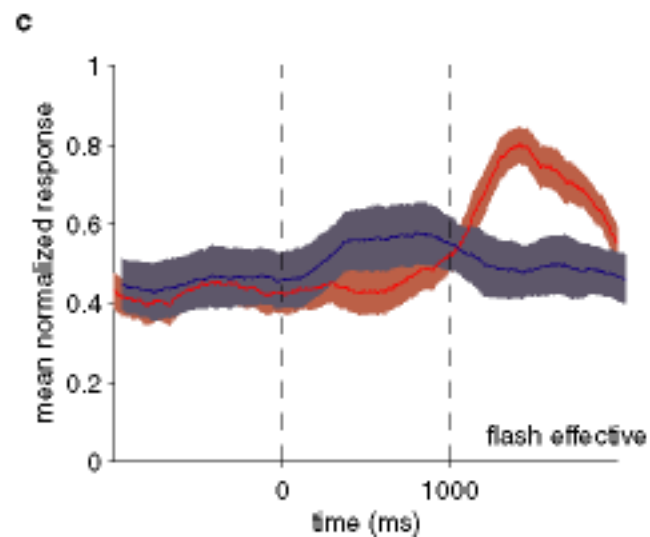
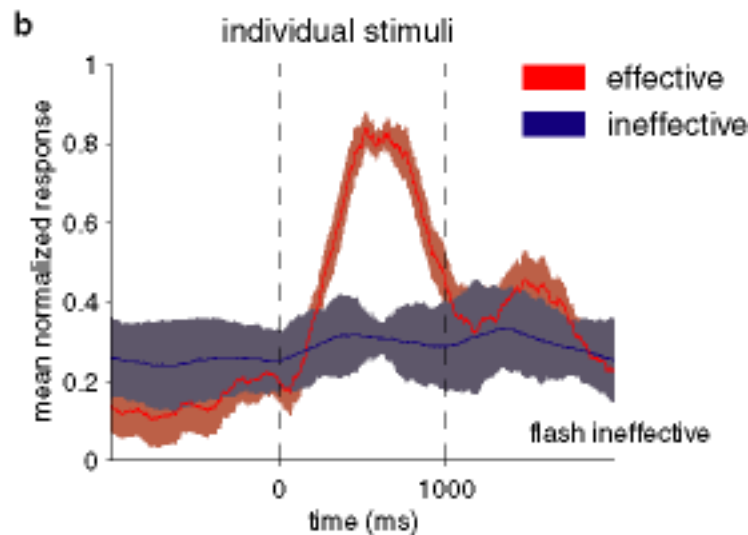
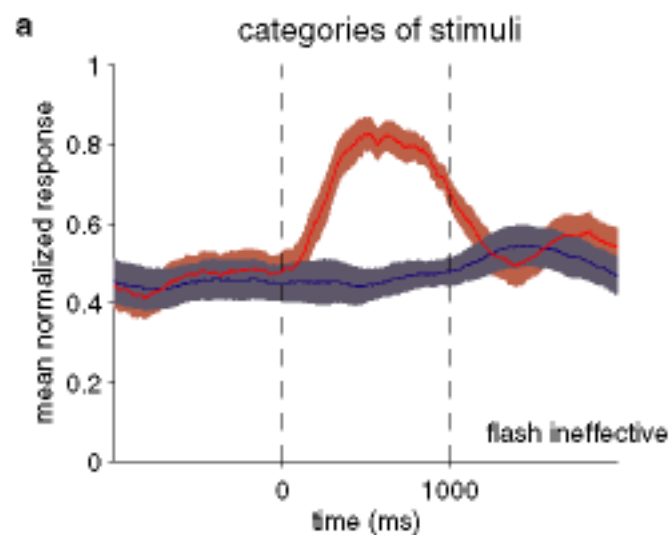
flashed
effective

Gabriel Kreiman
Thesis - Chapter 6
Figure 13

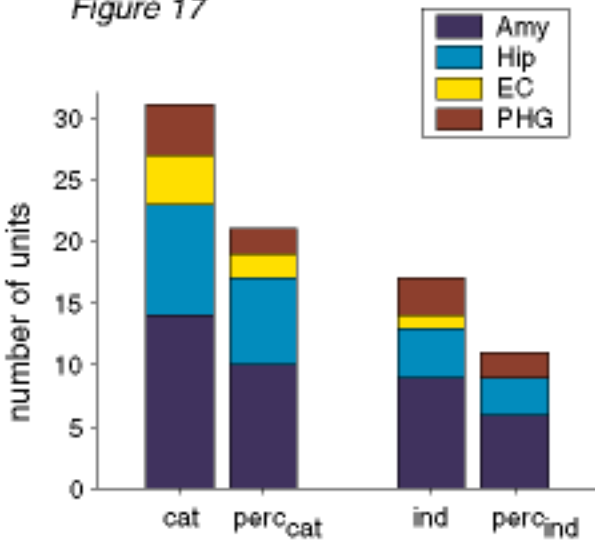


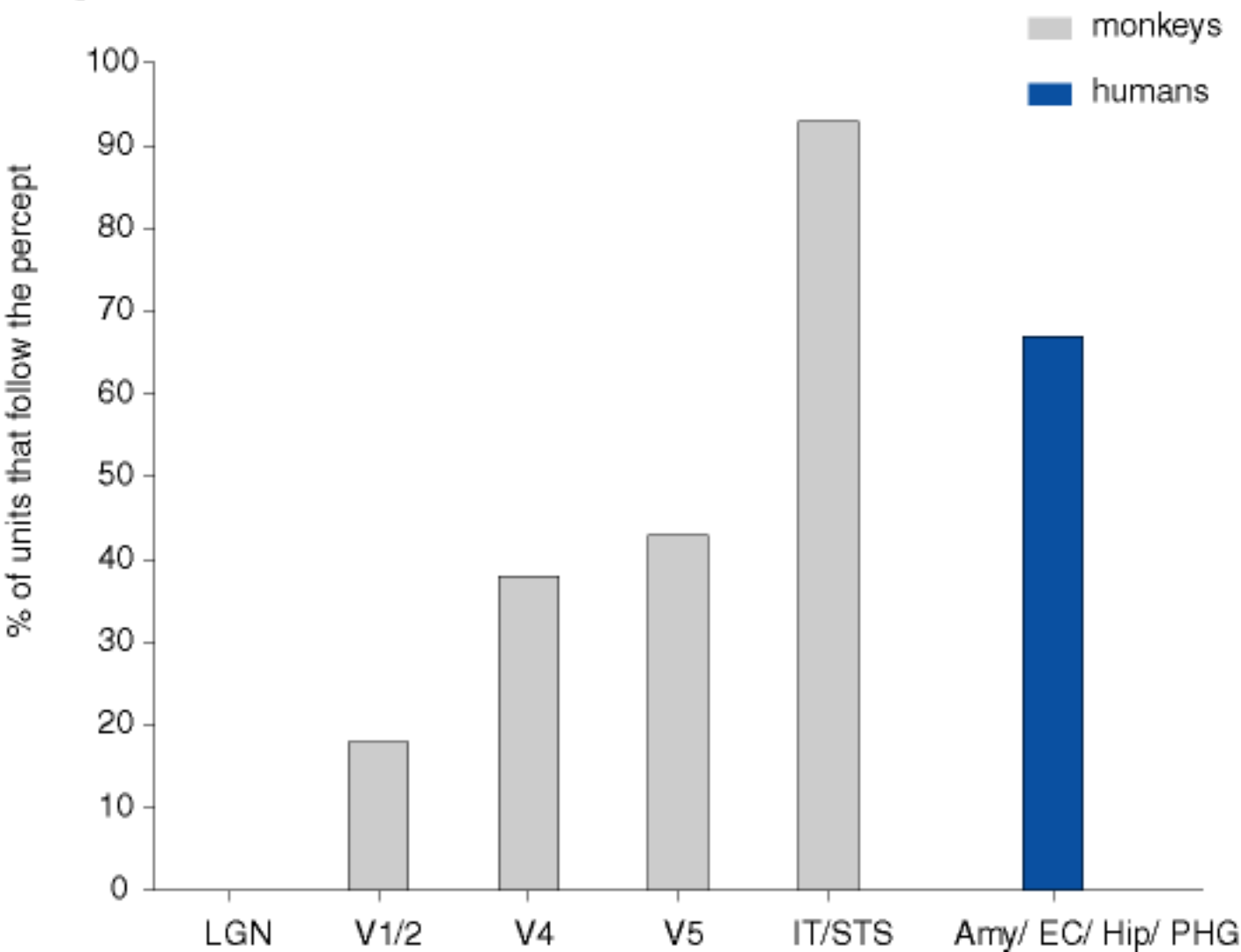




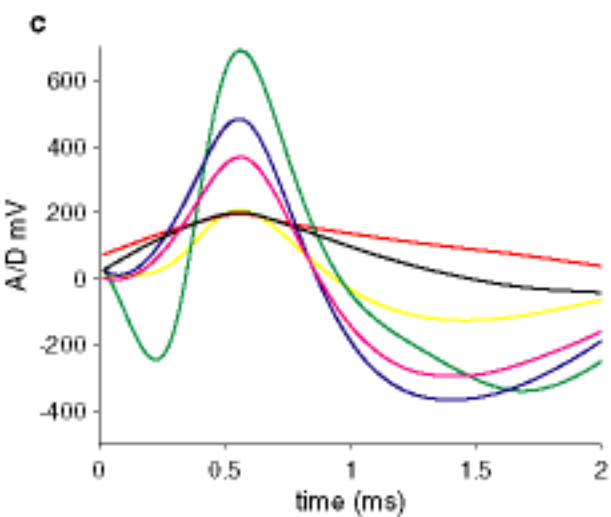
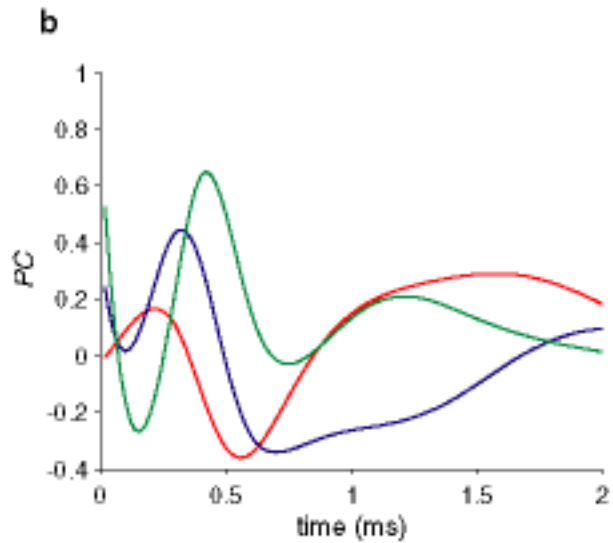
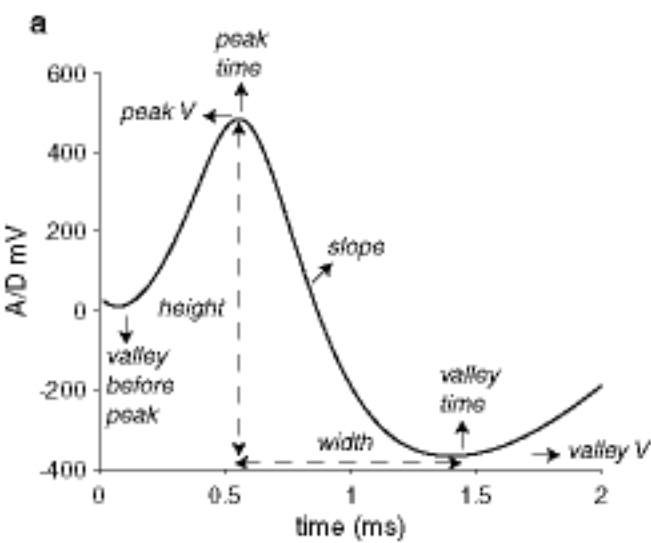


Gabriel Kreiman
Thesis - Chapter 6
Figure 17

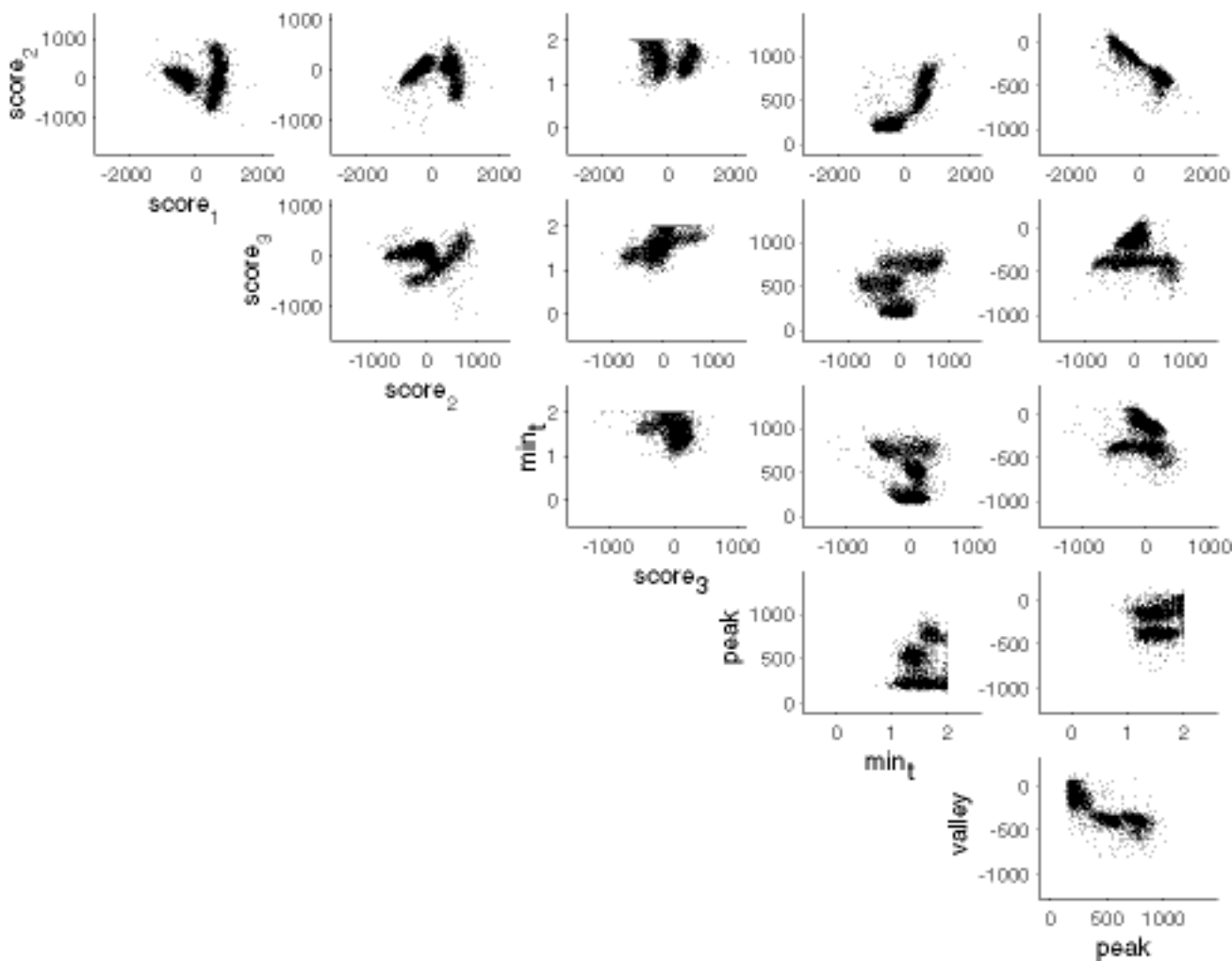




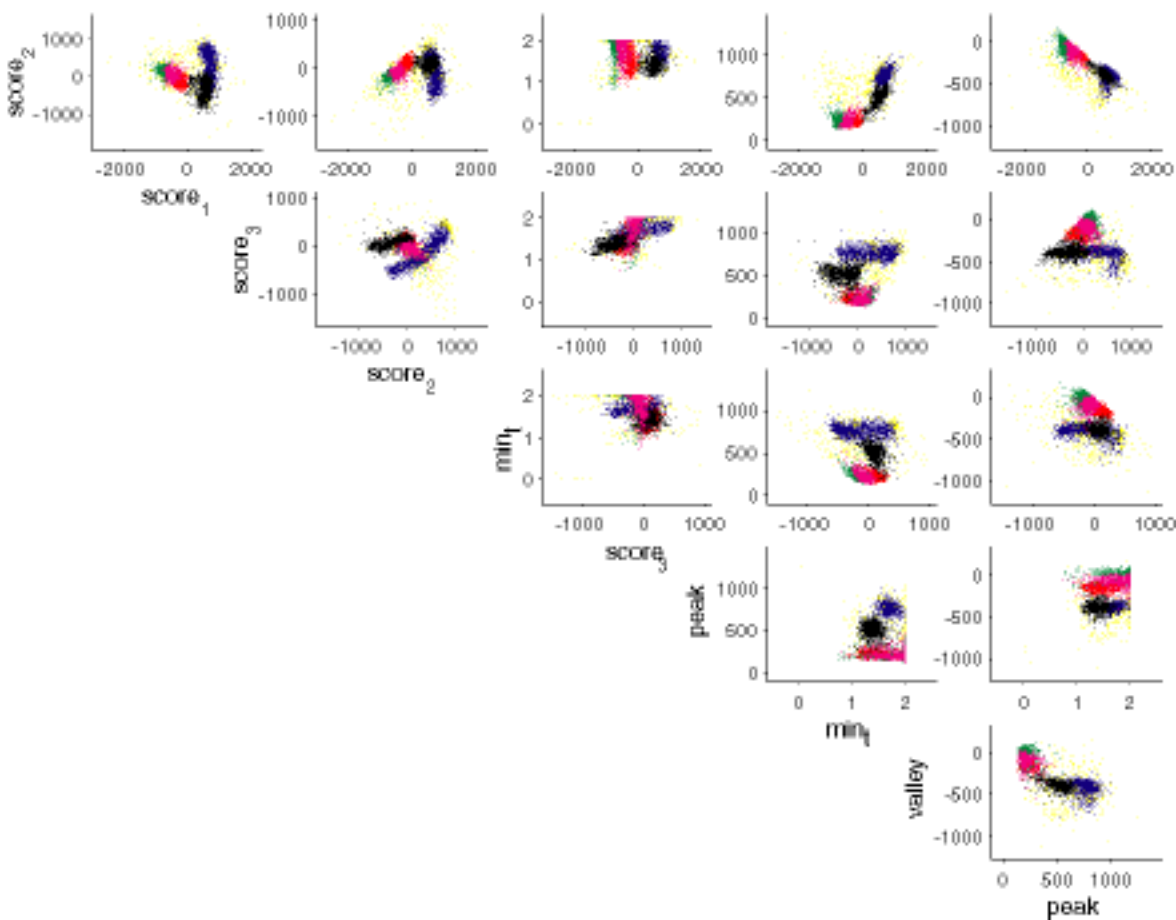
Gabriel Kreiman
Thesis - Appendix 1
Figure 1

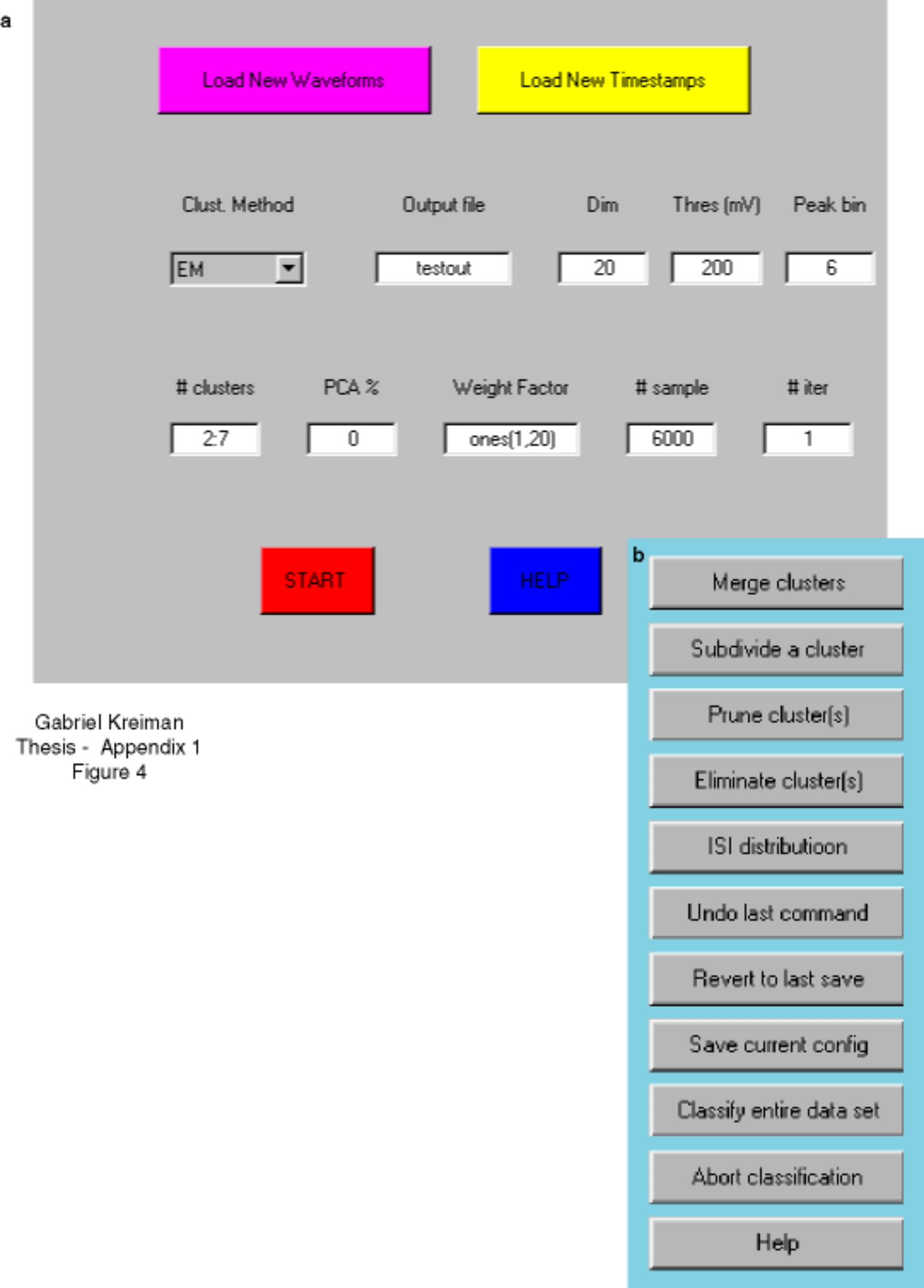


Gabriel Kreiman
Thesis - Appendix 1
Figure 2



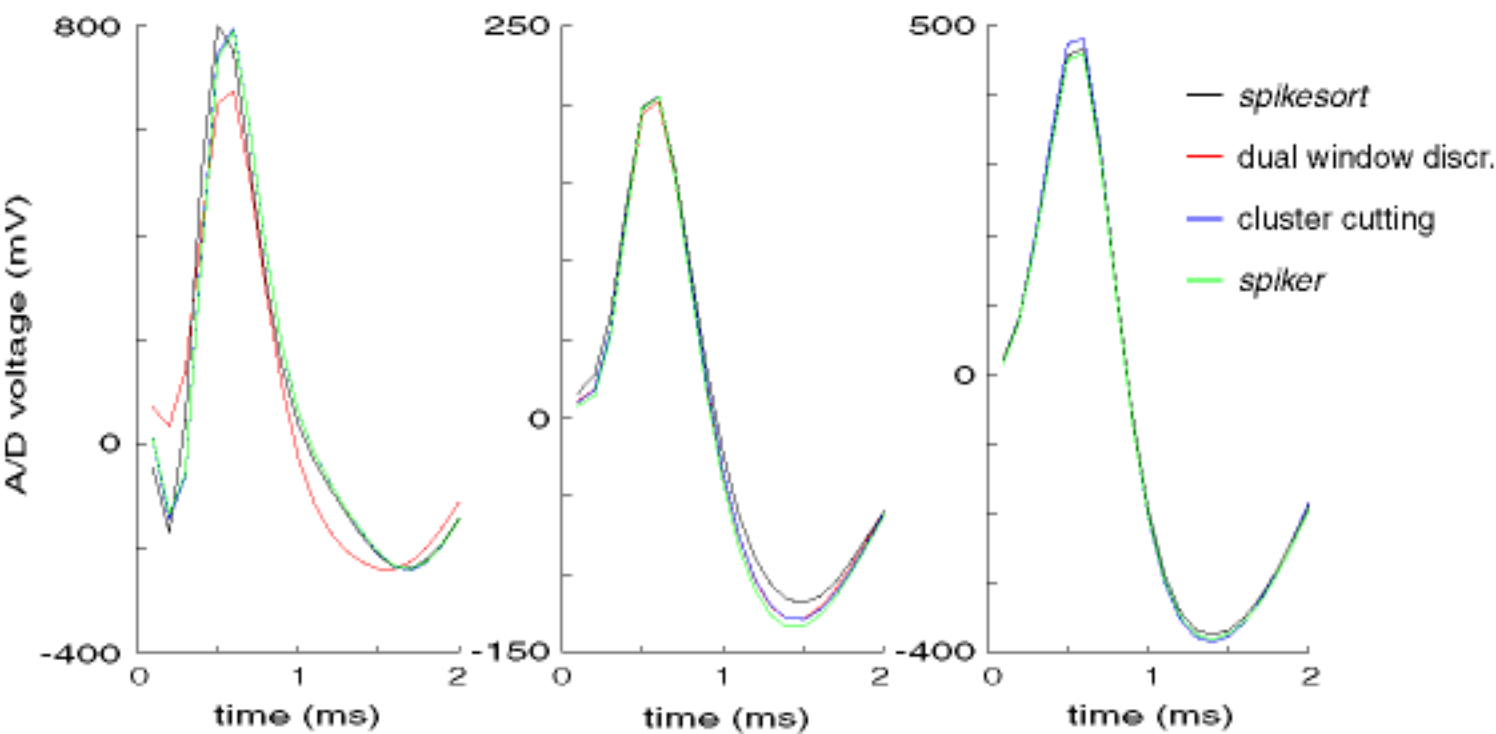
Gabriel Kreiman
Thesis - Appendix 1
Figure 3



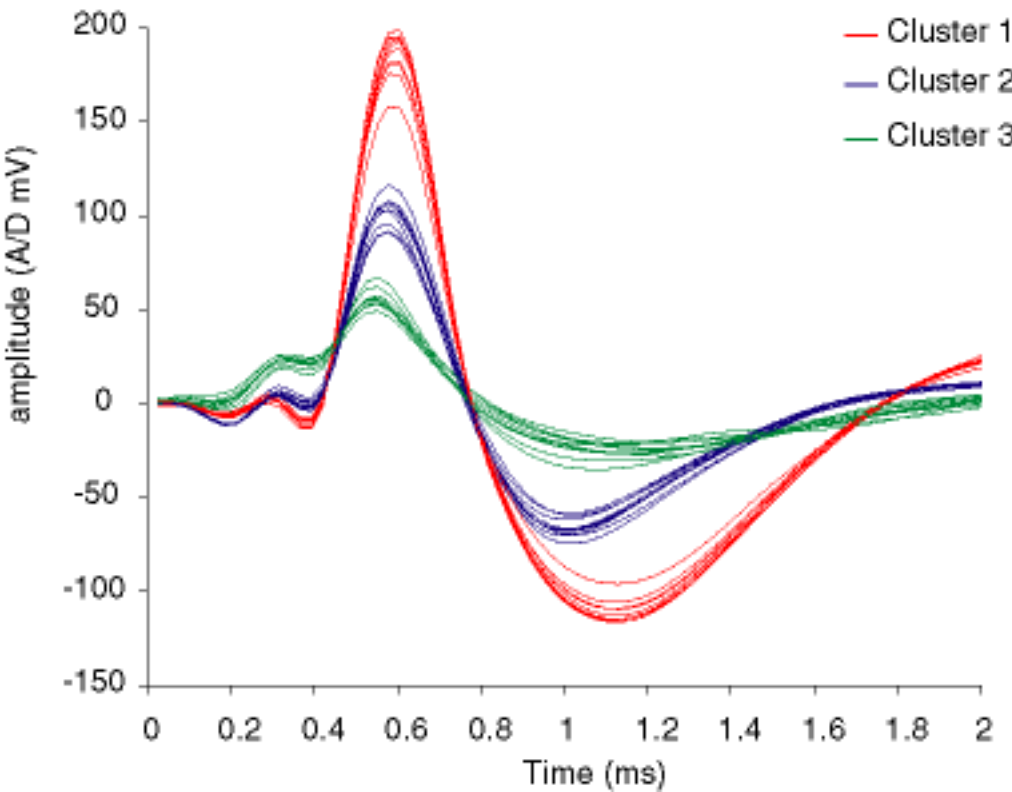


Gabriel Kreiman
Thesis - Appendix 1
Figure 4

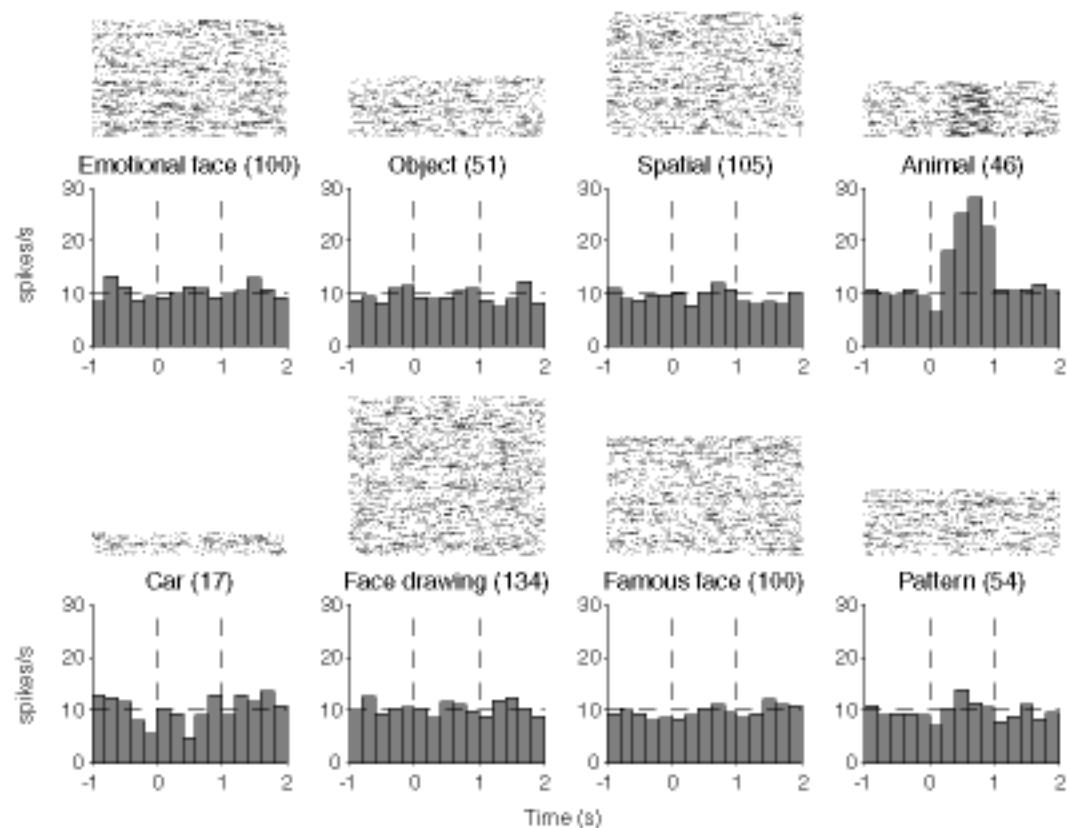
Gabriel Kreiman
Thesis - Appendix 1
Figure 5



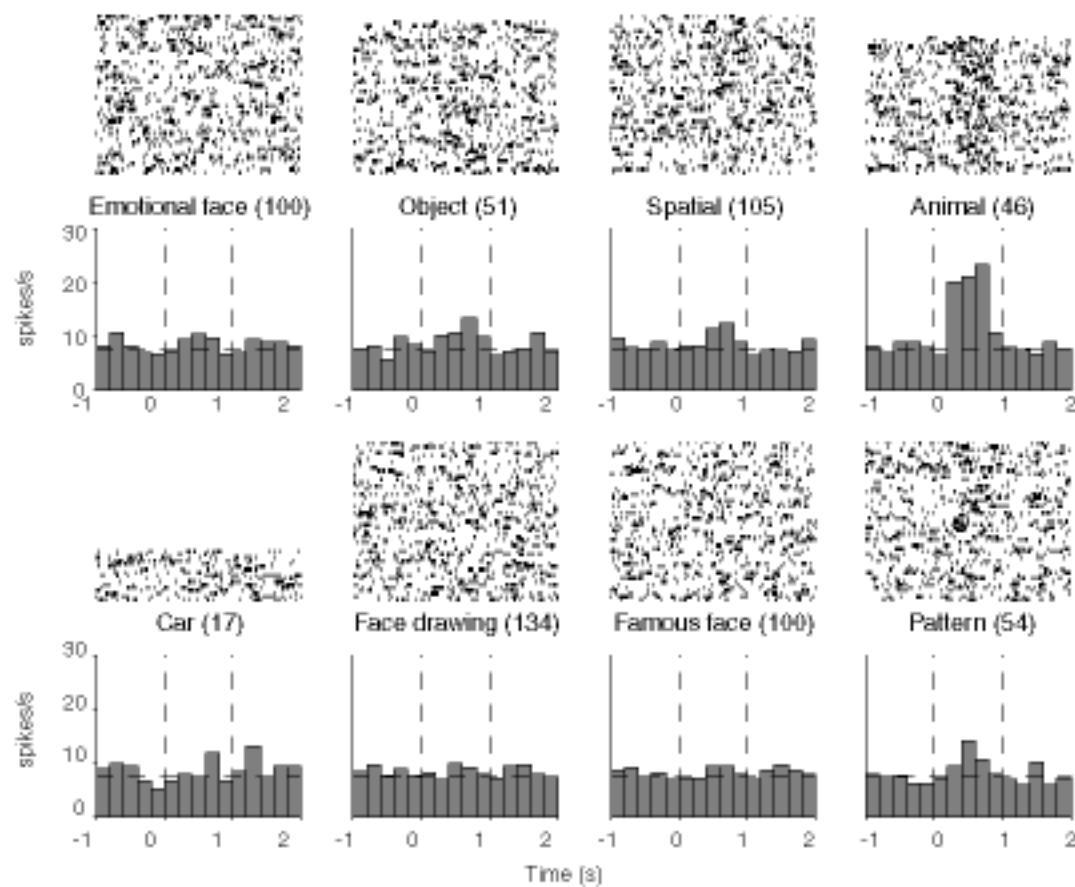
Gabriel Kreiman
Thesis - Appendix 1
Figure 6

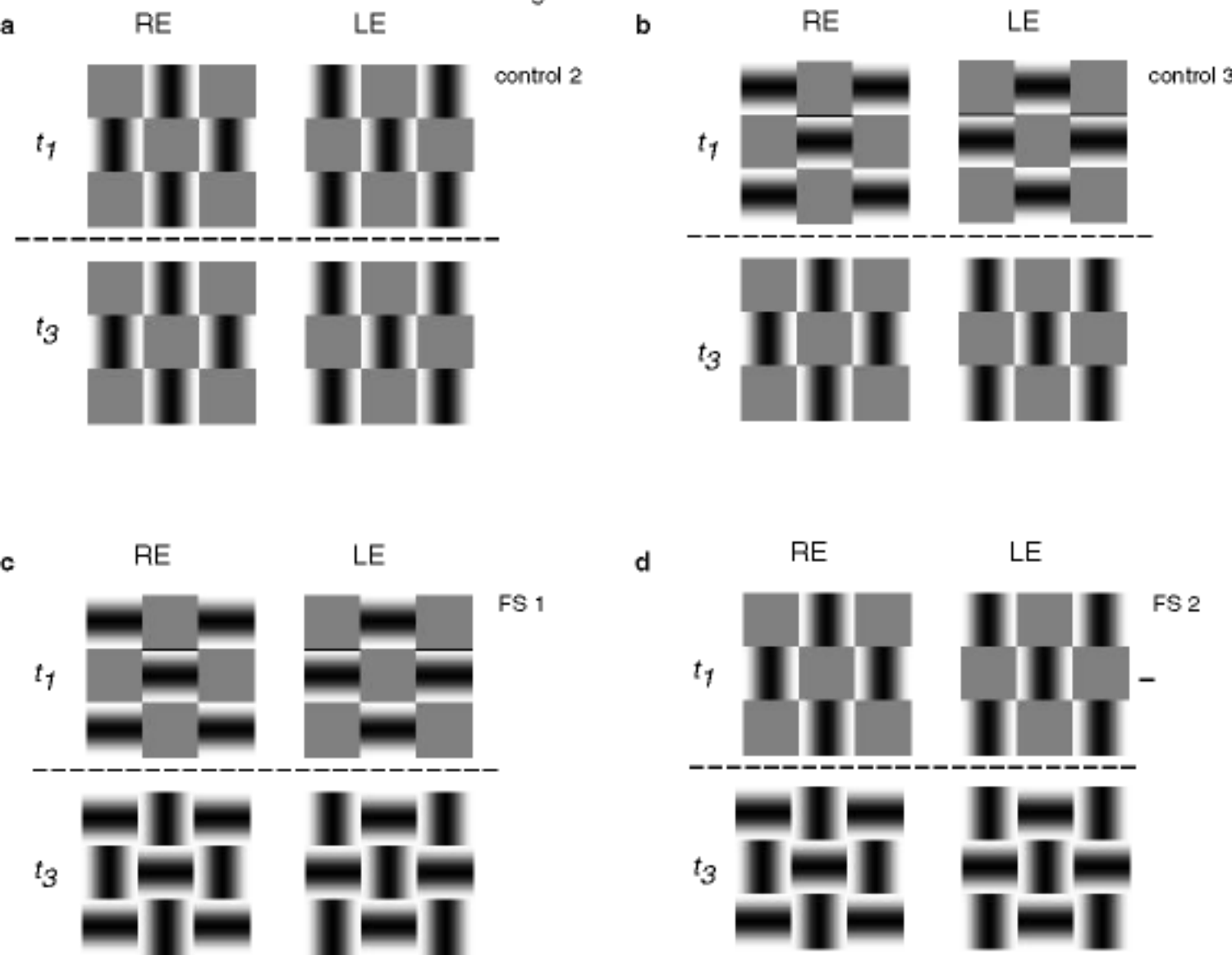


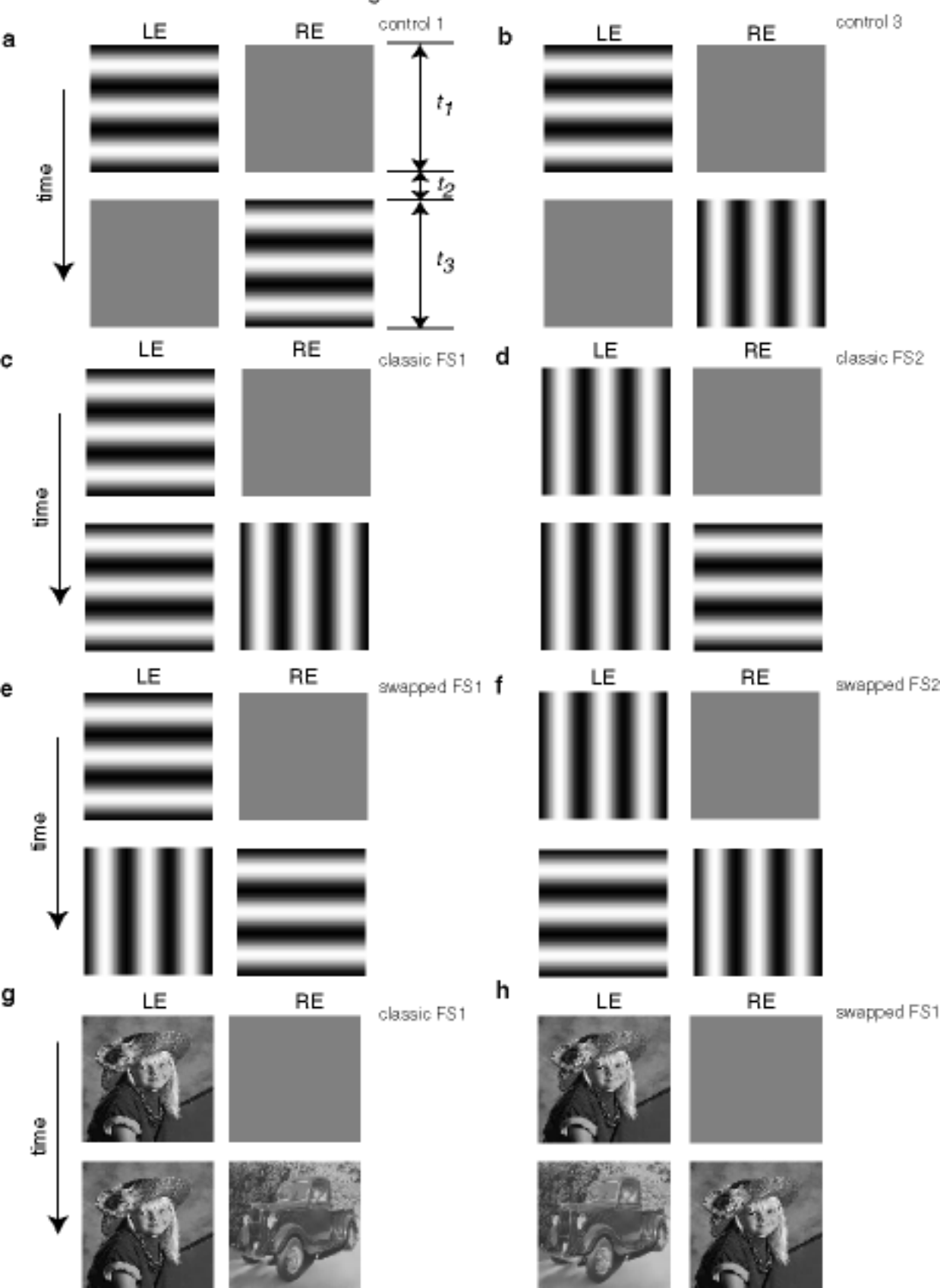
a



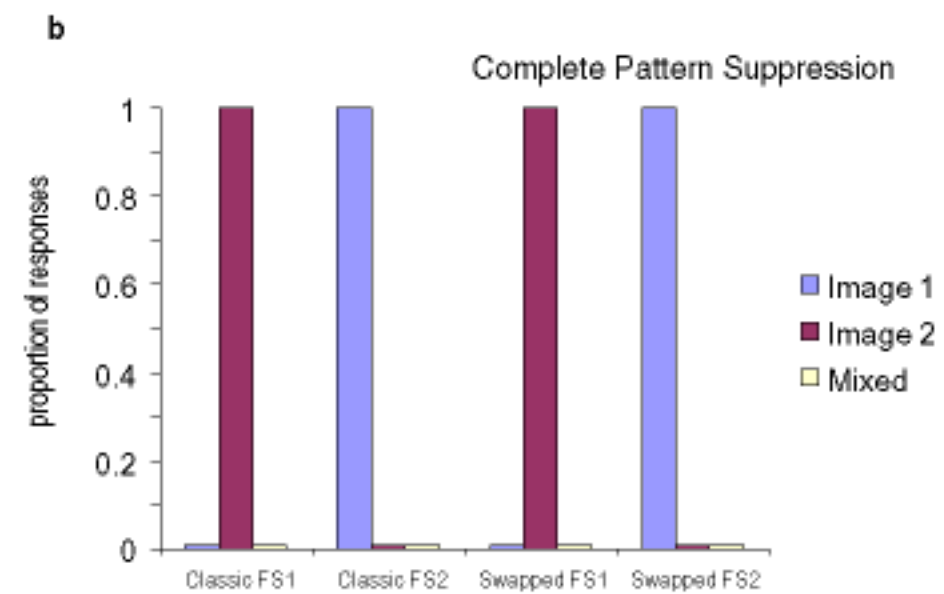
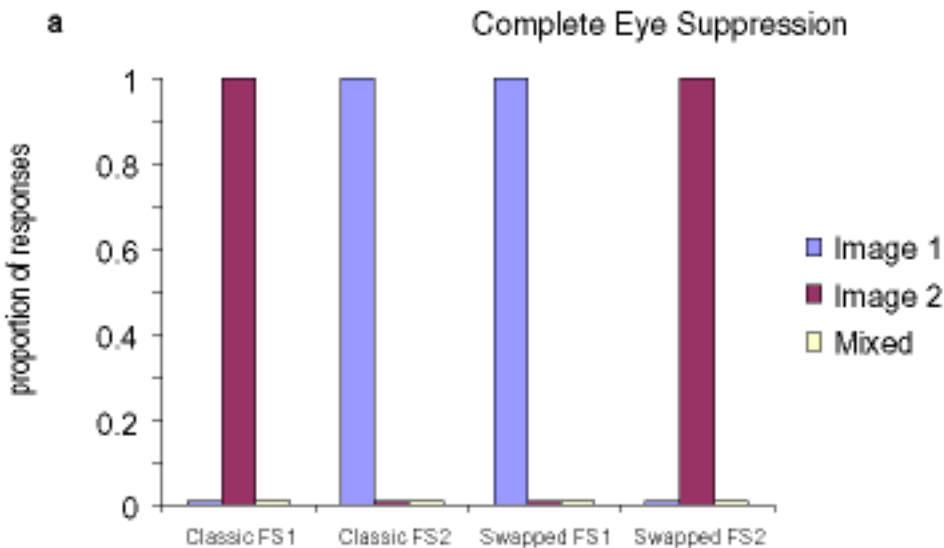
b

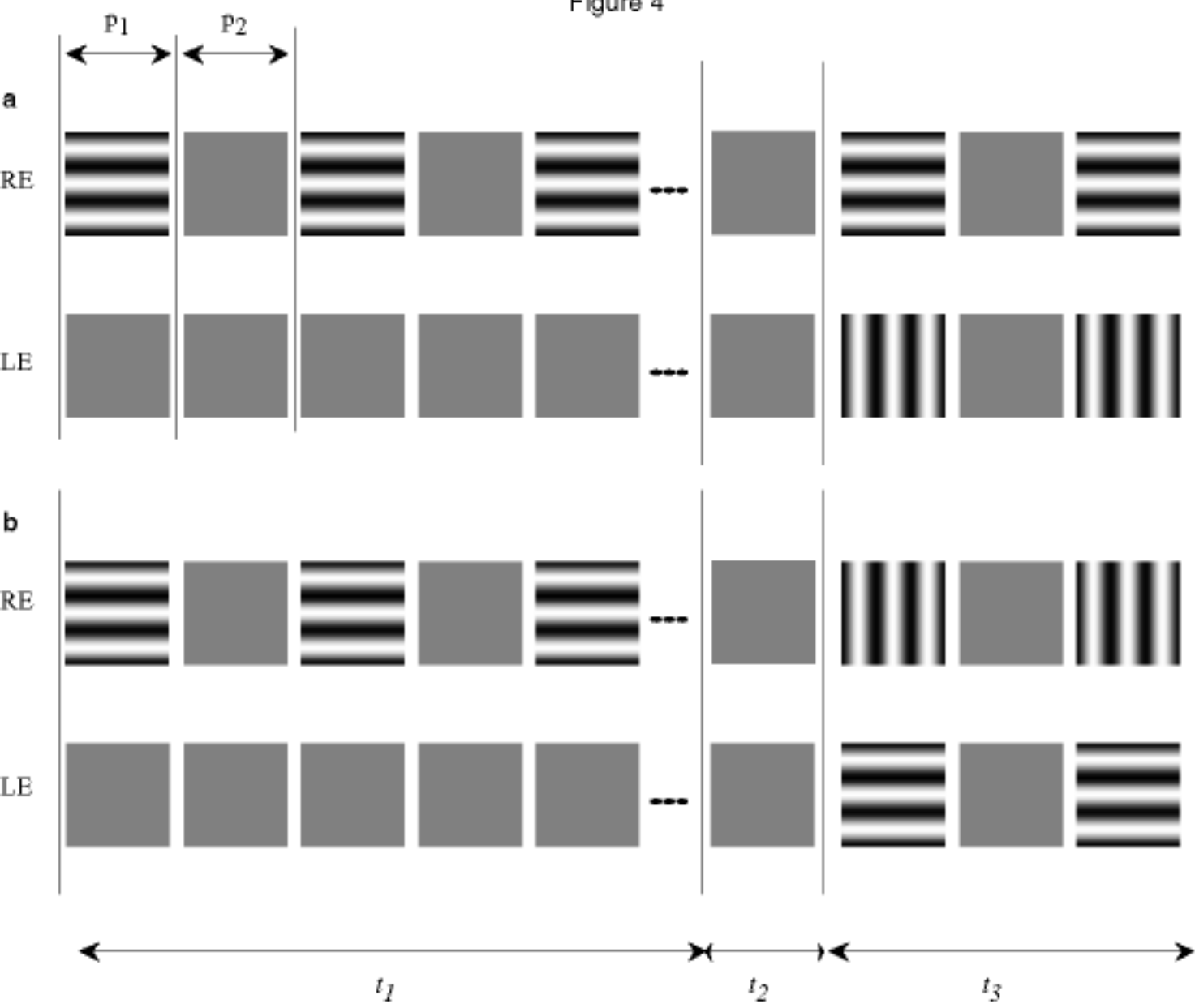




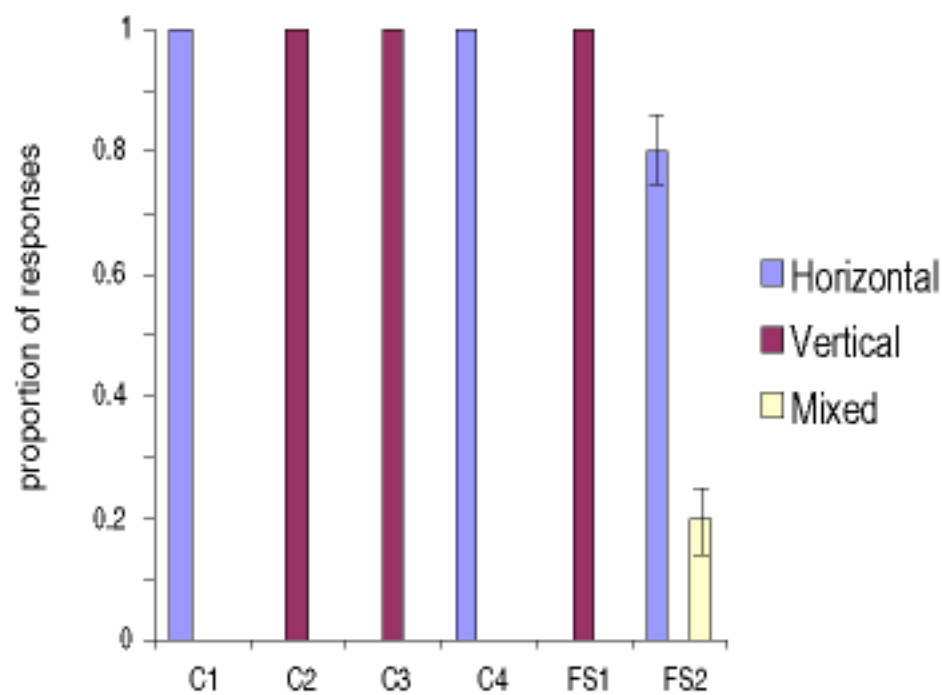


Gabriel Kreiman
Thesis - Appendix 2
Figure 3

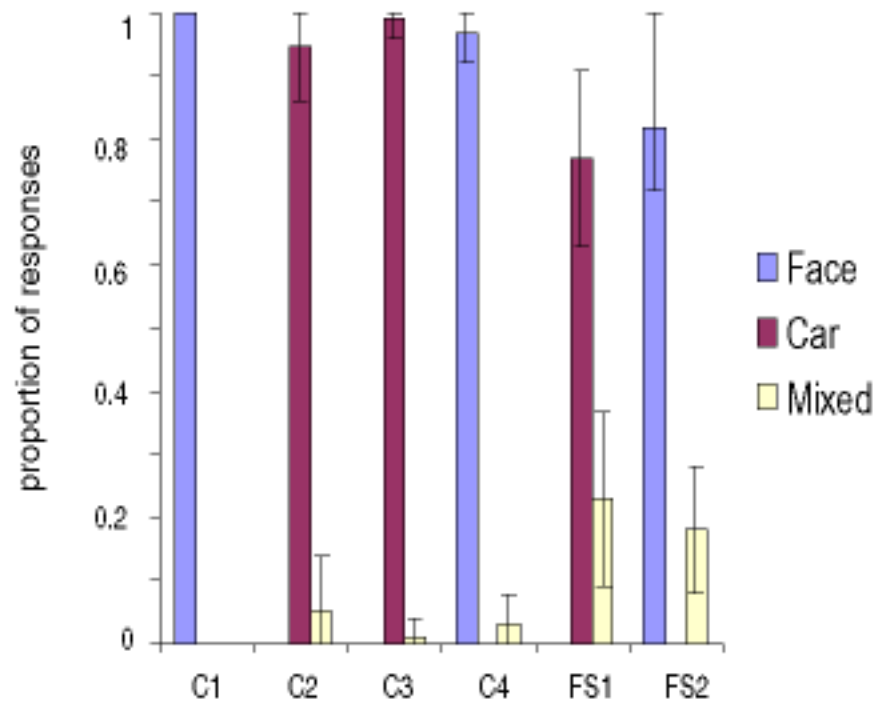


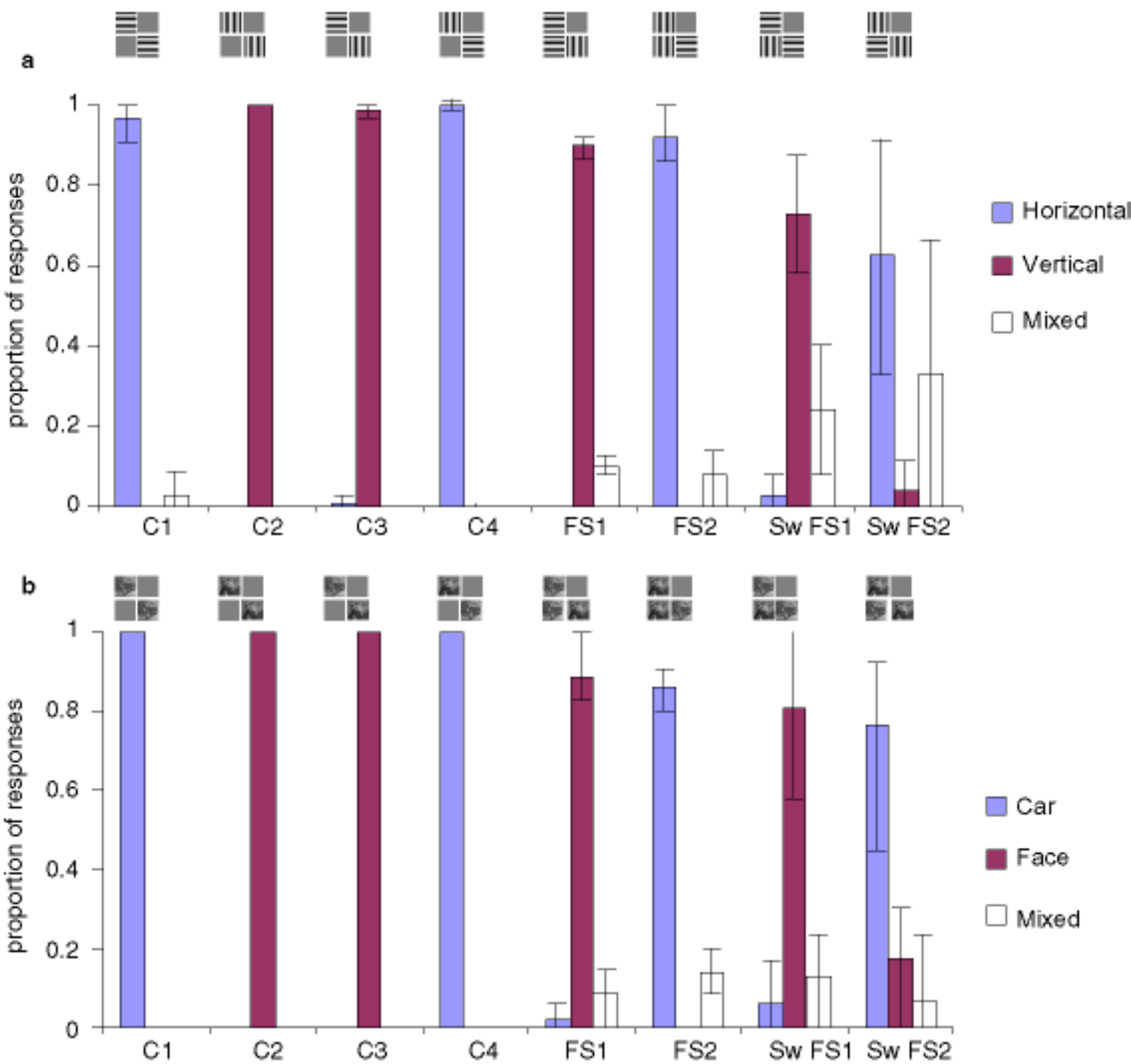


a



a



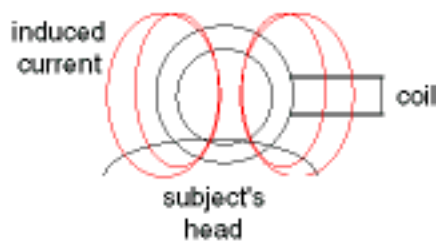


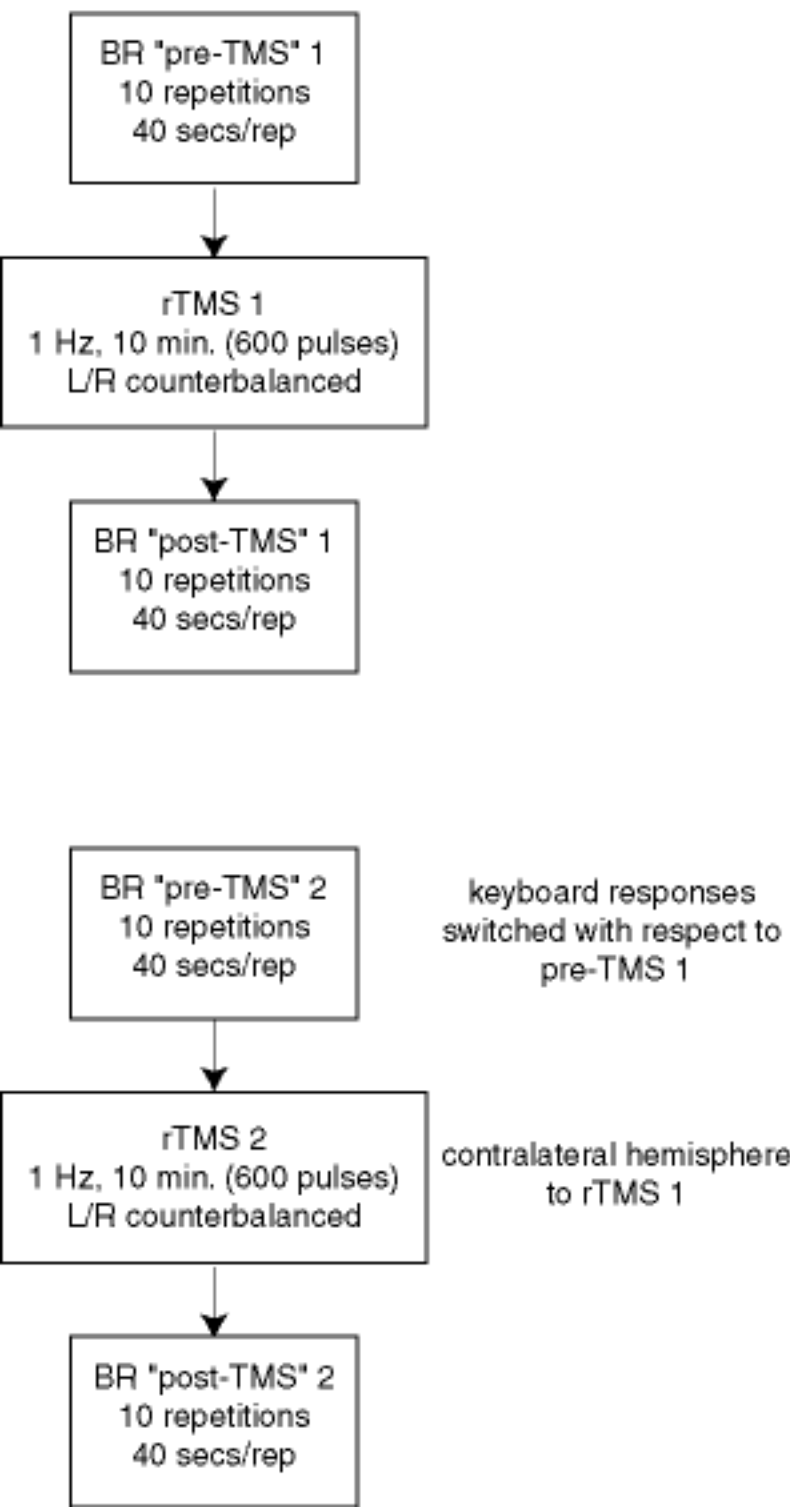
Gabriel Kreiman
Thesis - Appendix 3
Figure 1

a

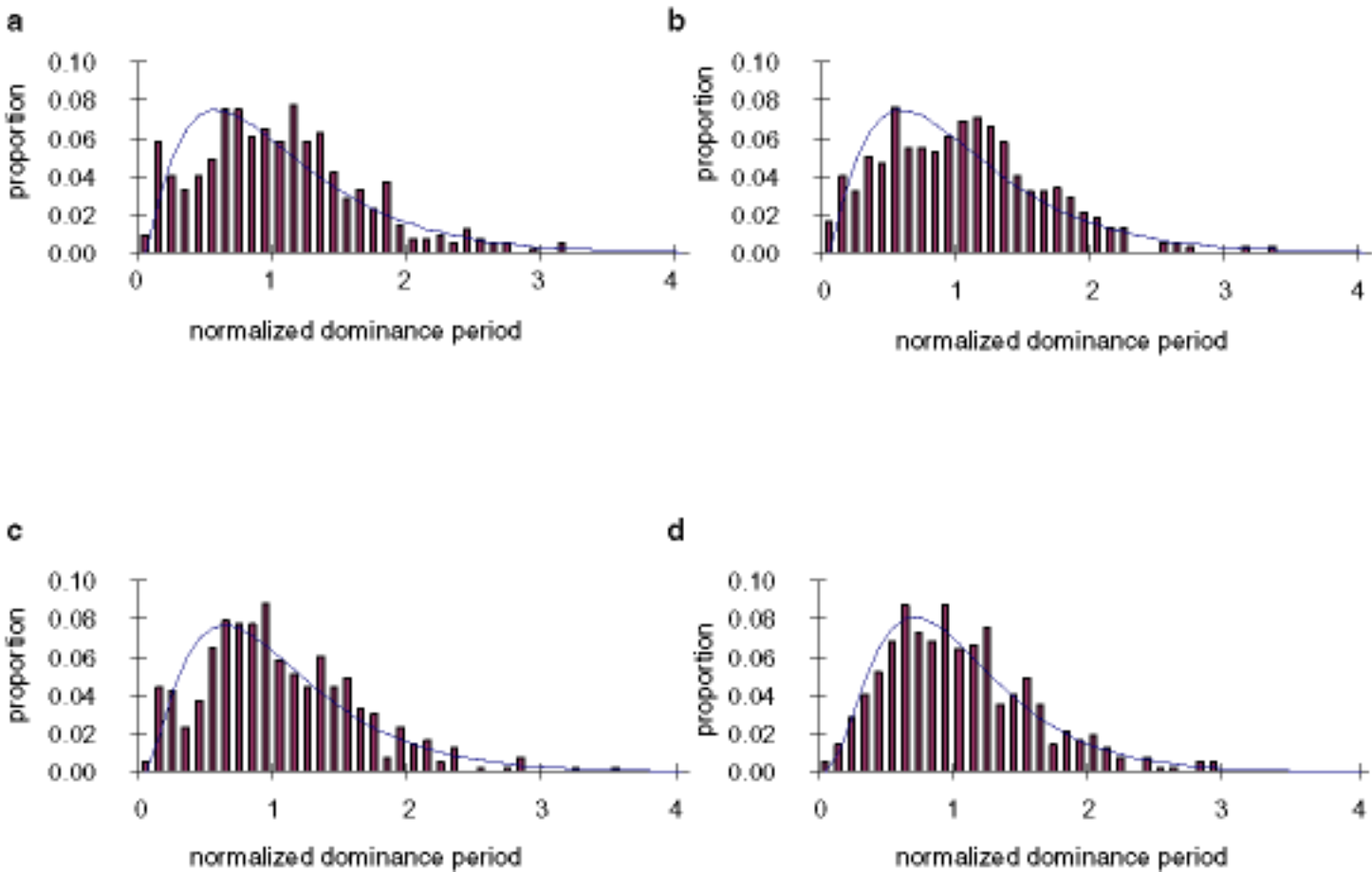


b

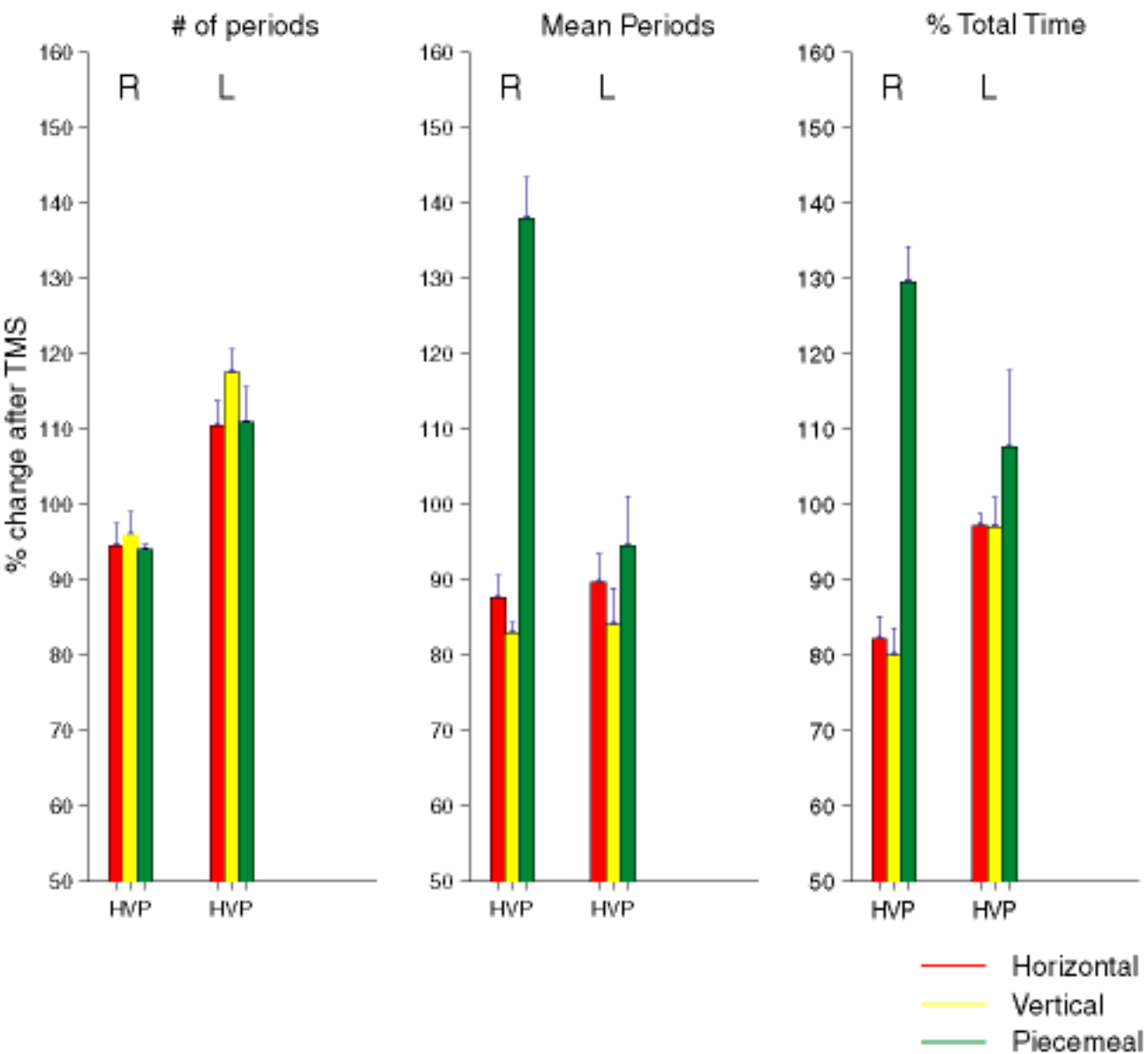




Gabriel Kreiman
Thesis - Appendix 3
Figure 3



Gabriel Kreiman
Thesis - Appendix 3
Figure 4



$t \sim \text{Gamma}(g_{1h}, g_{2h})$



p_{hv}

$1-p_{hv}$

$t \sim \text{Gamma}(g_{1v}, g_{2v})$



p_{vh}

$1-p_{vh}$

$t \sim \text{Gamma}(g_{1p}, g_{2p})$



p_{ph}

$1-p_{ph}$



Contact Information

Caltech 139-74
Pasadena, CA 91125
(626) 840-6194

<http://www.klab.caltech.edu/~gabriel/index.html>
gabriel@klab.caltech.edu

Education

- | | |
|-----------|--|
| 1996-2001 | Ph.D. California Institute of Technology. <i>Biology Division</i> .
Advisor: Professor Christof Koch |
| 1998-2001 | M.Sc. California Institute of Technology. <i>Computation and Neural Systems</i> . Advisor: Professor Christof Koch |
| 1991-1996 | B.Sc. With Honors. University of Buenos Aires. <i>Physical Chemistry</i> |

Journal publications

1. **Kreiman G**, Fried I, Koch C. Single neuron responses in the human brain during flash suppression *Proceedings of the National Academy of Science USA*, *In Press*
2. Krahe R., **Kreiman G.**, Gabbiani F., Koch C. and Metzner W. (2002) Stimulus encoding and feature extraction by multiple pyramidal cells in the hindbrain of weakly electric fish. *Journal of Neuroscience*, **22**:2374-2382.
3. Rees G., **Kreiman G.** and Koch C. (2002) Neural correlates of consciousness in humans. *Nature Reviews Neuroscience*, **3**:261-270.
4. Zirlinger M., **Kreiman G.** and Anderson D. (2001). Amygdala-enriched genes identified by microarray technology are restricted to specific amygdaloid sub-nuclei. *Proc. Nat. Acad. Sci.*, **98**:5270-5275.
5. **Kreiman G.**, Koch C. and Fried I. (2000). Imagery neurons in the human brain. *Nature*, **408**:357-361. {see also *Highlight*: Collins P. (2001). *Nature Reviews Neuroscience*, **2**, 9-9}
6. **Kreiman G.**, Krahe R., Metzner W., Koch C. and Gabbiani F. (2000). Robustness and variability of neuronal coding by amplitude sensitive afferents in the weakly electric fish *Eigenmannia*. *Journal of Neurophysiology*, **84**:189-204.
7. **Kreiman G.**, Koch C. and Fried I. (2000). Category-specific visual responses of single neurons in the human medial temporal lobe. *Nature Neuroscience*, **3**:946-953. {see also *News and Views*: Gross, C. (2000). *Nature Neuroscience*, **3**, 855-856}
8. Ouyang Y., Rosenstein A., **Kreiman G.**, Schuman E. M. and Kennedy M. B. (1999). Tetanic stimulation leads to increased accumulation of Ca²⁺ calmodulin-dependent protein kinase II via dendritic protein synthesis in hippocampal neurons. *Journal of Neuroscience*, **19**:7823-7833.
9. Inon de Iannino N., Briones G., **Kreiman G.** and Ugalde R. (1996). Characterization of the biosynthesis of $\beta(1-2)$ cyclic glucan in *R. Freddii*. *Cellular and Molecular Biology*, **42**:617-629.

Selected abstracts and other publications

1. **Kreiman, G.** (2001). On the neuronal activity in the human brain during visual recognition, imagery and binocular rivalry. Thesis. Department of Biology, California Institute of Technology, Pasadena.
2. **Kreiman, G.** (2001). Neural coding and feature extraction of time-varying signals. Thesis. Computation and Neural Systems, California Institute of Technology, Pasadena.
3. **Kreiman, G.** (2001) Moveo ergo sum. *BioEssays* **23**:662 (Book Review).
4. **Kreiman, G.**, Fried, I., Koch, C. (2001) In *Annual Meeting Visual Sciences Society*, Sarasota. Single neuron responses in humans during binocular rivalry and flash suppression. *Vision Sciences, In Press*
5. Krahe, R., **Kreiman, G.**, Gabbiani, F., Koch, C. and Metzner, W. (2001). Information transmission by multiple pyramidal cells in the electrosensory lateral line lobe (ELL) of the weakly electric fish, *Eigenmannia*. In *International Society of Neuroethology*. Bonn. *Journal of Physiology, In Press*.
6. **Kreiman, G.**, Staba, R., Wilson, C. and Fried, I. (2001). Synchrony between single neurons in the human brain during sleep. In *Annual Meeting of the Society for Neuroscience*, San Diego, Vol. 27
7. **Kreiman, G.**, Fried, I. and Koch, C. (2000). Category-specific visual responses of single neurons in the human medial temporal lobe. In *Annual Meeting of the Society for Neuroscience*, New Orleans, Vol. 26.

8. **Kreiman, G.**, Fried, I. and Koch, C. (2000). Imagery neurons in the human brain: A proposed substrate for visual experiential phenomena in epilepsy. In *American Epilepsy Society, Annual Meeting*, Los Angeles *Epilepsia* **41S**:32-32.
9. **Kreiman, G.** and Koch, C. (1999). Flash Suppression: Competition Between Eyes or Patterns? In Annual Meeting of the Association for Research in Vision and Ophthalmology, Fort Lauderdale, *Investigative Ophthalmology and Visual Science* **40S**:421-421.
10. **Kreiman, G.**, Gabbiani, F., Metzner, W. and Koch, C. (1999). Robustness, variability and modeling of amplitude modulation encoding by P-receptor afferent spike trains of weakly electric fish. In *Neural Information and Coding Workshop*. Big Sky, Montana.
11. **Kreiman, G.**, Gabbiani, F., Metzner, W. and Koch, C. (1999). Robustness and Variability of the neural code in P-receptor afferents of *Eigenmannia* electric fish. In *Frontiers in Computational Neuroscience*, Eilat, Israel.
12. **Kreiman, G.**, Gabbiani, F., Metzner, W. and Koch, C. (1998). Noise in *Eigenmannia* P-receptors. In *Dynamical Neuroscience Satellite Symposium*, Los Angeles.
13. Ouyang, Y., Rosenstein, A. J., **Kreiman, G.**, Kantor, D. B., Schuman, E. M. and Kennedy, M. B. (1998). Tetanic Stimulation increases both autophosphorylation and synthesis of CAMKII in area CA1 of hippocampal slices in 5 minutes. In *Annual Meeting of the Society for Neuroscience*, Los Angeles, Vol. 24.

Selected awards and honors

2000	Everhart Distinguished Graduate Student Lecture Award. Caltech.
1997	Argentine Chemistry Association Honor Award
1995	A. Luna Honor Award. Dow Chemical Company.
1994	Outstanding Youth Award. Buenos Aires Stock Exchange Foundation.
1990	Sub-champion. National Math Olympiads. Argentina.
1984	Metropolitan Chess Champion. Buenos Aires. Argentina.

Teaching experience

Instructor

1999-2000 CNS/Bi 163, Created and taught Caltech course (undergrad/grad. level).

Teaching assistant

1997-1998 Bi 8, CNS/Bi162. Caltech.

1993-1996 Physical Chemistry. UBA.

Additional information

Professional societies

Society for Neuroscience, Cognitive Neuroscience Society, American Association for the Advancement of Science, Vision Sciences Society, American Physiology Society

Software and computing experience

- MATLAB, C/C++, Java, HTML, Perl, Windows, Linux, Unix
- *Designed Spiker*: clustering software for separation of neurons from extracellular recordings.
- *Designed Gene_screen*: software for analysis of DNA microarray data.
- *Spike metrics*: algorithm for measuring distances between spike trains

Extra-curricular activities

- Volunteer Scientist. Science-By-Mail Program. 1998-2000.
- Graduate Student Council. Biology Representative. Caltech. 1997-1998
- Leloir Institute Research Board. Chemistry Undergrad. Representative. 1995-1996