Learning and representation of declarative memories by single neurons in the human brain

Thesis by

Ueli Rutishauser

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2008

(Defended May 8th, 2008)

Acknowledgments

The research presented in this thesis would not have been possible without the help, advice, and encouragement of a large group of people. While this thesis bears my name, this is partially their work too. I feel exceedingly fortunate in having such a great group of friends, colleagues, and mentors, all of whom contributed to this work.

First, I would like to thank my advisors Erin Schuman, Christof Koch, and Adam Mamelak. Their continued support, guidance, and intellectual freedom allowed me to pursue my interests. The work presented in this thesis was primarily done under the mentorship of Erin Schuman and Adam Mamelak. While initially intimidating, I'm grateful that I was given the opportunity to build, from scratch, a human single-unit recording operation. This was a large endowment of trust, given that I had little experience in neuroscience, let alone electrophysiology or medicine. It was a formative experience, which could not have succeeded without the continued guidance and support of Erin Schuman. The members of her laboratory created a nurturing environment, despite me working on a topic far removed from the Lab's mainstream. Alana Rathbun provided much needed assistance with numerous administrative tasks.

While the work I produced under the mentorship of Christof Koch is not part of this thesis, it had an unquestionable impact on my current thinking as reflected in this thesis. I admire his deep and intense love of science and the search for truth. I would like to thank Christof for all his support, encouragement, enthusiasm, and the nurturing environment that "K-lab" provided for me.

I also received advice from the other members of my committee: John O'Doherty and Gilles Laurent. In addition, I had the pleasure of interacting with Ralph Adolphs, John Allman, Henry Lester, Thanos Siapas and Pietro Perona. My first home at Caltech was Pietro Perona's lab, of which I have fond memories. I'm grateful for the freedom he gave me to explore my interests, which eventually led to this work.

Much of the experimental work I performed took place at Huntington Memorial Hospital (HMH). It was a pleasure to interact with people at HMH, in particular the EBMP (Epilepsy&Brain Mapping) nurses, staff, and physicians. Their help is greatly appreciated—many of these experiments would not have been possible without the continued support. I appreciate help with the Neuropsychological evaluations by Linda Philpott. Most of all, I would like to thank the many patients (whom I not allowed to list by name) for their participation. I'm very grateful for their willingness to participate during a time of great distress and uncertainty in their lives.

Over the years, I have greatly benefited from discussions about my research with Costas Anastassiou, Moran Cerf, Tobi Delbrueck, Daniela Dieterich, Rodney Douglas, Wolfgang Einhaeuser, Michael Fink, Alan Hampton, Jonathan Harel, Alexander Huth, Constanze Hofstoetter, Alex Holub, Hiroshi Ito, Vivek Jayaraman, Andreas Kotowicz, Sally Kim, Gabriel Kreiman, Sotiris Masmanidis, Kevan Martin, Florian Mormann, Dirk Neumann, Zoltan Nadasdy,

Kerstin Preuschoff, Johannes Schwarz, Bryan Smith, Michael Sutton, Chin-Yin Tai, and Nao Tsuchiya (in alphabetical order).

I had the opportunity to mentor and work with Caltech undergraduates, SURF students and master students Arjun Basal, Matt McKinley, Anthony Chong, Davhyd Wing and Andreas Kotowicz. It was a pleasure to work with them.

I would not be at Caltech today without the help and encouragement of my previous advisor, Rodney Douglas. I'm grateful for the opportunity he gave me by exposing me to the most interesting research problems. I would also like to thank the faculty members at my undergraduate institution HSR that have supported me in this endeavor—Josef Joller, Gabor Hardy and Andreas Mueller. Thanks also to Gaston Wolf of ZHW for advice in transfering to graduate school.

My way into graduate school was not as direct as it perhaps could have been—I spent several years in the software industry. Several people from this time have been influential in my personal development and I would not be where I'm today without them. Rene Berchten was my first mentor and taught me many of the fundamental software skills I still use today. Enzo Giannini always supported my efforts and was very helpful in the transition between business and academia. I would also like to thank my business partner Alain Schaefer of many years—It was a pleasure running easc together. The results in this thesis are the product of an estimated 60'000 lines of code (mostly Matlab). The skills I was fortunate to acquire have thus been tremendously useful.

Big thanks go to my friends, both here at Caltech and in Switzerland. I'm grateful for their friendship, support, and time spent away from the lab; much of this work could not have succeeded without them: Maik Berchten, Oliver Bandle, Moritz Delbrueck, Nicolas Foong, Patrick Gaus, Kasia Gora, Jonathan Harel, Alexander Huth, Prashant Joshi, Rene Lavanchy, Christof Moser, Tony Oertli, Bobby Rohrkemper, Bryan Smith, Alain Schaefer, and Aaaron Wittmann (in alphabetical order). Thanks to my fellow class-of-2008 members: Anusha Narayan, Hilary Gliden, Matthew Nelson, Dirk Neumann, Ming Gu, Casimir Wierzynski, Will Ford, and Jason Rolfe.

The California Institute of Technology and the Computation & Neural Systems (CNS) program has been an ideal home institution, providing the best possible environment to explore my diverse research interest. Much of my academic thinking has been shaped here. I'm grateful to have been part of Caltech. The group of people that make up the Caltech community is truly remarkable. I specifically would also like to thank International Student Programs, the Office of Technology Transfer, the Center for Neuromorphic Engineering, and the Library for help over the course of my stay. It was very rewarding to be at a place where intense dedication to science and technology is the norm rather than exceptional.

Last, but not least, I would like to thank my family who has continually supported me unconditionally, despite the long distance to Switzerland.

Abstract

Episodic memories allow us to remember not only that we have seen an item before but also where and when we have seen it (context). Neurons in the medial temporal lobe (MTL) are critically involved in the acquisition of such memories. Since events happen only once, the ability to distinguish novel from familiar stimuli is crucial in order to rapidly encode such events after a single exposure. Theoretically, this is a hard learning problem (single-trial learning). Yet, successful detection of novelty is necessary for many types of learning. During retrieval, we can sometimes confidently report that we have seen something (familiarity) but cannot recollect where or when it was seen. Thus episodic memories have several components which can be recalled selectively. We recorded single neurons and local field potentials in the human hippocampus, amygdala, and anterior cingulate cortex while subjects remembered, and later retrieved, the identity and location of pictures shown. We describe two classes of neurons that exhibit such single-trial learning: novelty and familiarity detectors, which show a selective increase in firing for new and old stimuli, respectively. The neurons retain memory for the stimulus for at least 24 h. During retrieval, these neurons distinguish stimuli that will be successfully recollected from stimuli that will not be recollected. Similarly, they distinguish between failed and successful recognition. Pictures which were forgotten by the patient still evoked a non-zero response. Thus, their response can be different from the decision of the patient. Also, we demonstrate that listening to these neurons (during retrieval) enables a simple decoder to outperform the patient (i.e., it forgets fewer pictures). These data support a continuous strength of memory model of MTL function: the stronger the neuronal response, the better the memory (as opposed to a dual-process model). I also describe specific power increases in specific frequencies of the local field potential that are predictive of later retrieval success. These neural signatures, recorded during learning, thus indicate whether plasticity was successful or not.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ACC	anterior cingulate cortex
AMPA	alpha-amino-3-hydroxy-5-methyl-4-isoxazolepropionic
AED	anti-epileptic drug
BOLD	blood-oxygen-level-dependent (see fMRI)
BDNF	brain-derived neurotrophic factor
Ca^{2+}	calcium
CA1-3	Cornu ammonis fields (of the hippocampus)
CT	computerized tomography
Cl	chloride
CS	conditioned stimulus
CR	conditioned response
CTA	conditioned taste aversion
CV	coefficient of variation
CMA	cingulate motor area
DA	dopamine
DG	dentate gyrus
DLPFC	dorsolateral prefrotal cortex
EC	entorhinal cortex
EEG	electroencephalography
EPSP	excitatory postsynaptic potential
EPSC	excitatory postsynaptic current
ERP	event-related potential
ERN	error-related negativity

FDR	false discovery rate
FEF	frontal eye fields
fMRI	functional magnetic reasonance imaging
Glu	glutamate (neurotransmitter)
GABA	γ-aminobutyric acid (neurotransmitter)
IPSP	inhibitory postsynaptic potential
IPSC	inhibitory postsyanptic current
IT	inferotemporal cortex (monkey)
ISI	interspike interval
kstest	Komogorov-Smirnof goodness-of-fit test
K^+	potassium
LFP	local field potential
LGN	lateral genculate nuclei
LTP	long-term potentiation
LTD	long-term depression
LIP	lateral intraparietal area
MRI	magnetic reasonance imaging
MEG	magneto-encephalographic
MTL	medial temporal lobe
MLREG	multiple linear regression
mPFC	medial prefrontal cortex
MT	middle temporal area of the cortex
MUA	multi-unit activity
MLE	maximum likelihood estimate

NMDA	N-methyl-D-aspartic acid (an amino acid)
Na^+	sodium
OCD	obsessive-compulsive disorder
OFC	orbitofrontal cortex
PCA	principal component analysis
PET	positron emission tomography
PFC	prefrontal cortex
PPF	paired-pulse facilitation
RLSC	regularized least-square classifier
RT	reaction time
ROC	receiver operator characteristic
RMS	root-mean-square
SNR	signal-to-noise ratio (but also see below)
SNr	nigra pars reticulata, substantia nigra
SVM	support vector machine
STDP	short-time dependent plasticity
SUA	single-unit activity
STD	standard deviation
SE	standard error
STN	subthalamic nucleus
TE	anterior part of inferior temporal cortex (monkey)
TEO	posterior part of inferior temporal cortex (monkey)
TLE	temporal lobe epilepsy
UR	unconditioned response

US	unconditioned stimulus
V1	primary visual cortex, Brodman area 17.
V2	part of the extrastriate visual cortex , Brodman area 17.
V3	part of the extrastriate visual cortex , Brodman area 17.
V4	part of the extrastriate visual cortex , Brodman area 17.
VTA	ventral tegmental area

Chapter 1. Introduction

In this chapter I introduce what we know about the function and anatomy of the brain areas discussed in this thesis (medial temporal lobe and cingulate cortex). I further discuss some features of epilepsy, with an emphasis on temporal lobe epilepsy and its treatment. All data presented in this thesis has been acquired from epilepsy patients. This thesis is not about epilepsy as such, but it is beneficial for the reader to understand the basics of epilepsy to better appreciate the clinical situation in which this research was conducted. The aim of this chapter is to set the stage for the results reported in this thesis so as to enable the reader to place them into context.

1.1 Memory

The capacity to learn and remember a seemingly infinite amount of information is one of the key facilities that makes us human. To illustrate this, ask yourself the following question: “Where where you on the following day: September 11th, 2001?”. Chances are, you not only know where you were, but also how you heard about what happened that day, who told you about it, what you felt and what you thought was going to happen. This example illustrates the many components of memory — not only does this day mean something to you, but you can also retrieve a large amount of associated attributes. These attributes are not neutral facts. Rather, many of them have an emotional component. Thinking about the past does (introspectively) not just bring up a list of facts but rather each attribute is remembered with many of its significant autonomic attributes still attached. Have you ever met somebody, knew with high confidence that you knew the person but could not remember who the person is nor where you last met? This example illustrates a further example of memories: they are not all-or-nothing monolithic entities.

Rather, it is possible to retrieve some aspects of a memory (e.g., that it exists, “I have met this person before”) without any other of the attributes associated with it. Nevertheless the not-remembered information is often not lost — after some time, it might very well possible to retrieve the missing information.

What is most remarkable about memories is that they can be acquired very quickly. Most events for which we have memories happen only once and often last a very short time. This is nevertheless sufficient to build a representation that can last a lifetime. Life events are special in that they occur at a certain time to you personally (egocentric, i.e., they are episodic memories). This is distinct from other types of memories, such as memories for facts (semantic memories). Not only are fact memories not acquired from a single learning experience (usually one rehearses or studies facts) but also usually no attributes are associated with them — like, where and when did you learn this fact? Together, episodic and semantic memories build the class of all explicit long-term memories referred to as declarative memories. Another distinguishing feature of a declarative memory is a strong sense of confidence about whether one remembers something or not — performance and confidence are highly correlated (Bayley and Squire, 2002).

The other major class of memories are procedural, non-declarative memories. These include motor skills such as riding a bike which we can do effortlessly, but without the ability to articulate how we do it. That is, procedural memories are expressed by performance rather then recollection. Such memories require hundreds of learning trials (with feedback) to acquire. Procedural memories rely, for the most part, on brain structures distinct from those involved in declarative memories. These structures, such as the cerebellum and the basal ganglia, are not discussed here.

1.2 Anatomy, connectivity, and function of the medial temporal lobe

The neuronal structures necessary for the acquisition of declarative memories are situated in the medial temporal lobe (MTL). The MTL consists of a cortical and subcortical part. The cortical parts include the perirhinal, entorhinal, and parahippocampal cortices. Subcortically, the MTL includes the hippocampal formation (CA fields, dentate gyrus, subiculum) as well as the Amygdala (Squire et al., 2004; Squire and Zola-Morgan, 1991; Suzuki and Amaral, 2004). The MTL exhibits remarkable evolutionary consistency across several major mammalian species such as rodents, monkeys, and humans. While the absolute size differs, the gross anatomical and neurophysiological features are remarkably similar.

Destruction, inactivation or surgical removal of the MTL due to accidents, surgery, or stroke results in severe anterograde amnesia, manifested by profound forgetfulness. Some retrograde amnesia occurs as well, but memories of events that happened more than a few months before the injury are typically well preserved (but see below). Other cognitive capabilities are not impaired. In particular, short-term memory (working memory) is not impaired. Also, intelligence is not impaired. This is well illustrated by the well-studied patient H.M., who had bilateral MTL removal for treatment of epilepsy (Corkin, 2002; Milner et al., 1968; Scoville and Milner, 1957). H.M. has profound anterograde amnesia, but can remember events of his childhood. Also, he can learn new motor skills (procedural memory) such as mirror drawing but does not remember having done so. Even if the MTL is only deactivated temporarily, such as in global amnesia, the events that happened during this period are not remembered later.

The role of the MTL (and in particular the hippocampus) in memory is time limited. The loss of parts of the MTL causes temporally graded retrograde amnesia. In humans, such loss of memory can be cover up to 15 years back in time (Corkin, 2002; Squire and Alvarez, 1995) . The extent of retrograde amnesia depends on how much of the MTL is damaged. For example, selective damage to CA1, an area of the hippocampus, resulted in 1–2 years of retrograde amnesia whereas more extensive damage can erase up to 15 years of memories (Rempel-Clower et al., 1996). Very remote autobiographical and factual memories beyond this time period appear unimpaired (Kirwan et al., 2008; Squire and Bayley, 2007). In animals, hippocampal damage induces retrograde amnesia lasting days to weeks only (rather then years in humans, see (Frankland and Bontempi, 2005) for a review). Thus, the hippocampus is not necessary for the retrieval of such memories. Rather, it is responsible for acquiring the memory. Over time, memories become independent of the hippocampus. While these facts are well established, it is not clear what the mechanisms are that make the hippocampus only necessary initially and lead to gradual independence. One framework originally formulated by Marr proposes that memories are first stored in the hippocampus and are then gradually transferred to cortical areas (Marr, 1970, 1971). He proposed that such transfer would occur through replay of activity during offline states (such as inactivity or sleep). While there have been several reports of such “replay” of activity (Buzsaki, 1998; Diba and Buzsaki, 2007; Foster and Wilson, 2006; Wilson and McNaughton, 1994), it remains to be demonstrated whether this indeed serves the purpose of transferring memories from the hippocampus to the cortex.

1.2.1 Anatomy of the hippocampal formation

The hippocampal formation consists of the hippocampus proper (cornu ammonis fields (CA), divided into CA3, CA2, and CA1) as well as the dentate gyrus (DG), subiculum, presubiculum, parasubiculum, and entorhinal cortex (Andersen et al., 2007; Duvernoy, 2005). Note that in the human literature, the entorhinal cortex is often referred to as the "parahippocampal area". The hippocampal formation (hippocampus and dentate gyrus) is about 100x bigger in humans than in rats (rat 32mm^3, monkey 340mm^3, human 3300mm^3). Nevertheless, the basic anatomical features are remarkably similar between these 3 species. Not all areas show a similar increase in size from rats to monkeys (and humans). One noteworthy difference is the thickness of the pyramidal layer in CA1: it is about 5 cells thick in rats compared to 10–15 cells in monkeys. In humans, it is as much as 30 cells thick. It is estimated that the human hippocampal formation contains about 60 million neurons, compared to ~ 4 million in the rat (Andersen et al., 2007).

The hippocampus is tightly interconnected with the rest of the brain. It is distinct from cortical areas in that the connections with other brain areas are largely unidirectional. Cortical areas, on the other hand, are connected reciprocally (Felleman and Van Essen, 1991). Most input to the hippocampus first reaches the entorhinal cortex. From the entorhinal cortex, two major input pathways project to the hippocampus: the perforant path (to the dentate gyrus) and the temporoammonic alvear pathway (to CA1). The dentate gyrus does not project back to the entorhinal cortex (unidirectional). The dentate gyrus projects exclusively to CA3 (via the mossy fibers). All granule cells in the DG project to CA3. Their axons terminate in a stereotypical region of the CA3 cell body layer (stratum lucidum). CA3 pyramidal neurons send axons onto themselves (recurrent) as well as to CA1 (Schaffer collaterals). This circuit makes up what is referred to as the trisynaptic circuit. Synapse 1 is EC->DG, synapse 2 DG ->CA3, and synapse 3 is CA3->CA1. CA1 projects back to the entorhinal cortex as well as to the subiculum. There are no known direct connections between CA3/CA1 and the neocortex.

The entorhinal cortex receives input from a large number of cortical areas. The primate EC receives substantially more diverse cortical input than the rat EC. Prominent connections are from (and to) various high-level unimodal visual areas such as areas TE and TEO (through perirhinal cortex), area V4 (through parahippocampal cortex), numerous polysensory regions in the superior temporal gyrus, frontal areas such as the orbitofronal cortex, and the cingulate, as well as the insula. The EC also receives input from subcortical structures, such as the amygdala or the claustrum (see below).

1.2.2 Computational principles of hippocampal function

Due to the recurrent nature of CA3 pyramidal cell connectivity it has long been proposed that CA3 is the site of implementation of a large randomly connected recurrent network (Hasselmo et al., 1995; Kanerva, 1988; Marr, 1970, 1971; Rolls, 2007; Treves and Rolls, 1994). Such networks can be used to rapidly establish new attractors as well as for pattern completion (as demonstrated by Hopfield networks (Hopfield, 1982)). Pattern completion is crucial to implement content-addressable memories. However, the role of CA3 as a pattern completion engine has remained largely theoretical. A recent study using a genetically modified mouse strain that allows a selective and specific knockout of area CA3 function after animals reach adulthood reveals selective deficits that support this hypothesis (Nakazawa et al., 2002). Rather then being unable to learn at all, this mouse was unimpaired at a number of tasks such as the Morris water maze. There were no behavioral impairments in learning and retrieval of spatial memory.

Crucially, however, the mouse was unable to learn from single trials (such as a novel location of the platform) nor was it able to retrieve the platform location if only a partial set of the previous sets of cues were presented (Nakazawa et al., 2002; Nakazawa et al., 2003).

The dentate gyrus (DG), on the other hand, has been proposed to implement a complementary function: pattern separation (O'Reilly and McClelland, 1994; Rolls, 1996; Shapiro and Olton, 1994; Treves and Rolls, 1994). This suggestion is motivated by the following observations: i) Connectivity between DG dentate cells and CA3 pyramidal cells is very sparse. This results in a small degree of divergence (~ 14 pyramidal cells per granule cell) (Acsady et al., 1998). ii) There are large differences in the number of neurons in the EC (200’000), DG (1’000’000) and CA3 (300’000). All estimates are for the rat (Amaral et al., 1990; Amaral and Lavenex, 2007; Boss et al., 1985; Henze et al., 2002) but similar proportions are valid for the monkey and human DG (Amaral and Lavenex, 2007; West and Slomianka, 1998). This leads to an expansion (EC->DG) followed by a contraction (DG->CA3) of effective dimensionality (see below). Thus, every CA3 neuron receives (with high probability) input from a different subset of DG granulate cells. iii) Only a small fraction of granule cells is activated in any given task (Chawla et al., 2005; Jung and McNaughton, 1993; Witter, 1993). From theoretical studies it is known that such a projection effectively increases the distance between every possible pattern, thus making the patterns more dissimilar to a downstream region. This is due to the increase in the effective dimensionality. The powerful computational properties of such a construct are well demonstrated by “liquid state machines”, which essentially consist of random sparse projections from a low-dimensional space into a high-dimensional space and back (Maass and Markram, 2004; Maass et al., 2002). Several experimental studies support this hypothesis. Until recently, the only experimental support for this hypothesis has come from behavioral observations after DG lesions in rats (Gilbert et al., 2001). In match-to-sample tasks, DG lesioned rats show deficits in distinguishing different objects (some of which indicate food and others don't) if they are close together in space. They have little deficits if objects are far apart (spatially). A more direct demonstration of the role of the DG in pattern separation comes from genetic NR1 knockouts restricted to DG granule cells (McHugh et al., 2007). Since NR1 is a necessary subunit of the NMDA receptors, this mutation prevents NMDA dependent plasticity in granule cells. No perforant path (EC->DG) potentiation could be invoked in these mice. Behaviorally, these mice had difficulty distinguishing between different contexts as measured by inappropriate freezing in contexts which are similar to, but slightly different, from the context in which conditioning took place. Mice with intact DG had no problem in distinguishing the two contexts. Interestingly, the deficit was only temporarily: more training (experience) could overcome the deficit. Thus, pattern separation is important for the rapid acquisition of new experiences. In spatial tasks, DG granule cells have place fields which are similar to CA3 and CA1 pyramidal cells. Place fields rapidly change their firing preference if the external environment (size, color, shape) is changed ("remapping"). Interestingly, remapping of DG place fields occurs more rapidly (for smaller environmental differences) than for CA3 cells. This is demonstrated by the finding that correlations between the firing activity of populations of DG cells decay rapidly as a function of small changes of the environment, whereas CA3 cells decorrelate only after large changes (Leutgeb et al., 2007).

These experimental findings (for both CA3 and DG) are the first to offer direct support for the long-standing theoretical proposal that one of the functions of the hippocampus is pattern separation (DG) followed by pattern completion (CA3).

1.2.3 Amygdala

The amygdala receives direct inputs from all sensory systems as well as the hippocampus (Aggleton, 2000). Its projections, amongst others, to different areas in the hypothalamus and brain steam and can thus directly influence the autonomic nervous system. The amygdala consists of several separate nuclei (central, basal, lateral, and medial). Each nucleus is either mainly an input or an output structure. The lateral nucleus receives input from all sensory systems. The central nucleus projects to the brainstem and hypothalamus, whereas the basal nucleus projects to cortical areas as well as the striatum. There is very substantial reciprocal connectivity between the hippocampus and the amygdala. The lateral and basal nuclei of the amygdala project prominently to the entorhinal cortex. Feedback connections from the EC terminate mostly in the basal nucleus of the amgydala. The amygdala is necessary for some kinds of rapid learning such as Pavlovian fear conditioning (Fanselow and LeDoux, 1999). Forms of synaptic plasticity like long-term potentiation (LTP) can be induced both in vivo and in vitro (Chapman et al., 1990; Rogan et al., 1997). While the amygdala is not necessary for declarative memory formation, it nevertheless modulates the strength of such memories (Phelps, 2004; Richardson et al., 2004). The amygdala is thus crucially involved in the acquisition of some types of memories. In the following chapters I will show that many of the single-neuron responses related to single-trial learning can be observed similarly in both the amygdala and the hippocampus.

1.2.4 Adult neurogenesis

Most neurons in the adult brain are postmitotic. Remarkably, there are a few exceptions: there are progenitor cells (stem cells) in the subgranular and subventricular zone that continuously divide and send new neurons to the adult dentate gyrus and the olfactory bulb, respectively. There are indications that other regions of the brain contain new neurons as well (Garcia et al., 2004; Gould et al., 1999). This adult form of neurogenesis occurs throughout life and has been shown to be modulated by numerous environmental factors such as stress, learning and exercise (van Praag et al., 1999; van Praag et al., 2002). Of interest to the results of this thesis, some report a relationship between the sensitivity to novelty and the number of new neurons in the dentate gyrus (Lemaire et al., 1999). The discovery of postnatal neurogenesis in the MTL suggests the intriguing possibility that it is related to our capacity to learn. It will be very interesting to explore the computational implications of the incorporation of new neurons into existing circuits, because new neurons have very distinctly different electrophysiological properties (van Praag et al., 2002), such as increased plasticity (Schmidt-Hieber et al., 2004). It is unclear how these different single-cell properties affect circuit function (Lledo et al., 2006). Evidence has recently accumulated that inappropriate incorporation of new neurons is implicated in epiloptogenesis (Buhl et al., 1996; Covolan et al., 2000; Parent et al., 1997). Whether this is a cause or an effect, however, is unclear, but it is a very promising route for further experimental investigation (Parent, 2007).

1.3 Mechanisms of plasticity—Circuit and single-cell properties

Establishing new memories (learning) is thought to require structural changes (plasticity) for long-term storage (Martin et al., 2000). Here, any mechanism that changes the composition, shape, size, or configuration of a cell is referred to as structural plasticity. Examples are insertion or removal of proteins such as neurotransmitter-activated ion channels, the removal or growth of new spines, as well as the growth of new axons. This form of memory is distinct from activity-based memories such as working memory, iconic memory, adaptation, or priming which (at least in theory) do not require permanent structural changes. Such memories only last as long as the neuronal activity (in the form of spiking or subthreshold processes such as deactivation) remains active—typically only a few seconds (see (Wang, 2001) for a review and (Romo et al., 1999) for an example of pre-frontal working memory activity). Longer-lasting memories, however, do not require constant activity. We do not lose our memories after we sleep, go into deep anesthesia, or have a severe epileptic seizure. A mechanism must exist that transforms information about the external environment (represented as neural activity) into changes in the brain. The exact nature of these changes is a matter of great debate and is largely unknown (see below). Understanding this mechanism at all levels involved has been one of the goals of neuroscience research since its beginning (Bliss and Lomo, 1973; Hebb, 1949; Pavlov, 1927). The results presented in this thesis contribute to this understanding by demonstrating that there are single neurons in the brain that function as generic novelty/familiarity detectors. It is hypothesized that these detectors are part of the system that initiates learning for a novel stimulus.

Long-term potentiation (LTP) and long-term depression (LTD) are a class of molecular and cellular mechanisms that can trigger such long-lasting changes in synaptic strength (Bliss and

Lomo, 1973). Here, I will briefly summarize what we know about LTP as well as its relevance to behavioral changes related to learning.

The amount of postsynaptic current influx evoked by presynaptic release depends on many different factors, such as the number of vesicles released, the number of channels in the postsynaptic terminal, internal Ca2+ stores, and the number, types, and composition of voltage-dependent channels (among many others). The modification of any of these factors potentially leads to changes in synaptic strength, i.e., the degree and amount of influence presynaptic release has on the postsynaptic neuron. Similarly, the number of synaptic contacts between two neurons can change as well (formation and destruction of synapses). It has been observed consistently that the synapses connecting two neurons get strengthened when the presynaptic neuron fires shortly before the postsynaptic neuron (LTP). Reversal of the temporal order (postsynaptic neuron fires before presynaptic) leads to a decrease of the synaptic weight (LTD). The existence of LTP/LTD has by now been shown in a wide variety of species and brain structures including the hippocampus, amygdala, and the neocortex. This is one of the fundamental principles of synaptic plasticity— loosely summarized "what fires together wires together". Hebb originally postulated that "when an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased" (Hebb, 1949). LTP/LTD is a candidate mechanism that implements this rule (see below).

This principle (Hebb's law or Hebbian learning) is commonly expressed as a correlation-based learning rule that describes the incremental change of the synaptic weight as the product of the pre and postsynaptic firing rates: $\Delta w = v_j v_i$. Since this implies that weights can grow infinitely, a maximum weight $w = \min(w + \Delta w, w_{\max})$ is usually imposed. Also, in this simple form, weights never decrease (see below for more detailed discussion). This principle can be observed at many different levels of organization, starting with a single synapse between two neurons all the way to behavioral observations (for example, Pavlovian conditioning). While we have some understanding of the detailed molecular mechanisms, the intermediate levels are much less clear and are poorly understood. For example, during Pavlovian conditioning, how is it that plasticity can be induced selectively at exactly the right synapses while not influencing all the other existing synapses? After all, acquisition of a new memory does not require that an existing memory be overwritten. Also, whether the mechanism of LTP is sufficient and/or necessary for learning of new memories remains to be demonstrated (Martin et al., 2000; Shors and Matzel, 1997; Stevens, 1998). One strategy to demonstrate that LTP is indeed sufficient for learning would be to monitor a large number of synapses during memory acquisition. If reverting these very same synapses after learning back to their original strength erases the memory it is demonstrated that these synaptic changes did indeed result in the observed behavioral change (Neves et al., 2008). A large number of studies have been conducted that attempt to demonstrate that LTP is indeed the mechanism underlying learning (see (Martin et al., 2000) for a review). For example, it has recently been demonstrated that behavioral single-trial learning of inhibitory avoidance by rats induces molecular and electrophysiological changes in the hippocampus very similar to those induced by artificial LTP induction in CA1 (Whitlock et al., 2006). While these and others suggest a close link between LTP and learning, this has not been convincingly demonstrated at this point of time.

The amount and direction of plasticity at a single synapse depends on many factors (see above). One of the most important factors, however, is time. For a long time, it was not clear what exactly qualifies as “fires together”. What qualifies as “together” ? Within 1 sec or within 1 ms ? Also, does the order matter (first presynaptic, then postsynaptic or the opposite)? One of the fundamental principles that has emerged is that both the temporal order as well as temporal distance matter on the order of milliseconds. Technical advances allowed the first experimental demonstration by dual intracellular patch recordings from two synaptically connected neurons (Bi and Poo, 1998; Markram et al., 1997). The remarkable finding was that evoking a spike in the presynaptic neuron 10 ms before evoking a spike in the postsynaptic neuron lead to strengthening of the synapse. The reverse (presynaptic spike follows the postsynaptic spike) leads to weakening of the synapse. Spacing the spikes closer together in time evokes stronger changes. Spikes that occur too far apart in time (> 40 ms) fail to induce any changes in synaptic strength. This mechanism is referred to as spike-timing-dependent plasticity (STDP). The time window for induction of STDP is at best ± 40 ms. In classical conditioning, the conditioned stimulus (CS) can be separated by up to several seconds from the unconditioned stimulus (US) (Pavlov, 1927). This is thus much bigger then the timescale over which STDP can occur. Additional mechanisms (such as working memory) must thus exist to bridge this time gap. Many computational learning algorithms such as reinforcement can be implemented with STDP as a mechanism. Such learning is usually referred to as correlation-based learning (or “Hebbian learning”). The existence of STDP has been shown in a large number of different species, brain areas, and cell types (see (Caporale and Dan, 2008) for a review).

The induction of long-lasting plastic changes is often dependent on changes in intracellular calcium (Ca^{2+}). There are many different types of neurotransmitter or voltage-gated ion channels which are permeable to calcium. This accumulation of intracellular Ca^{2+} triggers many molecular events which eventually lead to long-lasting changes of synaptic strength. Without calcium influx, LTP can not occur. This is convincingly demonstrated by the absence of LTP if the NMDA channels are blocked pharmacologically during LTP induction. The induction of plasticity itself is not sufficient to ensure a long-lasting structural change. If certain molecular processes are disrupted, the change does not last long. For long-lasting LTP ("late LTP"), gene transcription and the synthesis of new proteins is required (Huang et al., 1996; Kelleher et al., 2004; Schuman, 1999; Squire, 1992; Sutton and Schuman, 2006). This is also true in vivo: memories do not last if synthesis is inhibited (Davis and Squire, 1984; Flexner et al., 1963; Squire, 1992). The blockage of protein synthesis (during induction) thus prevents the conversion of a short-term to a long-term synaptic change. Rather, it decays back to baseline within a few hours. Synaptic changes can, however, occur without new proteins, i.e., the modification and relocation of existing proteins is sufficient. While protein synthesis blockers present during the induction of LTP only affect the late phase of LTP using traditional LTP induction protocols, there are other induction protocols for LTP where protein synthesis is required also for the early phase. Application of neurotrophic factors such as BDNF can induce potentiation without electrical stimulation (Kang and Schuman, 1995; Levine et al., 1995; Lohof et al., 1993). Application of synthesis inhibitors prevents this kind of induction of LTP (Kang and Schuman, 1996). There are also situations where synthesis is required for early LTP, such as high synaptic background activity (Fonseca et al., 2006). In the hippocampus, synthesis can also be modulated by changes in extracellular dopamine (Smith et al., 2005).

1.4 Temporal lobe epilepsy

Epilepsy is one of the most common forms of neurological impairment. The lifetime risk of experiencing at least one seizure is 3% (Chang and Lowenstein, 2003). About 1% of all people develop unprovoked seizures without any obvious reason (Steinlein, 2004). In the USA, an estimated 1.1 - 2.3 million people have epilepsy. Epilepsy is a medical condition characterized by the presence of recurrent seizures. Clinical manifestations of seizures can include loss of consciousness, involuntary twitching of muscles, brief periods of amnesia, sleep disturbances as well as other sensory, cognitive, psychic, or autonomic disturbances.

Seizures are fundamentally a circuit-level phenomenon. Thus, the study of seizures requires a systems perspective. They are thought to occur due to hypersynchronous neuronal discharges that lead to uncontrolled spread of excitatory activity to other areas of the brain. Any complex neuronal circuit relies on a tight balance between inhibition and excitation to function properly. This is particularly true for the cortex as well as the hippocampus due to extensive recurrent excitation. Reasons for synchronous discharges can either be loss of inhibition, an increase of excitation, or a mixture thereof. In some cases, the causes of epilepsy are clearly attributable to a specific component of the circuit, such as specific genetic mutations of voltage-gated ion channels. Such mutations have been identified for potassium (K), sodium (Na+) and chloride (Cl) channels. These mutations directly affect the excitability of the circuit. One particular example is a loss-of-function mutation in the Na+ channel Beta1 subunit (SCN1B) which leads to slower inactivation and thus more current influx (Wallace et al., 1998).

Epileptic seizures are classified on several different dimensions. Complex seizures result in loss of consciousness whereas simple seizures do not. Generalized seizures arise simultaneously in the entire brain (bilateral) whereas partial seizures arise from a local (unilateral) area of the brain. Partial seizures can spread and progress to a generalized seizure (secondary generalization). Generalized seizures include absence (petit mal) and grand-mal (tonic-clonic) seizures. Absence seizures are special because they are very brief and occur without warning. Patients cease normal activity and stare for only a few seconds and then return to normal immediately afterwards. They have no memory of the epileptic episode. These events can occur hundreds of times a day.

One of the most common forms of epilepsy in humans is temporal lobe epilepsy (TLE) (Engel, 2001; Ojemann, 1997). TLE seizures are usually complex partial and thus manifest themselves with alteration of consciousness. Often, a simple partial seizure precedes the complex partial. This phenomenon is referred to as an aura, as patients get a physical awareness of the onset of seizure before it progresses to cause an alteration in consciousness. In contrast to other types of epilepsies, medial temporal lobe epilepsies are often difficult to control with antiepileptic drugs (AEDs). It has been speculated that this might be due to biased criteria for the pre-clinical evaluation of drug candidates, which are usually screened only for effectiveness for petit mal absences and tonic-clonic seizures. As a result, about 50% of patients (Ojemann, 1997) require other treatment to achieve control of their seizures. For many, surgical removal of the epileptogenic parts of the MTL is an option. 70–90% of patients become free of disabling seizures after surgical treatment (Engel, 2001). This indicates that an underlying cause of TLE is a structural abnormality restricted to the MTL. Post-surgical histology of the removed tissue often shows marked loss of principal neurons of the hippocampus (sclerosis).

Planning for surgery requires extensive preparation to evaluate the location and extent of the resection, as well as an evaluation of possible loss of function resulting from the resection. Accurate localization can be very difficult and time consuming. This is particularly true for one of the most common pathologies that results in temporal lobe epilepsy: hippocampal sclerosis. The neuronal loss in the hippocampus is hard to detect using conventional structural MRI (see below) unless it is severe.

Several non-invasive indicators can be used to identify possible seizure origin areas: structural magnetic reasonance imaging (MRI), computed tomography (CT), positron emission tomography (PET), surface electroencephalography (EEG), or the behavioral symptoms accompanying a seizure. Only about 30% of all seizures are caused by tumors or lesions that are visible on MRI or CT (Engel, 2001). Others can be identified from surface EEG recordings of multiple seizures. If these methods fail to clearly localize the seizure (or if the methods contradict each other), invasive recording techniques can be used (Spencer et al., 2007). These include subdural grids of electrodes placed on the surface of the cortex as well as depth electrodes (see below). Together with video monitoring and surface EEG, such recordings allow accurate and high-resolution tracking of the evolution (pre-ictal, ictal, post-ictal) of a seizure (the “ictal” event) Intraoperative recordings on the surface of the cortex typically only allow the recording of interictal spikes (epileptoform EEG) but not spontaneously occurring seizures. It is thus frequently necessary to implant semi-chronic electrodes to record activity continuously until a seizure occurs. Electrodes are implanted bilaterally at likely epileptogenic sites using a lateral approach. We always implanted electrodes in the hippocampus, amygdala, anterior cingulate cortex, orbitofrontal cortex, and supplementary motor cortex. The exact implantation site was determined based on clinical criteria with the help of co-registered CT, structural MRI, and angiogram. Implantation was guided by a stereotactic frame fixed to the head of the patient (Spencer et al., 2007). Monitoring can take up to several weeks of continuous observation and recording. If activity preceding a seizure has a clear unilateral origin at a restricted set of electrodes (for example, the anterior hippocampus) surgical removal of that part of the brain controls seizures in a large fraction of cases (70–90%) with little or no functional deficits (Spencer et al., 2007; Vives et al., 2007). Since activity has to be recorded continuously (while waiting for a seizure to occur), there are large periods of time where brain function is normal but intracranial signals can be recorded. This gives us (scientists) the unique opportunity to directly observe the electrical activity of the awake human brain during behavior. The data reported in this thesis are all recorded during these periods of time.

It should also be mentioned that as non-invasive diagnostic technologies improve, the need for invasive electrode implantation will decrease. Thus, the window of opportunity to record from epilepsy patients for research purposes could eventually close. While this seems far into the future one should nevertheless keep this in mind when planning a new research program on the approach described here. There are, however, several other surgeries such as deep brain stimulation (DBS) implantation and small resection for treatment of severe psychological problems such as OCD (Williams et al., 2004) that will open up new opportunities to apply this approach.

1.5 Electrophysiology in epilepsy patients

1.5.1 Clinical

For clinical electrophysiology, two primary types of electrodes are used: subdural grids/strips and depth electrodes. Depth electrodes have 4–12 regularly spaced contacts (2–6 mm spacing) along the entire length and can thus be used to record intracranial EEG along the entire depth of the cortex and subcortical structures if implanted perpendicular to the cortex. Grid electrodes have arrays of disk electrodes (3–4 mm diameter) imbedded in a thin sheet of plastic so that the uninsulated side of the electrode rests on the pial surface of the cortex (Wilson, 2004).

1.5.2 Research

Two types of signals acquired from epilepsy patients with implanted electrodes can be used for research purposes. First, the signals originating from the clinical contacts (on grids as well as depth electrodes) can be utilized to record low-frequency (typically < 100Hz) local field potential (LFP). Due to their low impedance (< 1 kOhm) and large size, however, these electrodes do not allow the recording of local, small and fast extracellular currents such as those evoked by spikes. Second, microwires embedded in the depth electrode (the so-called "hybrid depth electrode", AD-Tech Medical Instrument Corp, Racine WI), can be used to record well localized LFP as well as single-unit activity. The wires are made of isolated Platinum/Iridium and are 40 μm in diameter (Fried et al., 1999). We used electrodes with 8 embedded micro wires. Their tip is exposed by cutting the wires to appropriate length (5 mm typical) during surgery. Due to their much smaller exposed tip, microwires have much higher impedance (several 100–500 kOhm at 1 kHz; see methods for detailed measured values). This allows the measurement of single-unit activity with high reliability.

1.5.3 Previous human single-neuron studies

Recording from single neurons in awake, behaving humans offers the unique opportunity to address questions about the function and structure of our brains– addressing questions that have proven difficult or impossible to address using animal models. While it is possible to investigate a large number of mechanisms using animals, there are capabilities which are either very difficult to assess in animals or are unique to humans. These include language, episodic memory, rapid learning, emotions, remote memory, planning, and subjective experience. In the following I will review what has been discovered so far using single-cell recordings from humans (Engel et al., 2005; Kreiman, 2007; Wilson, 2004), with an emphasis on studies reporting findings that would have been hard or impossible to achieve with animals. Also, the focus will be mainly on learning and memory and thus on medial temporal lobe recordings. Such recordings are predominantly from epilepsy patients who are being evaluated for surgery. The many studies reporting single-cell recordings from sub-corticial structures such as the subthalamic nucleus or the thalamus will not be reviewed here. These recordings are made interoperatively from Parkinson's patients and are mostly related to motor responses rather then memory.

Human in-vivo single-unit recordings have a long history. The earliest recordings were made interoperatively before resection of tissue (Verzeano et al., 1971; Ward and Thomas, 1955). Some of the earliest comprehensive studies that used semichronically implanted depth electrodes in the MTL already identified neurons that are selective to specific words, faces, stimulus on/offset, or motor responses (keypresses) (Halgren et al., 1978a; Heit et al., 1988, 1990). This has started a long debate as to whether these responses are visual or memory responses. The ventral visual stream contains consecutively more invariant and selective neurons that respond to very abstract concepts with the highest level of abstraction observed in the inferior temporal cortex. It is thus conceivable that responses in the hippocampus and closely connected cortical areas represent a continuation of such responses. On the other hand, they could be completely different in that high-level ventral stream responses are the fixed “vocabulary” of very well-known entities (such as animals, cars, trees) and MTL responses reflect whether these particular objects had been seen before or not. This strict dichotomy between memory and visual responses seems somewhat artificial, however. After all, any visually selective response that is not genetically innate is a “memory”. I propose that a more natural way to look at this distinction is as a continuous gradient of recency: abstract, long-term, and stable category-type memories (like “animal”) are represented in the ventral stream (inferior temporal areas such as TE and IT in monkeys) whereas more recent or specific memories such as “this particular animal” are represented in the MTL. This gradient could continue even further down the ventral stream to areas such as V4, where very long-term knowledge about basic visual features (such as colors) is represented (Gallant et al., 1996; Gallant et al., 2000). It is conceivable that such responses are plastic as well over the long-term. As a memory becomes more permanent and abstract (such as learning a new category), responses gradually emerge in the inferior temporal lobes. Available data about object selectivity and its emergence in non-human primates is well compatible with this view.

Studies by Fried et al. revealed a much better understanding of the responses evoked by presentation of visual stimuli such as objects and faces (Fried et al., 2002; Fried et al., 1997). First, neurons were identified that distinguished between faces and objects. However, neurons also responded selectively to attributes of faces (such as gender and the emotions happy, surprise, fear, disgust, angry, sad, and neutral). Additionally, many neurons respond differently when being exposed to a stimulus that has never been seen before by the patient (novel) compared to a familiar stimulus. Further recordings by Kreiman et al. revealed that a widespread feature of neurons in the amygdala, hippocampus, and entorhinal cortex is the selectivity to visual categories (Kreiman et al., 2000a). Such neurons show a highly invariant response to any instance of a broadly defined visual class such as animals, cars, objects, or faces. Also, these neurons follow the actual percept rather then the physical (retinal) input in a rivalry design (Kreiman et al., 2002; Reddy et al., 2006). Even when patients are asked to imagine a previously seen picture (such as of the well-known politician Bill Clinton), the same neurons respond both to a picture of Bill Clinton and to an imagined image of Bill Clinton (Kreiman et al., 2000b). Thus the response of these neurons follows the actual visual percept rather then the physical input. Extending this finding, the group of Koch et al. identified neurons which are highly selective (such as for a particular person) as well as highly invariant (Kraskov et al., 2007; Quiroga et al., 2007; Quiroga et al., 2005; Waydo et al., 2006). That is, their response is sparse in the sense that any given neuron only responds to a very small subset of all tested stimuli (response sparseness). Other studies of single-unit activity in relation to learning, particularly comparisons between viewing a stimulus the first and second time, are discussed in the following chapters (Cameron et al., 2001; Viskontas et al., 2006).

Another line of investigation that has benefited greatly from human single-unit recordings has been the study of language (Creutzfeldt et al., 1989a, b; Ojemann et al., 1988; Ojemann et al., 2002). Taking advantage of the possibility to record from temporal cortex that is later resected, Ojemann et al. has recorded a large variety of responses evoked by language comprehension and production. While not part of the MTL, the lateral temporal cortex is known to be crucial for declarative memory of verbal material (Ojemann et al., 1988; Ojemann and Dodrill, 1985; Perrine et al., 1994). Language function is strongly lateralized in the dominant hemisphere and the ability to record from both the dominant and non-dominant hemisphere has contributed to this understanding. Interestingly, only few neurons sampled from the surface of the lateral temporal cortex respond to visual stimulation (in contrast to the MTL responses summarized above). Rather, neurons responded to either silent (reading without pronouncing) or overt speech. Other neurons responded to the memorization or retrieval of verbal memory material. Superior temporal gyrus neurons respond very prominently while subjects listened to spoken language and were often selective to specific combinations of consonants (Creutzfeldt et al., 1989a). Also, neurons were found that respond preferentially to the patient's own voice.

A series of electrical stimulation studies by Halgren et al. revealed crucial insights into the results of direct temporary disruption of the MTL (Halgren et al., 1978b; Halgren and Wilson, 1985; Halgren et al., 1985). While it was well known that bilateral structural damage caused severe amnesia, the causal functions of the MTL can only be established by selective (and reversible) disruption. This has been achieved by injecting a short pulse (100 μs) of current into a number of depth electrodes simultaneously with the onset of a novel or repeated stimulus (Halgren et al., 1985). This single pulse of stimulation disrupted normal ongoing activity for 400 ms. Patients where shown a series of stimuli and had to indicate whether the stimulus had been shown before or not (old/new). Stimulation was applied either during learning, retrieval, or both. This stimulation protocol did not disrupt performance if the delay between presentation of the two stimuli (new and old) was short (2 s). Performance was severely impaired if stimulation was applied during both learning and retrieval (16% correct vs. 66% without stimulation). Most interestingly, performance was severely impaired if stimulation was selectively applied either during learning (no acquisition) or retrieval. The same stimuli that could not be retrieved if stimulated during retrieval could be retrieved if there was no stimulation. This demonstrates a direct causal role of the human MTL in both memory acquisition as well as retrieval (at least of relatively recent memories). It also demonstrated, together with other studies (Chapman et al., 1967; Halgren and Wilson, 1985; Ojemann and Fedio, 1968), that MTL stimulation does not disrupt perception, decision making, response execution, retrieval of remote memories, or otherwise severely alter the cognitive state. It is also interesting to note that patients reported with confidence that they had not seen the stimulus before and not that they did not know or could not answer the question (Halgren and Wilson, 1985).

Electrical stimulation of the temporal lobe (using similar techniques as described above) during the absence of external visual input can evoke a series of phenomena such as deja-vu (a strong sense of familiarity), complex hallucinations, alimentary sensations, fear or anxiety, and amnesia (Bancaud et al., 1994; Halgren et al., 1978b; Penfield, 1958; Penfield and Perot, 1963). Authors have also remarked on the extreme variability of the type of effects evoked (stimulation sites and patients). This confirms that the temporal lobes are directly involved in the retrieval of memories.

1.6 The anterior cingulate cortex

The cingulate cortex is located directly above the corpus callosum (Allman et al., 2001; Paus, 2001). Its anterior part is referred to as the anterior cingulate cortex (ACC). It is thought to play a crucial role in many higher cognitive functions such as error monitoring, attentional control, conflict resolution and reward processing. Similarly, ACC dysfunction has been attributed to several major pathologies such as obsessive-compulsive disorder (OCD), bipolar affective disorder (BAD), chronic pain, and major depression. Surgical removal of parts of the ACC has proven to be a successful treatment of last resort for these pathologies (Jung et al., 2006a; Williams et al., 2004). The ACC also plays a major role in cue-induced craving in drug addiction (Kalivas and Volkow, 2005). Despite this, the function of the ACC remains poorly understood and controversial. One of the common elements of the above pathologies is a major learning deficit. In OCD, for example, inappropriate actions are repeated over and over despite explicit knowledge of their negative consequences. Similarly in cue-induced craving, drug consumption is induced because of a strong association between a cue and rewarding behavior.

A variety of functions have been attributed to the ACC; however, it has proven difficult to study with animal models (lesions, electrophysiology; but see (Frankland et al., 2004; Han et al., 2003)). While this is certainly partly due to the poorly understood function of the ACC, another likely reason is that the ACC is important for cognitive functions that are difficult or impossible to study and quantify in animals. Additional difficulty is added in that the non-human primate homologue of ACC is not clearly anatomically defined and overlaps with the cingulate motor area (CMA). In non-primate mammals the anterior cingulate exists but the anterior part (subgenual) is referred to as the prelimbic and infralimbic areas (Uylings et al., 2003). Also, the

ACC (Brodmann's Area 24) contains especially large spindle cells ("von economo neurons") that exist only in humans and great apes, but not other mammals including monkeys (Allman et al., 2001; Nimchinsky et al., 1999). This suggests that parts of the function of the ACC might be unique to humans and very closely related species.

Here I synthesize the conclusions from a number of studies and critically review what is known about the function of the ACC. I will start by reviewing what has been learned from event-related potential studies and will proceed by discussing how these findings have been extended and revised based on fMRI studies. Finally, I will compare those findings to human lesion studies and point out a number of discrepancies with the fMRI studies.

1.6.1 Reward and Dopamine

Dopamine (DA) is a crucial modulator of learning. It is released by DA neurons located in the ventral midbrain and the striatum. One of the prominent projections of midbrain DA neurons is the ACC (Gaspar et al., 1989). These DA neurons fire in short bursts in response to unexpected rewards, the expectation of reward, as well as novel items (Schultz, 2000). Similarly, many drugs of abuse induce the release of massive amounts of dopamine. Learning-induced plasticity such as long-term potentiation (LTP) is profoundly modulated by the presence of DA. Changing behavior in response to external feedback such as reduced reward is signaled by single neurons in the monkey ACC (Shima and Tanji, 1998). The reversal learning task requires subjects to associate arbitrary stimuli with actions. From time to time, the reward contingencies reverse. Thus, the optimal behavior in this task is to switch to the other action in response to receiving reduced reward. Similarly, if there is no error, the beneficial action should be sustained.

Deactivation of the ACC by lesions (Kennerley et al., 2006; Williams et al., 2004) or temporary deactivation (Shima and Tanji, 1998) profoundly impairs performance in this task in both monkeys and humans.

1.6.2 Event-Related Potentials

Before the advent of fMRI studies, the function of the ACC had mostly been studied with event related potentials (ERPs). Using reaction time (RT) tasks that require a fast response to sometimes conflicting stimuli, subjects make errors. Subjects are usually aware that the response was an error before the feedback signal. An example of such a task is the Stroop interference task (Kerns et al., 2004): subjects are shown a word on a screen, printed in a particular color. The task is to respond, as fast as possible, by pressing a button that indicates the color the words are shown in. For example, the word could either be red or green. These two words are printed either in red or green. Of the 4 possible combinations, 2 are congruent (red, green) and 2 incongruent (red, green). Incongruent trials require a significantly longer time to respond than do congruent trials (typically 40–80 ms on average). Additionally, if speed is more important than accuracy, incongruent trials evoke more erroneous responses.

A prominent observation during such tasks is a negative potential referred to as the error-related negativity (ERN). The potential peaks over frontal-parietal electrodes at 100 to 150 ms after the response (Paus, 2001). But because ERPs are recorded on the scalp, it is impossible to localize the source of the signal precisely. However, dipole models can be utilized to predict possible configurations of electric sinks and sources that could account for the data. One possible model is a dipole located in the ACC. These observations motivated a number of imaging studies that attempted to definitively localize the source of this potential.

1.6.3 Neuroimaging studies: PET and fMRI

Early imaging studies, comparing blocks of trials between which only the variable of interest changed ("block design"), have suggested that the ACC is generally involved in the executive control of cognition. These studies have used tasks that require selective attention, working memory, and self-monitoring for errors (Bush et al., 1998). None of these tasks was found to elicit ACC activity that was specific to a particular function and it was thus proposed that the ACC generally responds to task difficulty, irrespective of the specific task. PET studies have generally reached the same conclusion (Paus et al., 1998): ACC activity is most strongly correlated with task difficulty. Since these results are achieved by subtracting the activity of two different blocks (easy and difficult), task difficulty refers to any possible variable that can make a task more difficult. This, for example, includes different demands for working memory, attention, difficulty of cognitive analysis (e.g., reading versus telling the font), and increased demand for motor commands (e.g., precision of movement). Because many experimental manipulations have been observed to change ACC neural activity, a number of different hypotheses for ACC function have been advanced. Two influential theories are the "conflict monitoring" and the "error detection theory". It has, however, proven difficult to clearly disambiguate the predictions that different theories make by using blocked designs.

1.6.4 Conflict monitoring, error monitoring, and cognitive control

One of the prevalent views is that one of the functions of the dACC (dorsal part of the ACC) is conflict monitoring. This interpretation was primarily developed because of earlier blocked studies of the Stroop interference task that showed enhanced activation of a part of the dorsal ACC when comparing blocks with and without interference effects. However, interference can cause multiple effects. Firstly, it creates a conflict between a fast, overtrained automatic process (reading the word) with a slower process (telling the color of the ink the word is printed in). Secondly, the very same effect increases attentional load, task difficulty, and behavioral errors. It is thus important to attempt to dissociate between these effects. One possibility is to use an event-related design where the degree of interference is modulated separately from task difficulty. One study (Carter et al., 1998) applies this approach. Carter et al. use a task where the subject is first presented with a cue (A or B) and than later with a probe (X or Y). The subject is instructed to only respond if the cue is an A and the probe an X. All other combinations are presented with low probability. They can be divided into low (BY) and high (BX,AY) interference trials. Task difficulty was modulated by removing pixels from the cue and probe stimuli (harder to read). The authors found that activity within ACC was higher for interfering vs. non-interfering trials (compatible with conflict monitoring hypothesis). However, they also found increased activity for error trials vs. correct trials. Interestingly, the authors found that ACC activity was also increased (relative to baseline) for the correct and non-interfering trials, but less than in error or interference trials. This finding resolves the previous debate in demonstrating that ACC is actually activated by both error and interference. However, this finding does not rule out that both types of activation are the result of a common underlying cause. There are multiple possibilities that could cause this pattern of activity without any relationship to conflict or error monitoring. One example is increased attentional load, which could be caused both by an error as well as by interference. Also, errors are not independent from interference: more errors are made if interference is higher. The conclusions that can be drawn regarding the function of the ACC from these types of design are thus limited.

Two follow-up studies (Botvinick et al., 1999; Kerns et al., 2004) have shed light on this issue by using a Stroop interference task (color naming, see above). In an attempt to disambiguate conflict monitoring from error-related activity, the authors compared BOLD ACC activity of incongruent trials (I) which were either followed by a congruent (c) or an incongruent (i) trial (cI or iI). If ACC activity is related to interference itself, activity on both types of trials should not differ. If, however, ACC activity relates to the monitoring of conflict and the subsequent induction of control, activity on iI trials should be lower than activity on cI trials. Also, the reaction time for the I trials which follow an incongruent trials should be inversely correlated with the strength of ACC activity on the preceeding i trial. This is indeed what the authors found. This strengthens the hypothesis that one function of the ACC is the monitoring of conflict and the ensuing induction of control. This argument is supported by the observation that there was a trial-by-trial correlation with strength of ACC activity in the first incongruent trial with the reaction time on the following incongruent trial.

Another area which is frequently seen to increase activity with task difficulty is the dorsolateral prefrontal cortex (DLPFC). In many tasks the ACC as well as the DLPFC increase activity under the same conditions. It has thus remained unclear whether the two areas have different functions. Mainly based on human lesions, the prefrontal cortex has long been implicated in cognitive control whereas ACC has been implicated in the inhibition of control. Also, DLPFC has been observed in the absence of ACC activity in working memory tasks (Fletcher et al., 1998), whereas ACC activity without DLPFC activity has been observed for incongruent response situations like in the Stroop task. To combine these two types of tasks, (MacDonald et al., 2000) have used a modified version of the Stroop color naming task. Before the start of a trial, an instruction was displayed as to whether the color of the word (ink) or the meaning of the word should be reported in this given trial. BOLD activity was examined both during the display of the instruction as well as during the response period. As previously reported, greater ACC activity was found for incongruent vs. congruent trials during the response period. DLPFC activity was elevated but equal for both types of trials. However, left DLPFC activity was different during the instruction period for color vs. word instructions. The authors conclude that this pattern of activation is indicative of a role for the DLPFC in cognitive control (setting of task) and the ACC in conflict monitoring. This finding is also supported by the study discussed in the previous paragraph (Kerns et al., 2004), which found that ACC activity predicted the extent of later PFC activity. That is, ACC signals a conflict, and due to this PFC takes the necessary actions ("control").

1.6.5 Error likelihood and reward

Another study (Brown and Braver, 2005) has challenged the above finding by proposing that ACC activity represents error likelihood rather than a conflict monitoring signal. The authors directly compared these two hypotheses using a go/nogo task with high- and low-error trials. Subjects were instructed to respond with a left or right button press to a left or right arrow (go cue) appearing on the screen. However, in some trials, the go cue was reversed shortly after displaying it (no-go). The incidence of the no-go signal appearing was low in low-error trials and high in high-error trials. The subject was instructed by a color cue whether the current trial was a high- or low-error trial. Following the error monitoring hypothesis, activity should not differ between correct low- and high-error trials that were not followed by the no-go cue. In contrast, following the error likelihood hypothesis, activity in the high error-likelihood trials should be higher than the low error-likelihood trials, regardless of whether the no-go cue was actually displayed or not. This is because both trials have a higher likelihood of error, regardless of whether the error actually happened or not. Also, the comparison between high and low error-likelihood error trials that were not aborted (go cue) allows excluding effects of conflict (there is none) as well as error monitoring (there is none). The only parameter that differs is error likelihood. Indeed, the authors found that the fMRI signal measured in the ACC follows the error-likelihood hypothesis. That is, the signal was positively correlated with the potential negative reinforcement associated with a given trial. This is also in line with another study that found that fMRI ACC activity was positively correlated with the false alarm rate (Casey et al., 1997), because the higher false alarm rate was presumably related to trials where more errors could be made.

Another possible function of the ACC is the representation of some function of reward. This is in addition to uncertainty, as discussed in the previous paragraph. ACC modulation by reward is expected because it is known that the ACC has the highest density of innervation by the mesocortical dopamine system originating in the midbrain (Schultz and Dickinson, 2000). The midbrain dopaminergic system responds strongly to expectancy mismatches of reward. While these dopaminergic connections itself are not capable of exciting neurons in the ACC (they are modulatory), dopamine is known to have a strong modulatory influence on synaptic plasticity and could thus influence ACC firing indirectly. Indeed, (Critchley et al., 2001) found that the extent of ACC activity (fMRI BOLD) is positively correlated with the amount of uncertainty (risk) as well as arousal (measured by skin conductance).

1.6.6 Lesions and human intracranial recordings

Given the contested nature of the function(s) of the ACC it is instructive to consult studies of the human ACC that look at causal rather than correlative effects. A rare opportunity of doing so are human patients that have lesions restricted to the ACC. One study investigated the performance in Stroop interference and go/nogo tasks in 4 patients with damage to the dorsal ACC and compared it with 12 healthy controls (Fellows and Farah, 2005). Overall reaction time in both the go/nogo task as well as the Stroop task were higher (slower) for the lesion patients. However, both the error rate as well as the size of the Stroop effect (percentage RT difference congruent vs. incongruent) were not different. Also, the modulation of the error rate and the Stroop effect size by high vs. low conflict was equal to controls. Slowing of the RT following an incongruent trial is considered one of the effects of cognitive control induced by ACC activity. However, it was observed at the same rate as in controls. Surprisingly, this study thus finds no difference in all measures of conflict that are traditionally considered functions of the ACC. Other studies (Stuss et al., 2001; Vendrell et al., 1995), however, agree with this finding. This finding is also in line with a number of non-human primate single-unit recordings that have generally failed to find neurons that respond to Stroop-like incongruence tasks. Rather, the neurons were found to respond to functions of reward expectancy (Shidara and Richmond, 2002). The same has been found by a human study of selective cingulotomy patients (Williams et al., 2004) that allowed behavioral and electrophysiological measurements of performance on a task before and after resection. They found that performance was only impaired on trials that are related to reward, but not without.

The difficulty of studying ACC at the single-unit or LFP level is also illustrated by a recent study (Wang et al., 2005): the authors administered several different tasks to the same subjects and recorded LFPs from depth electrodes implanted for purposes of epileptic seizure localization. Tasks included auditory oddball detection, new/old word recognition (memory), and a reaction time task. Interestingly, the authors found activity that distinguishes between the different task elements (new/old, wrong/correct, rare/frequent) at each site—indicating that the function of the ACC at this particular site could not be attributed to one of these very different processes exclusively.

Chapter 2. Online detection and sorting of extracellularly recorded action potentials

2.1 Introduction[1]

Recent technological advances have made it possible to simultaneously record the activity of large numbers of neurons in awake and behaving animals using implanted extracellular electrodes. In densely packed neuronal structures such as the cortex and the hippocampus, the activity of multiple neurons can be recorded from a single extracellular electrode. A complete understanding of neural function requires knowledge of the activity of many single neurons and it is thus crucial to accurately attribute every single spike observed to a particular neuron. This task is greatly complicated by uncertainties arising from noise caused by firing of nearby neurons, inherent variability of spike waveforms due to bursts or fast changes in ion channel activation/deactivation, uncontrollable movement of the electrodes, and external electrical noise from the environment.

There are two different ways to acquire and analyze electrophysiological data: i) store the raw electrical potential observed on all electrodes and perform spike detecting and sorting later (offline sorting), or ii) detect and sort spikes immediately (during acquisition) and only store the sorted spikes (realtime online sorting). A combination of the above approaches is to detect spikes

[1] The material in this chapter is based on Rutishauser, U., Schuman, E.M., and Mamelak, A.N. (2006b). Online detection and sorting of extracellularly recorded action potentials in human medial temporal lobe recordings, in vivo. J Neurosci Methods *154*, 204-224.

online and only store the detected spikes for later offline sorting. While it is reasonable to use offline sorting methods in certain cases, it is becoming increasingly necessary to develop realtime online sorting methods. There are three main reasons to use such methods: i) Realtime online decoding allows "closed-loop" experiments, e.g., the adaptation of the experiment to the specific neural responses observed (compare to dynamic clamp on the single-cell level, see (Prinz et al., 2004) for a review); ii) Fast data analysis: sophisticated offline spike sorting methods require extensive amounts of computation, whereas online sorting allows immediate data analysis. iii) massive reduction in data transmission and storage. Moving from offline sorting to realtime online sorting requires two separate technological advances: i) developing an online spike detection and sorting algorithm, and ii) developing a realtime implementation of this algorithm. The first condition is strictly necessary before a realtime version can be implemented and presents the main methodological challenge that needs to be addressed. An algorithm that is online only uses information available at the current point in time and not information available in the future. Applied to our approach, “online sorting” means that a spike observed at time t is sorted only using all information observed prior to and including point of time t. This is in contrast to offline sorting algorithms, which require that all spikes are available before sorting can start and thus require that all data is acquired and stored beforehand. Removing this requirement for total spike availability presents a formidable challenge and we focus exclusively on doing so in this paper. Note that it will be possible to implement the algorithm presented here for realtime analysis of many channels in parallel; this will be the focus of our future efforts.

While the problem of offline sorting has been intensively investigated (for a review see (Lewicki, 1998), but also see (Abeles and Goldstein, 1977; Fee et al., 1996a; Harris et al., 2000;

Pouzat et al., 2004; Pouzat et al., 2002; Quiroga et al., 2004; Redish, 2003; Sahani et al., 1998; Shoham et al., 2003)), relatively little work has been done on online sorting. Early attempts at online sorting focused on techniques which require manual definition of each cluster before sorting commences (Nicolelis et al., 1997). Other online classification approaches require a learning phase, after which neurons are classified in realtime (Aksenova et al., 2003; Chandra and Optican, 1997). The disadvantage of this class of online methods is that only neurons which fire during the learning phase can be classified. In addition, if the spike shapes change during the experiment, the neuron can no longer be recognized. In this paper, we present and demonstrate an online spike detection and sorting method. Spikes originating from different neurons are distinguished based on spike waveform shape and amplitude differences, features which are unique for individual neurons. The algorithm iteratively updates the model and assigns spikes to clusters. It thus does not require a separate learning phase and is capable of detecting new neurons during the experiment. This feature is particularly crucial for experiments with human subjects because firing is very sparse and the "optimal" stimuli for recorded neurons are often unknown. As a result, it is not possible to excite all neurons during a learning phase that precedes the experiment. We will discuss this point further at a later stage in the paper.

We demonstrate our method by applying it to data recorded from arrays of single-wire depth electrodes that are semi-chronically implanted in the medial temporal lobe of human epilepsy patients. This analysis is particularly challenging because the data were acquired in an electrically noisy clinical setting without the option of re-positioning the electrodes to optimize spike detection. As a result, the data are compromised by low signal-to-noise ratios (SNR) as well as non-stationarities in the noise levels. Additionally, electrodes are implanted in densely packed neuronal structures (for example, the hippocampus), which complicates separating single-unit activity. These neurons generally have very low basal firing rates and can respond very selectively to certain stimuli.

Our experimental setup allows us to conduct long-term recordings simultaneously with complex behavioral experiments which can only be done with awake behaving humans. In these experiments, fast data analysis is highly desirable. Our patients are extremely rare (< 8 a year) and our recording sessions are short (1–4 hours). Although we can record for 1–5 days, the same neuron cannot be obtained with any reliability on subsequent recording days. There is always a trade-off between sorting quality and fast data analysis, but in this kind of experiment it is crucial to know as fast as possible to what a neuron responded, so that the experiment can be adapted immediately. One possible compromise to achieve this is to use a simple, but online, algorithm which is capable of detecting most neurons and correctly sorting their spikes. This approach is reasonable for recordings from chronically implanted arrays of electrodes that do not allow for the individual movement of the electrodes to optimize response properties. Additionally, implanted arrays allow the simultaneous recording of many neurons over a long period of time and thus yield large amounts of data. However, it has proven difficult to store, process, and analyze these large data sets because efficient methods for processing and analysis are lacking (see (Buzsaki, 2004) for a discussion of these issues). An online spike detection and sorting algorithm, such as the one described below, will enable experimenters to process complex and large amounts of data in an efficient and effective way.

2.2 Methods

2.2.1 Glossary of mathematical symbols and notation

Symbol	**Definition**
$\vec{S}_i$	The raw waveform of spike i
$\vec{M}_j$	Mean waveform of cluster j
$\lvert\vec{M}_k\rvert$	Number of spikes assigned to cluster k
m	Total number of mean waveforms
C	Number of spikes used to calculate mean waveforms (last N spikes assigned to each cluster)
T_S, T_M	Threshold for sorting (S) and merging (M)
N	Number of datapoints of a single waveform
$\vec{D}$	Vector of distances
$\vec{Z}$	Matrix of noise traces (with N datapoints each, each row is a noise trace)
$\vec{C}$	Noise covariance matrix (dimensions: NxN)
$\vec{P}_i$	Prewhitenend raw waveform of spike i
d_S, d_M	Distance between 2 clusters for sorting (S) and merging (M)
d	Distance between 2 clusters (projection test)

All population measurements are specified as mean ± standard deviation.

The raw waveform of spike i is referred to as $\vec{S}_i$. A waveform is a vector that consists of N=256 datapoints. For every spike i, $\vec{S}_i(l)$ refers to the amplitude of the waveform at the sampling point l (l can take any value between $1...N$). T denotes the threshold and is always a scalar. $f(t)$ and $p(t)$ refer to the bandpass filtered raw signal amplitude and the local energy at time point t respectively.

2.2.2 *Filtering and spike detection*

Spikes are detected using threshold crossings of a local energy measurement $p(t)$ of the bandpass filtered signal (Bankman et al., 1993; Kim and Kim, 2003), which allows more reliable spike detection than thresholding the raw signal (Appendix A). If $p(t)$ is locally bigger than five times the standard deviation of $p(t)$, (or another factor, referred to below as the extraction threshold), a candidate spike is detected (Csicsvari et al., 1998). For each threshold crossing (Figure 2-1C,D), a sample of 2.5 ms (64 samples at a 25 kHz sampling rate) is extracted from the filtered signal. This sample is upsampled 4 times using interpolation (Bremaud, 2002), that is, by transforming the sample to Fourier space using FFT and back with more data points. After upsampling, the spike is sampled at 100 kHz and consists of N= 256 data points, with the maximum realigned at position 95: $\arg\max_l(S_i(l)) = S_i(95)$. Upsampling eliminates the roughness in the waveform introduced by undersampling the signal and the high-pass filtering, and also allows a more accurate determination of the real peak of the waveform. Note that the peak of the waveform is typically not measured accurately because it is only reached for a very short time and thus often falls between points of time at which the signal is sampled.

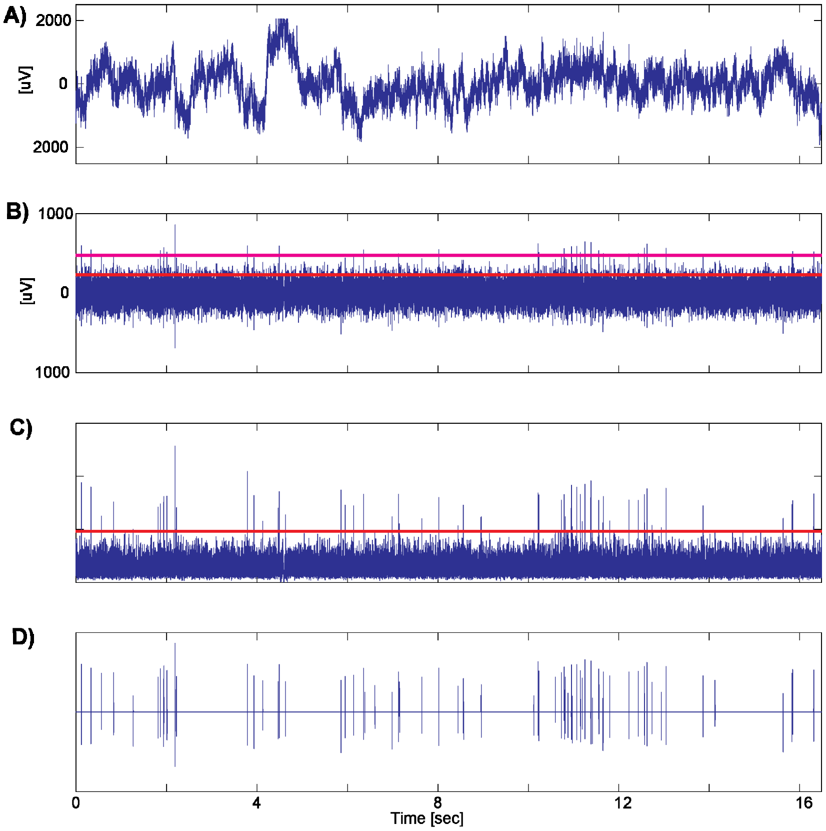


Figure 2-1. Filtering and detection of spikes from continuously acquired data. Shown are 412000 timepoints, corresponding to 16.48 sec at a sampling rate of 25000 Hz. **A)** Raw signal. The amplitude is in units as measured after amplification, not corrected for gain. **B)** Bandpass filtered signal 300–3000 Hz. The two lines indicate possible thresholds for direct spike extraction (see text). **C)** Average square root of the power of the signal, calculated with a running window of 1 ms and thresholded (line). The y axis is arbitrary. **D)** Position and amplitude of detected spikes (detected in C), but extracted from B).

2.2.3 Distance between the waveforms of two spikes

The estimation of the number of neurons present, as well as the assignment of each spike to a neuron, is based on a distance metric between two spikes (Appendix A). Based on this distance, a threshold is used to decide i) how many neurons are present, and ii) to assign each spike uniquely to one neuron or to noise, if unsortable. A crucial element of this approach is the threshold, which is calculated from the noise properties of the signal (Appendix A) and is equal to the squared average standard deviation of the signal, calculated with a sliding window. The threshold is thus not a parameter as it is automatically defined by the noise properties of the recording channel and is equal to (in a theoretical sense) the minimal signal-to-noise ratio required to be able to distinguish two neurons. It is assumed that the background noise is additive (see results) and the presence of a spike does not influence the noise properties (Fee et al., 1996b). It can thus be assumed that the variance of the noise of all waveforms of the same neuron is approximately constant (Pouzat et al., 2002). One concern is that the estimation of the threshold is strictly valid only if it is independent of the number of neurons and their spiking frequency on a specific channel. It is worth noting, however, that even if there exist multiple neurons, each with high spiking frequency, most data points of the raw signal will not belong to a spike (but see (Quiroga et al., 2004)). We are thus assuming that the variance of the raw signal is approximately independent of the number of neurons (Fee et al., 1996b).

2.2.4 Online sorting

Each newly detected spike is sorted as soon as it is detected (Figure 2-2). The raw waveform of a newly detected, as of yet unsorted spike, is used to calculate the distance to all already known mean waveforms (clusters). The spike is assigned to the existing cluster to which it has minimal distance if the distance is smaller than a threshold value. If the minimal distance is larger than the threshold, a new cluster is automatically created. Every time a spike is assigned to a cluster, the mean waveform of that cluster is updated by taking the mean of the last C spikes that were assigned to this cluster. This causes the mean waveforms of each cluster to change as well, which might result in two clusters which have mean waveforms whose distance is less than the threshold. In this case, the two clusters become indistinguishable and they are thus merged. The spikes assigned to both clusters will be assigned to the newly created cluster (see Appendix B for details of the algorithm). Note that not every cluster created in this manner will represent a single unit. In fact, many small clusters will be created which represent noise. These can easily be discarded by requiring a minimal number of spikes for a valid cluster. However, noise of a stereotypic shape will create large clusters; these are also discarded. See the section below on how to evaluate potential single-unit clusters below for a discussion of this issue.

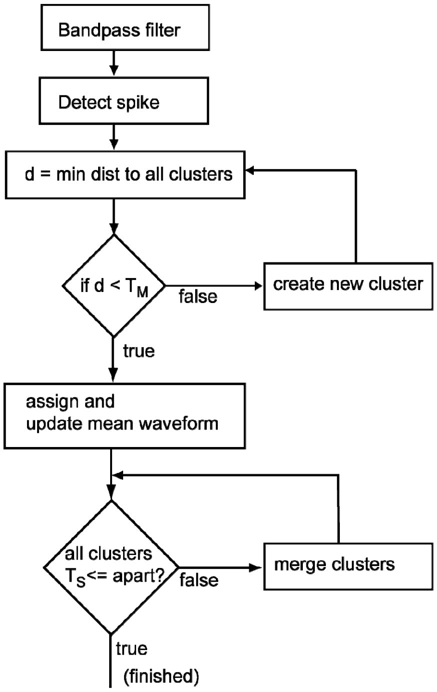


Figure 2-2. Schematic illustration of spike detection and sorting.
The signal is (continuously) bandpass filtered 300 - 3000Hz. Spikes are detected by thresholding a local energy signal that is continuously calculated from the raw filtered signal. After detection and appropriate re-alignment, a distance metric is used to calculate the distance to all known clusters at the current point in time. If the minimal distance is smaller than a threshold T_M , the spike is assigned to this cluster. Otherwise, a new cluster is created and the new spike is assigned to it. The thresholds are automatically and continuously calculated from the noise properties of the raw filtered signal. After assigning a spike to a cluster, that cluster's mean waveform is updated accordingly. This enables tracking of moving electrodes as well as short-term changes due to bursts. After updating the mean waveform, clusters might overlap. If this is the case, they are merged and the spikes assigned to the cluster are reassigned. Periodically, the statistical evaluation criteria (ISI distribution, power spectrum, and autocorrelation) as well as the projection test for each pair of clusters are calculated. This allows us to continually discard noise and multi-unit activity.

2.2.5 Calculating the threshold

There are two thresholds used in the algorithm: The threshold for considering a new spike part of an existing cluster T_S and the threshold for considering two clusters apart T_M . We considered two possible ways of estimating these two thresholds from the background noise of the raw signal. Common to both are that they are calculated automatically from the data.

The first (exact) approach is to pre-whiten the waveforms of detected spikes using the covariance matrix of the noise (see Appendix D). In this way, the datapoints of a given waveform can be considered uncorrelated and the noise is white and of standard deviation 1 in each dimension (by design). The summed squared residuals of the difference between two waveforms (Eq 3b) can thus be considered χ^2 distributed with the number of degrees of freedom equal to the number of datapoints that constitute a waveform. The threshold of the distance calculated as such can be estimated from the χ^2 distribution (Eq 5). The distance between the mean waveforms of two clusters can be calculated as the square root of the summed squared residuals, which is, by definition, the standard deviation multiplied by the number of datapoints. The threshold for merging can thus be set in terms of number of standard deviations by which clusters should be separated until they are considered equal. This procedure allows us to estimate the two thresholds T_S and T_M automatically by using the covariance of the noise. While this is the statistically optimal estimate of the thresholds, it requires an accurate estimate of the covariance. This turns out to be a non-trivial task for real data and its iterative computation is computationally expensive. Additionally, pre-whitening requires computation of the inverse of the covariance matrix. Unfortunately, the determinant of the covariance matrix is often small (close to singularity), which makes this operation numerically unstable in some situations. To circumvent this problem we also tested the algorithm by using an approximated version of the threshold which does not require pre-whitening of the waveforms. The approximated thresholds (both T_S and T_M) are equal to the variance of the raw signal (Eq 4a). The distance between two waveforms, both for sorting and merging, is calculated as the sum of the squared residuals of the difference between two waveforms (Eq 3a). Here, the raw waveforms (after upsampling and re-alignment) are used. No pre-whitening is performed. In the results section we present performance estimates for both the exact as well as the approximation method for estimating the threshold.

2.2.6 Simulation of synthetic data

Simulated raw data traces were generated by using a database of 150 mean waveforms taken from well-separated neurons recorded in previous experiments. To generate random background noise, a large number of those waveforms were randomly selected, randomly scaled, and added to the noise traces. Executing this procedure many times resulted in realistic background noise, as judged by comparing the raw signal, the filtered signal, and its autocorrelation (Figure 2-3) to the real data. This random background noise trace can be arbitrarily rescaled to a pre-specified standard deviation to simulate different noise situations. Noise is scaled to a standard deviation of 0.05, 0.10, 0.15, and 0.20.

Identifiable neurons are added by simulating a number of neurons (between 3 and 5 in the following cases) with a renewal Poisson process with a refractory period of 3ms and a fixed firing rate between 1 and 10 Hz (which corresponds to the typical firing rate of real neurons in our data). For each neuron, one pre-defined mean waveform was used. Mean waveforms were re-scaled such that they were bounded in the range [-1..1] (arbitrary units). By systematically varying the noise levels, signal-to-noise ratios (SNR) comparable to those observed in real data were simulated. We calculate the SNR ratio (Eq 6 in Appendix A) as the root mean square value of the mean waveform divided by the standard deviation of the noise (Bankman et al., 1993). The average SNR is calculated by averaging the SNR of each waveform. To aid comparison, this method of generating simulated raw data traces was intentionally chosen to be essentially the same as the one used by Quian Quiroga et al. 2004.

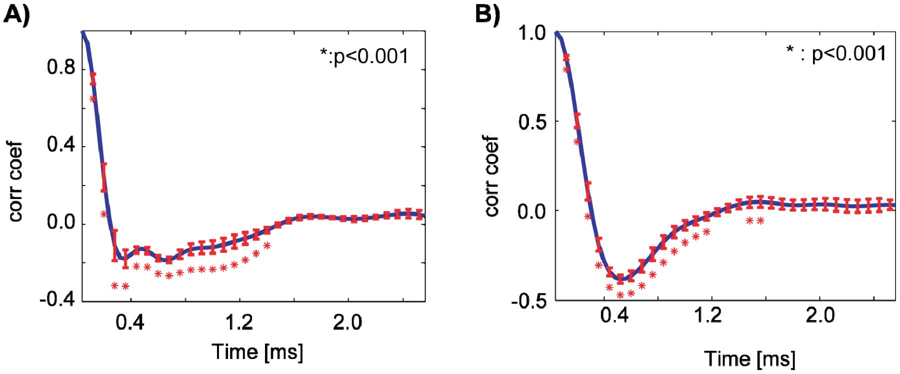


Figure 2-3. Autocorrelation of raw data.
Autocorrelation of real (A) and simulated (B) data. The autocorrelation is calculated from noise traces (which do not contain spikes). **A):** Autocorrelation of the raw signal from real data. Notice that the signal is strongly autocorrelated untill approximately 1.2 ms. **(B)** Autocorrelation of simulated data. The autocorrelation remains significant up to 1.2 ms (stars indicate $p < 0.001$, t-test for null hypothesis mean = 0). Error bars shown are ± s.d. (n = 8542 noise traces).

2.2.7 Extracellular recordings

We use data recorded from human patients implanted with hybrid chronic depth electrodes to treat drug-resistant epileptic seizures. The electrodes contain an inner bundle of eight 50 μm microwires that extend approximately 5 mm beyond the tip of the depth electrode (Fried et al., 1999). The clinical reason for implanting electrodes is to record electrical activity during epileptic seizures to locate the anatomical locus of seizure onset.

Electrodes were surgically removed approximately 2–4 weeks after implantation. Recording sessions, each 1–2h long, started approximately 48 hours after electrode implantation and lasted up to 4 days. We recorded extracellularly from 3 macroelectrodes with a total of 24 single channels (each connected to a single wire). One wire of each macroelectrode (with low impedance) was used for local grounding. Electrodes were implanted in the amygdala and hippocampi of subjects and data was recorded while subjects performed visual psychophysical experiments, similar to those reported in (Kreiman et al., 2000a), as well as other behavioral experiments, such as navigating in a virtual world. Data were acquired continuously with a low-pass cutoff of 9 kHz, sampled at 25 kHz, and stored for later analysis. The gain of the amplifiers (Neuralynx Inc) was set individually on a case-by-case basis (based on electrode impedance and noise) in the range of 20000 to 50000, with an additional A/D gain of 4.

All subjects gave informed consent to participate in the research, and the research was approved by the Institutional Review Boards of both Huntington Memorial Hospital and the California Institute of Technology. The location of the implanted electrodes was solely determined by clinical requirements for locating the seizure onset and the research team had no influence on electrode placement. The exact location of the electrodes was determined from high-resolution structural MRI images taken immediately before and after electrode implantation.

2.2.8 Criteria to identify clusters representing single-units

A collection of spikes is well separated if the following criteria are met: i) a small (e.g., < 3.0 %) percentage of all spikes have an ISI of less than 3 ms (refractory period), ii) the power spectrum is within ± 5 standard deviations in the range of 20–100Hz (excluding < 20 Hz because of theta/gamma oscillations) and does not go to zero for high frequencies (Poisson process). Note that at low frequencies (< 40Hz), a dip is expected due to the refractory period (Franklin and Bair, 1995; Gabbiani and Koch, 1999).

2.2.9 Quality of separation evaluation criteria

We use a statistical tool commonly called a *projection test* to quantify both the degree of overlap between the clusters and the goodness-of-fit to the theoretically expected distribution of spikes around the cluster center. In the context of spike sorting this test was originally proposed by (Pouzat et al., 2002). We only summarize the procedure here and mention some additional problems associated with it (see also Discussion and Appendix D): The raw waveforms are first pre-whitened (e.g., decorrelated) using the known autocorrelation (Figure 2-3) of pure noise segments (where no spikes were detected). Mathematically, this implies that the noise must be of full bandwidth and the covariance matrix of the noise traces is thus invertible. However, this is not always the case. See appendix D for further discussion of this issue. After this step, each datapoint of the raw waveform is independent of all the others, with white noise of standard deviation 1. This is done for the waveform of each detected spike. Afterwards, each waveform (with N datapoints) can be regarded as one point in N-dimensional space. The center of a cluster is represented by the point in N dimensional space that corresponds to the mean of all waveforms assigned to the cluster. Since the noise is white with a known standard deviation of 1, the theoretically expected distribution of spikes of the same cluster around this center is known (a multivariate Gaussian with a standard deviation of 1).

For any pair of clusters found on a single wire, the projection test can be applied to quantify the overlap between the two clusters. This is done by projecting the difference of every spike and the center of the cluster it is assigned to (residuals) onto the vector that connects the two centers of the clusters. This results in two distributions of a single one-dimensional quantity, centered on the two centers (Figure 2-5D and Figure 2-7D). The distance between these two centers can conveniently be used as a measure of separation. If the distance is too small, one or both of the clusters have to be discarded. If the goodness-of-fit of the two clusters to the expected distribution is reasonably good (see below), then the overlap can be estimated: a distance of >5 guarantees an overlap of less than 1%, a distance > 3.2 an overlap less than 5% and a distance of > 2.8 an overlap of less than 7.5%. Please see the discussion for an application to our data.

For any given pair of clusters, the theoretically expected distribution (normal with standard deviation = 1) of the projected residuals can be compared against the empirically observed distribution. We use a R^2 goodness-of-fit between the empirically- estimated probability density function and the theoretically expected probability density function to quantify this. Note that the empirically estimated distribution of the same cluster can look different if compared to different (other) clusters, since the residuals are a projection of the residuals onto the vector connection the two centers (e.g., Figure 2-5D the first 2 subplots, where cluster 1 is compared against cluster 2 and 3). The projection test can either be applied posthoc after sorting is finished or periodically (e.g., every few minutes) during the recording session. If it is applied periodically, clusters that don't qualify can be discarded automatically.

2.2.10 Implementation

We implemented the proposed system in MATLAB (Mathworks, Natick, MA) to assess its usefulness and evaluate its properties. The implementation is split into two parts: spike detection and sorting. Spike detection reads a raw data stream either from the network (broadcast by the acquisition system) or from a file and detects spikes. The raw data stream is in the Neuralynx (Neuralynx Inc, Tucscon, AZ) NCS format. The detected spikes are passed on to the online sorting part, which sorts the spikes one-by-one, as they become available. The results of the sorting are stored and later analysed using the statistical methods described. Our implementation is not optimized for speed at this time. All running time measurements were made on the same machine (Intel Xeon 3Ghz) with MATLAB version R14SP1.

2.3 Results and Discussion

2.3.1 Signal acquisition and filtering

The continuously recorded signal (with a sampling rate of 25 kHz, Figure 2-1A) is bandpass filtered by a 4-pole Butterworth filter with a high-pass frequency of 300 Hz and a low-pass cutoff of 3000 Hz (Figure 2-1B) to exclude both the low-frequency components, e.g. local field potentials (LFP), and high-frequency components (noise) of the signal.

2.3.2 Spike detection

Spike detection from raw data with high noise levels (Figure 2-1B) was reliably achieved using the local energy thresholding method (see Methods). Figure 2-1C demonstrates the advantage of the method: whereas the spikes between 8 s and 10 s (x axis) cannot be detected in the filtered signal (Figure 2-1B), they are reliably picked up by the local energy signal (Figure 2-1C).

2.3.3 Waveform extraction and re-alignment

For every spike detected, 64 data samples are extracted, with the peak at sample 25. The waveform is then upsampled 4x and re-aligned again, such that the peak is at sample 95 (see methods for details). Re-aligning twice, once before extraction and once after upsampling, is crucial because the upsampling will change the location of the peak. The position of the peak is estimated more accurately after upsampling. The accurate determination of where the peak of the waveform is located is crucial. This is, however, difficult and great care needs to be taken to avoid the erroneous splitting of one cluster into two because of re-alignment issues. This situation arises because we observe many very different waveforms in our recordings. Often the waveform has a dominant peak in either the positive or negative direction, but sometimes the situation is less obvious. Consider, for example, the 3 waveforms shown in Figure 2-4C. Whereas the blue and the red waveform have a dominant peak on the positive and negative side respectively, the situation for the green waveform is less clear. It has a peak of approximately the same amplitude in the negative and positive direction and either could be used for re-alignment. This situation is not artificial and arises often in our recordings (e.g., Figure 2-7A). If the simplest re-alignment procedure is chosen, e.g., re-align all spikes at their absolute maximal amplitude, the spikes originating from the green neuron shown would artificially be split into two clusters. This is because variance caused by noise would sometimes make the negative peak maximal and sometimes make the positive peak maximal. The strategy we have found to avoid this problem as best as possible is to use the order in which the peaks occur. If the peak in the negative direction appears before the peak in the positive direction, the waveform is re-aligned at the negative peak. If, on the other hand, the positive peak appears before the negative peak, the positive peak is used to re-align. Exceptions to this procedure are used if only one or none of the peaks are significant, that is, their peak amplitude is less than the standard deviation of the noise (see Algorithm 3 in Appendix C). Using this procedure, we can accurately re-align and sort spikes such as the one shown in Figure 2-4C. However, there are still situations in which this method is not able to correctly re-align spikes. For example, if the waveform of a neuron has a first peak which is barely significant and a peak which is highly significant, the cluster will be artificially split. This will only be the case for neurons which are close to the distinguishable signal-to-noise level and in our experience this case is rather rare. But in the rare occurrence, this problem is detected by the projection test and this cluster is then discarded.

2.3.4 Evaluation of sorting— synthetic data

We performed spike detection and online sorting on synthetic data to evaluate the online algorithm's performance. Data were simulated to resemble the real data as closely as possible. Specifically, we observe that the noise in our data is strongly autocorrelated (Figure 2-3) and thus we do not assume independent Gaussian noise. Rather, the noise itself likely consists of many randomly mixed waveforms of unidentifiable neurons. Identifiable neurons are simulated as independent Poisson renewal processes with a pre-set firing rate (see Methods). Every time the simulated Poisson neuron fires, its waveform is added to the noise trace. The waveforms, both for the simulated background noise and the simulated neurons, are chosen such that they closely resemble waveforms we have observed in previous experiments.

Since the mean waveform is added to the already generated noise trace, the added waveform will be corrupted by the strongly correlated background noise. As Poisson neurons fire independently, it is possible that there are overlapping spikes. Since the background noise and the neuronal firing are independent, it will be the case that some of the spikes will not be detectable and thus the number of sortable spikes could be less than the number of spikes originally inserted. In addition, for real datasets, low sample rates, compared to the frequency of spike waveforms, can cause problems in spike sorting due to misaligned peaks (the real peak was not sampled). We include this effect in our simulated data by originally simulating the data at 4 times the sampling rate (100 kHz) and then downsampling the data afterwards (to 25 kHz) before it is used for detection. This reproduces the misalignment of peak values that can be observed in real datasets. We used the approximation method for estimating the thresholds for sorting and merging. See the next section for a performance comparison of the two methods (exact and approximate) of estimating the threshold.

N **#**	**Spikes** **#**	**# Detected** [*1] **1 / 2 /3 /4**				**TP** [*2] **1 / 2 /3 /4**			
1 ***red***	475	475	475	448	**366**	459	455	414	**328**
2 ***blue***	718	718	718	701	**568**	693	694	674	**521**
3 ***green***	383	383	383	377	**306**	361	354	319	**245**
Tot	1576	1576 100%	1576 100%	1526 97%	**1240** **79%**	1513 100%	1503 100%	1407 97%	1094 89%
Thr		4	4	4	4				
N **#**	**FP** [*2,3] **1 / 2 /3 /4**					**Misses (Sorting)** **1 / 2 /3 /4**			
1 ***red***	0		0	15 (12/3)	**54** **(51/** **3)**	16	20	34	38
2 ***blue***	0		1 (0/1)	29 (7/22)	**101** **(59/** **42)**	25	24	27	47
3 ***green***	0		1 (0/1)	2 (0/2)	**8** **(2/6)**	22	29	58	61
Tot	0 0%		2 0%	46 3%	**163** **11%**	63	73	119	146
Thr									

Table 2-1. Simulation 1.

Simulation 1, consisting of 3 neurons with a peak amplitude of 1 and a firing rate of 5 Hz, 7 Hz, and 4 Hz, respectively, simulated for 100 s. The colors in column 1 refer to Figure 2-4. The 4 noise levels are as follows: 1) s.d = 0.05 and SNR = 6.7 2) s.d. = 0.10 and SNR = 3.4, 3) s.d. = 0.15 and SNR = 2.2, 4) s.d = 0.20 and SNR = 1.2. The case with the lowest SNR is marked bold because it is the situation we most commonly observe in our real data. Abbreviations: Thr: Extraction threshold, TP: True positive, FP: False Positive. [*1] Percentages for # detected are in terms of % theoretically detectable. [*2] Percentages for TP and FP are in terms of % of all spikes assigned to the sorted cluster. [*3] The numbers in parentheses represent a split up of the FP into false positives due to noise (first number) and false positives due to assignment to wrong cluster (second number).

Simulated Dataset 1: This dataset contains 3 neurons (Figure 2-4), each simulated by a renewal Poisson process with a refractory period of 3 ms and a mean firing rate of 5, 7, and 4 Hz, respectively. To provide equal SNR ratios for all waveforms, the mean waveforms of the 3 neurons were rescaled so that their peak amplitude was 1 (Figure 2-4C). A 100 s background noise trace was simulated as described (see Methods) and scaled so that it had a standard deviation of 0.05, 0.10, 0.15, or 0.20. Neuronal firing was simulated for 100 s each and the point of time at which each neuron fired was stored. For each of the 4 noise levels, the noise trace is rescaled appropriately and then the mean waveforms of the neurons are added to the trace at the timepoints the Poisson neuron fired. Using this procedure, there will be 4 traces with different noise levels that contain exactly the same noise (same signal, but different amplitude) and exactly the same neuronal firing (In Figure 2-4A,B, the noise trace with added firing for noise level 0.20 is shown).

The simulated raw data traces were processed exactly as real data is processed (bandpass filter, spike detection, spike extraction, online sorting). The different noise levels (1, 2, 3, and 4) were processed and evaluated independently (Table 2-1). They correspond to an SNR of 6.7, 3.4, 2.2, and 1.2, respectively. No parameters were modified or specified manually except the extraction threshold (row “Thr” in Table 2-1). The results of the algorithm were evaluated independently for both detection and sorting.

To illustrate how to read the detailed results in Table 2-1, we consider the results of one particular noise level (level 3, noise standard deviation = 0.15, SNR of waveforms 3.4). Theoretically, there were 475, 718, and 383 spikes, respectively, generated by the 3 neurons. Of those, 97% were correctly detected (448, 701, and 377). This implies that 3% of the generated spikes were not detectable, either because they were corrupted by noise and hence failed to cross the threshold or they were inappropriately aligned. Of the 1526 correctly detected spikes, 1407 were correctly assigned to one of the 3 clusters. 46 spikes were incorrectly assigned to one of the 3 clusters (false positives (FP)). False positives can be either true spikes which are assigned to the wrong cluster (misses) or noise waveforms inappropriately detected as spikes and then assigned to one of the clusters. Both forms of FP are shown in the table. In this case, 119 spikes were misses. The number of misses plus the number of correctly assigned (TP) equals the number of detected spikes. The number of TP plus FP equals the number of spikes assigned to a cluster. TP and FP are specified as percent (%) of total number of spikes assigned to a cluster.

This dataset demonstrates that the algorithm is capable of correctly sorting 3 distinguishable neurons with equal SNR. Even in the worst case, where the SNR equals 1.2, 79% of all spikes could be detected correctly and 89% of all spikes assigned to one of the 3 clusters were assigned correctly. Figure 2-4D illustrates the result for all 4 levels of noise and also indicates for each noise level the variance of individual waveforms. Figure 2-4A and Figure 2-4B show an extract of a raw data trace with the most difficult noise level (SNR=1.2). This is a situation we commonly observe in our real data (see Figure 2-9A).

The results of dataset 1 thus demonstrate the basic capabilities and limits of the algorithm and the parametric choices made. With the following two datasets we will address more specific elements of the algorithm: the limits of detectability (spike detection) and the limits of discriminability (spike sorting).

Simulated Dataset 2 — Limits of detectability: This second set of data addresses the limits of detectability, that is, under what conditions will the spiking of a neuron become undetectable due to background noise. To address this issue, a more realistic situation is simulated: we simulated 3 neurons with mean waveforms of different peak amplitude and thus different SNR. The 3 waveforms are illustrated in Figure 2-5A. All other conditions of the simulation were the same as in dataset 1. The average SNR of the 4 noise levels is 5.2, 2.6, 1.7 and 1.3. However, the SNRs of the individual waveforms are not equal and some will thus be harder to detect (see Table 2-2 for details). An additional difficulty presented by the 3 mean waveforms in Figure 2-5A is that they all have approximately equal peak amplitudes in the negative and positive direction. This makes this task more difficult and where a spike should be re-aligned is sometimes ambiguous.

The algorithm's performance on dataset 2 is shown in Table 2-2. Looking at the case of noise level 3, with mean waveform SNRs of 1.4, 1.4 and 2.3 (average 1.7), 56%, 56%, and 98% of the spikes of each unit could be detected, respectively. Compared to noise level 2, this presents a substantial drop in the percent detected for the first two units. Further, looking at noise level 4, where the SNR of the first 2 neurons drops to 1.1, only 21% and 15% of the spikes were detected. The limits of our spike detection and re-alignment technique are thus between an SNR of 1.1 and 1.4 for waveforms which are difficult to re-align. Detectability is limited because low SNR spikes do not cross the spike detection threshold or, if they do cross the threshold, they cannot be correctly re-aligned and are discarded (see section on re-alignment). For waveforms (e.g., unit 3 in this dataset) that possess an easily detectable peak, a substantial number of spikes can be correctly detected and re-aligned at relatively low SNR values (e.g., 70% for an SNR of

1.7). The extraction threshold (column labeled “Thr” in Table 2-2) used for the 4th noise level was 4.5, which is a conservative value compared to the value of 4.0 used in dataset 1. This value was chosen to diminish the false positive rate. The choice of the extraction threshold is always a trade-off between missed detections and false detections, but as can be seen in this simulation, a value of 4.5 seems to provide a good balance between these two opposing factors.

N **#**	**Spikes** **#**	**# Detected** *1 **1 / 2 /3 /4**				**TP** *2 **1 / 2 /3 /4**			
1 ***blue***	470	470	466 99%	**263** **56%**	101 21%	442	384	**184**	25
2 ***green***	706	706	700 99%	**395** **56%**	105 15%	644	523	**235**	45
3 ***red***	392	392	392 100%	**384** **98%**	274 70%	374	344	**343**	242
Tot	1568	1568 100%	1558 99%	**1042** **66%**	480 31%	1460 100%	1251 99%	762 82%	312 76%
Thr		3.0	3.0	4.0	4.5				

N **#**	**FP** *2,3 **1 / 2 /3 /4**				**Misses (Sorting)** **1 / 2 /3 /4**			
1 ***blue***	0	12 (11/1)	**40** **(19/21)**	4 (0/4)	28	82	**79**	76
2 ***green***	0	1 (1/0)	**11** **(8/3)**	15 (12/3)	62	177	**160**	60
3 ***red***	0	1 (0/1)	**151** **(32/119)**	115 (38/77)	18	48	**41**	32
Tot	0	14 1%	**202** **18%**	134 24%	108	307	**280**	168
Thr								

Table 2-2. Simulation 2.

Simulation 2, consisting of 3 neurons with varying amplitude with a firing rate of 5 Hz, 7 Hz, and 4 Hz, respectively, simulated for 100 s. The colors in column 1 refer to Figure 2-5. The 4 noise levels are as follows: 1) s.d. = 0.05 and SNRs of the 3 neurons 4.3, 4.3, 6.9. 2) s.d = 0.10 and SNRs 2.1, 2.1, 3.5. 3) s.d = 0.15 and SNRs 1.4, 1.4, 2.3. 4) s.d. = 0.20 and SNRs 1.1, 1.1, 1.7. The results for the third noise level correspond most closely to what we observe in our data and are marked bold. Abbreviations: Thr: Extraction threshold, TP: True positive, FP: False Positive. *1 Percentages for # detected are in terms of % theoretically detectable. *2 Percentages for TP and FP are in terms of % of all spikes assigned to the sorted cluster. *3 The numbers in parentheses represent a split up of the FP into false positives due to noise (first number) and false positives due to assignment to wrong cluster (second number).

N **#**	**Spikes** **#**	**# Detected** *1 **1 / 2 /3 /4**				**TP** *2 **1 / 2 /3 /4**			
1 *blue*	509	508	474	**191**	112	463	408	**0** **(m)**	0 (m)
2 *green*	672	671	586	**186**	100	446	318	**65**	27
3 *red*	375	329	163	**31** **8%**	26	296	110	**0** **(-)**	0 (-)
4 *l-blue*	591	591	590	**394**	225	539	532	**349**	182
5 *mag.*	839	839	839	**817**	678	787	777	**779**	611
	2986	2938 98%	2652 89%	**1619** **54%**	1141 38%	2531 100%	2145 87%	**1193** **82%**	820 58%
Thr		3	3	4	4				

N **#**	**FP** *2,3 **1 / 2 /3 /4**				**Misses (Sorting)** **1 / 2 /3 /4**			
1 *blue*	6 (0/6)	55 (13/42)	**n/a**	n/a	45	66	**191**	111
2 *green*	0	10 (3/7)	**12** **(2/10)**	61 (28/33)	225	268	**121**	73
3 *red*	1 (0/1)	28 (4/24)	**n/a**	n/a	33	53	**31**	25
4 *l-blue*	0	215 (10/205)	**97** **(14/83)**	128 (56/72)	52	58	**45**	43
5 *mag.*	0	9 (0/9)	**161** **(9/152)**	119 (40/79)	52	62	**38**	67
	7	317 13%	**270** **18%**	308 42%	407	507	**426**	319
Thr								

Table 2-3. Simulation 3.

Simulation 3, consisting of 5 neurons with varying amplitude with a firing rate of 5 Hz, 7 Hz, 4 Hz, 6 Hz, and 9 Hz, respectively, simulated for 100 s. The colors in column 1 refer to Figure 2-5B. The 4 noise levels are 1) std = 0.05, SNRs of 5 neurons 4.3, 3.8, 2.8 4.9, 7.9. 2) std=0.10, SNRs 2.1,1.9,1.4,2.4,3.9. 3) std=0.15, SNRs 1.4, 1.3, 0.9, 1.6, 2.6. 4) std=0.20, SNRs 1.1, 0.9, 0.7, 1.2, 1.9. The results for the third noise level correspond closest to what we observe in our data and are marked bold. Notice in noise level 3 that neuron #3 becomes undetectable and in level 4 neurons 1 and 2 merge, which can be seen by the high percentage of false positives in the one remaining cluster. Abbreviations: (m): merged, (-): not detected, Thr: Extraction threshold, * : only detected clusters considered, TP: True positive, FP: False Positive. *1 Percentages for # detected are in terms of % theoretically detectable. *2 Percentages for TP and FP are in terms of % of all spikes assigned to the sorted cluster.

Simulated Dataset 3 — Limits of Discriminability: This dataset combines the factors addressed by dataset 1 and 2 and adds difficulty by using 5 simulated neurons (Figure 2-5B), some of which have very similar waveforms (basically just scaled versions of each other). This will, at high noise levels, lead to merging of similar neurons because they can no longer be distinguished from one another. Additionally, all 5 neurons have similar firing rates (5, 7, 4, 6, and 9 Hz respectively). The detailed results are listed in Table 2-3. Figure 2-5C shows part of the raw data trace for all 4 noise levels.

Consider noise level 2, with an average SNR of 2.3 (individual SNRs of 2.1, 1.9, 1.4, 2.4, 3.9). Detection as well as sorting of all 5 units works reliably: 89% of all spikes were correctly detected and 87% of all sorted spikes were assigned to the correct cluster. Noise level 3 has an average SNR of 1.6 (individual SNRs of 1.4, 1.3, 0.9, 1.6, 2.6). Unit 3 becomes very hard to detect in this scenario and thus only 8% of all unit 3's spikes were correctly detected. However, due to additional difficulties presented by this waveform (red mean waveform in Figure 2-5B) in terms of re-alignment, none of them could be sorted. This is because both peaks of the mean waveform have an amplitude that is less than the noise standard deviation, and thus, due to precautions taken in the re-alignment procedure, the spikes have been discarded. Also, the false positive rate increased markedly, indicating that clusters started to merge. Units 1 and 5, for example, were partially merged, with most of the spikes of unit 1 missclassified as belonging to unit 5. Note that the two waveforms are very similar to each other (magenta and blue waveforms in Figure 2-5B). This makes it hard to discriminate between these two units at high noise levels. Figure 2-5C illustrates the difficulties of detecting units with small SNRs in high levels of noise. Shown is the same data segment (length 1 s) for all 4 levels of noise.

The merging of neurons poses a unique problem: Can we detect merging without knowing the true number of neurons (as is the case in real recordings)? To accomplish this, the projection test can be used. As illustrated in Figure 2-5D, the projection test quantifies the overlap between every pair of clusters. For each cluster, the distribution of the residuals around the mean projected onto the line between the two mean waveforms in high-dimensional space is shown. Due to transformations applied to the data to calculate this test (see Methods), the residuals distribute (if sorting is perfect) around the mean with standard deviation = 1. This knowledge can be used to estimate two important factors: i) Do spikes which were assigned to one cluster really belong to one cluster? and ii) Are two clusters separate enough so as to be considered independent? The answer to the first question can be addressed by evaluating the goodness-of-fit of a normal distribution with standard deviation = 1. We use an R^2 value to do so. The closer to 1.0 this value is, the better the fit. In case of corrupted clusters, the distribution will start to be skewed to one side and the R^2 value will be lower (for example, the combination 1 -> 4 in Figure 2-5D). The second question can be addressed by measuring the distance between two neurons (in terms of standard deviations). If two clusters are too close to each other to be accurately separated, they overlap (e.g. 1 -> 5 and 3 -> 4 in Figure 2-5D, where the distance between the means is 4.6 and 5.0 standard deviations, respectively). If both clusters that are compared are well fit by a normal distribution, a theoretical minimal distance can be calculated by setting an upper bound of overlap between the two normal distributions (e.g., distance >= 5 equals less than 1% overlap).

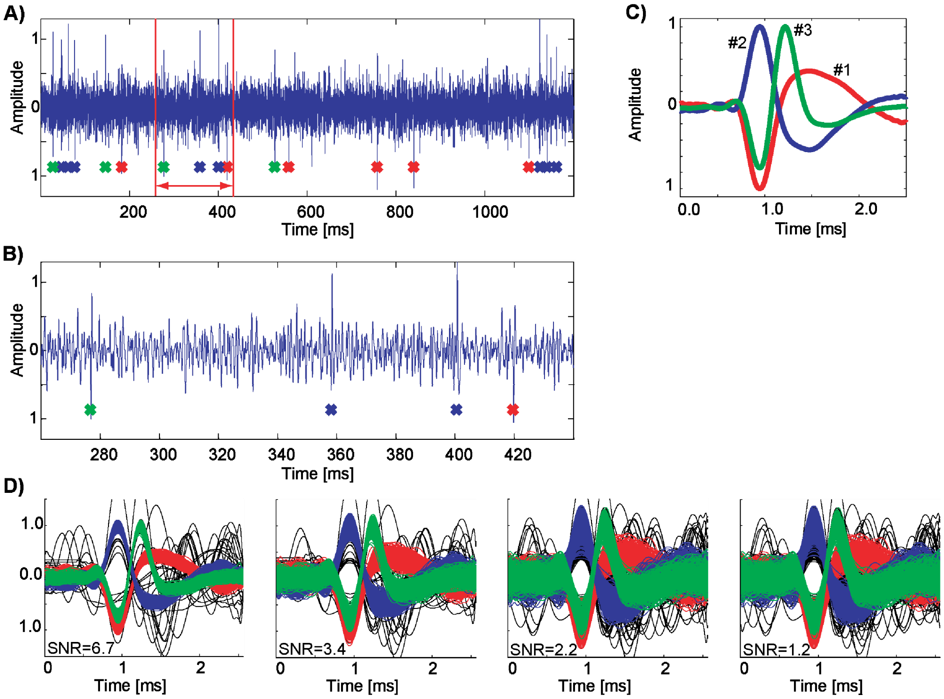


Figure 2-4. Simulated raw signal (dataset 1).
Simulated raw signal (dataset 1) from a model extracellular electrode with 3 distinguishable single-units (total length 100 s). **A and B)**: Simulated raw signal (bandpass filtered 300–3000 Hz) with a noise standard deviation of 0.20 (Level 4 in Table 2-1). Shown are 1.2 s (A) and a zoom-in of 0.3 s (B). The colored crosses indicate spikes fired by the randomly firing neurons superimposed on noise. **C)** The mean waveforms of the three single-units. The peak amplitude of each mean waveform is rescaled to 1 (of arbitrary units) to normalize the signal-to-noise ratio. The units fire with a mean frequency of 7, 5, and 4 Hz, respectively (blue, red, green). **D)** Result of detection and sorting for different noise levels (indicated by the respective signal-to-noise (SNR) ratios). The length of the simulated raw data trace was 100 s. Correctly sorted spikes are colored (compare to C) while all detected waveforms not associated with any of the 3 units are plotted in black.

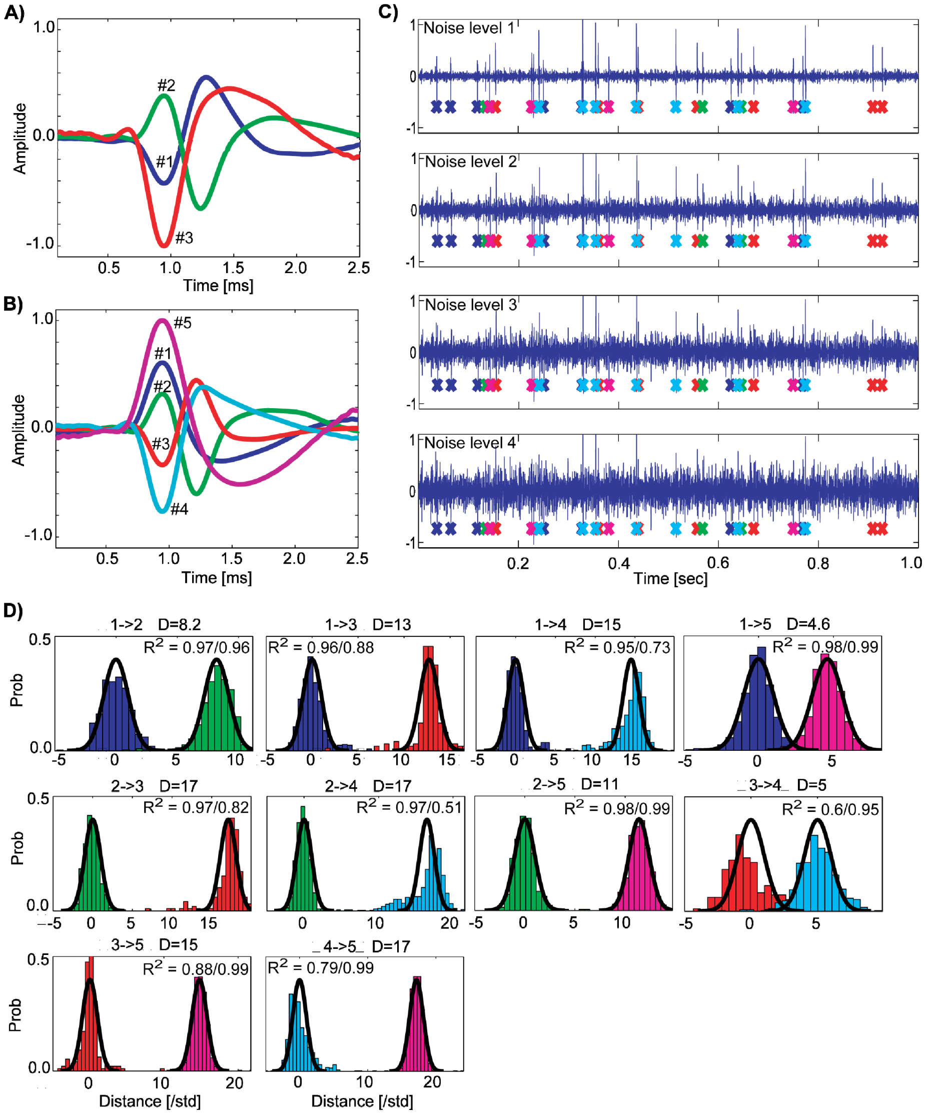
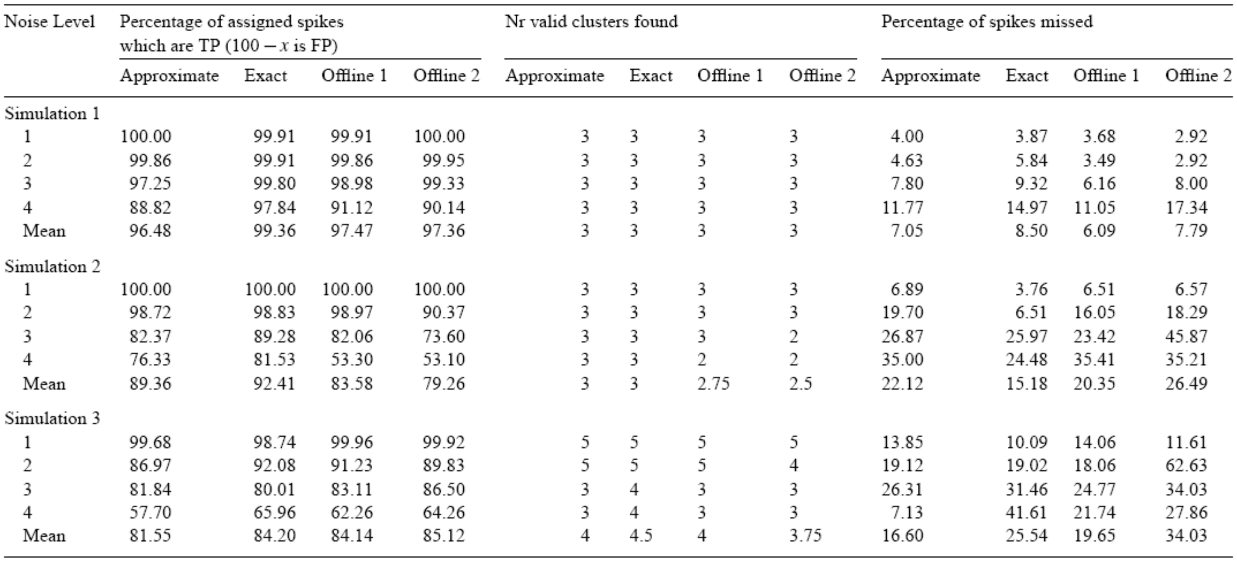


Figure 2-5. Mean waveforms used for simulated dataset 2 and 3.
Simulated data set 2 (A) and 3 (B). In contrast to dataset 1 (Figure 2-4), the peak amplitudes of each waveform are scaled randomly, with only one waveform possessing a maximal amplitude of 1. The amplitude is of arbitrary units. **C)**: Raw bandpass-filtered data segment of simulated dataset 3 for all 4 levels of noise (from top to bottom). Each segment shown contains spikes of the same 5 neurons. Notice, for example, the two spikes at the right side of the trace (red crosses), which become hard to detect in noise level 3 and 4. **D)**: Projection test for simulated dataset 3. Shown are all combinations of the 5 neurons shown in B) for noise level 2, matched with color of the histogram and the waveform, as well as by number. The histograms depict the probability density function

estimated from the residuals of all spikes associated with one cluster. Fit to each distribution is a normal density function with standard deviation = 0. The goodness-of-fit is shown using R^2 values. For each combination of neurons, the distance between the two distributions is described by how many standard deviations they are apart (D = in the title of the plots). It can clearly be seen that neurons 1 and 5 as well as 3 and 4 overlap. Also, some of the units are corrupted by noise and thus the R^2 value is low. Note that the form of the histogram for the same cluster changes as it is compared to different clusters because the residuals are projected on the line between the two clusters (see text for further discussion).

2.3.5 Comparison between exact and approximate threshold calculation methods

In the Methods section, we compare two different ways of calculating the threshold: a computationally cheap method that approximates the threshold, and a computationally more demanding method that calculates the statistically optimal threshold. In the previous section we used the approximation method to calculate the threshold. We repeated the same analysis for all 3 simulated datasets using the exact threshold calculation method. The results are illustrated in Table 2-4 and Figure 2-6. The mean improvement in true positive rates for the 3 simulations is 2.9%, 3.1%, and 2.6%. By definition, false positives are lowered by the same percentages. Also, in simulation 3 the exact threshold estimation method found 4 of the 5 existing clusters for the 2 most difficult noise levels. The exact threshold estimation method had its biggest advantage for the most difficult noise levels, where it lead to an average true-positive increase (and therefore false-positive reduction) of 7.5%. On the other hand, the performance increase for the first 2 noise levels was only minor. It is thus only advantageous to use the exact estimation method if neurons are hard to distinguish and/or background noise is high. In those cases the removal of correlations caused by the background noise results in a remarkable performance increase. The information contained in the background noise is thus useful for improving performance, as others have demonstrated before for offline sorting algorithms (Pouzat et al., 2002).



Noise Level	Percentage of assigned spikes which are TP (100 − x is FP)				Nr valid clusters found				Percentage of spikes missed			
	Approximate	Exact	Offline 1	Offline 2	Approximate	Exact	Offline 1	Offline 2	Approximate	Exact	Offline 1	Offline 2
Simulation 1												
1	100.00	99.91	99.91	100.00	3	3	3	3	4.00	3.87	3.68	2.92
2	99.86	99.91	99.86	99.95	3	3	3	3	4.63	5.84	3.49	2.92
3	97.25	99.80	98.98	99.33	3	3	3	3	7.80	9.32	6.16	8.00
4	88.82	97.84	91.12	90.14	3	3	3	3	11.77	14.97	11.05	17.34
Mean	96.48	99.36	97.47	97.36	3	3	3	3	7.05	8.50	6.09	7.79
Simulation 2												
1	100.00	100.00	100.00	100.00	3	3	3	3	6.89	3.76	6.51	6.57
2	98.72	98.83	98.97	90.37	3	3	3	3	19.70	6.51	16.05	18.29
3	82.37	89.28	82.06	73.60	3	3	3	2	26.87	25.97	23.42	45.87
4	76.33	81.53	53.30	53.10	3	3	2	2	35.00	24.48	35.41	35.21
Mean	89.36	92.41	83.58	79.26	3	3	2.75	2.5	22.12	15.18	20.35	26.49
Simulation 3												
1	99.68	98.74	99.96	99.92	5	5	5	5	13.85	10.09	14.06	11.61
2	86.97	92.08	91.23	89.83	5	5	5	4	19.12	19.02	18.06	62.63
3	81.84	80.01	83.11	86.50	3	4	3	3	26.31	31.46	24.77	34.03
4	57.70	65.96	62.26	64.26	3	4	3	3	7.13	41.61	21.74	27.86
Mean	81.55	84.20	84.14	85.12	4	4.5	4	3.75	16.60	25.54	19.65	34.03

Table 2-4. Comparison of sorting results for the two different threshold estimation methods, as well as other algorithms.

Percentages of true positives (TP) are specified in terms of percent of all spikes assigned to the cluster. False positives (FP) are thus by definition 100-TP. The column “nr valid clusters found” specifies how many of the original clusters were found. The right column “percentage of spikes missed” specifies what percentage of all correctly detected spikes (spikes which are known to belong to one of the simulated neurons, excluding noise detections) were not assigned to the correct cluster. This number includes both spikes assigned to background noise and those assigned to the wrong cluster.

2.3.6 Comparison with offline sorting algorithms

We used the same simulated datasets as described in the previous section to evaluate how the performance of our algorithm compares to other algorithms. We used two commonly used algorithms. Both algorithms are offline sorting algorithms, that is, they require all data to be available before sorting starts. The first algorithm (referred to as Offline 1) that we compared against is the well known *KlustaKwik* clustering algorithm (Harris et al., 2000). We used the first 10 principal components, computed using PCA (Jolliffe, 2002), as features. The minimum number of clusters was set to 3 and the maximum number clusters to 30. Otherwise, all parameters were set to the default values. All parameters were the same for all simulations and noise levels. The second algorithm we compared against is the *WaveClus* algorithm developed by (Quiroga et al., 2004), referred to as Offline 2. This algorithm is particularly relevant for our comparison because it has been used to sort data similar to ours. Since this algorithm selects its own features (wavelets) directly from the data, we used the waveforms as input features. For both algorithms, we used the publicly available version of the code written by the authors. To exclude influences on sorting performance of different detection methods, we used our detection method to detect spikes. Spikes were upsampled and re-aligned before processing. Both algorithms thus had the exact same input data. The clusters generated by the two algorithms were manually matched to the clusters which originally generated the data. Clusters which do not exist in the original data (overclustering, noise) were assigned to noise.

The results of the comparison are summarized in Figure 2-6 and Table 2-4. The performance of a given algorithm can not be reduced to a single number because, depending on the experimental situation, different criteria of performance are most crucial for the experimenter. To allow a fair comparison, we calculated 4 performance measurements: true positives (TP), false positives (FP), number clusters found, and misses. We calculated the TP/FP in terms of the percentage of all spikes assigned to a given cluster that actually belong to this cluster (true positives, TP). The false positives (FP) are thus by definition the difference between the TP and 100%. Misses are in percent of all detected spikes which were missassigned. This includes spikes which were assigned to background noise. Overall, we find that all algorithms perform remarkably similar on all datasets. This is particularly true for the first two noise levels (Figure 2-6A–C, levels 1 and 2). Performance differences are larger for the more difficult noise levels 3 and 4. While all algorithms show a drop in performance for these two levels, the two offline algorithms identify fewer clusters than our online algorithm. This is because in the high noise situations, some of the clusters become very small and partially overlap with other clusters. The differences between these clusters cannot be resolved if correlations introduced by the background noise are not taken into account. This explains why in the case of noise level 4 in simulation 2 (Figure 2-6B, red line) the online algorithm using the exact threshold clearly has the best performance of all algorithms compared. Generally we observed that the offline algorithms appear to artificially merge clusters earlier than our algorithm. This causes an increase in the number of false positives, which then decreases the number of true positives. This does not imply that fever spikes were correctly assigned but is a consequence of our definition of true positives, which we believe is the most relevant for experimental purposes. We also observed that the offline sorting algorithms generally tend to overcluster — that is, they generate ficticious clusters. As these artificial clusters also tend to be small, they typically do not violate the refractory period condition of no ISIs < 3 ms. One possibility to avoid this problem is to use the projection test as a post-hoc test after sorting with one of the offline sorting algorithms.

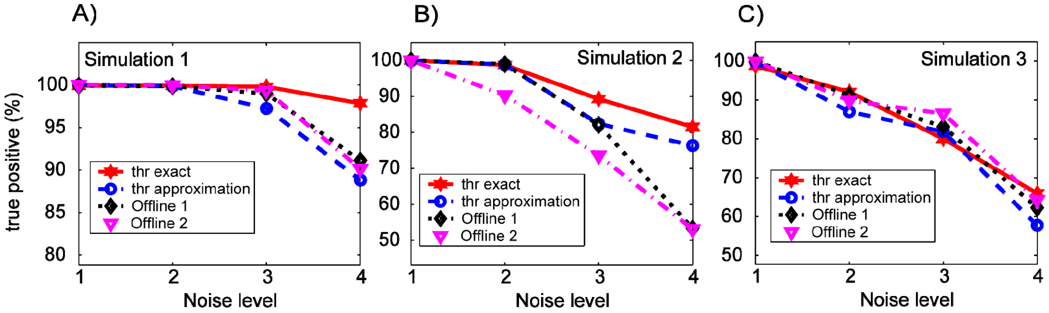


Figure 2-6. Performance comparison of different spike sorting algorithms. We compared the performance of our algorithm to two other offline sorting algorithms (Offline 1 is the Klustakwik Algorithm, and Offline 2 the WaveClus Algorithm, see text) , examining the true positives (% of spikes assigned to a given cluster that actually belong to the cluster). For our algorithm we used the two different threshold estimation methods (thr exact and thr approximation). Please see Table 2-4 for details. The false positive rate is by definition 100-TP.

2.3.7 Evaluation of sorting — real data

We chose 2 datasets from 2 different recording sessions to demonstrate the application of the algorithm to real datasets. In both sessions, we recorded from the right and left hippocampus (RH, LH) and from either the right or left amygdala (RA, LA). These two recording sessions were chosen because the first one represents an example with a high number of neurons per channel (on average, 3.7 ± 1.7 neurons per active channel, range 1–7) and the second a more typical case of fewer, but hard to distinguish, neurons (On average 2.0 ± 0.8 neurons per active channel, range 1–3). Using these two examples demonstrates that the algorithm works reliably in both cases.

Using our algorithm as described, with all parameters automatically estimated from the data and the extraction threshold set to 5 (see simulations for how to find this value), we found a total of 76 well-separated single neurons that pass all statistical tests and visual inspection.

Figure 2-7 shows the result and the statistical criteria used for one particular channel (a single wire, implanted in the RA). A total of 9096 raw waveforms were detected, 7237 (80%) of which were assigned to one of the 5 well-separated single units (1682, 3669, 210, 142, and 1534 for each cluster, respectively). In Figure 2-7A (from left to right), an overlay of all raw waveforms, the mean waveforms, and the decorrelated raw waveforms and means are shown. Each neuron is color-matched across the whole figure (1=cyan, 2=yellow, 3=green, 4=red, 5=blue). For the first two neurons detected, the raw waveforms, the interspike interval histogram (ISI), the powerspectrum of the ISI and the autocorrelation of the ISI are shown in Figure 2-7B and C (from left to right). The pertinent features for evaluation that are used are as follows: the fraction of ISIs shorter than 3 ms (specified in % of all ISIs), the absence of peaks in the power spectrum and an approximately zero autocorrelation for small (< 3 ms) timelags. We find that only the combination of all 3 criteria allow a sufficient classification of clusters as single unit or not. We, for example, often observe clusters which have a perfect ISI (no < 3 ms) but with large peaks in the powerspectrum caused by noise (e.g., 60 Hz and harmonics). Such clusters have to be discarded. Other indications of potential problems are an autocorrelation which does not return to 0 at long (> 100 ms) timelags.

Applying the above criteria allows us to identify all well-defined clusters that might represent single units, but it is not sufficient For example, special concern is warranted if two mean waveforms appear to be linearly scaled versions of each other, without any other distinguishing features (e.g., neuron 1 and 2 in Figure 2-7). In contrast, some neurons (e.g., neuron 4 and 5 in Figure 2-7) are very similar on some, but importantly not all, indices. Two waveforms that are linearly scaled versions of each other could be the result of spike height attenuation during a burst or electrode movement. The artificial splitting of a single unit into multiple clusters as well as erroneous merging of two single units into one cluster can be detected using the projection test. There are two indicators of the projection test that can be used to assess splitting and merging: the distance between the two means of the clusters and the goodness-of-fit of the empirical to the theoretical distribution. If the distance between the two means is not sufficiently large (e.g., > 5 for less than 1% overlap) and/or the goodness-of-fit to the distribution is bad, one or both of the clusters has to be discarded. Figure 2-7D illustrates this method for the 4 pairs of neurons in which overlap might be suspected. As the left panel in Figure 2-7D shows, the distance between neuron 1 and 2 is sufficiently large (6.6) and the fit to the distributions is very good. In contrast, the fit of neuron 4 (3rd panel, red) is less good but still sufficient. Also, a few outliers can be identified which represent missalignments (far right of red distribution). Another reason for poorly separated single units is the merging of two clusters representing unique units. This can also be detected by the projection test. In this case, the distribution of spikes around the mean will be too broad (long, fat tails), which is an indication for merged clusters. Such clusters represent multi-unit activity and can be used as such in the further analysis. It is also helpful to look at a post-hoc PCA plot of the first two principal components (Figure 2-8). The principal components are computed from the raw, not pre-whitened, waveforms. The color is assigned by the clustering algorithm. In this plot it is also evident that cluster 1 and 2 are indeed separate. From the PCA plot it is less clear whether clusters 4 and 5 are indeed separate. Consultation of the projection test (Figure 2-7D) confirms that the clusters are separate but also indicates that there is some degree of overlap, as can also be seen in the PCA plot.

For comparison, we repeated the sorting of the same detected waveforms as shown in Figure 2-7 with the *WaveClus* offline sorting algorithm (see offline algorithm section for details). The algorithm identified a very similar number of spikes for each cluster (same order as above: 1529, 3452, 197, 113, and 1513). No other clusters were found except for the noise cluster. In total it assigned 75% of the total 9096 detected waveforms to one of the 5 clusters.

Population data for all 76 sorted neurons is shown in Figure 2-9. The average SNR of all mean waveforms, calculated by using the noise standard deviation for each channel, was 2.12 ± 0.85 (Figure 2-9A). This measurement defines the SNR typically observed in experiments and thus serves as a guideline for the estimation and verification of parameters using the simulated data. A good general indicator of separation quality is the percent of ISIs which are shorter than 3 ms (on average 0.21 ± 0.27%, Figure 2-9B). For all channels on which there was more than one neuron we calculated the distance between all pairs of neurons on each channel. The average distance was 12 ± 5 (Figure 2-9C).

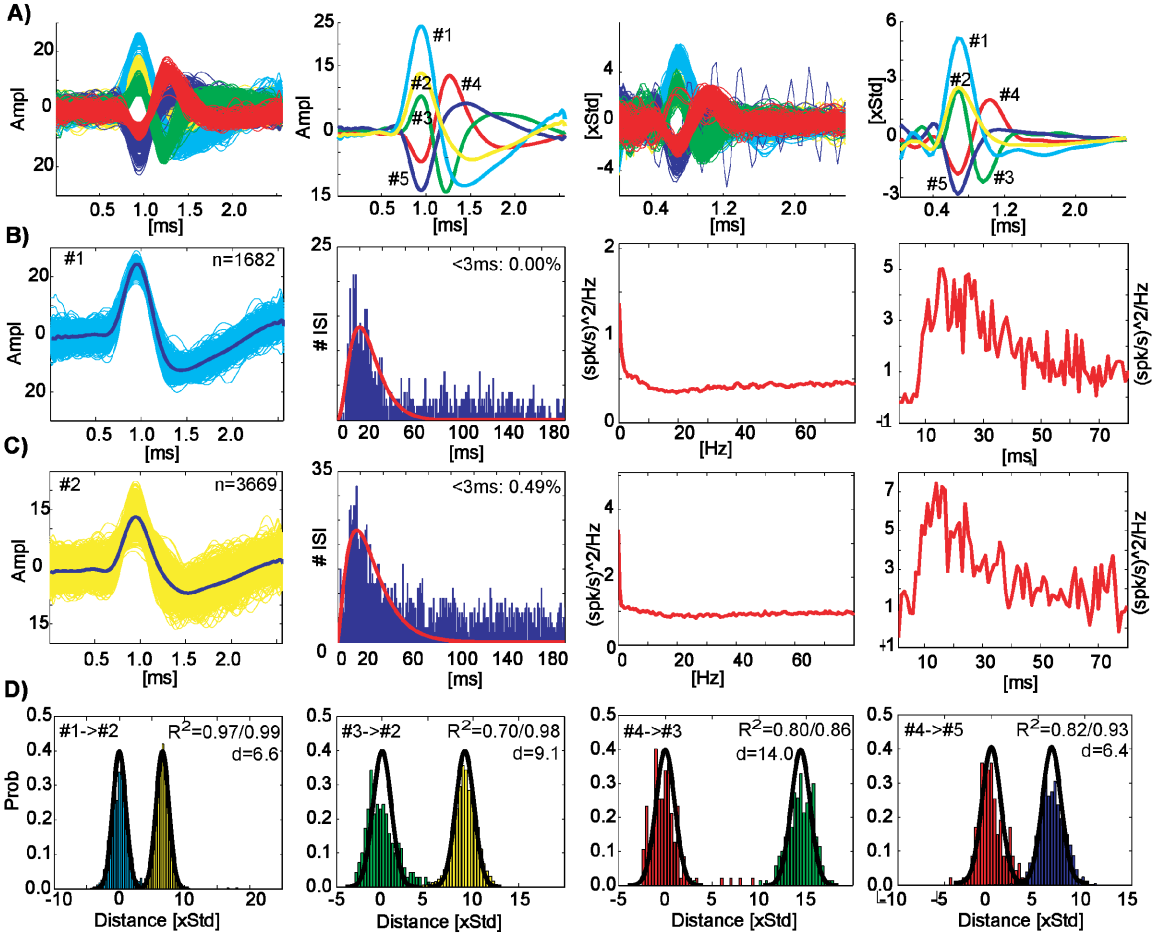


Figure 2-7. Illustration of tools for evaluation of the sorting result, using real data. All data shown is from the same channel, which was recorded from the right amygdala. Five well-separated neurons could be sorted, with 1682, 3669, 210, 142, and 1534 spikes, respectively (neurons are numbered 1–5 in this order). All subfigures are color matched. **A)** From left to right, all raw waveforms, mean waveforms, decorrelated raw waveforms, and mean decorrelated raw waveforms (see text for discussion of decorrelation). **B and C)** Details for two of the neurons (#1 and #2, cyan and yellow). From left to right: raw waveforms, ISI histogram, powerspectrum of the ISI and autocorrelation of the ISI. Note that the gamma distribution fitted to the ISI is for illustration purposes only and is not used for evaluation. **D)** Projection test for the 4 combinations of mean waveforms which are "closest" and could possibly overlap/be not well separated. For example, take mean waveforms #1 and #2. They appear to be scaled versions of each other, and clear separation is thus difficult to achieve. It might thus be suspected that they overlap. Consulting the projection test probability density functions shown in the first panel of D), however, allows us to conclude with confidence that these two sets of spikes are well

separated and thus likely represent two unique neurons. The distance (6.6) is big enough and the fit to the theoretical distribution is reasonable.

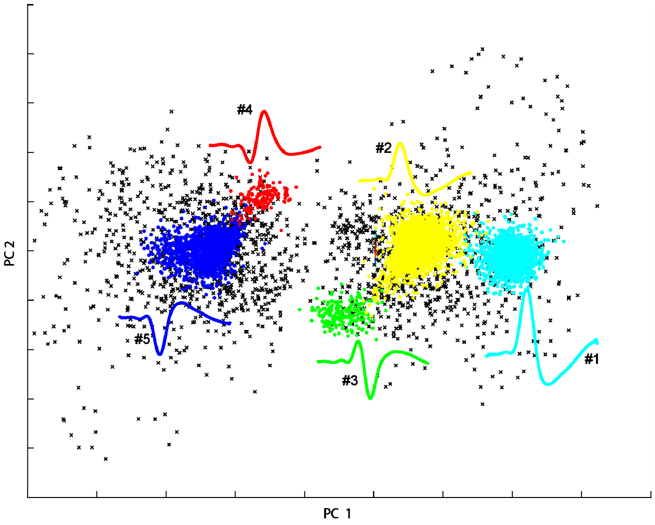


Figure 2-8. Illustration of PCA analysis for one channel of real data.
PCA analysis for one channel of real data together with data obtained using our algorithm to sort. Shown is the projection of the first 2 principal components for all waveforms detected on the channel. The colors refer to the same 5 neurons as identified in Figure 2-7. Black points are detected waveforms which are not assigned to any of the 5 clusters (noise or unsortable). The numbers refer to Figure 2-7A. This data represents approximately 45 minutes of continuous recording.

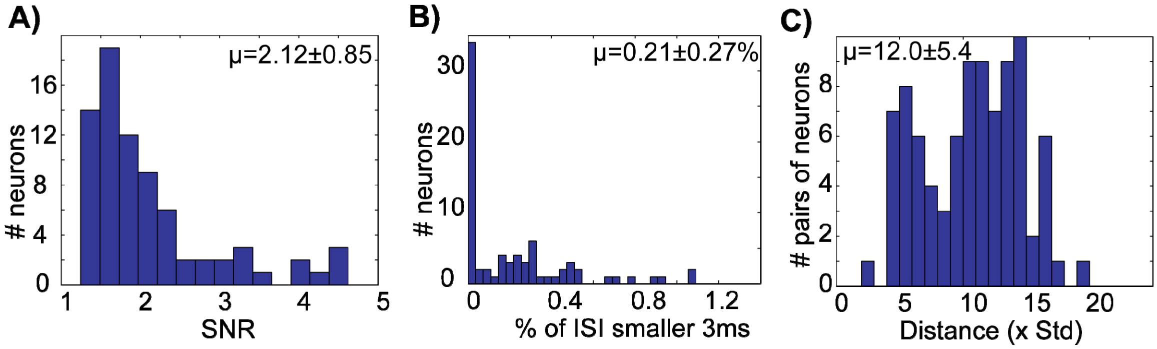


Figure 2-9. Population statistics from the 76 neurons obtained from in vivo recordings.

A) Histogram of the SNR of all 76 neurons. The SNR is calculated from the mean waveform. The mean SNR was 2.12 ± 0.85 (±s.d.). **B)** Histogram of the percent of all interspike intervals (ISI) which are shorter than 3 ms. The threshold for accepting a neuron is 3%. The mean of all 76 neurons was 0.21 ± 0.27% (± s.d.). **C)** Histogram of the distance between pair of neurons, calculated using the projection test. This test can only be calculated for channels which have at least one neuron. The mean distance was 12.0 ± 5.4 (± s.d.). The distance is expressed as the number of standard deviations of the distribution of waveforms around the mean waveform, which is 1 (by design) for each neuron.

2.3.8 Bursts

The calculation of the threshold for sorting (minimum distance between clusters required) thus far only takes into account variance due to extracellular sources. However, the waveforms of a single neuron also vary due to intracellular reasons, mainly due to spikes which follow each other with an interspike interval of less than 100 ms (Fee et al., 1996b; Harris et al., 2000; Quirk and Wilson, 1999). This additional variance needs to be accounted for. As such, it is necessary to assume a slightly higher threshold than is estimated from the background noise. If it is known that the data which is sorted does not contain bursts, this correction does not need to be applied. A rough estimate whether there are bursts or not can be made by looking at a plot of the first two principal components of all detected raw waveforms. If there are distinct elongated clusters, bursting neurons are probably present and a correction needs to be applied.

The extracellular waveform during short ISIs is changed in a characteristic way. Most features of the spike remain the same, but the amplitude changes. That is, the waveform is linearly scaled. This will mainly affect the peak region of the spike. In our case, the peak region occupies approximately 0.5 ms. The overshoot region will also be scaled, but the increase in variance due to this is minor because of its smaller amplitude relative to the spike peak. Peak spike amplitudes can be attenuated by up to 40% (Quirk and Wilson, 1999). To account for this, the variance used to calculate the threshold has to be increased by 40% for the 0.5 ms region of the peak region. See Equations 4b and 4c in Appendix A for the calculation, which results in a correction factor for the threshold of approximately 1.2. The fact that short ISIs cause scaling of the extracellular waveform also has important implications for the evaluation of the sorting results. Cases where two seemingly well-separated clusters have mean waveforms which appear to be linearly scaled versions of each other can be further evaluated manually.

2.3.9 Non-stationarities of noise levels

Depending on the environment, the levels of background noise can change over time. Whereas this problem is manageable for recordings done in a controlled research environment, it is not possible to control external noise levels in clinical or other uncontrolled (e.g., behavioral studies) environments. The ability to dynamically adapt to non-stationary noise levels is thus crucial. We adapt to changing noise levels on two timescales: for fast, high-powered bursts of noise, we immediately stop extracting waveforms until the burst is over (usually far less than 200 ms). To slowly changing levels of noise we adapt by calculating the threshold (which is calculated from the standard deviation of p(x), see methods) for spike extraction as a running average over a long time window (e.g., 1 minute).

2.3.10 Computation cost

Our implementation (details in the methods) serves as a proof of principle and is not optimized for speed. We nevertheless report approximate running times for the different stages of the algorithm to enable a comparison against other algorithms, but it should be noted that careful optimization and more efficient implementation in a compilable programming language such as C++ will provide substantial improvements over the numbers reported here. We measured the running times while sorting a session consisting of 21 active channels, each recorded in parallel over a duration of 35 min. Raw data was read from data files from the harddisk (one file per channel) A total of 143947 spikes were detected (average 6854 ± 5234 spikes per channel). Detection took on average 194 ± 13 sec per channel. This includes detection, extraction of pure noise sweeps, calculation of the noise autocorrelation, and pre-whitening of each spike detected. Per channel, approximately 100000 noise traces (40 per second) were extracted. Sorting took on average 18.24 ± 13.9 sec per channel. Considering the number of spikes on each channel, this results in a sorting speed of 376 spikes/s. In total, this allows processing of a single channel at approximately 10 times the duration of data acquisition (on average 3.5 minutes for each channel). Optimizing this implementation will allow the processing of many hundreds of channels in realtime.

2.3.11 Future improvements

There are multiple ways in which the procedure presented here could be improved. One issue that is currently not addressed in our implementation[2] is overlapping spikes, which are caused by two nearby neurons firing in synchrony or by neurons firing closely together by chance. If two close-by neurons are synchronized such that they always fire together in a systematic and consistent way, the overlapping spike becomes detectable because a distinct cluster will be created. However, in the more common situation where spikes overlap in widely different situations, such spikes would be disregarded and classified as noise. It is imaginable to also test for linear combinations of mean waveforms to allow classification of such combined spike events. Indeed such an approach has been proposed (Atiya, 1992; Takahashi et al., 2003).

The proposed algorithm has so far only been applied to the sorting of data from single wire electrodes but it would be straight forward to extend its usage also to tetrode data (Harris et al., 2000). Instead of one mean waveform per identified source there would be four mean waveforms. This would further enhance performance and reliability while still using the same principle.

The re-alignment procedure we have described allows the accurate re-alignment of many difficult cases, but sometimes it still fails. Accurate re-alignment is necessary because our distance measurement for comparing two spikes requires that the two spikes are accurately re-

[2] Our implementation (Matlab), as well as the real+simulated data we used for testing is available at http://emslab.caltech.edu/software/spikesorter.html or http://www.urut.ch/ . Our opensource implementation “Osort” can also process other file formats (such as Medtronic or Neuralynx Digital Cheetah) not discussed here. It also implements more advanced spike detection methods not discussed here. It includes a graphical user interface (GUI) for ease of use.

aligned (at the same position). If this is not the case, the procedure fails. There are two possible improvements that could be made to remedy this situation. One would be to enhance the distance measurement so that it does not rely on realignment (e.g., re-positioning the two waveforms on a case-by-case basis for each distance measurement or using a translation-invariant distance measurement). The second improvement could utilize a combined spatial and frequency space measurement, as has been proposed (Rinberg et al., 2003).

Our algorithm assigns each spike to one cluster only. This decision is taken at the point of time the spike is detected ("hard clustering"). An alternative approach would be to assign each spike a probability to which cluster it belongs and update this probability as the model (mean waveforms) change over time ("soft clustering"). While we have not taken this approach, it is imaginable that it could be implemented in the framework we present here. Because we build and update our model iteratively over time, it is indeed possible that the model converges to the wrong solution. This is rather unlikely, though, because if a cluster slowly converges towards another cluster, the two cluster centers eventually get too close and they are merged. However, merges are never reversed. If two clusters are very close by and are merged erroneously, this situation will never be resolved. Soft clustering could possibly deal with this situation.

2.4 Conclusions and relevance

Here, we propose a general online sorting algorithm and demonstrate and evaluate its sorting ability by applying it to a challenging dataset recorded in a clinical environment. There are a wide variety of applications made possible by online sorting which we are only starting to explore. The experimental approach taken in most animal single-unit recordings involves first the design of an experiment and then the search for neurons that respond appropriately to the experimental task. Obviously, this type of experimental design requires that electrodes can be moved freely by the experimenter; this is not possible in human studies. Of the many limitations posed by a clinical environment, the most constraining one is that chronically implanted electrodes are at a fixed position that can not be moved (Fried et al., 1999). Thus, only the neurons that can be recorded in the vicinity of the electrode can be analyzed. While it is still possible to design a static experiment and observe a neuronal response, it is the case that most neurons will not react in any systematic way to the stimuli presented. As one does not have access to the response properties of neurons during the experiment, these (non-stimulus-related) spike events are recorded and then during offline analysis discovered to be essentially useless. Electrodes in epilepsy surgery patients are implanted in higher-level brain structures such as the medial temporal lobe (MTL), including the hippocampus or the amygdala, and prefrontal cortex. Unlike the response properties of neurons in the primary sensory cortices, MTL neuron responses are multi-sensory and complex (Brown and Aggleton, 2001), and hence possess less-predictable response properties.

Thus, to make the most of the information obtainable with chronic implants in humans the traditional approach has to be reversed: the experiment needs to adapt itself to the neuronal response observed. Creating an adaptive experiment poses significant technological challenges which need to be addressed. The work presented in this paper is one of the main required techniques to be able to conduct adaptive experiments. Online sorting for the first time allows the experimenter to conduct real "closed-loop" experiments in awake behaving animals, similar to what is already possible with dynamic-clamp in single-cell experiments (Prinz et al., 2004). Such experiments will be designed to immediately react to the neuronal response observed to a certain stimulus.

Additionally, online sorting is tremendously useful for conducting extracellular recordings in a noisy environment, like the hospital room. It is very hard and often impossible to judge manually (by visual inspection) whether the signals visible in the raw data trace are of sortable neurons or not. This can make the decision on which amplifier settings to use and from which electrodes to record arbitrary and often wrong. We, for example, often face the situation that there are more electrodes implanted than we can record from simultaneously. As such, we have to make an on-the-spot decision about which subset of electrodes to record from. Using offline data analysis, it sometimes becomes clear that the best available electrode was not chosen because it was not possible to identify the spikes by visual inspection alone. On the other hand, channels which look active and interesting often turn out to be corrupted by noise, so that they can't be used. Online spike sorting, implemented in realtime, will enable the experimenter to make the best-informed choices about which electrodes to include during an experiment.

Another possible area of application is brain-machine interfaces. It has been demonstrated that it is possible to decode intended movements using chronically implanted electrodes in non-human primates using single-cell spike data from motor cortex (reviewed in (Mussa-Ivaldi and Miller, 2003)) and higher cortical areas, e.g. (Musallam et al., 2004). Combined with the recent development of microdrive-driven chronically implanted arrays of electrodes this will ultimately allow online control of cortically controlled neural prosthetics (Schwartz, 2004). The algorithms for decoding intentions of movements (Chapin, 2004) depend on the ability to simultaneously record the activity of many single neurons over a long time and it is thus crucial that spikes can be detected and sorted reliably in realtime. This presents a particular challenge in the uncontrolled and noisy environments in which such devices will have to function. Moving from the well-controlled laboratory environment to a noisy real-world environment will increase the difficulty of spike detection and sorting tremendously. Our algorithm could be of use for such applications.

2.5 Appendix A — Signal processing and spike detection

2.5.1 Spike detection

The local energy, or power, $p(t)$ (Eq 1) of the signal is the running square root of the average power of the signal $f(t)$ using a window size of 1 ms (n = 20 samples at 25 kHz sampling), the approximate duration of a spike (Bankman et al., 1993). $\bar{f}(t)$ is the running average, going back n samples in time. $p(t)$ can be efficiently calculated for a signal of arbitrary length using a convolution kernel or a running window in online decoding.

$$P(t) = \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(f(t-i) - \overline{f}(t) \right)^2 \right\}^{\frac{1}{2}} \qquad (1)$$

$$\overline{f}(t) = \frac{1}{n} \sum_{i=1}^{n} f(t-i) \qquad (2)$$

2.5.2 Distance between waveforms

The distance between two spikes $\vec{S}_i$ and $\vec{S}_j$ is calculated as a residual-sum-of-squares (Eq. 3a) for the approximated threshold method. For the exact threshold estimation method, the same equation applies because the covariance matrix Σ in Eq 3b is equal to I for pre-whitened waveforms (by definition). Note that this distance is generally used to calculate the distance between a spike and a mean waveform of a neuron, and not between two spikes.

$$d_S(\vec{S}_i, \vec{S}_j) = \sum_{k=1}^{N} \left(S_i(k) - S_j(k) \right)^2 \tag{3a}$$

$$d_S(\vec{P}_i, \vec{P}_j) = (\vec{P}_i - \vec{P}_j)\Sigma^{-1}(\vec{P}_i - \vec{P}_j)^T \tag{3b}$$

Calculating the distance between the means of two clusters is achieved differently for the two methods of estimating the threshold: i) for the approximated threshold, $d_M = d_S$, and ii) for the exact threshold, $d_M = \sqrt{d_S}$ (equal to Eq 11 in the projection test).

2.5.3 Calculation of the threshold

There are two thresholds which need to be calculated: T_S (sorting) and T_M (merging). In the case of the approximated threshold method, $T_S = T_M = T$, whereas T is calculated as shown in Eq 4a. $\langle \sigma_r \rangle$ is the average standard deviation of the filtered signal f(x), calculated continuously with a long (e.g., 1 minute) sliding window. For efficiency reasons, the distance calculated in Eq 3a is not divided by N to normalize for the number of datapoints, but rather the threshold is multiplied by N in Eq 4a. This is mathematically equivalent, but Eq 3 can be calculated more efficiently in matrix notation in this form.

$$T = N\langle\sigma_r\rangle^2 \qquad \text{(Eq 4a)}$$

In the case of the exact threshold estimation method, the two thresholds are calculated differently: Since d_S is χ^2 distributed (Johnson and Wichern, 2002), the distance that includes all points belonging to the cluster with probability $1-\alpha$ can be calculated from the χ^2 distribution (Eq 5). The threshold d_M for merging is simply the number of standard deviations clusters need to be apart to be considered separate, which we assumed to be 3. α is typically set to 0.05 or 0.10 (5%, 10%) and p is the number of degrees of freedom (see text).

$$P[(\vec{P}_i - \vec{P}_j)\Sigma^{-1}(\vec{P}_i - \vec{P}_j)^T \le {\chi_p}^2(\alpha)] = 1-\alpha \qquad (5)$$

2.5.4 Correction factor for bursts

The distance as calculated by Eq 4 does not take into account systematic variability of the waveform for reasons other than extracellular noise. To account for systematic waveform changes, particularly in spike amplitude, a correction factor is applied to increase T appropriately (Eq 4b).

$$T_C = cT \qquad (4b)$$

The correction factor c is calculated as following (here, N is assumed to be 256 datapoints): A burst is going to scale the peak region of the spike, which occupies approximately $B = 50$ datapoints (0.5 ms). Correcting T for B datapoints using a higher variance and leaving the other $N - B$ with the baseline variance is calculated using Eq 4c.

$$T_C = T(\frac{N - B}{N} + \frac{b_c B}{N}) \qquad (4c)$$

The correction factor b_c specifies how much the variance is assumed to increase due to this. A conservative estimate is $b_c = 2$. Using above numbers, this results in a correction factor of 1.2, as is used throughout this paper. This correction factor is only applied if the threshold is calculated using the approximation method.

2.5.5 Signal-to-noise ratio

The signal-to-noise ratio is calculated as the root-mean-square (rms) of a spike divided by the standard deviation (Bankman et al., 1993) of the raw data trace (Eq 6).

$$SNR = \frac{\left\| \vec{S}_i \right\|}{\sqrt{N\sigma^2}} . \qquad (6)$$

2.6 Appendix B — Online sorting

For each detected spike $\vec{S}_i$ the distance of $\vec{S}_i$ to all mean waveforms is calculated. Using algorithm 1, a spike is associated to cluster j if it meets the following criteria: i)

$d(\vec{S}_i, \vec{M}_j)$ is minimal compared to all other mean waveforms, and ii) $\min(d(\vec{S}_i, \vec{M}_j)) < T$.

If these conditions are met, Algorithm 2 is used to assign $\vec{S}_i$ to the existing cluster that meets the conditions. Also, the mean waveform of the cluster is updated using the last C spikes that were associated to this cluster. This change could potentially create overlapping clusters (and will do so, especially when not many spikes have been processed), which are automatically merged by Algorithm 2 (see below).

2.6.1 Algorithm 1

Task: Assign newly detected spike $\vec{S}_i$ to cluster or create new cluster if necessary.

```
1: $d_j = d_S(\vec{M}_j, \vec{S}_i)$ for $j = 1...m$        {distance to all known clusters}
2: **if** $\min(d_1, d_2, ..., d_m) \le T_S$ **then**
3:         assignSpike( $\vec{S}_i$ ) {call Algorithm 2}
4: **else**
5:         $m \Leftarrow m + 1$
6:         $\vec{M}_m \Leftarrow \vec{S}_i$
7: **end if**
```

2.6.2 Algorithm 2

Task: Assign spike $\vec{S}_i$ to cluster and merge clusters if necessary.

```
1: $j \Leftarrow \arg\min(d_1, d_2, ..., d_m)$
2: assign $\vec{S}_i$ to cluster $j$
3: $\vec{M}_j \Leftarrow \langle \vec{S}_k \rangle$, for $k = |\vec{M}_j| - C ... |\vec{M}_j|$        {update mean waveform as average of last C
assigned spikes}
$\vec{D} = d_M(\vec{M}_j, \vec{M}_i)$, for $i = 1 \ldots j-1, j+1 \ldots m$        {distance of update mean waveform to all other mean waveforms}
while $\min(\vec{D}) < T_M$
    $k \Leftarrow \arg\min(\vec{D})$
    merge cluster $j$ with cluster $k$
    remove cluster $k$
    reassign all $\vec{S}_i$ assigned to cluster k to cluster $j$
    $\vec{D} = d_M(\vec{M}_j, \vec{M}_j)$, for $j = 1 \ldots m$        {distance between all mean waveforms}
end while
```

2.7 Appendix C — Spike realignment

2.7.1 Algorithm 3

Task: Decide where the peak of $\vec{S}_i$ is that is to be used for realignment.

```
sigLevel $\Leftarrow 2 * \langle \sigma_r \rangle$        {twice the std of the raw signal, see Eq4}
if $abs(\min(\vec{S}_i)) >= sigLevel$ and $abs(\max(\vec{S}_i)) >= sigLevel$ then
    {Align according to temporal order of peaks}
    if $find(\vec{S}_i == \max(\vec{S}_i)) < find(\vec{S}_i == \min(\vec{S}_i))$ then
        peakInd = $find(\vec{S}_i == \max(\vec{S}_i))$        {realign at positive peak}
    else
        peakInd = $find(\vec{S}_i == \min(\vec{S}_i))$        {realign at negative peak}
    end if
else
```

```
    if $(abs(\min(\vec{S}_i)) >= sigLevel$ and $abs(\max(\vec{S}_i)) < sigLevel)$ or
    $(abs(\min(\vec{S}_i)) < sigLevel$ and $abs(\max(\vec{S}_i)) >= sigLevel)$ then
        {only one peak is significant, realign at it}
        if $abs(\max(\vec{S}_i)) > abs(\min(\vec{S}_i)$ then
            peakInd = $find(\vec{S}_i == \max(\vec{S}_i))$
        else
            peakInd = $find(\vec{S}_i == \min(\vec{S}_i))$
        end if
    else
        {This spike can't be re-aligned, discard}
    end if
end if
```

2.8 Appendix D — Projection test

2.8.1 Pre-whitening of waveforms

The raw waveform, consisting of N datapoints, is corrupted by strongly correlated noise. To de-correlate the noise, that is, to make each datapoint statistically independent of the others, a pre-whitening procedure (Kay, 1993) is applied as follows. A large number of noise traces (usually many thousand) is extracted from the same raw data signal as the spike waveforms but from the parts where no spike is detected. Each noise trace has the same number of datapoints as a spike waveform (N). Arranging all these traces in a large matrix $\vec{Z}$ (each row is one noise trace), the covariance matrix $\vec{C}$ of the noise can be calculated (Eq 7). Using the Cholesky decomposition (Eq 8), this matrix can be decomposed such that the product of the resulting matrix multiplied by its inverse results in the original matrix $\vec{C}$ (Eq 9).

$$\vec{C} = \mathrm{cov}(\vec{Z}) \qquad (7)$$

$$\vec{R} = chol(\vec{C}) \qquad (8)$$

$$\vec{C} = \vec{R}'\vec{R} \qquad (9)$$

By multiplying each raw spike waveform $\vec{S}_i$ by the inverse of $\vec{R}$ from the right side, all correlations are removed (Eq 10). After this operation, all datapoints of $\vec{P}_i$ uncorrelated.

$$\vec{P}_i = \vec{S}_i\vec{R}^{-1} \qquad (10)$$

The Choleksy decomposition (Eq 8, 9), however, requires that the covariance matrix $\vec{C}$ is invertable, that is, of full rank. But this is generally only the case for full bandwith noise. Various other forms of noise, for example narrow-band noise, result in a rank deficiency of the covariance matrix $\vec{C}$. Unfortunately we commonly observe this situation in our data. There exist methods for prewhitening of signals with rank-deficient noise (Doclo and Moonen, 2002; Hansen, 1998), but this is beyond the scope of this paper. Since all significant covariance values are usually very large, it is technically sufficient to add a very small amount of white noise to the covariance matrix (e.g., with a mean that is only 0.0001% of the covariance values) to make it full rank. While this is theoretically incorrect, it works sufficiently and we have not observed any noticeable differences in the decorrelated data with a rank-deficient pre-whitening method and the above method. We are thus using this approach to maximize efficiency.

An alternative approach for whitening is to design a whitening filter and whiten the signal itself before detecting and extracting spikes. This can for example be done by using the matlab function *lpc* to design a filter, and using this filter to whiten the signal. This way of processing is less susceptible to the numerical problems mentioned above but is harder to implement in a realtime environment. We used this method of whitening for the results reported in this paper (simulations with exact threshold estimation method).

2.8.2 Projection test

The projection test is entirely calculated on the basis of the pre-whitened waveforms $\vec{P}_i$ as described above. In the following, a waveform associated to cluster j is denoted as $\vec{P}_i^{(j)}$ and the center of cluster j is $\vec{P}^{(j)}$.

$$d = \left\| \vec{P}^{(j)} - P^{(k)} \right\| \qquad (11)$$

$$r_i = (\vec{P}_i^{(j)} - P^{(j)}) \bullet \frac{\vec{P}^{(j)} - P^{(k)}}{\left\| \vec{P}^{(j)} - P^{(k)} \right\|} \qquad (12)$$

The distance between two clusters is calculated by taking the norm of the difference between the two centers of cluster j and k (Eq 11). The residual r_i (scalar) for each spike $\vec{P}_i^{(j)}$ that is assigned to cluster j against cluster k (pairwise comparison between clusters j and k) is calculated by the dotproduct of the difference vector between the center and the spike $\vec{P}_i^{(j)}$, projected onto the vector that connects the two cluster centers (Eq 12).

Chapter 3. Single-trial learning of novel stimuli by individual neurons of the human hippocampus-amygdala complex

3.1 Introduction[3]

One prominent feature of nervous systems is the ability to distinguish novel from familiar stimuli. A rapid assessment of stimulus novelty is a prerequisite for certain kinds of learning (Davis et al., 2004; Kohonen and Lehtio, 1981; Li et al., 2003; Stark and Squire, 2000; Yamaguchi et al., 2004). For instance, conditioned taste aversions (CTA) and some forms of conditioned fear can be acquired in a single learning trial. Crucially, successful conditioning depends on the novelty of the conditioned stimulus (CS) (see Welxl, 2000 for a review). Pre-exposure to the CS severely diminishes associative learning (a.k.a. "latent inhibition"). Further, conditioning is also reduced if only some aspects of the CS are novel while others are familiar. The sensitivity to CS novelty, but not the taste aversion itself, is blocked by hippocampal lesions (Gallo and Candido, 1995). The novelty dependence of single-trial learning in the CTA paradigm points to the importance of a rapid assessment of stimulus novelty or familiarity.

The medial temporal lobe (MTL) is crucial for the acquisition of declarative memories and some functional imaging techniques have shown activation of MTL structures associated with either novel or familiar stimuli (Stark and Squire, 2000; Stern et al., 1996; Tulving et al., 1996; Yamaguchi et al., 2004). Lesion studies have repeatedly demonstrated that

[3] The material in this chapter is based on Rutishauser, U., Mamelak, A.N., and Schuman, E.M. (2006a). Single-trial learning of novel stimuli by individual neurons of the human hippocampus-amygdala complex. Neuron *49*, 805-813.

MTL damage impairs or abolishes behavioral, electrographic, and skin responses to novel stimuli (Kishiyama et al., 2004; Knight, 1996; Yonelinas et al., 2002). While these studies suggest a role of the MTL in novelty detection, the cellular basis for this discrimination has yet to be described. We report here that single neurons in the human MTL can alter their firing behavior to discriminate between novel and familiar complex stimuli following a single trial, exhibiting rapid plasticity as a result of single-trial learning.

3.2 Results

3.2.1 Task paradigm and behavioral results

We recorded single neuron activity using microwires implanted in the human hippocampus-amygdala complex (Figure 3-1A,B; see Table 3-1 for electrode locations), while subjects performed a object learning and recognition task. The delay between the learning and the initial recognition period was approximately 30 min, during which time the subject performed a different, cognitively demanding task. During learning, subjects were shown 12 different visual images. Each image was presented once, randomly in one of four quadrants on a computer screen (Figure 3-1C). Subjects were instructed to remember both the identity and the position of the image(s) presented. During the recognition period, subjects saw either previously viewed (familiar) or new images (novel) presented at the center of the screen (Figure 3-1D). For each image, the subject was asked to indicate whether the stimulus was new (novel) or old (familiar). Note that the novelty of a stimulus is only defined by whether it has been seen before or not (contextual). No other attributes of the stimulus changed. For each image identified as familiar, the subject was also asked to identify the quadrant in which the stimulus was originally presented (spatial recollection). Subjects correctly identified, on average, 88.5 ± 2.8% of all familiar and novel items during recognition (Figure 3-5). Subjects correctly recalled the quadrant location for 49.5 ± 8.0% of the familiar stimuli.

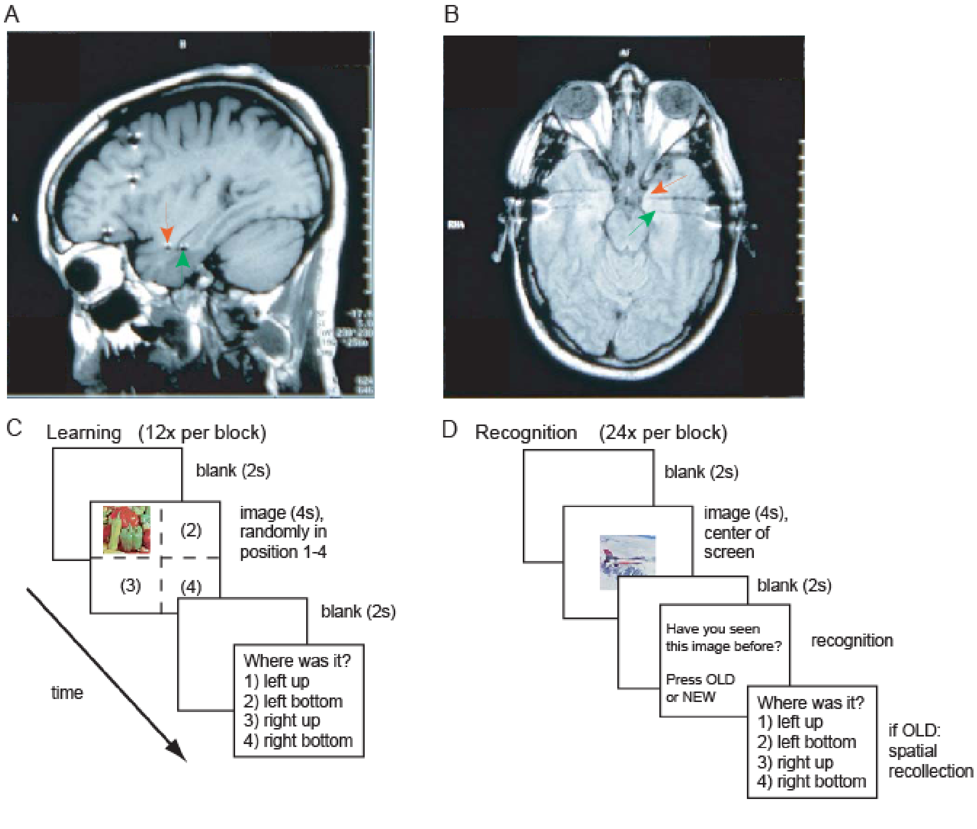


Figure 3-1. Electrode placement and task design.
(**A**) Saggital and (**B**) axial post-implantation structural MRI of one patient. The electrodes implanted in the amygdala (red) and the hippocampus (green) are indicated with arrowheads. The experiment has a learning (**C**) and a recognition block (**D**). Learning trials consisted of 12 images presented in one of 4 quadrants on the screen. 2 seconds after the stimulus was removed and replaced by a blank screen, the subject was asked to report in which quadrant the stimulus was presented. During recognition trials (30 min later), the subject was shown the 12 old images mixed with a set of 12 new images and asked to indicate whether the image had been viewed before (old) or not (new). After classifying an image as "old", the subject was also asked to indicate where the picture was during learning (spatial recognition).

3.2.2 Neural representations of single-trial learning, novelty, and familiarity

We analyzed the response of every neuron recorded (total number of neurons across all subjects = 244) during the baseline, stimulus presentation, and post-stimulus delay period. A neuron was considered selective if it exhibited an altered firing rate as a function of the stimulus (novel vs. familiar) ($p < 0.05$, bootstrap, see methods) and as a function of the task (learning vs. recognition phase). Neurons that increased their firing when exposed to novel vs. familiar stimuli were classified as signaling "novelty", whereas neurons that increased their firing to familiar stimuli were classified as signaling "familiarity" (Figure 3-2). Additionally, we classified responding neurons according to *when* they increased their firing: during the stimulus presentation of the stimulus or during the post-stimulus period (Figure 3-6D). Note that neurons signaling "novelty" increased their firing to new stimuli during the learning phase *and* also increased their firing to new stimuli presented during the recognition phase.

Are individual neurons capable of signaling that learning has occurred? If this is the case, then once the subject learns something about a stimulus (e.g., that it has been seen before) the firing properties of the neuron should reflect this knowledge. In our task, any knowledge about whether the specific stimulus presented has been seen before must result from a single trial experience. We indeed found subsets of neurons that showed enhanced or depressed firing rates on the second of two stimulus presentations, indicating the capacity for single-trial learning of familiarity. There are two different patterns of responses we observed that indicate single trial learning. One set of neurons ("familiarity detectors") exhibited enhanced firing when previously viewed stimuli were presented a second time during the recognition phase of the experiment. An example of this type of response is shown in Figure 3-2, where the neuron does not exhibit any appreciable response to the stimuli when first presented (Figure 3-2B) but when these same stimuli are presented a second time a dramatic increase in firing rate was observed (Figure 3-2D). These cells, which form a class of “familiarity” detectors, thus exhibit single-trial learning, exhibiting memory for a stimulus that was presented only one time. The other class of cells increased firing only for the first presentation of the stimulus ("novelty detectors”, see Figure 3-8 for an example). All told, 40 neurons consistently signaled either novelty (n = 23) or familiarity (n = 17) (Figure 3-3A,B). To characterize the firing differences of all neurons, we used two measures: i) average firing rate increase relative to baseline for new or old stimuli (depending on type of neuron), and ii) the average firing rate difference between new vs. old stimuli. For both measures, spikes were counted in the entire 6 s period following stimulus onset. We find that neurons increase firing on average 47% relative to baseline and the average firing difference between old vs. new stimuli is 76% (Figure 3-3C). The larger difference when comparing new vs. old firing indicates that in addition to increasing firing to the preferred stimulus (e.g., familiar), neurons decrease firing for the other stimulus type (e.g., novel). The large change in firing rate observed was induced by a single presentation of the stimulus and as such, these neurons provide a potential source for the rapid single-trial memory exhibited behaviorally by the subjects.

Do the observed neuronal changes reflect either a priming or a habituation response, or alternatively, do they reflect a form of long-term memory? If the former is the case, one would expect that, if presented with the same familiar stimuli (as well as new stimuli) 24 h later, the neuronal response to the familiar stimulus would be diminished. On the other hand, if the response reflects long-term memory, the altered firing pattern should still be observed the next day. To address this, we conducted a recognition session on the second and/or third day of recording, presenting subjects with the stimuli learned the previous day (4 sessions total in 3 patients) as well as a new set of stimuli. The time delay between the learning and the second recognition session was approximately 24 h (including one night of sleep). The behavioral performance (recognition and recollection) of these 3 patients did not differ significantly after a 30 min or 24 h time delay. Unfortunately, single-unit microwire recordings do not allow one to unambiguously determine whether the same individual neurons can be recorded on two sequential days. As such, we asked whether individual neurons, recorded 30 min or 24 hrs after the stimulus presentation, showed differences in firing to old vs. new stimuli. We then compared the average response strength per neuron after 30 min and 24 h time delays. We found that neither the average response strength per neuron nor the average increase in firing rate relative to baseline (Figure 3-3D) differed significantly for the two different time delays (2-way ANOVA with groups neuron type (Novelty/Familiarity) and time delay (30 min/24 h), $p < 0.05$). These neurons thus reflect the memory of the stimulus learned 24 h earlier but do not exhibit any further increases in firing rate (see discussion). The majority of neurons (37 of 40) exhibited a significant response within the first 2 s after stimulus onset (Figure 3-7C). Does the response strength decrease as a function of trial number? We found that neither novelty nor familiarity neurons significantly reduce their response strength over the duration of the experiment, during either learning or recognition (1-way ANOVA with block-nr and $p < 0.05$ reveals no significant effects for blocks of 1, 2, 3, or 4 trials). In addition, we found both types of neurons, familiarity and novelty detectors, in the amygdala as well as the hippocampus (Figure 3-6). However, the overall incidence of these neurons was significantly less in the amygdala when compared to the hippocampus: 19.7 ± 4.9% (n = 11) of all hippocampal neurons and 8.3 ± 2.7% (n = 12) of all amygdala neurons were classified as either novelty or familiarity neurons (n is number sessions, $p < 0.05$).

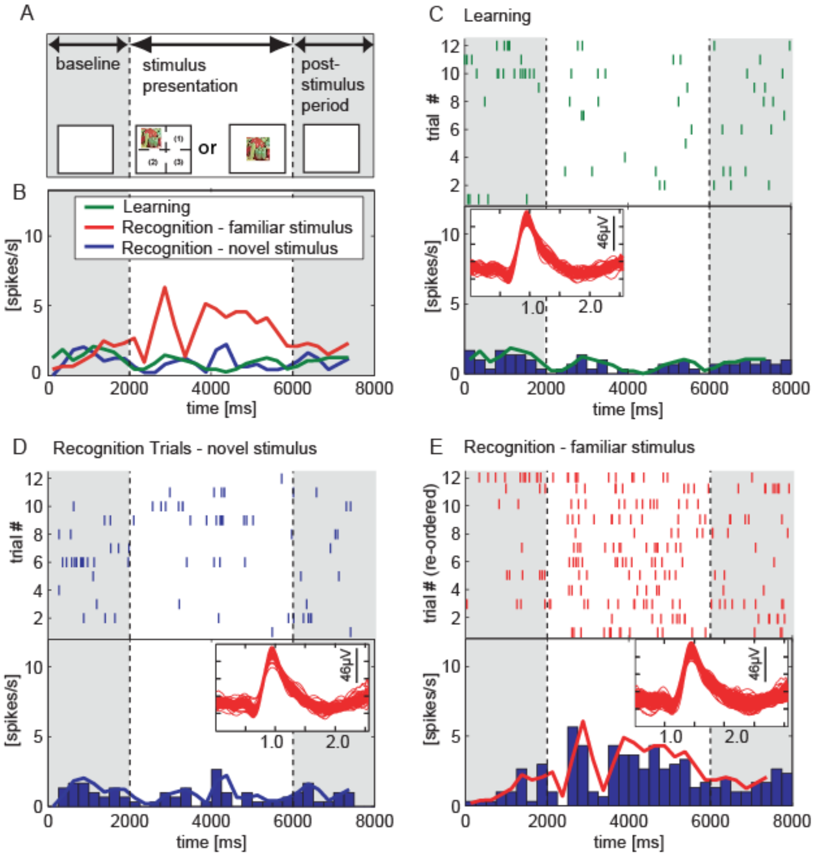


Figure 3-2. Example of a single hippocampal neuron during learning and recognition.

(**A**) Schematic representation of the experiment. Baseline (blank screen) from 0 to 2s, stimulus presentation from 2 to 6s, and post-stimulus period (blank screen) from 6 to 8s. (**B**) Average responses (spikes/sec). **(C-E)** The top portion of each figure shows the rasters depicting individual spikes. The stimulus was presented during the epoch defined by the dashed vertical lines. The bottom portion of each figure shows the binned histograms across all trials. Insets show overlays of all spike waveforms during the phase of the experiment depicted. (**C**) Responses during each learning trial. (**D**) Responses during the recognition phase for all new (not previously viewed) stimuli. (**E**) Responses during the recognition phase for all previously viewed (old) stimuli. Trials were randomly ordered during the experiment but are shown in (E) in the same order as during learning (C). This neuron increases its firing rate for stimuli seen before (E) but not for stimuli viewed for the first time (novel during both learning and recognition) (C and D).

Note that in C and E, the exact same visual stimuli are presented to the subject (12 images). When the stimuli are presented the first time (C), the neuron does not respond, whereas for the second presentation (E) it responds strongly.

3.2.3 Single neuron and population decoding

We analyzed how reliably these neurons can signal novelty or familiarity with an ideal-observer model. The model has access to the number of spikes fired during the 6 s period following stimulus onset. Using this information, a “decision” is made as to whether the subject is viewing a novel or a familiar stimulus. By parametrically varying the threshold (number of spikes) above which a single trial was considered novel or familiar, we conducted a receiver operator characteristic (ROC) analysis for each single neuron (Figure 3-7) and compared the true and false positives ratio at different thresholds. As a summary measure, we computed the area under the curve (Britten et al., 1996), which is the probability of correctly predicting whether the subject is currently viewing a novel or familiar stimulus (probability is between 0 and 1.0; 0.5 represents chance performance). We found that our neurons have an average single-trial single-neuron prediction probability of 0.72 ± 0.02. The population average is significantly above the chance level, which is determined by randomly shuffling the novel/familiar labels while keeping the spike trains intact. An observer that only has access to a single neuron can thus predict with on average 72% success whether a subject is seeing a familiar or novel stimulus.

How much information does the population of all recorded neurons contain about the familiarity of a stimulus? While ROC analysis quantifies how much information a single neuron conveys about the stimulus, it remains to be investigated how well this information can actually be decoded from a population of neurons on a single-trial basis. Single trials are highly variable and noisy. Does combining multiple neurons allow more accurate decoding than observing only a single neuron? Only if the signal or the noise were uncorrelated among neurons would one expect an improvement in decoding accuracy.

To address these questions, we used a simple population decoder which has access to all simultaneously recorded neurons that were previously identified as signaling novelty or familiarity. The decoder does not know the identity (novelty or familiarity detector) of the neurons. The only information available to the decoder is the number of spikes each neuron fired in the 6 s period following stimulus onset. The weighted sum (Figure 3-4A) of all spike counts is used to predict whether, for a given trial, an Old or New stimulus was presented. The weights are estimated from a set of labeled trials (Old or New) using multiple linear regression (see Methods).

We evaluated the properties of the classifier by considering only behaviorally correct recognition trials. For each recording session, we trained the classifier with all trials except a randomly chosen one (the "left-out trial"). Afterwards, we tested the classifier's performance by using it to predict whether the "left-out trial" was Old or New. Repeating this procedure many times for each session gives an accurate estimate of classifier performance (leave-one-out cross validation, see Methods). Additionally, we restricted the number of neurons that the classifier has access to. We found that the average single-trial classification performance increases from 67% correct for one neuron to 93% when 6 simultaneously recorded neurons are considered (Figure 3-4B, red line). A 1-way ANOVA reveals a significant effect of number of neurons (F = 6.6, p = 0.0001). Repeating the same procedure using randomly scrambled labels for the test trial results in a chance (50%) level performance (Figure 3-4B, black line). This analysis shows that it is beneficial for an "ideal" decoder to look at multiple neurons simultaneously. This indicates that the spikes fired by individual neurons signaling familiarity are uncorrelated in the sense that each of them contributes additional information that can be used to increase the accuracy of decoding.

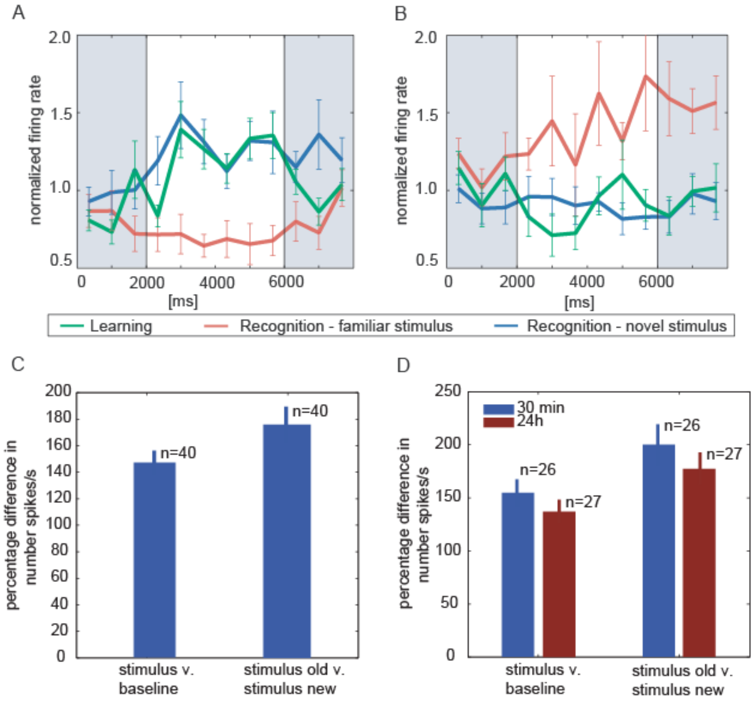


Figure 3-3. Population summary of all responding neurons.
Learning trials are in green, recognition old (familiar) trials are in red and recognition new (novel) trials are in blue. Neurons were classified according to which stimulus (old or new) they exhibited an increased firing rate and when they increase their firing (during either the stimulus or post-stimulus period or both). (**A,B**) Population average of all novelty (n=18) and familiarity neurons (n=10) which signal during the stimulus period. (**C**) Summary of response, quantified either as percentage firing rate difference during the 6 s post-stimulus period for old vs. new stimuli (right) or as percentage rate change relative to baseline (left). Note that the average rate increase of 75% is the result of a single stimulus exposure — the stimulus is learned after one trial. (**D**) Comparison of response for different time delays between learning and recognition. Shown is the average response strength with 30 min and 24 h delay. There is no significant difference in response strength for 30 min and 24 h delay (ttest, $p < 0.05$) nor is there a difference for novelty and familiarity detectors (not shown, 2-way ANOVA, $p < 0.05$). All errorbars are ±s.e. and n specifies number neurons.

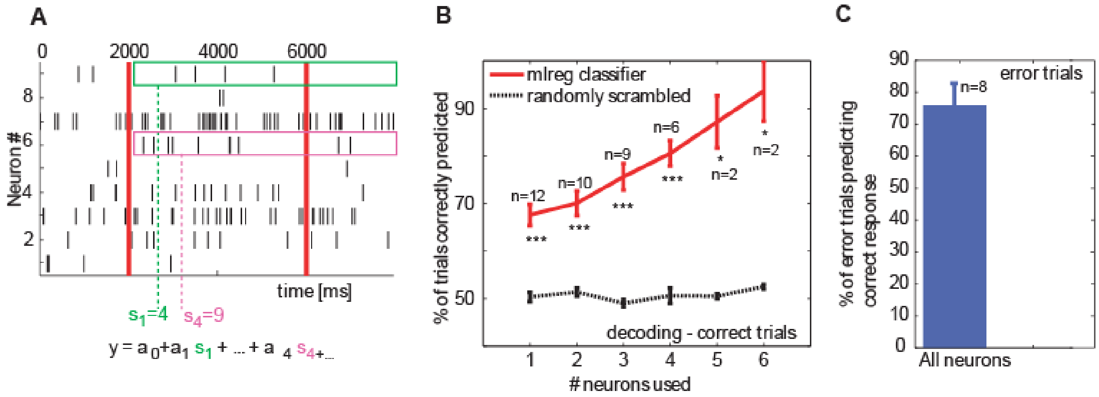


Figure 3-4. Population decoding from simultaneously recorded neurons. (**A**) Illustration of the decoding approach. Spikes of each neuron that signals novelty/familiarity (9 neurons in this example) are counted in the 6 s period following stimulus onset (first red line). Each neuron is assigned a weight determined by multiple linear regression. For a given trial, y predicts whether the trial is "Old" or "New". (**B**) Performance of the single trial-predictor as a function of number of simultaneously recorded neurons. Decoding performance increases when information from multiple recorded neurons is considered. The number of neurons used for decoding has a significant effect on performance of the decoder (1-way ANOVA, $p < 0.001$). n indicates the number of recording sessions. (**C**) The population decoder as trained in (B) applied to error trials. For 75% of all error trials in each session it predicts the correct response, that is, the neurons have better memory than the patient has behaviorally. The maximum number of available neurons is used for each session (mean number of neurons = 4.5). Only sessions that have at least 2 error trials are included (8 sessions). Errorbars are s.e. per session ($n = 8$) and the mean per session is significantly different from chance ($p < 0.01$).

3.2.4 Relations between neural responses and behavior

What is the relationship between the familiarity/novelty responses of individual neurons and the behavioral performance of the subject? The neuronal activity associated with behavioral errors allows us to answer this question. In our experiments, there were two kinds of error trials: i) recognition (novel vs. familiar) errors and ii) spatial recollection (which quadrant) errors. Below we investigate each type of error separately, beginning with spatial recollection errors.

There have been conflicting accounts as to whether retrieval-related activity in the hippocampus is related to familiarity recognition or recollection (Cameron et al., 2001; Stark and Squire, 2000; Yonelinas et al., 2002). One hypothesis states that the hippocampus is not involved in the retrieval of pure recognition memory, that is, memory without a recollective component. To investigate this issue, we examined neural activity during trials with successful recognition but failed recollection (spatial location of stimulus). We found that the subsequent successful spatial recollection is not required for neurons to exhibit familiarity responses. In fact we observe novelty and familiarity selective neurons in subjects who perform at chance levels for spatial recollection: In 4 (of 12) sessions, spatial recollection performance was at chance (21.7 ± 15.8%) and yet we found that 12 of the total 68 recorded units (17%) signaled novelty or familiarity. Thus, despite the fact that these patients weren't able to correctly recollect the spatial location in any of the trials, the same percentage of cells signaled novelty as in the other sessions. Also, for the sessions in which spatial recollection performance was above chance, we repeated our analysis including only trials associated with failed spatial recollection. Of the original 30 neurons, 26 remained significant (see Methods for details). We thus conclude that successful recollection is not required to observe a novelty/familiarity response in the hippocampus.

How is the neuronal activity during the stimulus presentation related to errors in recognition? Recognition of pictures is a highly automatic and reliable form of memory and subjects are usually very confident in their responses. This results in a small number of errors even when a large stimulus set is used, which has prevented analysis of such error trials in the past (Xiang and Brown, 1998). In our experiments, however, we record from many neurons simultaneously and can thus use a population decoder that allows accurate single-trial decoding (see discussion above). For each recording session, we trained the population decoder using all behaviorally successful trials. Afterwards, we used it to investigate what it would predict for the spiking activity observed during error trials. What might the population decoder (classifier) predict for an error trial? The classifier could: i) be at chance, ii) mimic the subject's (incorrect) response, or iii) predict the (correct, but not chosen) response. Each outcome would be informative: i) if it is at chance, these neurons do not contain any information about the stimulus on error trial; ii) if it predicts the behavioral response given, these neurons would likely represent some form of decision taken by the patient or motor planning activity related to the key the patient used to indicate the response; iii) if it predicts the correct response, these neurons would likely represent some form of high-fidelity memory. The third possibility is intriguing because it would suggest that these neurons exhibit "better memory" than the subject's behavioral response indicated. Since we are interested in the fraction of error trials per session that predict a certain outcome, we consider only sessions which contain at least 2 error trials (8 out of 12 sessions with a total of 33 error trials). For each session, we trained a classifier with all available neurons (on average 4.5) that signaled novelty/familiarity using all behaviorally correct trials and used it to predict the outcome of each error trial. We find that the classifier predicts the actual correct response for 75±7% of all error trials. The classifier is thus able to correctly predict the correct response in 75% of all cases even when the subject responded incorrectly (Figure 3-4C). These neurons thus have better memory than the patient exhibited behaviorally. This also suggests that the neuronal activity reported here does not represent some form of motor activity related to the subject's intended or actual response.

3.3 Discussion

3.3.1 Novelty and familiarity detectors in the human brain

We identified single neurons in the human hippocampus and amygdala that signal novelty or familiarity with an increase in firing rate. Several other groups have described non-human primate neurons that gradually (over many trials) decrease their response magnitude as specific stimuli become more familiar (Asaad et al., 1998; Fahy et al., 1993; Li et al., 1993; Rainer and Miller, 2000; Rolls et al., 1993). These types of neurons have also been observed in rodents (Berger et al., 1976; Vinogradova, 2001). The opposite pattern, neurons that increase their response magnitude for familiar stimuli, have largely not been observed in the primate brain (Fahy et al., 1993; Heit et al., 1990; Rolls et al., 1993; Xiang and Brown, 1998), and only rarely in humans (Fried et al., 1997). Also, studies investigating the relative proportion of novelty/familiarity- selective neurons in different areas of the MTL have usually failed to find any such neurons in the non-human primate hippocampus (Riches et al., 1991; Xiang and Brown, 1998) or, in one case, found only a very small proportion of such cells (Rolls et al., 1993). In contrast, we found a large proportion (17%) of familiarity/novelty-sensitive neurons, with an approximately equal number of neurons that increased firing for novelty or familiarity in the human hippocampus and amygdala. It has been speculated that the apparent absence of novelty/familiarity neurons in the primate hippocampus can be attributed to the lack of a spatial component in the tasks used (Riches et al., 1991; Xiang and Brown, 1998). To address this point, we used a non-spatial (old/new) and spatial recollective component in our task and find that the responses observed do not depend on successful spatial recollection. Another crucial difference is the behavioral task. Our task consists of a learning and recognition block with an interposed time delay. During the delay, other tasks are conducted. Others have used a serial recognition task where learning and recognition trials are intermixed and as such, there is no time delay that would permit a diversion of cognitive resources. It is possible that the emergence of the neuronal response requires time to develop. In our experiments, the firing rate increase can be observed after an initial delay of 30 min and remains equally strong for at least 24 h. This indicates that these neurons represent some form of long-term memory. Also note that the response strength does not increase further between 30 min and 24 h delays. The ability to correlate neuronal responses with human behavior may also be critical: we used an abstract task that can be rapidly learned thus facilitating the detection of these rapidly changing neuronal responses. In contrast, in non-human primates a simple associative memory task can take many trials for animals to reach criterion and learning-induced changes in hippocampal activity show a similar prolonged temporal profile (Wirth et al., 2003).

Could it be that the different findings are caused by eye movements? Most primate studies require the animal to fixate. In our experiments, subjects are free to move their eyes as they like. This is to make the task as natural as possible. Owing to clinical constraints, we were unable to record eye movements but there are several pieces of evidence which argue that eye movements cannot explain our results. The first few fixations made on any picture are mostly dominated by the statistics of the stimulus and do not change as a function of the familiarity of the stimulus (Noton and Stark, 1971). Also, a previous study of human MTL neurons found no influence of the fixated location of the picture on the visual response properties (Kreiman et al., 2002).

Others have reported that some neurons in the human MTL (Kreiman et al., 2000a) and the primate cortex (Li et al., 1993) are sharply tuned to the visual category of stimuli. Here, we used stimuli from many different visual categories (e.g., planes, cars, bottles, animals, mountains, people, computers, cameras, houses, books, chairs, and trucks) with one example per category. While the small stimulus set required for this kind of memory experiment prevents us from testing large numbers of stimuli from different categories, the response observed is invariant to at least a majority of the visual categories we have used. Thus, the neurons we describe here are capable of signaling the familiarity of the stimulus regardless of its visual category. One possibility is that the neurons preserve their tuning to categories and additionally increase or decrease their firing to indicate familiarity in an additive way. If this were the case, we would only detect broadly tuned units because narrowly tuned units would respond to a very limited set of stimuli. The neuronal responses we describe could thus serve as "general" novelty detectors that serve to establish the significance of behavioral stimuli during the acquisition of new or consolidation of existing memories (Lisman and Otmakhova, 2001).

Recognition and recollection are two largely distinct memory processes. Here we study recognition memory, but to allow a comparison with earlier human studies of recall/recollection we have included a spatial recollective component. Importantly we find that the response to the second presentation of the stimulus does not depend on whether spatial recollection is successful. This is in agreement with an earlier study of recollective memory which found that recall success is not correlated with the response of hippocampal neurons (Cameron et al., 2001). Also note that (Cameron et al., 2001) used the same stimuli many times during learning, so that the resulting neuronal changes cannot be related to any specific stimulus presentation. Similar studies of associative memory in the monkey hippocampus (Wirth et al., 2003; Yanike et al., 2004) are also complicated by this issue: stimuli were presented a large (10–30) number of times in order for the monkey to achieve behavioral criterion. These studies generally find that hippocampal neurons only change their response after many learning trials and thus seem to represent some form of "well learned" information. In contrast, in our study of human MTL neurons we use a single-trial learning paradigm that reveals that neurons are capable of rapid, single-trial plasticity.

3.3.2 Neurons that remember better than subjects

The finding that the neuronal activity during a majority of the error trials predicts the correct response represents an interesting disassociation between behavior and neuronal activity. In theory, an error could occur because the subject did not pay attention (not see the stimulus), accidentally pressed the wrong button, or because the subject did not remember the image correctly. Since the population decoder was not at chance levels for error trials, the first possibility can be excluded. Whether the subject accidentally pressed the wrong button or did not remember the image correctly cannot be determined from the available data. However, given the generally very high performance in the task and the absence of pressure to respond fast, it is unlikely that a majority of the error trials are caused by accidental wrong responses. If one examines the successful recognition trials exclusively, one might conclude that the neuronal responses represent the outcome of the decision taken (Old or New) or a consequence of that decision, e.g., planning and/or pre- or post-motor activity. If this were the case, however, activity during error trials would have to predict the response that was actually observed. However, we observed the opposite: activity during error trials predicts the correct response. We thus conclude that the neurons reported here represent some form of memory. In addition, the proportion of trials correctly identified by the neuronal responses is higher than what we observed behaviorally. Our data do not address at what point in the circuit the accurate neuronal responses on error trials fail to translate into correct behavioral responses. However, it is likely that information from multiple brain areas must be integrated to decide about the novelty of a stimulus. Any system of this nature requires an internal threshold for what is considered sufficient cumulative evidence for a stimulus to be classified as familiar. One could thus imagine situations where some brain areas provide input indicating familiarity but the cumulative evidence does not pass this threshold. Such a system would be maximally robust because it integrates multiple sources of information, perhaps trusting some more than others (Pouget et al., 2003). While it seems puzzling to have neurons that have better memory than is behaviorally observable, it makes sense in light of resistance to noise and erroneous transmission.

It has previously been observed that the average firing rate of some MTL neurons differs for successful vs. non-successful retrieval (Fried et al., 2002; Fried et al., 1997). However, in these studies, activity of the same neuron was not recorded during learning and it has thus remained impossible to determine whether these neurons changed their firing as a function of previous stimulus exposure or as a function of the task. In contrast, here we demonstrate that these changes result from a single stimulus exposure.

3.3.3 Relationship to fMRI and ERP findings

It has proven difficult to find human MTL fMRI activity correlated with behavioral success in recognition memory tasks (Manns et al., 2003; Stark and Squire, 2000). Using single-unit recordings we find evidence for the coexistence of novelty and familiarity cells recorded at the same time in the same brain region. On half of all macroelectrodes (18 of 36), we detected both novelty and familiarity neurons. On 2 of 6 microwires with more than one novelty/familiarity neuron both types were found. Since fMRI methods have limited spatial and temporal resolution and often rely on subtractive techniques, it is likely that the presence of both classes of neurons prevented their detection (Logothetis et al., 2001). The coexistence of MTL neurons that signal novelty or familiarity is likely an important feature used in establishing the significance of environmental events during learning.

Scalp and intracranial event-related potentials (ERP) recorded during serial recognition tasks have revealed a prominent potential (P300) to novel as well as target stimulus items (McCarthy et al., 1989; Sutton et al., 1965). That is, there is a potential to both novel as well as familiar (task relevant) items, but not to distractors. In hippocampal lesion patients it has been observed that the P3a component of the P300 is reduced (Knight, 1996). While we did not record ERPs in this study, the P300 response has been observed previously with intracranial electrodes in similar locations (McCarthy et al., 1989). It is thus of interest to note that the identified subpopulations of novelty and familiarity neurons we identified here could contribute to the P300.

3.3.4 Interaction with other brain systems

What is driving the response of these neurons? Neurons from multiple other brain areas can signal novelty or, more generally, the behavioral relevance of stimuli encountered in the environment. These include noradrenergic neurons in the locus coeruleus, cholinergic neurons in the basal forebrain as well as dopaminergic neurons in the midbrain (see (Schultz and Dickinson, 2000) for a review). Their response to novel events habituates with brief delays, evidence for short-term memory. Common to all these areas is the modulatory nature of their output — it is thus unlikely that their output is sufficient to account for the MTL responses we observe. These modulatory systems are known to regulate the strength of hippocampal-dependent learning, however (Frey et al., 1990; Neuman and Harley, 1983; Williams and Johnston, 1988), raising the possibility that the rapid plasticity we describe is related to the simultaneous release of neuromodulators that help induce long-lasting memories.

It is well known that animal behavior can be modified by a single exposure to a relevant stimulus (Sokolov, 1963). One instance of such memory is episodic memory, which is, by definition, memory of a single experience (Tulving et al., 1996). Other instances of single-trial learning include object recognition (Standing et al., 1970), spatial learning, and food caching (Clayton et al., 2001). In contrast, other forms of learning, like classical conditioning or rule learning (Wirth et al., 2003), require many learning trials. The neurons that underlie or participate in the rapid behavioral plasticity have, for the most part, evaded detection. Here we find that MTL neurons exhibit remarkable plasticity: a single exposure to a stimulus was sufficient to induce a dramatic and significant change in the spiking pattern. The observation of single-trial learning in MTL neurons indicates that, at least in principle, the rapid learning that human subjects exhibit has an electrophysiological correlate that occurs at the level of individual neurons.

3.4 Experimental procedures

3.4.1 Subjects and electrophysiology

Subjects were 6 patients (3 male, 3 female; mean age 37.5 ± 5.5 years; all native English speakers) diagnosed with drug-resistant temporal lobe epilepsy and implanted with intracranial depth electrodes to record intracranial EEG and single-unit activity. Patients underwent stereotactic placement of hybrid depth electrodes containing both clinical field potential contacts and microwire (50 µm) single-unit contacts, as described by (Fried et al., 1999). Briefly, electrodes were placed using orthogonal trajectories through the dorsolateral cortex, with the tip of the electrode targeting the amygdala, anterior hippocampus, orbitofrontal region, supplementary motor area, or anterior cingulate gyrus. The commercially available electrodes (Behnke hybrid depth electrode, Adtech Inc, Racine, MN), contain 4–6 platinum-iridium 5 mm long circular electrodes, with a hollow center. After insertion of the electrode in the target, the inner cannula was removed and a bundle of microwires was passed through the center of the electrode, extending 5 mm beyond the tip of the electrode in a “flower spray” design. The electrodes were secured in place via a skull anchor bolt. All electrodes were placed based on clinical criteria alone. Patients were recruited for the research study after surgery was completed and EEG monitoring was initiated. Participation was voluntary and patients could withdraw from the study at any time. Informed consent was obtained and the protocol was approved by the Institutional Review Boards of the Huntington Memorial Hospital and the California Institute of Technology. For further details regarding the electrophysiological recordings, please see the supplemental material.

3.4.2 Data analysis

Spikes were sorted with a template-matching method (Rutishauser et al., 2006b). Only well-separated single neurons were used (see supplemental methods for details). We used a nonparametric bootstrap statistical test (Efron and Tibshirani, 1993) to assess significance at $p < 0.05$ (see supplement for discussion why not a t-test). To determine whether a neuron responds to new or old stimuli we compared the number of spikes fired for old vs. new stimuli during the stimulus on (4 s) and the post stimulus (2 s) period. For bootstrapping, 10,000 randomly re-sampled (with replacement) sets of spike counts were generated and tested for equality of means (Efron and Tibshirani, 1993). A second statistical test was performed to determine whether the firing of a neuron between old stimuli during recognition and all stimuli during learning (which are, by definition, new) was different. Only if both statistical tests were passed with $p < 0.05$ was the neuron determined to function as a novelty or familiarity detector. We randomly shuffled the start/endpoints of trials (in time) while keeping everything else the same to establish chance performance for this statistical procedure. We repeated this procedure 10 times and found a chance performance of 4.4% of all neurons (Figure 3-6D). Error trials during learning (incorrect position) and recognition (New/Old wrong) were excluded from this analysis.

All errors are standard error (s.e.), unless noted otherwise.

3.4.3 Population analysis

To quantify how well we were able to decode information about the novelty of the stimulus for a single trial, we used a population decoder. This also allowed us to analyze whether and how the decoding performance depends on the number of simultaneously recorded neurons. We used a simple weighted sum classifier of the form $y = a_0 + a_1 s_1 + ... + a_n s_n$, where s_x represents the number of spikes in the 6 s period following stimulus onset for neuron x, and a_x is the weight of this neuron. The weights are determined from labeled training data using multiple linear regressions (Johnson and Wichern, 2002). The label y is either set to 1 (New) or -1 (Old). Only neurons which were previously found to be signaling novelty/familiarity were considered for this analysis.

For verification purposes, we trained the classifier on behaviorally correct trials using leave-one-out cross validation. The performance of this classifier was then verified by evaluating its prediction for the left-out trial. Repeating this procedure many times gives an accurate estimate of the true performance of the estimator. We repeated the same analysis by restricting the number of neurons the classifier had access to. In cases where more neurons were available than the classifier could consider, a random subset of the available neurons was chosen and the procedure was repeated multiple times so that all possible combinations were explored. All error bars in the population analysis are given as s.e., with n being the number of sessions, to demonstrate the variance over multiple patients and recording sessions rather than over multiple neurons.

3.5 Supplementary material

3.5.1 Electrophysiology

Recordings were conducted using a commercial (Neuralynx Inc, Arizona) acquisition system with specially designed, head-mounted pre-amplifiers. Signals were filtered and amplified by hardware amplifiers before acquisition. The frequency band acquired was either 1–9000Hz or 300–9000Hz, depending on the noise levels. Great care was taken to eliminate noise sources. This included using batteries to power the amplifiers, experimental computers, IV machines and heartbeat monitors. Recordings commenced the second day after surgery and continued for 2–4 days for about 1 hour per day. The experiments reported in this paper were done on two consecutive days for all 6 patients (12 sessions in total).

The amplifier gain settings, set individually for each channel, were typically in the range of 20000–35000 with an additional A/D gain of 4 (2 in some cases). The raw data was sampled at 25 kHz and written to disk for later filtering (300–3000Hz bandpass), spike detection, and spike sorting. Spikes were detected using a local energy method (Bankman et al., 1993) and sorted by a template-matching method (Rutishauser et al., 2006b). Great care was taken to ensure that the single units used passed stringent statistical tests (projection test (Pouzat et al., 2002)) . It is thus likely that we underestimate the number of single units present. Only neurons with mean firing rates ≥ 0.25 Hz were included in the analysis.

3.5.2 Electrodes

In each macroelectrode, 8 microwires were inserted (Fried et al., 1999). One microwire was used as local ground and the other 7 were used for recordings. The impedance of a total of 56 microwires in 2 patients was, on average, 135 ± 62kOhm (± s.d.) with a range of 38–245 kOhm.

Electrode position was determined by an experienced neurosurgeon (ANM) from structural MRIs taken 1 day after electrode implantation on a clinical 1.5 Tesla MRI system (Toshiba, Inc). We always recorded from 3 macroelectrodes simultaneously: left/right hippocampus and either left or right amygdala (total of 24 channels, 8 channels for each macroelectrode with 1 channel used as local ground).

3.5.3 Localization of electrodes

We localized the position of each macroelectrode in a standardized stereotactic coordinate system (Talairach) in a subset of 4 patients for which high-resolution structural MRIs were available (Table 3-1). We transformed each structural 1.5 T MRI scan to Talairach space by manually identifying the anterior and posterior commisure as well as the anterior, posterior, superior, and inferior points of the cortex. We used BrainVoyager (Brain Innovation B.V.) for this procedure. After co-registration we identified the Talairach coordinates by finding a consensus from the different structural scans. For each patient, we performed 4 different scans with 1x1 mm resolution in the following plane: coronal, sagittal, and 2 axial with different pulse sequences (2TW and FLAIR).

Patient	Amygdala (r/l)	Hippocampus (r/l)
P2	-20,1,-19 26,-2,-20	-26,-9,-11 28,-11,-20
P3	-20,-3,-15 18,-4,-15	-23,-13,-12 33,-12,-16
P4	-19,4,-26 28,7,-26	-21,-9,-25 27,-7,-26
P6	-23,-2,-14 23,-6,-13	-25,-13,-12 29,-18,-12

Table 3-1. Electrode position in stereotactic coordinates (Talairach)

3.5.4 Implementation of behavioral task

The task was implemented using Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) in Matlab (Mathworks Inc) and ran on a notebook PC placed directly in front of the patient. Distance to the screen was approximately 50 cm and the screen was approximately 30 by 23 degrees of visual angle. The pictures used were approximately 9 by 9 degrees. Specially marked keys ("New", "Old") on the keyboard were used to acquire subject responses. We chose to use natural pictures as stimuli rather than words or faces because it has been shown that pictures reliably result in bilateral fMRI activation of the MTL, whereas words and faces result in primarily unilateral (left) activation (Kelley et al., 1998).

3.5.5 Data analysis

We conducted all statistical analysis using bootstrap tests (see Methods of main text). To be thorough, we repeated the same analysis using a two-tailed t-test ($p < 0.05$) and found reasonable overlap with the pool of neurons determined to signal novelty or familiarity using the above bootstrap method. We found, however, that using the t-test more neurons were classified as novelty/familiarity detectors, some of which (by visual inspection) were likely false positives. Also, the chance performance determined by random shuffling was high (~ 10%). We thus decided to exclusively use the bootstrap method since it yielded the most consistent and conservative results. Post-stimulus histograms (PSTH) were created by binning the number of spikes into 250 ms bins. To convert the PSTH to an instantaneous firing rate, a Gaussian kernel with standard deviation = 300 ms was used to smooth the binned representation. Population averages (Figure 3-3C and D) were constructed by averaging the normalized firing rate of each neuron. Firing rates were normalized to the mean firing rate of the neuron during the particular part of the experiment (learning block or recognition block). We averaged the raw normalized PSTH of each neuron (above PSTH smoothing is not applied to normalized PSTH of each neuron, nor to the population average).

3.5.6 Spatial recollection analysis

To investigate whether the response observed during familiarity/novelty recognition required later successful spatial recollection we conducted additional data analyses. Based on several pieces of evidence we find that successful spatial recollection is not required for emergence of novelty/familiarity cells: i) In 4/12 sessions spatial recollection performance was at chance levels (mean 21.7 ± 7.9%) and yet we found that 14.8% of the recorded neurons in these sessions signaled novelty/familiarity during recognition and showed single-trial learning. This percentage is remarkably similar to the percentage of all neurons that signal novelty or familiarity (Figure 3-6). Thus despite the fact that these patients weren't able to correctly recollect the spatial location in any of the trials the same percentage of cells signaled novelty as in the other sessions. ii) In the 8 sessions with above-chance spatial recollection performance (mean 63.91±7.02%), 28 neurons were found (17.2% of all recorded neurons). Repeating the analysis as described above, but only including trials with successful recollection, results in 26 of those 30 neurons remained significant. The number of selective neurons is thus decreased if only trials with successful spatial recollection are included and error trials are thus contributing valuable information. iii) In 9 sessions there were at least 4 spatial recollection error trials (correctly recognized as Old, but location wrong). Considering only these error trials (disregarding trials with correctly remembered locations), 20 out of originally 26 (77%) neurons remain significant. A high proportion of all originally identified neurons thus signal novelty/familiarity even in the absence of successful spatial recollection.

3.5.7 Single-neuron ROC analysis

To determine how well the response of a single neuron during recognition predicts whether the patient is currently viewing a familiar or novel stimulus we conducted an ROC (receiver-operator characteristic) analysis (Britten et al., 1996; Green and Swets, 1966). This analysis assumes that an ideal observer, who only has access to the number of spikes fired by a single neuron during the presentation of the stimulus and the post-stimulus period (6 s period), should be able to correctly classify individual neurons as signifying novelty vs. familiarity. Only trials where the subject correctly replied with "Old" or "New" were used for this analysis (this was 88.5% of all trials). We quantify the ROC for each neuron recorded by integrating the area under the curve (AUC) of the ROC. This number equals the probability of correctly predicting, on a single-trial basis, whether the “subject” has viewed a novel or familiar stimulus. An AUC of 0.5 equals chance. We confirmed the validity of our analysis by randomly shuffling the labels "New" and "Old" while leaving the spike trains intact. Repeating this procedure 50 times for each neuron resulted in AUC values clustered around 0.5 (Figure 3-7A,B).

We conducted this ROC analysis without preclassifying neurons into novelty/familiarity detectors. This results in a cluster of neurons with a prediction probability significantly below 0.5 and one significantly above 0.5. Since Old/New is a binary state, this contributes equal information and we thus subtracted 1-x for all ROC values x < 0.5 to get an unimodel distribution, as shown in Figure 3-7A.

We repeated the analysis above for different time bins following stimulus onsets (step size 500 ms), e.g. counting spikes in bins 2000–2500 ms, 2000–3000 ms, 2000–3500 ms, etc. Using this analysis we defined for each neuron when its ROC value became significantly above chance the first time (Figure 3-7C).

3.5.8 Epileptic vs. non-epileptic tissue

One concern regarding the neurons described in this paper is that they were recorded from epilepsy patients. To confirm that our findings are also valid for "healthy" tissue, we repeated our analysis but excluded all electrodes which were in tissue that was later resected

(Table 3-2). Of the total 244 recorded neurons, 138 were in tissue which was not resected. Of these 138 neurons, 22 signalled novelty or familiarity (15.9%).

Patient	**Side of temporal lobe lobectomy**
P1	left
P2	left
P3	right
P4	left
P5	left
P6	right

Table 3-2. Location of resected tissue (temporal lobe lobectomy in each case).

3.6 Supplementary figures

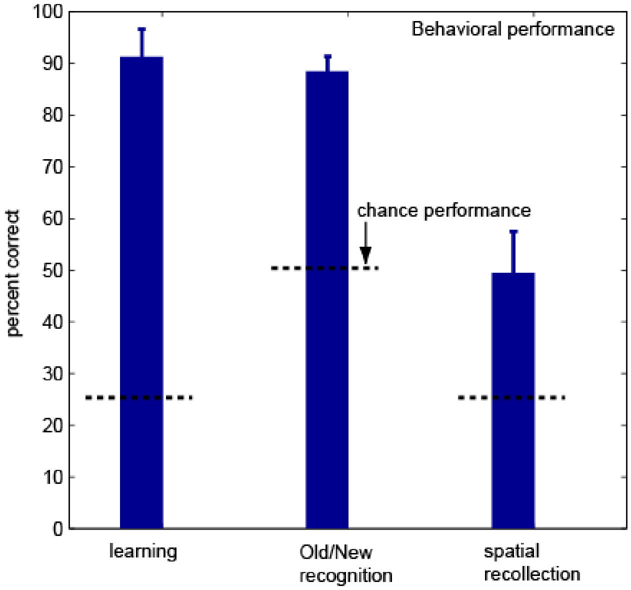


Figure 3-5. Behavioral performance of all subjects.
Recognition performance (Old/New) was close to 90% (chance 50%) whereas spatial recollection, in which the subject reports the quadrant in which the images was presented for all images classified as "Old", was 49%. All performance levels are significantly different from chance ($p < 0.05$).

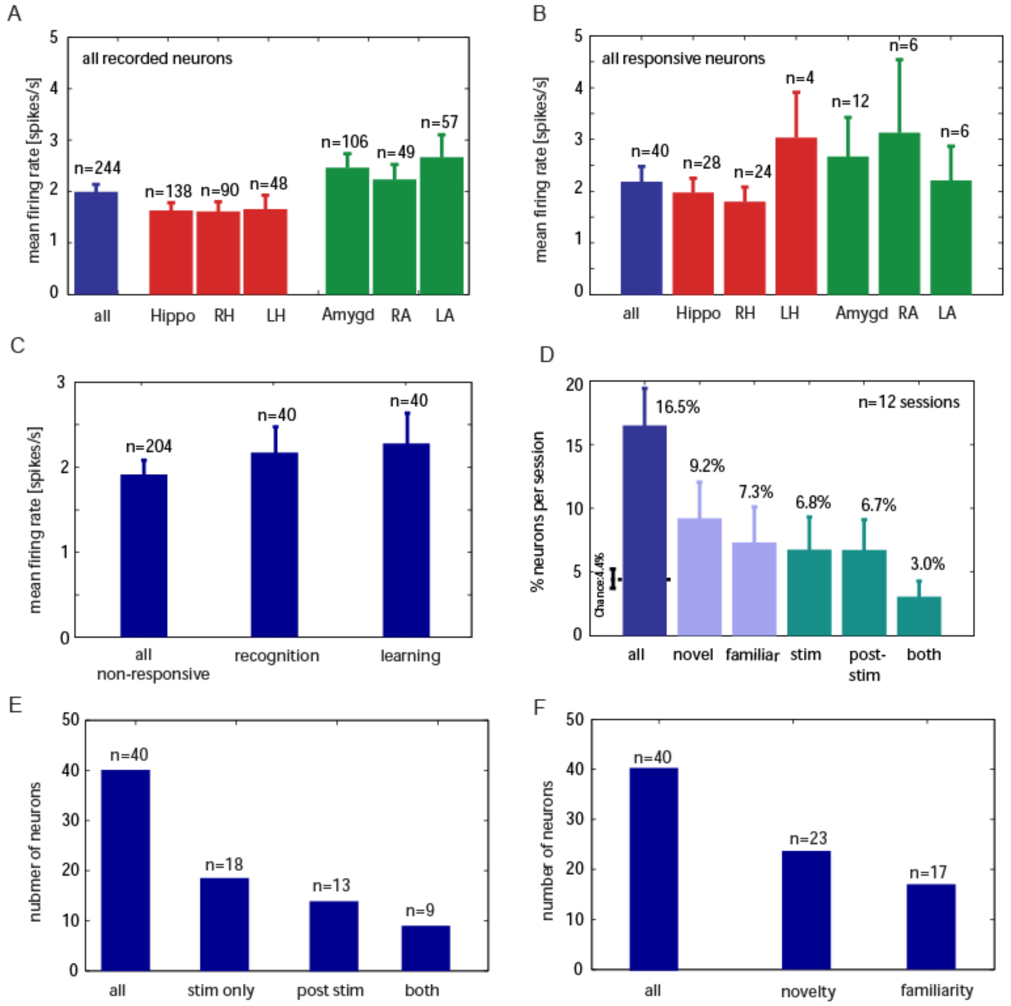


Figure 3-6. Population statistics for all neurons.

(**A**) as well as the subset of significantly responsive neurons (**B-F**). (**A**) The mean firing rates of all neurons recorded (n = 244) was 1.96 ± 0.14 Hz. The mean firing rate was not significantly different among different brain areas (1-way ANOVA, $p < 0.05$). (**B**) The mean firing rate of all responsive neurons (n = 40) was 2.17 ± 0.30 Hz, with no significant difference amongst different brain areas. (**C**) The mean firing rate for novelty and familiarity neurons was not statistically different from all other neurons recorded (1-way ANOVA, $p <0.05$) during either learning or recognition. (**D**) Considering all sessions, 16.5% of all recorded neurons indicated novelty or familiarity in every session (2 sessions each in 6 patients). There were slightly more novelty neurons (9.2%/per session) than familiarity neurons (7.3%/per session). (**E**) We found a total of 40 significant neurons, 18 of which signaled during the stimulus period, 13 during the post-stimulus period, and 9 during both; (**F**) There were 24 novelty and 18 familiarity neurons.

Abbreviations: RH, right hippocampus; RA, right amygdala, LH, left hippocampus; LA, left amygdala; hippo, hippocampus; amygd, amygdala. All error bars are ±s.e and n always specifies number of neurons.

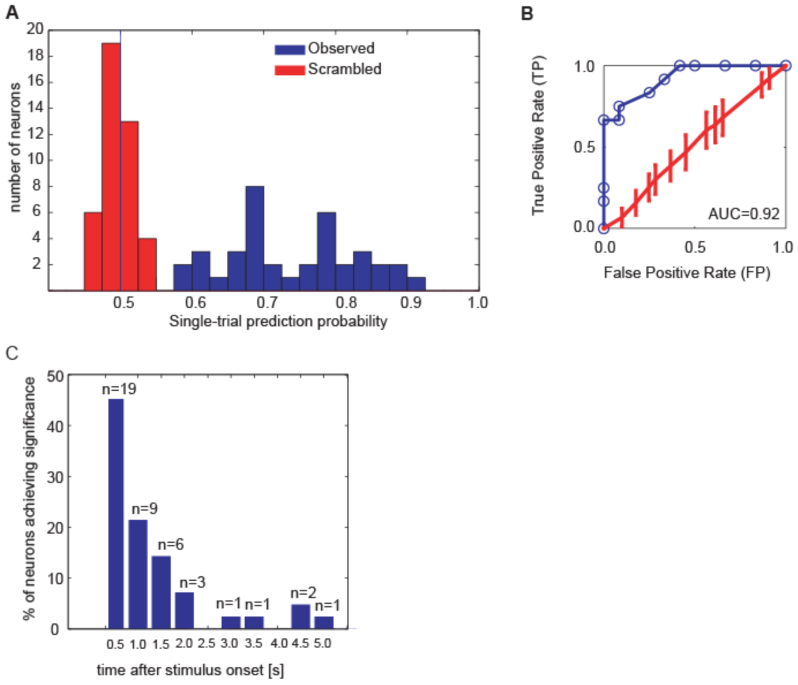


Figure 3-7. Single-neuron prediction probabilities.
(A) Histogram of the single-trial prediction probabilities for all 40 significant neurons. The mean probability was 0.72±0.02. The prediction probability is equal to the area under the curve of the ROC of each neuron and specifies the ratio of recognition trials in which novelty or familiarity is successfully predicted on a trial-by-trial basis by observing a single neuron. Randomly shuffling (scrambled) the spike counts of new and old trials results in a mean of 0.5 (red in A, error bars are s.d.). The ROC for the same neuron as shown in figure 2 is shown in (**B**) (blue = real trials, red = randomly shuffled). (**C**) Latency of response for all neurons. Shown are, for each time following stimulus onset, the percentage of neurons which became significant for the first time in this time bin.

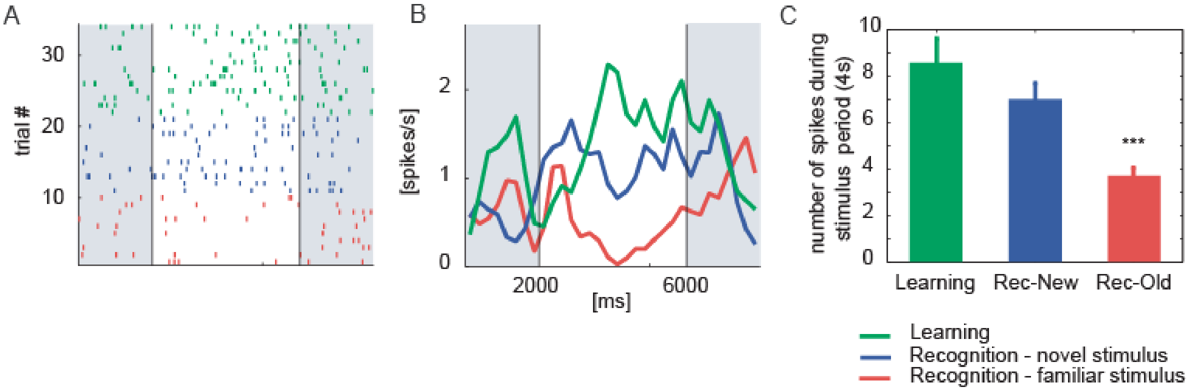


Figure 3-8. Example of a novelty-sensitive neuron.
Neuron which increases firing to novel stimuli during both learning and recognition. (A) Raster for all spikes during learning (green), recognition old (red), and recognition new (blue). (B) Histogram summarizing the response. Note the decrease to familiarity. (C) Comparison of the number of spikes fired during the 4 s stimulus period (white in B). The number of spikes fired for familiar items is significantly different from the number of spikes fired during learning and recognition of new items. (p < .001 for both comparisons, 1-way ANOVA with posthoc multiple comparison. n = 12 (number of trials)).

Chapter 4. Activity of human hippocampal and amygdala neurons during retrieval of declarative memories

4.1 Introduction[4]

Episodic memories allow us to remember not only whether we have seen something before but also where and when (contextual information). One of the defining features of an episodic memory is the combination of multiple pieces of experienced information into one unit of memory. An episodic memory is, by definition, an event that happened only once. Thus, the encoding of an episodic memory must be successful after a single experience. When we recall such a memory, we are vividly aware of the fact that we have personally experienced the facts (where, when) associated with it. This is in contrast to pure familiarity memory, which includes recognition, but not the "where" and "when" features. The MTL, which receives input from a wide variety of sensory and prefrontal areas, plays a crucial role in the acquisition and retrieval of recent episodic memories. Neurons in the primate MTL respond to a wide variety of stimulus attributes such as object identity (Heit et al., 1988; Kreiman et al., 2000a) and spatial location (Rolls, 1999). Similarly, the MTL is involved in the detection of novel stimuli (Knight, 1996; Xiang and Brown, 1998). Some neurons carry information about the familiarity or novelty of a stimulus (Rutishauser et al., 2006a; Viskontas et al., 2006) and are capable of changing that response after a single learning trial (Rutishauser et al., 2006a). The MTL, and in particular the

[4] The material in this chapter is based on Rutishauser, U., Schuman, E.M., and Mamelak, A.N. (2008). Activity of human hippocampal and amygdala neurons during retrieval of declarative memories. Proc Natl Acad Sci U S A *105*, 329-334.

hippocampus, are thus ideally suited to combine information about the familiarity/novelty of a stimulus with other attributes such as the place and time of occurrence.

The successful recall of an experience depends on neuronal activity during acquisition, maintenance, and retrieval. The MTL plays a role in all three components. Here, we focus on the neuronal activity of individual neurons during retrieval. The MTL is crucially involved in the retrieval of previously acquired memories: brief local electrical stimulation of the human MTL during retrieval leads to severe retrieval deficits (Halgren et al., 1985). Two fundamental components of an episodic memory are whether the stimulus is familiar and if it is, whether information is available as to when and where the stimulus was previously experienced (e.g., recollection). How these components interact, however, is not clear. A key question is whether there are distinct anatomical structures involved in these two processes (familiarity vs. recollection).

Some have argued that the hippocampus is exclusively involved in the process of recollection but not familiarity (Eldridge et al., 2000; Yonelinas, 2001). Evidence from behavioral studies with lesion patients, however, seems to argue against this view (Manns et al., 2003; Stark et al., 2002; Wais et al., 2006). Rather than removing the capability of recollection while leaving recognition (familiarity) intact, hippocampal lesions cause a decrease in overall memory capacity rather than the loss of a specific function. Lesion studies, however, do not allow one to distinguish between acquisition vs. retrieval deficits.

Recollection of episodic memories is difficult to study in animals (but see (Hampton, 2001)) but can easily be assessed in humans. Recordings from humans offer the unique opportunity to observe neurons engaged in the acquisition and retrieval of episodic memories. We recorded from single neurons in the human hippocampus and amygdala during retrieval of episodic memories. We used a memory task that enabled us to determine whether a stimulus was only recognized as familiar or whether an attribute associated with the stimulus (the spatial location) could also be recollected. We hypothesized that the neuronal activity evoked by the presentation of a familiar stimulus would differ depending on whether the location of the stimulus would later be recollected successfully or not. We found that the neuronal activity contains information about both the familiarity and the recollective component of the memory.

4.2 Results

4.2.1 Behavior

During learning, subjects (see Table 4-1 for neuropsychological data) were shown 12 different pictures presented for 4 seconds each (Figure 4-1A). Subjects were asked to remember the pictures they had seen (recognition) and where they had seen them (position on the screen). After a delay of 30 min or 24 h, subjects were shown a sequence of 12 previously seen ("Old") and 12 entirely different ("New") pictures (Figure 4-1B). Subjects indicated whether they had seen the picture before and where the stimulus was when they saw it the first time. We refer to the true status of the stimulus as *Old* or *New* and the subject's response as *Familiar* or *Novel*. With the exception of error trials the two terms are equivalent. Subjects remembered 90 ± 3% of all old stimuli and for 60 ± 5% of those they remembered the correct location (Figure 4-1C). Some subjects were not able to recollect the spatial location of the stimuli whereas others remembered the location of almost all stimuli. For each 30 min retrieval session, we determined whether the patient exhibited, on average, above chance (R^+) or at chance (R^-) spatial recollection and then calculated the behavioral performance separately (Figure 4-1D,E). Patients with good same-day spatial recollection performance (30 min R^+) remembered the spatial location of on average 77±6% (significantly different from 25% chance, $p < 0.05$, z-test) of stimuli they correctly recognized as familiar whereas at-chance patients (30 min R^-) recollected only 35±4% of stimuli (approaching but not achieving statistical significance, $p = 0.07$). There were thus two behavioral groups for the 30 min delay: one with good and one with poor recollection performance.

We also tested a subset of the subjects that had good recollection performance on the first day with an additional test 24 h later (4 subjects). Subjects saw a new set of pictures and were asked to remember them overnight. Overnight memory for the spatial location was good (66±1%, $p < 0.05$). All 3 behavioral groups (30 min R^+, 30 min R^-, 24 hr R^+) had good recognition performance (Figure 4-1E) that did not differ significantly between groups (ANOVA, $p = 0.24$). The FP rate was on average 7±3% and did not differ significantly between groups (ANOVA,$p = 0.37$).

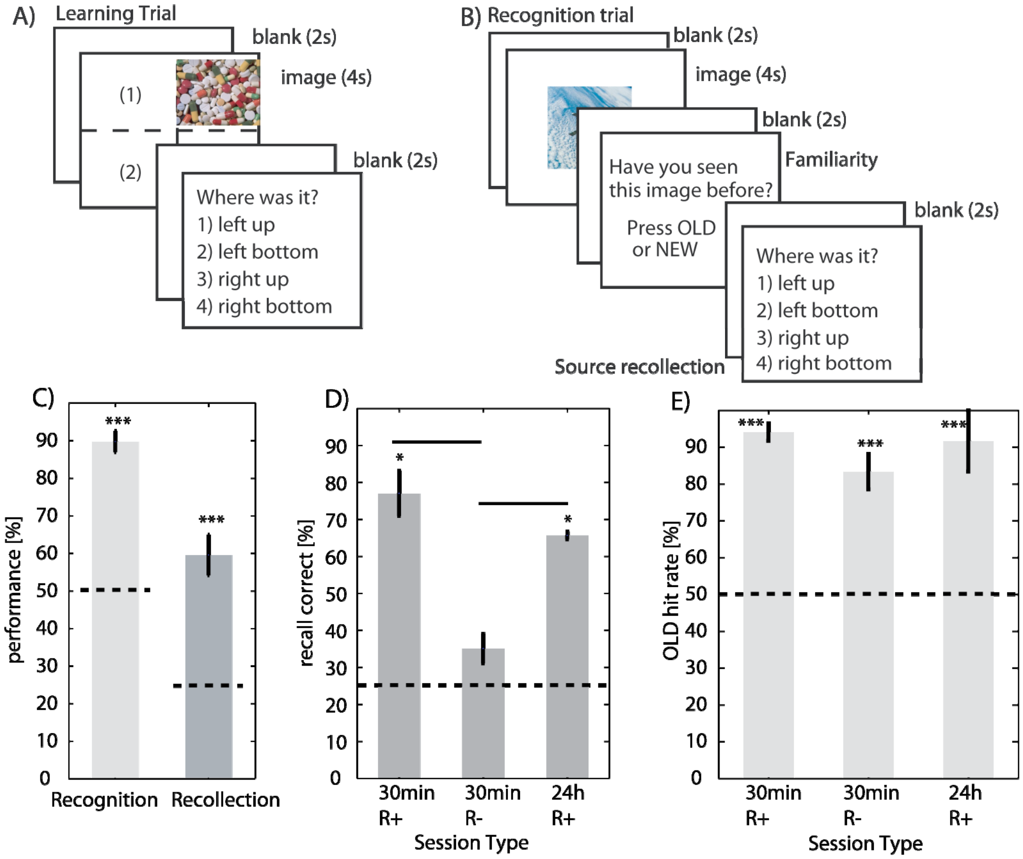


Figure 4-1. Experimental setup and behavioral performance.
The experiment consists of a learning (A) and retrieval (B) block. (C) Patients exhibited memory for both the pictures they had seen (recognition) as well as where they had seen them (recollection). n = 17 sessions. (D) Two different time delays were used: 30 min and 24 h. 30min delay sessions were separated into two groups according to whether recollection performance was above chance or not. (E) For all groups, patients had good recognition performance for old stimuli, regardless of whether they were able to successfully recollect the source. n = 7,5,4 sessions, respectively. Errors are ± s.e.m. Horizontal lines indicate chance performance. R^+ = above chance recollection, R^- at chance recollection.

4.2.2 Single-unit responses during retrieval

We recorded the activity of 412 well separated units in the hippocampus (n = 218) and amygdala (n = 194) in 17 recording sessions from 8 patients (24.24±11.51 neurons (±s.d.) per session). The mean firing rate of all neurons was 1.45±0.10 Hz and was not significantly different between the amygdala and the hippocampus (Figure 4-5A). For each neuron we determined whether its firing differed significantly in response to correctly recognized old vs. new stimuli. Note that “old” indicates that the subject has seen the image previously during the learning part of the experiment. Thus, the difference between a novel and old stimulus is only a single stimulus presentation (single-trial learning). We found a subset of neurons (114, 6.7±4.7 per session, see Table 4-2) that contained significant information about whether the stimulus was old or new. Because error trials were excluded for this analysis, the physical status (old or new) is equal to the perceived status (familiar or novel) of the stimulus. Neurons were classified as either familiarity (n = 37) or novelty detectors (n = 77) depending on the stimulus category for which their firing rate was higher (see methods). The analysis presented here is based on this subset of neurons. The mean firing rate of all significant neurons (1.6±0.2Hz, n=114) did not differ significantly from the neurons not classified as such (1.4±0.1Hz, n = 298). Similarly, the mean firing rate of neurons that increase firing in response to novel stimuli was not different from neurons that increase firing in response to old stimuli (Figure 4-5C,D).

The response of a neuron that increased firing for new stimuli is illustrated in Figure 4-2A–C. This neuron fired on average 1.1±0.2 spikes/s when a new stimulus was presented and only 0.6±0.1 spikes/s when a correctly recognized, old stimulus was presented (Figure 4-2C). Of the 10 old stimuli (2 were wrongly classified as novel and are excluded), 8 were later recollected whereas 2 were not. For the 8 later recollected items (R+) the neuron fired significantly less spikes than for the not recollected items (0.5±0.1 v. 0.9±0.3, $p < 0.05$, Figure 4-2C). Thus, this neuron fired fewer spikes for items which were both recollected and recognized than for items which were not recollected. We found a similar, but opposite pattern for neurons that increase their firing in response to old stimuli (see below). We thus hypothesized that these neurons represent a continuous gradient of memory strength: the stronger the memory, the more spikes that are fired by familiarity-detecting neurons (Figure 4-2D). Similarly, we hypothesized that the opposite relation would hold for novelty neurons: the fewer spikes, the stronger the memory.

We analyzed 3 groups of sessions separately: Same day with good recollection performance (30 min R^+), same day with at chance recollection performance (30 min R^-) and overnight with above-chance recollection (24 h R^+). Sessions were assigned to the 30 min R^+ or 30 min R^- groups based on behavioral performance. We hypothesized that if the neuronal firing evoked by the presentation of an old stimulus is purely determined by its familiarity, the neuronal firing should not differ between stimuli which were only recognized and stimuli which were also recollected. On the other hand, if there is a recollective component, then a difference in firing rate should only be observed for recording sessions in which the subject exhibited good recollection performance.

First we examined the novelty (Figure 4-2E) and familiarity neurons (Figure 4-2F) in the 30 min R+ group. The pre-stimulus baseline was on average 1.7±0.4 Hz (range 0.06–9.5) and 2.6±1.0 Hz (range 0.2–12.9) for novelty and familiarity neurons, respectively, and was not significantly different. Units responding to novel stimuli increased their firing rate on average by 58±5% relative to baseline. Similarly, units responding to old stimuli increased their firing by 41±8% during the second stimulus presentation. We divided the trials for repeated stimuli into two classes: stimuli that were later recollected (R+) and not recollected (R-). A within-neuron repeated measures ANOVA (factor trial type: new, R- or R+) revealed a significant effect of trial type for both novelty ($p < 1e-12$) as well as familiarity units ($p < 1e-6$). This test assumes that neurons respond independently from each other. For both types of units we performed two planned comparisons: i) New vs. R- and ii) R- vs. R+. For novelty neurons, the hypothesis was that the amount of neural activity would have the following relation: New > R- and R- > R+. For familiarity, the hypothesis was the opposite: New < R- and R- < R+ (Figure 4-2D). For novelty as well as familiarity neurons, each prediction proved to be significant (one-tailed t-test. Novelty: New vs. R- $t = 4.3$, $p < 1e-4$ and R- vs. R+ $t = 2.2$, $p = 0.01$. Familiarity: New vs. R- $t = -1.7$, $p = 0.05$ and R- vs. R+ $t = -2.0$, $p = 0.02$). Thus both novelty- and familiarity-detecting neurons signaled that a stimulus is repeated even in the absence of recollection (New vs. R-) and whether a stimulus was recollected or not (R- vs. R+).

The same analysis applied to the remaining groups (30 min R- and 24 h R+) revealed a significant main effect of trial type for novelty ($p < 1e-4$ and $p < 1e-5$, respectively) as well as familiarity neurons ($p < 0.001$ and $p < 0.001$, respectively). However, only the New vs. R- planned comparison was significant (Novelty: $p < 0.001$ and $p < 0.001$; Familiarity: $p < 0.001$ and $p < 0.001$) whereas the R- vs. R+ comparison was not significant for either group (Novelty: $p = 0.6$ and $p = 0.7$; Familiarity: $p = 0.68$ and 0.49). Thus, the activity of these units was different for new vs. old stimuli but the response to old items was indistinguishable for recollected vs. not recollected stimuli.

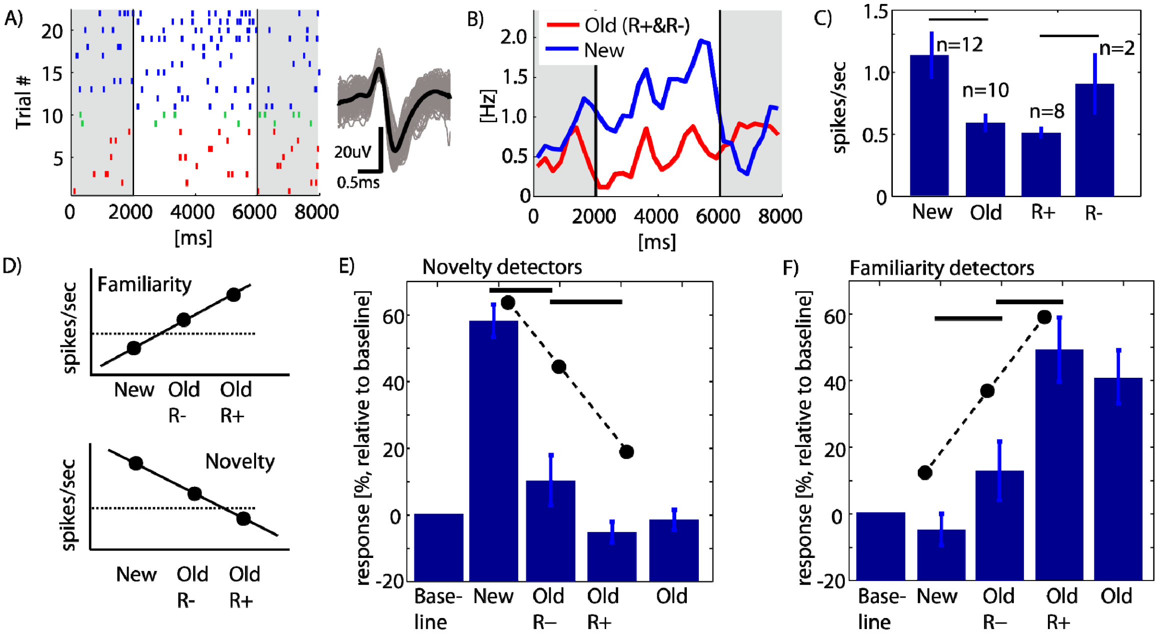


Figure 4-2. Single cell response during retrieval.
(A–C) Firing of a unit in the right hippocampus that increases its firing in response to new stimuli that were correctly recognized (*novelty detector*). (A) Raster of all trials during retrieval and the waveforms associated with every spike. Trials: New (blue), old and recollected (red, R+) and old and not recollected (green, R-). (B) PSTH. (C) Mean number of spikes after stimulus onset. Firing was significantly larger in response to new stimuli and the neuron fired more spikes in response to stimuli which were later not recollected compared to stimuli which were recollected. (D) The hypothesis: the less novelty neurons fire, the more likely it is that a stimulus will be recollected. The more familiarity-detecting neurons fire, the more likely it is that a stimulus will be recollected. The dashed line indicates the baseline. (E–F) Normalized firing rate (baseline = 0) of all novelty (E) and familiarity-detecting (F) neurons during above-chance sessions (30 min R+). Novelty neurons fired more in response to not recollected items (R-) whereas familiarity neurons fired more in response to recollected items (R+). Errors are ±s.e.m. nr of trials, from left to right, 388, 79, 259, 338 (E) and 132, 31, 96, 127 (F).

4.2.3 Quantification of the single-trial responses

Both groups of neurons distinguished recollected from not recollected stimuli, but the difference was of opposite sign. In the novelty case, neurons fire less for recollected items (Figure 4-2E) whereas in the familiarity case neurons fire more (Figure 4-2F). We thus hypothesized that both neuron classes represent a continuous gradient of memory strength. In one case, firing increases with the strength of memory (familiarity detectors) whereas in the other case firing decreases with the strength of memory (novelty detectors). Thus, a strong memory (R+) is signaled both by strong firing of familiarity units as well as weak firing of novelty neurons. Weak memory (R-) is signaled by moderate firing of familiarity and novelty neurons. No memory (a new item) is signaled by strong firing of novelty detectors and weak firing of familiarity detectors. Another feature of the response is that it is often bimodal (see also Figure 4-6). For example, familiarity neurons do not only increase their firing for old items but also decrease firing to new items (Figure 4-2F). This pattern can also be observed in the firing pattern shown in Figure 4-2A: Immediately after stimulus onset, this neuron reduces its firing if the stimulus is old.

We developed a response index R(i) that takes into account the opposite sign of the gradient for the two neuron types, the bimodal response as well as different baseline firing rates. This index makes use of the entire dynamic range of each neuron's response. R(i) is equal to the number of spikes fired during a particular trial i, minus the mean number of spikes fired to all new stimuli divided by the baseline (Eq 1). For example, if a neuron doubles its firing rate for an old stimulus and remains at baseline for a novel stimulus the response index would equal 100%.

By definition, R(i) is negative for novelty units and we thus multiplied R(i) by -1 if the unit was previously classified as a novelty unit.

First, we describe the response of the 30 min R^+ group. In terms of the response index, the average response was significantly stronger to presentation of old stimuli that were later recollected when compared to stimuli which were later not recollected. This was true for a pairwise comparison for every neuron (Figure 4-3A, 68% vs. 50%, n = 45 neurons from 4 subjects) as well as for a trial-by-trial comparison (Figure 4-3B, 67% vs. 45%, $p < 0.01$, n = number of trials). Note that the same difference exists if neurons from the hippocampus (n = 30, R+ vs. R-, $p < 0.05$) or the amygdala (n = 15, R+ vs. R-, $p < 0.05$) are considered separately (see Figure 4-7A and Table 4-2). The difference in response (of 22%) is entirely due to recollection of the source. Re-plotting the data as a cumulative distribution function (cdf) shows a shift of the entire distribution due to recollection (Figure 4-3C, green vs. red line; $p \leq 0.01$). The cdf shows the proportion of all trials that are smaller than a given value of the response index. It illustrates the entire distribution of the data rather than just its mean. We also calculated the response index for correctly identified new items. By definition the mean response to novel stimuli is 0, but it varies trial-by-trial (blue line). The shift in response induced by familiarity alone (blue vs. green, $p \leq 10\text{-}5$) lies in between the shift induced by comparing novel stimuli with old stimuli that were successfully recollected (Figure 4-3C, blue vs. red, $p \leq 10\text{-}19$). The response index is thus a continuous measure of memory strength. From the point of view of this measure, novel items are distractors and old items are targets. We fitted normal density functions to the three populations (distractors, R- and R+ targets). R+ targets showed a greater difference from the distractors than R- targets (Figure 4-3D).

Is there a significant difference between recollected and not recollected stimuli for patients whose behavioral performance was near chance levels? We found that the mean response to recollected and not recollected stimuli did not differ (Figure 4-3E,F. 45% vs. 46%, p = 0.93). This is further illustrated by the complete overlap of the distribution of responses to R+ and R- stimuli (Figure 4-3F, p = 0.53). (This is also true if hippocampal neurons are evaluated separately, Figure 4-7). Thus, the difference (22%) associated with good recollection performance was entirely abolished in the subjects with poor recollection memory.

Was the neuronal response still enhanced by good recollection performance after the 24 h time delay? Subjects in the 24 h delay group had good recollection performance (66%) that was not significantly different from their performance on the 30 min delay period. Thus, information about the source of the stimulus was available to the subject. Surprisingly, however, we found that the firing difference between recollected and not recollected items was no longer present (Figure 4-3G,H). Firing differed by 59% for recollected items compared to 61% for not recollected items (Figure 4-3G,H. p = 0.81). (This is also true if hippocampal neurons are evaluated separately; Figure 4-7C). This lack of difference between R+ and R- items is in contrast to the 30 min R+ delay sessions, where a difference of 22% was observed.

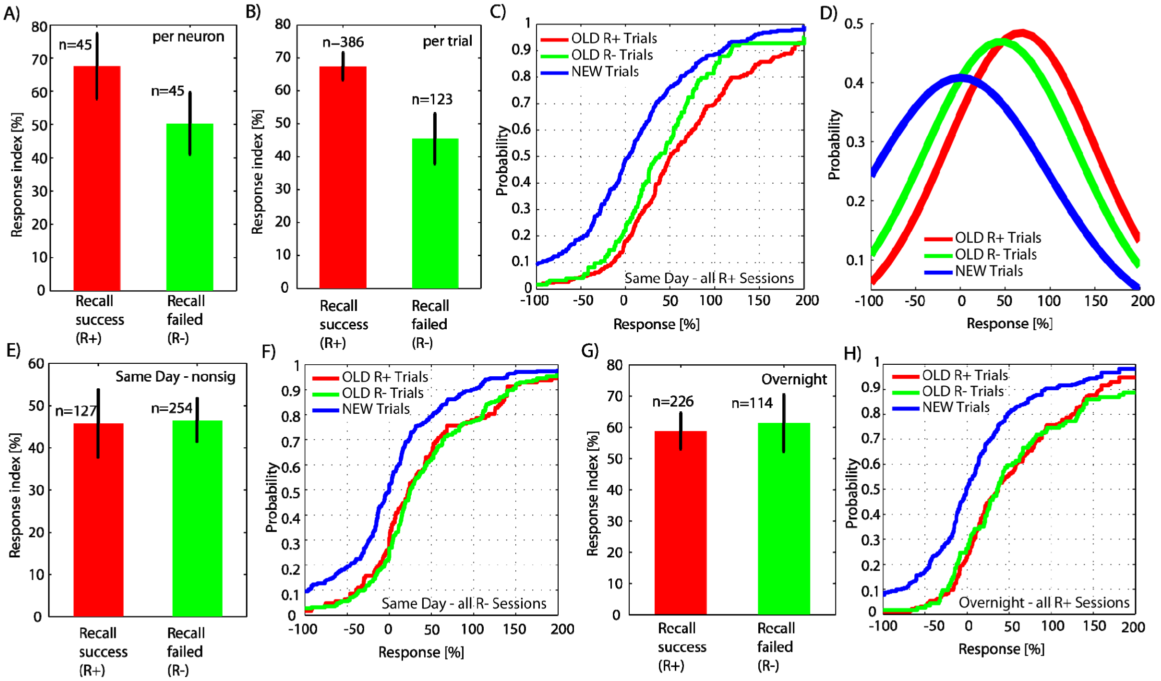


Figure 4-3. Neuronal activity distinguishes stimuli that are only recognized (R-) from stimuli that are also recollected (R+).

(A–E) Same day sessions with above-chance recollection performance (30 min R+). (A) Pairwise comparison of the mean response for all 45 neurons (paired t-test). (B) Trial-by-trial comparison. The response was significantly higher for stimuli which were recalled (R$^+$, n = 386) compared to the response to stimuli which were not recalled (R$^-$, n = 123). n is number of trials. (C) Cumulative distribution function (cdf) of the data shown in (B). The response to new stimuli is shown in blue (median is 0). The shift from new to R$^-$ (blue to green) is induced by familiarity only. (D) Normal density functions showing a shift of R$^+$/ R$^-$ relative to new stimuli. (E–F) Same plots for sessions with chance level performance. There is no significant difference. The cdfs of R$^+$ (n = 127) and R$^-$ (n = 254) overlap completely but are different from the cdf of new trials (blue v. red/green, p < 10-9). (G–H) activity during retrieval 24h later did not distinguish successful (n = 226) from failed (n=114) recollection. Errors are ±s.e.m.

4.2.4 Neural activity during recognition errors

What was the neural response evoked by stimuli that were incorrectly recognized by the subject? Patients could make two different types of recognition errors: i) not remembering an item (false negative, FN) and ii) identifying a new picture as an old picture (FP). Here, we pooled all same-day sessions (13 sessions from 8 patients) regardless of recollection performance. First, we focused on the FNs. We hypothesized that if the neuronal activity truly reflects the behavior, the response should be equal to the response to correctly identified novel stimuli. On the other hand, if the neurons we recorded from represent a general representation of memory strength, we expect to see a response that is smaller than that observed for correctly recognized items. Indeed, we found that the mean response during "forgot" error trials was 14±3% (Figure 4-4A, yellow), significantly different from the response to novel stimuli (Figure 4-4B, blue vs. yellow; p < 10-4, ks-test). It was also significantly weaker when compared to all correctly recognized items (Figure 4-4B, yellow v. green and red, $p \leq 0.05$, ks-test, Bonferonni corrected). What was the response to stimuli which were incorrectly identified as familiar? We hypothesized that if the FPs represent responses that were truly wrongly identified as old (rather than an accidental button press) we would observe a neuronal response that was significantly different from that observed for novel items. Indeed we found that the response to FPs was significantly different from 0 as well as from the response to novel stimuli (Figure 4-4B, blue v. gray; ks-test p = 0.007). The response to FPs and FNs was not significantly different (Figure 4-4B, gray vs. yellow; ks-test, p = 0.14). (For the previous analysis we pooled neurons recorded from the hippocampus as well as the amygdala. The same response pattern holds, however, if hippocampal units are evaluated separately; Figure 4-7D). This pattern of activity during behavioral errors is consistent with the idea that the neurons represent memory strength on a continuum.

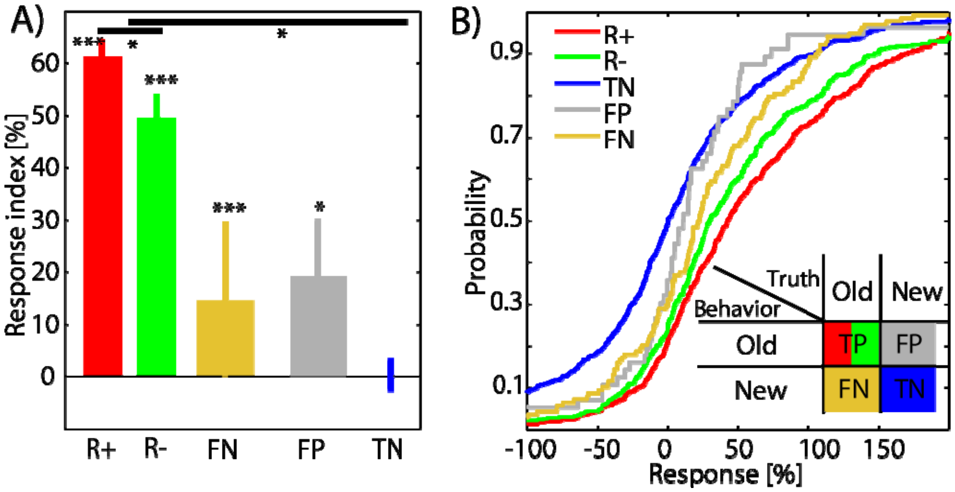


Figure 4-4. Activity during errors reflects true memory rather than behavior. All 30 min sessions are included for this analysis. (A) Neural response. (B) Response plotted as a cdf. Notice the shift from novel to false negatives (p < 10-4): the same behavioral response (novel) leads to a different neural response still differed significantly when compared to real novel pictures. The inset shows the different possible trial types. Errors are ±s.e.m, n is nr of trials (759, 521, 1372, 148, and 56, respectively; 13 sessions, 8 patients).

4.3 Discussion

We analyzed the spiking activity of neurons in the human MTL during retrieval of declarative memories. We found that the neural activity differentiated between stimuli that were only recognized as familiar and stimuli for which (in addition) the spatial location could be recollected. Further, we found that the same neural activity was also present during behavioral errors, but with reduced amplitude. This data is compatible with a continuous signal of memory strength: the stronger the neuronal response, the better the memory. Forgotten stimuli have the weakest memory strength and stimuli which are only recognized but not recollected have medium strength. The strongest memory (and thus neuronal response) is associated with stimuli which are both recognized and recollected.

We used the spatial location of the stimuli during learning as an objective measure of recollection. An alternative measure is the “remember/know” paradigm (Eldridge et al., 2000). However, this measure suffers from subjectivity and response bias. Alternative theories hold that remember/know judgments reflect differences in memory strength rather then different recognition processes (Donaldson, 1996). Thus we chose to use an explicit measure of recollection instead.

We tested 2 different time delays: same day (30 min) and overnight (24 h). Despite good behavioral performance on both days, the neuronal firing only distinguished between R+ and R- trials on the same day. Thus, while the information was accessible to the patient, it was not present anymore in the form of spike counts — at least in the neurons from which we recorded. In contrast, information about the familiarity of the stimulus was still present at 24 hrs and distinguished equally well between familiar and novel pictures (Figure 4-8). While the lack of recordings from cortical areas prevents us from making any definitive claims about this phenomena, it is nevertheless interesting to note that these two components of memory (familiarity and recollection) may be transferred from the MTL to other brain areas with different time courses. Indeed, recent data investigating the replay of spatial sequences by hippocampal units suggest that episodic memories could be transferred to the cortex very quickly. Replay starts in quiet (but awake) periods shortly after encoding and continues during sleep (Foster and Wilson, 2006).

We found that the responses described here can be found both in the hippocampus and the amygdala. Previous human studies have similarly found that visual responses can be found in both areas with little difference (Fried et al., 1997; Kreiman et al., 2000a). Similarly, recordings from monkeys have also identified amygdala neurons which (i) respond to novelty and (ii) habituate rapidly (Wilson and Rolls, 1993). It has long been recognized that the amygdala plays an important role in rapid learning. This is exemplified by its role in conditioned taste aversion (CTA), which is acquired in a single trial, is strongly novelty-dependent, and requires the amygdala (Lamprecht and Dudai, 2000).

The subset of neurons that we selected for analysis exhibited a significant firing difference between old and new stimuli during the stimulus presentation period. This selection criteria allows for a wide variety of response patterns. The simplest case is when a neuron increases firing to one category and remains at baseline for the other. But more complex patterns are possible: the neuron could *decrease* firing for one category and remain at baseline for the other. Or the response could be bimodal, e.g., increase to one category and decrease to the other. To further investigate this, we compared firing during the stimulus period to the pre-stimulus baseline (see supplementary discussion and Table 4-2). 54% of the neurons changed activity significantly for the trial type for which the unit was classified (i.e., old trials for familiarity neurons). 92% of the neurons change their firing rate relative to baseline for either type of trial (e.g., decrease in firing rate of familiarity neurons for new trials). Thus, 38% of the neurons signal information by a significant firing decrease and 8% of the neurons have a bimodal response which individually is not significantly different from baseline. We maintain that the firing behavior of this 8% group contains information about the novelty of the stimulus, even though the responses are not significantly different from baseline. Below we describe several scenarios by which this 8% population might contain decodable information. We repeated our analysis with only the remaining 92% of neurons to assess whether our previous conclusions, based on the entire data-set, still hold true. We found that all results remain valid: The within-repeated ANOVA for the 30 min R+ group revealed a significant difference of New vs. R- as well as R+ vs. R- for both novelty ($p < 1e-4$ and $p = 0.03$, respectively) as well as familiarity units ($p = 0.05$ and $p = 0.02$, respectively). Similarly, the per-neuron ($N = 42$ neurons, $p = 0.03$) as well as the per-trial comparison ($p = 0.01$) remained significant (compare to Figure 4-3A-C). Considering only hippocampal neurons that fire significantly different from baseline, the difference between R+ and R- ($p = 0.04$), R- and New ($p < 0.001$) and New vs. FNs ($p = 0.003$) remained significant (all are tailed ks-tests; compare to Figure 4-7A). All R+ vs. R- comparisons for the 30 min R- and 24 h sessions remained insignificant.

How might a neural network decode the information about a stimulus if it is signaled with no change or a decrease in firing rate? One obvious possibility is by altering excitatory-inhibitory network transmission: if the neuron that signals with a decrease in firing is connected to an inhibitory unit that in turn inhibits an excitatory unit, the excitatory neuron would only fire if the input neuron decreases its firing rate. A similar network could be used to decode information that is present in an unchanged firing rate. How can a network decode information from units that are significantly different new vs. old but not relative to baseline? One possibility is that the network gets an additional input that signals the onset of the stimulus. Thus, it knows which time period to extract. Also, while we can only listen to one single neuron, a readout mechanism gets input from many neurons and can thus read signals with much lower signal-to-noise ratios.

4.3.1 Models of memory retrieval

It is generally accepted that recognition judgments are based on information from (at least) the two processes of familiarity and recollection. How these two processes interact, however, is unclear. Here we have shown that both components of memory are represented in the firing of neurons in the hippocampus and amgydala. Clearly, the neuronal firing described here can not be attributed to one of the two processes exclusively. Rather, the neuronal firing is consistent with both components summing in an additive fashion.

This result has implications for models of memory retrieval. There are two fundamentally different models of how familiarity and recollection interact. The first (i) model proposes that recognition judgments are either based on an all-or-nothing recollection process ("high threshold") or on a continuous familiarity process. Only if recollection fails is the familiarity signal considered (Mandler, 1980; Yonelinas, 2001). An alternative (ii) model is that both recollection as well as familiarity are continuous signals that are combined additively to form a continuous signal of memory strength that is used for forming the recognition judgment (Wixted, 2007). Our data is more compatible with the latter model (ii). We found that the stronger the firing of familiarity neurons, the more likely that recollection will be successful. However, the ability to correctly decode the familiarity of the stimulus does not depend on whether recollection will be successful. This is demonstrated by the single-trial decoding (Figure 4-8): recognition performance only marginally depends on whether the stimulus will be recollected or not. Also, the familiarity of the stimulus can be decoded equally well in patients that lack the ability to recollect the source entirely. Thus, the firing increase caused by recollection is additive and uncorrelated with the familiarity signal. This is incompatible with the high-threshold model, which proposes that either the familiarity *or* the recollective process is engaged. The neurons described here distinguished novel from familiar stimuli regardless of whether recollection was successful. Thus the information carried by these neurons does not exclusively present either index. Rather, the signal represents a combination of both.

4.3.2 Neuronal firing during behavioral errors

What determines whether a previously encountered stimulus is remembered or forgotten? We found that stimuli which were wrongly identified as novel (forgotten old stimuli) still elicited a significant response. Previously we found that this response allows single-trial decoding with performance significantly better than the patient's behavior (Rutishauser et al., 2006a). Thus, information about the stimulus is present at the time of retrieval. This implies the stimuli were (at least to some degree) properly encoded and maintained. However, the neural activity associated with false negative recognition responses was weaker than the responses to correctly recognized but not recollected stimuli (about 60% reduced, Figure 4-4A). The response to false negatives fell approximately in between the response to novel and correctly recognized familiar stimuli (Figure 4-4B). The neuronal response can thus be regarded as an indicator of memory strength. The memory strength for not remembered items is less than for remembered items but it is still larger than zero. However, the memory strength was not strong enough to elicit a "familiar" response. Others (Messinger et al., 2005) have also found neurons that indicate, regardless of behavior, the "true memory" associated with a stimulus. Thus, the neurons considered here likely signal the strength of memory that is used for decision making rather than the decision itself.

False recognition is the mistaken identification of a new stimulus as familiar. The false recognition rate in a particular experiment is determined by many factors, including the individual bias of the subject as well as the perceptual similarity of the stimuli (gist) or their meaning (for words). Here, we found that neurons responded similarly (but with reduced amplitude) to stimuli that were wrongly identified as familiar when compared to truly familiar stimuli. Thus, from the point of view of the neuronal response, the stimuli were coded as somewhat familiar. As such, it seems that the behavioral error possesses a neuronal origin in the very same memory neurons that respond during a correct response — and can thus not be exclusively attributed to simple errors such as pressing the wrong button. MTL lesions result in severe amnesia, measured by a reduction in the TP rate and an increased FP rate relative to controls. However, in paradigms where normal subjects have high FP rates due to semantic relatedness to studied words, amnesics have lower FP rates than controls (Schacter and Dodson, 2001). Thus, in some situations, a functional MTL can lead to more false memory. Similarly, activation of the MTL (and particularly the hippocampus) during false memory has also been observed with neuroimaging (Schacter et al., 1996). This and our finding that neuronal activity does consider such stimuli as familiar suggests that FPs are not due to errors in decision making.

4.4 Methods

4.4.1 Subjects and electrophysiology

Subjects were 10 patients (6 male, mean age 33.7). Informed consent was obtained and the protocol was approved by the Institutional Review Board. Activity was recorded from microwires embedded in the depth electrodes (Rutishauser et al., 2006a). Single units were identified using a template-matching method (Rutishauser et al., 2006b).

4.4.2 Experiment

An experiment consisted of a learning and retrieval block with a delay of either 30 min or 24 h in between. During learning, 12 unique pictures were presented in random order. Each picture was presented for 4 s in one of the 4 quadrants of a computer screen. We asked patients to remember both which pictures they had seen and where on the screen they had seen them. To ensure alertness, patients were asked to indicate where the picture was after each presentation during learning.

In each retrieval session, 24 pictures (12 New, 12 Old, randomly intermixed) were presented at the center of the screen. Afterwards, the patient was asked whether he/she had seen the picture before or not. If the answer was "Old", the question "Where was it?" was asked (see Figure 4-1A). During the task no feedback was given.

4.4.3 Data analysis

A neuron was considered responsive if the firing rate in response to correctly recognized old vs. new stimuli was significantly different. We tested in 2 sec bins (0–2, 2–4, 4–6 s relative to stimulus onset). A neuron was included if its activity was significantly different in at least one of these 3 bins. We used a bootstrap test (p <= 0.05, B = 10000, two-tailed) of the number of spikes fired to New vs. Old stimuli. We assumed that each trial is independent, i.e. the order of trials does not matter. Neurons with more spikes in response to new stimuli were novelty neurons whereas neurons with more spikes in response to Old stimuli were familiarity neurons.

We also used an aggregate measure of activity that pools across neurons. For each trial we counted the number of spikes during the entire 6 s post stimulus period. The response index (Eq 1) quantifies the response during trial i relative to the mean response to novel stimuli.

$$R_i = \frac{nrSpikes_i - mean(NEW)}{mean(baseline)} 100\% \tag{1}$$

R(i) is negative for novelty detectors and positive for familiarity detectors (on average). R(i) was multiplied by -1 if the neuron is classified as a novelty neuron. Notice that the factor -1 depends only on the unit type. Thus, negative R(i) values are still possible.

The cdf was constructed by calculating for each possible value x of the response index how many examples are smaller than x. That is, F(x) = P(X ≤ x) where X is a vector of all response index values.

All statistical tests are t-tests unless stated otherwise. Trial-by-trial comparisons of the response index are Kolmogorov-Smirnov tests (abbreviated as ks-test). All errors are ± s.e. unless indicated otherwise.

4.5 Supplementary results

4.5.1 Behavior quantified with d'

d' was 3.11±0.08, 2.40±0.28 and 2.67±0.68 for the 30 min R^+, 30 min R^- and 24 h groups, respectively. Pairwise tests revealed a significant difference between the 30 min R^+ and R^- group (t-test, $p \leq 0.05$). Thus, in terms of d', patients that exhibited no recollection had significantly lower recognition performance.

4.5.2 Neuronal ROCs

Based on the response values as summarized in Figure 4-3 we constructed two neuronal ROCs (Macmillan and Creelman, 2005): one for trials with spatial recollection and one without (Figure 4-9). The z-transformed ROC was fit well by a straight line (R = 0.997 and R = 0.988 for R+ and R-, respectively). The slope for both curves was significantly different from 1, indicating that the variance of the targets and distractors was different (for a 95% confidence interval the slope was 1.11±0.03 and 1.16±0.07, respectively). The d' for recognized and recollected targets was 0.81 and for targets that were only recognized it was 0.55. Thus, the d' was increased by the addition of recollective information. This is in analogy to the behavioral recognition performance, which was also increased (Figure 4-1E, see above).

Interestingly, the slopes of the neuronal z-ROCs are bigger than 1 (see above). This indicates greater variability for distractors (here new items) compared to familiar items. z-ROC slopes derived from behavioral data are found to be smaller than 1 (Ratcliff et al., 1992).

This has been used as evidence that the target distribution has higher variance compared to the distractor distribution. Intriguingly, we found that the slopes of our z-ROCs are bigger than 1. This further indicates that the neuronal signals in the medial temporal lobe (which we analyze here) represents a memory signal that should be regarded as the input to the decision process, not its output. What is measured behaviorally is the decision itself and it is thus conceivable that the decision process adds sufficient variance to change the slope of the z-ROC.

4.5.3 Responses of novelty and familiarity neurons compared to baseline

The neurons used for our analysis were selected based on a significant difference in firing in response to new vs. old stimuli. This is the most sensitive test because it detects many different patterns in which activity could differ. Example patterns that are detected by this way of classifying units are: i) increase of firing only for one category (new or old) whereas the other remains at baseline, ii) decrease of firing only for one category, with the other remaining at baseline, iii) a bimodal response with an increase to one category and a decrease to the other category. One concern with this analysis is that the response itself might not be significantly different from baseline. This would primarily be the case if the response is bimodal, i.e., a slight increase to one category and a slight decrease to the other. To investigate this possibility we performed additional analysis by comparing the activity of neurons which are classified as novelty or familiarity detecting units against baseline (Table 4-2). We used two different methods: the first ("method 1") tests whether the unit increases its firing rate significantly for either the old (familiarity neurons) or the new trials (novelty neurons). However, there are several classes of units which this method misses. For example, a unit which remains at baseline for old trials and reduces its firing rate for new trials would be classified as a familiarity unit. However, it would not pass the baseline test since the response for old trials remains at baseline. To include such units we used a second method (“method 2”): for a unit to be considered responsive, the activity of either the new or the old trials needs to be significantly different from baseline. The unit in the above example would pass this test.

Using method 2, we found that 92% of all units which were classified as signalling a difference between new and old were in addition also firing significantly different relative to baseline (see Table 4-2 for details). Using method 1, 54% of all units pass this additional test. Thus approximately 40% of the units signal information by a decrease in firing rate rather then an increase.

4.5.4 Population activity

So far we have analyzed the spiking of single neurons which fired significantly different for new vs. old stimuli. However, the majority of neurons (72% of neurons; 298 of 412) did not pass this test and thus were not considered in our first set of analyses. Was there a difference in mean firing between new and old stimuli if neurons were not pre-selected? To address this, we calculated a mean normalized activity for all recorded neurons in all sessions, separately for new and old trials (Figure 4-10A). This signal reflects the overall mean spiking activity of all neurons and is thus similar to what might be measured by the fMRI signal (see discussion). Only trials where the stimulus was correctly recognized were included. The mean firing activity of the entire population was significantly different in the time period from 2–4 s relative to stimulus onset ($p \leq 0.05$, t-test, Bonferroni corrected for $n = 8$ comparisons). Thus, a difference in overall mean activity for novel vs. familiar stimuli can be observed even without pre-selecting neurons. However, the initial response (first 1 s, Figure 4-10A) did not differentiate between the two types of stimuli. Rather, a sharp onset in the response could be observed for both classes of stimuli. Did the population only differentiate because the novelty and familiarity detectors were included in the average? We also calculated the population average (as in Figure 4-10A) using only the units which were not classified as either novelty or familiarity detectors. The average population activity still exhibited a sharp peak for both types of stimuli after stimulus onset and significantly differentiated between novel and familiar items in subsequent time bins ($p \leq 0.05$, t-test, Bonferroni corrected for $n = 8$ comparisons).

Is the population response different for stimuli which are recollected compared to stimuli which are only recognized? The previous average included all old trials, regardless of whether the stimulus was recollected or not. Next, we averaged all trials from all neurons recorded for the 30 min delay sessions with good recollection performance (30 min R^+). We found a similar pattern of population activity (Figure 4-10B). Crucially, however, the neuronal activity in response to familiar stimuli which were later not recollected peaked earlier. Measured in time bins of 500 ms, the only significant difference between familiar stimuli that were recollected or not was in the first 500 ms after stimulus onset ($p \leq 0.05$, t-test, Bonferroni-corrected for $n = 16$ comparisons). Thus, the population activity peaks first for stimuli that are not recollected, followed by novel and recollected stimuli.

4.5.5 Decoding of recognition memory

Is the ability to determine whether a stimulus is old influenced by whether the stimulus was recollected or not? In the main text we have shown that the responses to recollected stimuli are stronger compared to items which are not recollected. Here, we investigate whether this increased response leads to an improvement in the ability to determine (based on the neuronal firing only) whether a stimulus is new or old. If the two types of information (familiarity and recollection) interact, one would expect that the ability to recollect would increase the ability to determine whether a stimulus has been seen before. Alternatively, recollection could be a process that is only triggered after the familiarity is already determined and these two types of information would thus be independent. Thus, one would expect no difference in the ability to determine the familiarity from the spiking of single neurons in cases of successful vs. failed spatial recollection. To answer this question, we used a simple decoder. It used the weighted linear sum of the number of spikes fired after the onset of the stimulus. The weights were determined using regularized least squares, a method very similar to multiple linear regression (see methods). The decoder had access to the number of spikes in the 3 consecutive 2 s bins following stimulus onset (3 numbers per trial).

First, we used the decoder to determine for how many trials we could correctly predict whether the stimulus was new or old, based only on the firing of a single neuron. For all sessions (n = 17), the decoder was able to predict the correct identity for 63 ± 1% of all trials. We repeated this analysis for each of the 3 behavioral groups (R^+ 30 min, R^- 30 min, and R^+ 24 hr). We found (Figure 4-8A) that the recognition decoding accuracy (chance 50%) did not depend on whether the subject was able to recollect the source of the stimulus or not (1-way ANOVA, p =

0.35). Thus, decoding of familiarity is equally effective, even in the group where patients were not able to recollect at all (Figure 4-8A, 30 min R- sessions).

Was there a difference in decoding performance in the same-day group where subjects had good recollection performance? We selectively evaluated the performance of the decoder for two groups of trials: trials with correct recollection and trials with failed recollection. We find that firing during trials with failed recollection does carry information about the familiarity of the stimulus (Figure 4-8B, R-). The ability to predict the familiarity of the stimulus was slightly improved for the behavioral group with good recollection performance on the first day (Figure 4-8B, right. p = 0.03, paired t-test).

4.6 Supplementary discussion

4.6.1 Differences between amygdala and hippocampal neurons

So far, we have analysed neurons recorded from the amygdala and the hippocampus as a single group. We pooled the responses from both groups because we previously found that both structures contain units which respond to novel and familiar items in a very similar fashion (Rutishauser et al., 2006a). Nevertheless we also analyzed the activity separately for both brain structures. We find that the previous finding still holds — while the response magnitude differs, the overall response pattern is very similar. In particular, all primary findings of our paper hold independently for the hippocampus as well as the amygdala (see below).

We found that the increased response to old stimuli which are recollected (R+) compared to stimuli which are not recollected (R-) is present in both hippocampal as well as amygdala neurons (Figure 4-11; 74.8±5.3% v. 61.3±8.6% for the hippocampus and 52.2±6.8% vs. 13.7±14.2% for the amygdala). The response magnitude (comparing all old trials, regardless of whether they are R+ or R-), however, is larger in the hippocampus (71.6±4.5% v. 42.8±6.3%, $p < 0.001$). While the amplitude of the response is different there is nevertheless a significant difference between R+ and R- trials in both areas.

This is further illustrated in Figure 4-7, where we replotted the response to old R+, old R, new, and false negatives (forgotten items) for all 3 behavioral groups only considering hippocampal units (Figure 4-7A–C). The relevant differences (R+ vs. R-, New vs. false negative) are the same as for the pooled responses (see Figure 4-7 legend for statistics). Similarly, the responses during the error trials (false negatives and false positives) are the same (compare Figure 4-7D to Figure 4-4B).

We also repeated the within-group ANOVA for only the hippocampal units of the 30min R+ session. The ANOVA was significant for novelty ($p = 4.1e-6$) as well as familiarity ($p = 1.3e-19$) units. The planned contrasts of R- v.s New and R+ vs. R- revealed a robust difference for novelty ($p = 5.1e-5$ and $p = 0.04$, respectively) units. For familiarity units, the R- vs. New contrast was significant ($p = 0.002$) whereas the R+ vs. R- contrast was only approaching significance ($p = 0.17$). This is because there were only 7 familiarity units that contribute to this comparison. Repeating the same comparisons while excluding all units that do not fire significantly different from baseline (see Table 4-2) reveals a similar pattern: the ANOVA for familiarity units remains unchanged (all units different from baseline) whereas the novelty units ANOVA still shows a significant difference between R- vs. New (p = 2.7e-5) as well as R+ vs. R- (p = 0.016).

4.6.2 Differences between epileptic and non-epileptic tissue

Was the neuronal response reported here influenced by changes induced by disease? All subjects for this study have been diagnosed with epilepsy and as such some of the effects may not extend to the normal population. Behaviorally, our subjects were comparable to the normal population (see Table 4-1). Also, we separately analyzed a subset of neurons which were in a non-epileptic region of the subject's brain. We found a comparable (but stronger) response to old stimuli in this "healthy" neuron population (Figure 4-11D). Similarly, we find that neurons from the "to be resected" tissue still exhibited a response to old stimuli (Figure 4-11E). This response was, however, weaker and there was no significant difference between recollected and not recollected stimuli. Thus, it is possible that the average difference between recollected and not recollected items in normal subjects will be larger than that observed in the epileptic patients in our study.

4.6.3 Relationship to previous single-cell studies

A previous human single-cell study (Cameron et al., 2001) concluded that the neuronal activity observed during retrieval is due to recollection. The task used was the repeated presentation of word pairs with later free recall and thus included no recognition component. Due to the choice of words and the repeated presentation of the same word pairs, the novelty/familiarity of the stimuli was not controlled for. It is thus not clear whether the activity observed was related to recollection or to the recognition of the familiarity of the stimuli. Here, we combine both components in the same task and thus demonstrate that the same neurons represent information about both aspects of memory simultaneously. Similar paired associates tasks have been used with monkeys (Sakai and Miyashita, 1991; Wirth et al., 2003). Changes in neuronal firing were, however, only observed after many learning trials (> 10). A neuronal correlate of episodic memory requires changes after a single learning trial. It thus seems possible that this study documented the gradual acquisition of well-learned associations rather than episodic memories.

4.6.4 Relationship to evoked potentials

Both surface and intracranial evoked potentials show prominent peaks in response to new stimuli. Scalp EEG recordings during recognition of previously seen items show an early frontal potential (~ 300 ms) which distinguishes old from new items, as well as a late potential (~ 500–600 ms) that is thought to reflect the recollective aspect of retrieval (Rugg et al., 1998). However, the signal origin of these scalp recordings is not known. These differences between evoked potentials in response to new and old items are reduced or absent in patients with hippocampal sclerosis (Grunwald et al., 1998). Intracranial EEG recordings from within the hippocampus as well as the amygdala show prominent differences between new and old items (around 400–800 ms) (Grunwald et al., 1998; Mormann et al., 2005; Smith et al., 1986), further suggesting the MTL as a potential source for the scalp signal. The latencies and nature of these potentials are also in agreement with the average population activity that we have analyzed (Figure 4-10). We find that the peak activity is within the 500–1000 ms timeframe (Figure

4-10B). Remarkably, the activity peaks first (within the first 500 ms) if recollection fails. If recollection is successful, the peak is in the second bin (500–1000 ms). This suggests that a recognition judgment based purely on familiarity occurs quicker. In addition, it is worth noting that the average population activity we recorded is compatible with the previous intracranial EEG findings but conflicts with BOLD signals obtained by others (Eldridge et al., 2000; Yonelinas et al., 2005) .

4.6.5 Relationship to fMRI studies

This is also in apparent conflict with previous functional magnetic resonance imaging (fMRI) findings (Eldridge et al., 2000; Yonelinas et al., 2005) that identified regions within the MTL that are selectively activated only for memories that are recollected. Crucially, however, these studies assumed *a priori* that model (*i)* above is correct by searching for brain regions which correlate with the components identified by that model. If model (*i)* is not correct, however, these results are subject to alternative interpretation. Also, these studies used the “remember/know” paradigm to identify memories which were recollected by the subjects. However, this paradigm requires a subjective decision (yes/no) as to whether the memory was recollected or not (as discussed above). It is thus possible that the brain areas identified using these paradigms reflect the decision taken about the memory rather than the retrieval process itself. In our study, no decision as to whether or not recollection succeeded was necessary. Also, our data analysis makes no assumptions about the validity of any particular model.

What is the appropriate baseline activity to consider in the MTL? The MTL is highly active during quiet rest. In fact it is often more active during rest than during memory retrieval

(Stark and Squire, 2001). Imaging studies can suffer from this undefined baseline and results may vary owing to different choices of representative baseline activity (Stark and Squire, 2001). This may also contribute to the apparently disparate findings regarding the involvement of the MTL in recognition memory.

To further investigate the discrepancy between fMRI and single-cell studies, we averaged the neuronal activity of all neurons recorded regardless of their behavioral significance, to approximate a signal that might be similar to an fMRI signal (Figure 4-10, see Results). We found that even under this condition, the overall population activity successfully distinguished between new and old items. The response to old items was not selective for recollected items and was clearly present even if the failed recollected trials were considered separately (Figure 4-10B). Clearly these data differ from previously measured hippocampal BOLD signals (e.g. (Eldridge et al., 2000)).

4.7 Supplementary methods

4.7.1 Electrophysiology

All patients were diagnosed with drug-resistant temporal lobe epilepsy and implanted with intracranial depth electrodes to record intracranial EEG and single units. Electrodes were placed based on clinical criteria. Electrodes were implanted bilaterally in the amygdala and hippocampus (4 electrodes in total). Each electrode contained 8 identical microwires, one of which we used as ground. We were able to identify single neurons in the hippocampus and/or amygdala in 9 of the 10 patients. One additional patient was excluded because he had no recognition memory (performance was at chance). Thus, this study is based on

8 patients (6 of which overlap with a previous study; (Rutishauser et al., 2006a)). We recorded a total of 21 retrieval sessions from these 8 patients. 4 of these sessions (from 4 different patients) were excluded due to insufficient recognition performance (see below). Thus, this study is based on 17 retrieval sessions from 8 different patients. The 17 retrieval sessions were distributed over 16 different days (on one day, 2 retrieval sessions were conducted). We recorded from 24–32 channels simultaneously (3 or 4 electrodes) and found, on average, 11.9±4.4 (±s.d.) active microwires (counting only microwires with at least one well-separated unit). The average number of identified units per wire was 2.0±1.0 (± s.d.). Inactive wires (no units identified) are excluded from this calculation (77 of 280). There were 130 wires with more than one unit (on average 2.6±0.8 for all wires with > 1 unit). For those wires, we quantified the goodness of separation by applying the projection test (Rutishauser et al., 2006b) for each possible pair of neurons. The projection test measures the number of standard deviations the two clusters are separated after normalizing the data such that each cluster is normally distributed with a standard deviation of 1 (see (Rutishauser et al., 2006b) for details). We found that the mean separation of all possible pairs (n=315) is 13.68±6.98 (± s.d.) (Figure 4-12A). We identified, in total, 412 well-separated single units. We quantified the quality of the unit isolation by the percentage of all interspike intervals (ISI) which are shorter than 3 ms. We found that, on average, 0.3±0.4 percent of all ISIs were below 3ms (Figure 4-12B). The signal-to-noise ratio (SNR) of the mean waveforms of each cluster relative to the background noise was on average 2.4±1.2 (Figure 4-12C).

For the purpose of comparing only neurons from the "healthy" brain side (left or right), we excluded all neurons from either the left or right side of the patient if the patient’s diagnosis (Table 4-1) included temporal lobe damage (Figure 4-11). No neurons were excluded if the diagnosis indicated that the seizure focus was outside the temporal lobe.

4.7.2 Behavior

Each session consisted of a learning and retrieval block. We quantified, for each session, the recognition rate (percentage of old stimuli correctly recognized), the false positive rate (percentage of new stimuli identified as old), and the recollection rate. The recollection rate was the percentage of stimuli identified as old for which the spatial location was correctly identified. Sessions with a recognition rate of ≤ 50% were excluded (3 sessions). Each session was assigned to either the 24 h or 30 min delay group.

For each session, we estimated whether spatial recollection rate was significantly different from chance (25%). Due to the small number of trials (maximally 12), the significance was estimated using a bootstrap procedure (see below). Based on this significance value, we further divided each of these two groups into a group with good spatial recollection performance ($p \leq 0.05$, above chance, R^+) and one with poor spatial recollection performance (not significantly different from chance, $p > 0.05$, R^-). For the 24 h group there was only one session with poor recollection performance and thus this analysis was not conducted. Thus, there were 3 behavioral groups which were used for the neuronal analysis: 30 min R^+ ($n = 7$), 30 min R^- ($n = 6$) and 24 h R^+ ($n = 4$). The assignment of sessions to groups was based entirely on behavioral performance. Neuronal activity was not considered.

4.7.3 Data analysis — behavioral

We labeled each retrieval trial during which a correctly recognized old stimulus was presented as either correctly or incorrectly recollected. For each session we then tested (bootstrap, $p \leq 0.05$, one-tailed, $B = 20000$) whether recollection performance was above chance level. We used the bootstrap test instead of the z-test because of the small number of samples. The resulting p values were more conservative (larger) compared to the p values obtained with the z-test. Only sessions which passed this test were considered to have “above chance” recollection performance. Trials which failed this test were considered as "at chance". This was to ensure that only neurons from patients that had a clearly demonstrated capability for source memory were included. Also, recording sessions with less than a 50% hit rate for old stimuli were excluded to ensure that only sessions with sufficient recognition performance were included. We verified for each group of sessions (Figure 4-1) whether performance was significantly above chance using a z-test. For this, we pooled all trials of a particular group and labeled each as either correct or incorrect. Then we used one z-test to test whether the ratio correct:incorrect was above chance. We used this instead of individual tests for each session to avoid artificially boosting performance due to the small sample size (e.g., 4 out of 12 correct) in each particular session.

4.7.4 Data analysis — response index

We compared, trial-by-trial, the response (quantified by the response index) to old stimuli which were successfully recollected (R^+) to old stimuli which were not recollected (R^-). For this comparison, trials with recognition errors were excluded (thus, all trials are familiar).

The error trials were analysed separately. There was one data point for every trial for every neuron (e.g., if there are 10 trials and 10 neurons, there are 100 data points). There were 1368 old stimulus trials (12 retrieval sessions with total 114 neurons), with 1230 trials with a correct recognition response (familiar, TP), and 138 trials which were errors (misses). We analyzed the error trials separately.

We compared the responses of the R^+ and R^- trials with a two-tailed t-test, as well as using a Kolmogorov-Smirnov test. Both were significant at $p \leq 0.05$. Paired comparisons were made with a t-test. Normal density functions were constructed by estimating the mean and standard deviation from the data (using maximum likelihood).

4.7.5 Data analysis — baseline comparison

To determine whether a unit was responsive relative to baseline we compared the firing during the 2 s period in which the new vs. old comparison is significant to the 2 s period before the stimulus onset. These comparisons were performed using a boostrap test as described in the main methods.

4.7.6 Neuronal ROCs

Neuronal ROCs (Figure 4-9) were constructed by considering all trials as old if the response R(i) was above a threshold T. The threshold T was varied in variable steps (see below) from the smallest to the largest value of R(i). Thresholds were varied such that each increase accounted for a 5% quantile of all available datapoints (the 0% and 100% quantiles were excluded). This procedure assured that the same number of datapoints was used for the calculation of each point in the ROC. The hit/false positive rate was calculated for each threshold value. d' was calculated for each pair of hit/false positive rates and averaged. We z-transformed the ROC and fit a line through all points using linear regression to find the slope of the curve. A slope of 1.0 indicates that the two distributions (distractors and targets) are of equal variance whereas a slope of unequal 1.0 indicates a difference in variance. The z transformed ROC was fit well by a straight line for both R^+ and R^- trials (Macmillan and Creelman, 2005).

4.7.7 Population averages

Population averages (Figure 4-6, Figure 4-10) were constructed by normalizing each trial to the mean baseline firing in the 2 s before stimulus onset. The number of spikes were binned into 1 s bins (non-overlapping) and averaged for all neurons. No smoothing was applied. To avoid normalization artifacts, only neurons with a baseline rate of at least 0.25Hz were considered for the population averages (346 of 412 neurons for Figure 4-5). Also, for Figure 4-10 only neurons with a significant response in the stimulus period (first two of the 2 s bins) were considered (this does not apply for the trial-by-trial analysis).

4.7.8 Decoding

We used a linear classifier to estimate how well the firing of a single neuron during a single trial can signal the identity (new or old) of the presented stimulus. The classifier was provided with the number of spikes fired in 3 consecutive 2 s bins after stimulus onset (0–2 s, 2–4 s, 4–6 s). The classifier consisted of a weighted sum of these 3 numbers. The weights were estimated using regularized least squares (RLSC) (Evgeniou et al., 2000; Rifkin et al., 2003). This method is equal to multiple linear regression with the exception of an added regularizer term λ (see below; we used λ = 0.01 throughout). The decoding accuracy of the classifier was estimated using leave-one-out crossvalidation for all training samples available. The estimated prediction error was equal to the percentage of correct leave-one-out trials. There were maximally 12 samples in each class (old or new). However, due to behavioral errors, fewer trials were sometimes available for analysis. Error rates for false positives and false negatives were approximately equal and the number of samples was thus approximately balanced in both classes. Of concern was whether a slight imbalance of the number of samples in one class could bias the results. We performed two controls to assess whether this was the case: we performed leave-one-out cross-validation with the label of the test sample randomly re-assigned with 50% probability. If the classifier was biased, the resulting error would be different from 50%. We found that this was not the case (Figure 4-8A). Also, we re-ran all analysis that used the decoder with a balanced number of samples (that is, equal number of samples in either class) and found no difference in the results.

The weights were determined by regularized least squares. Regularized least squares are very similar to multiple linear regression. In the following we would like to point out these differences because in a previous study we used a multiple linear regression (Rutishauser et al., 2006a).

With multiple linear regression (Eq S1), the weights w are determined by multiplying the inverse of data samples Z with the trainig labels y (Johnson and Wichern, 2002).

$$w = [Z'Z]^{-1} Z'y \qquad \text{(S1)}$$

In contrast, in regularized least squares (Evgeniou et al., 2000; Hung et al., 2005; Rifkin et al., 2003), an additional term is added to the data samples (Eq S2). Here, I is the identity matrix and λ is a scalar parameter (the regularizer).

$$w = \left[Z'Z + \lambda I\right]^{-1} Z'y \qquad \text{(S2)}$$

The value of the regularizer is arbitrary. The bigger it is, the more constraints are placed on the solution (the less the solution is determined by the data samples). A small value of the regularizer, on the other hand, makes the solution close to the multiple linear regression solution. Importantly, however, even a small value of the regularizer punishes unrealistically large weights and also guarantees full rank of the data matrix. Regularization becomes particularly important when there are a large number of input variables relative to the number of training samples. This is the case in our study because each neuron contributed 3 variables (3x 2 s time periods) and the number of training samples was small (on the order of 10). Thus, regularization was necessary. We found that performance was maximal for a small (but non-zero) regularizer and used $\lambda = 0.01$ throughout.

4.8 Supplementary figures

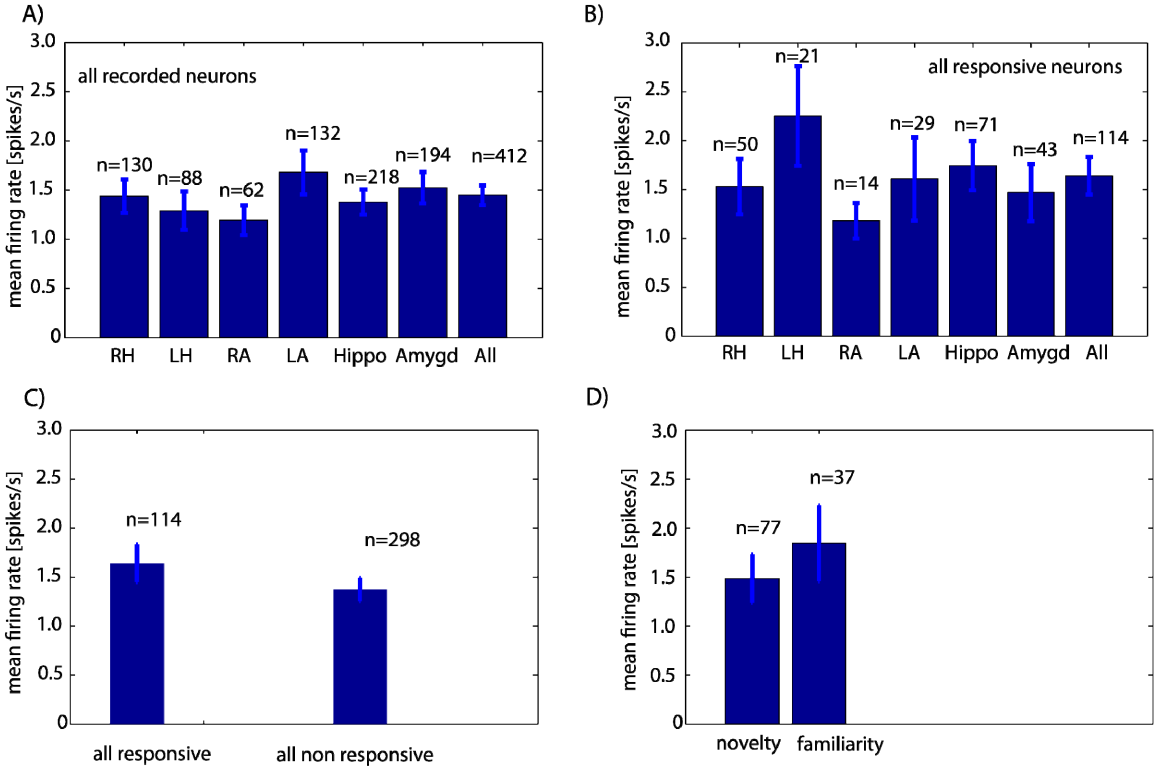


Figure 4-5. Population average of all recorded neurons.
(A) Population average of all recorded neurons that have a baseline firing rate of >0.25Hz (n = 346). While the firing of most neurons was not significantly different between new vs. old, a significant difference between new and old stimuli could still be observed in the population average. Errors are ±s.e.m and ** indicates significance of a one-tailed t-test at $p \leq 0.006$ ($p \leq 0.05$ Bonferonni-corrected for 8 multiple comparisons). (B) Population average of all neurons with recollected and not recollected familiarity trials shown separately. (C) Population average of all neurons recorded in the 30 min delay sessions with above chance recollection performance. The signal for the not recollected items peaked earlier than the signal for recollected items. ** indicates a significant difference between recollect (R^+) and not recollected (R^-) items at $p \leq 0.003$ ($p \leq 0.05$ Bonferonni-corrected for 16 multiple comparisons). The only difference was for the first time bin (0–500 ms after stimulus onset). n = 134 neurons.

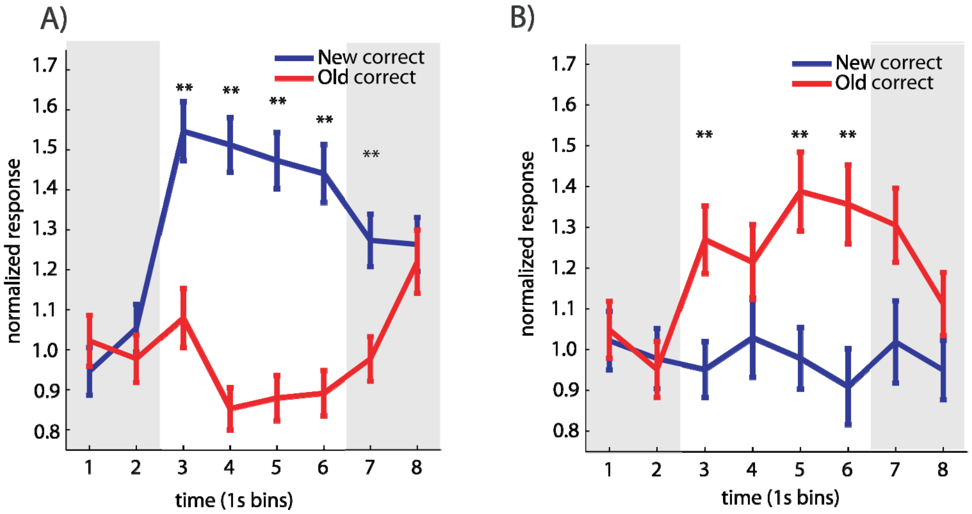


Figure 4-6. Population response.

(A-B) Population average of all neurons that responded significantly during the stimulus period. The stimulus was on the screen during the 4 s period marked in white. **(A)** Average of all neurons that increased firing to correctly recognized new items ("novelty detectors") (*n* = 48). **(B)** Average of all neurons that increased firing to correctly recognized old items ("familiarity detectors") (*n* = 26). Errors are ± SEM and ** indicates significance of a one-tailed *t* test at $P \leq 0.006$ ($P \leq 0.05$ Bonferroni corrected for multiple comparisons). Firing was normalized to the 2 s baseline firing before stimulus onset marked in gray. Note that this does not mean all neurons fired during the entire period; but rather represents the population average.

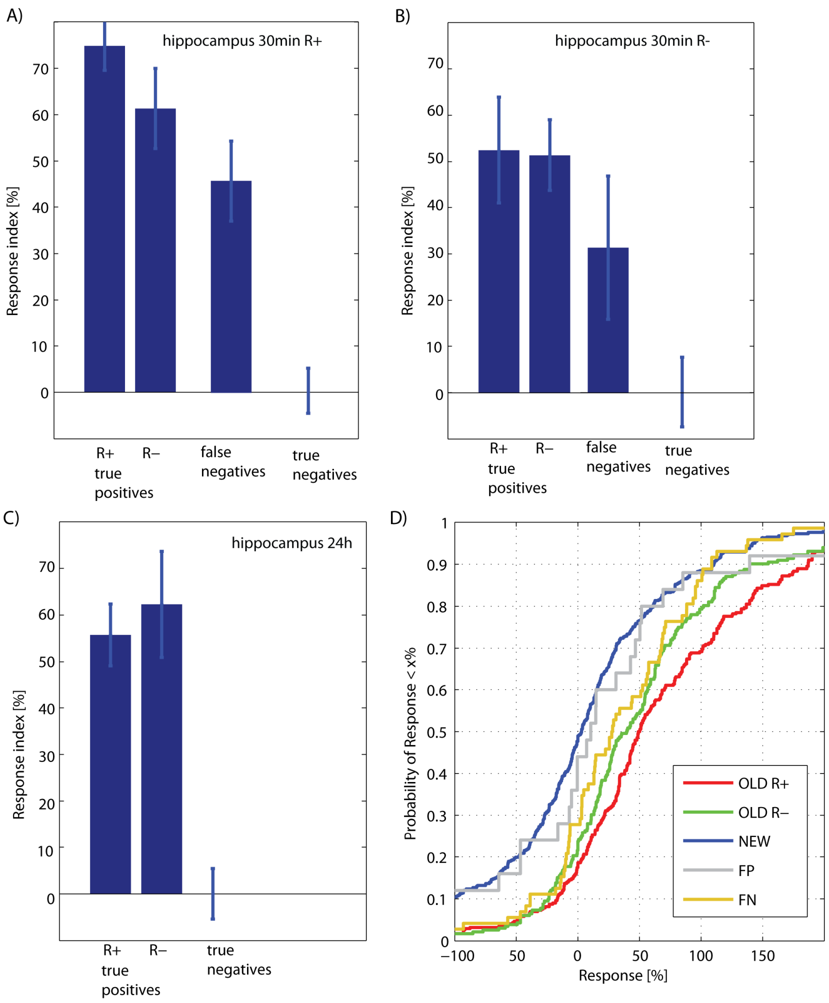


Figure 4-7. A continuous strength of memory gradient exists when the hippocampal neuronal population is considered in isolation.

In this figure, the same measures are replotted, but all units recorded from the amygdala are excluded. All findings remain valid. (A) Trials from the 30 min R+ sessions. There is a significant difference between R+ and R- trials ($P = 0.03$) as well as between new and false negatives ($P = 0.001$). Compare to Figure 4-3C. (B) Trials from the 30 min R- session. There is no significant difference between R+ and R- trials ($P = 0.93$) but false negatives are still significantly different from new trials ($P = 0.07$). Compare to Figure 4-3F. (C) Trials from the 24 h sessions. There is no significant difference between R+ and R- trials. Error trials are not shown (not enough for 24 h sessions). Compare to Fig. 4-3H. (D) cdf of response index of all hippocampal neurons recorded in all 30 min sessions. R+ and R- trials are significantly different (red v. green, $P = 0.01$) as are new and false negatives (blue vs. yellow, $P < 0.001$). Not enough false positive trials are available to allow statistical analysis of false positives. Compare to Fig. 4-4. All errorbars are ± SE.

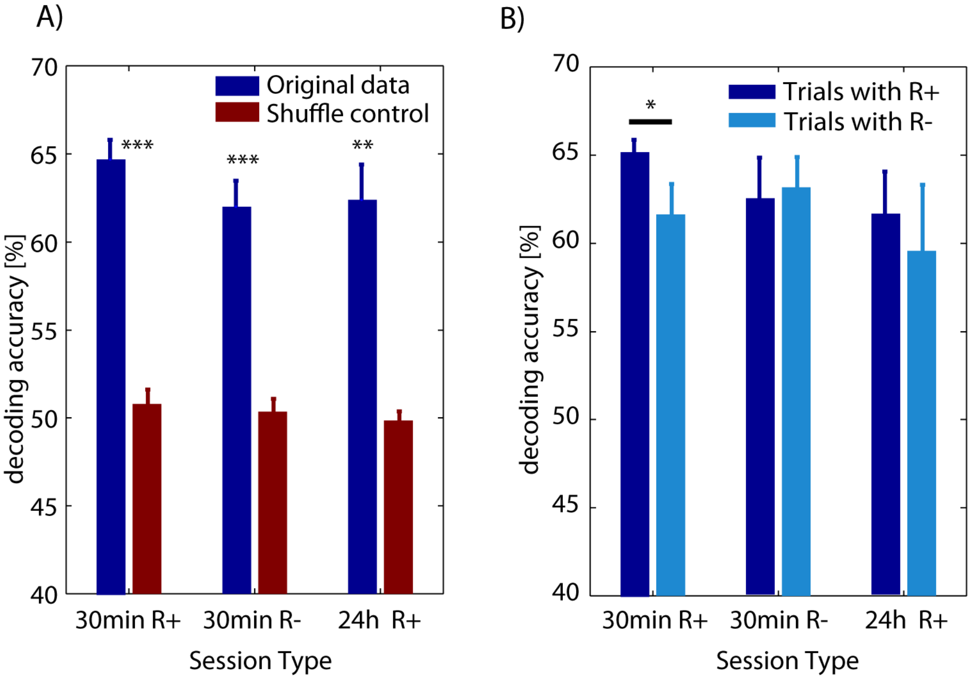


Figure 4-8. Whether a stimulus is new or old can be predicted regardless of whether recall was successful or not.

The decoder had access to the number of spikes fired in the 3 consecutive 2 s bins following stimulus onset (3 numbers total). (A) Session-by-session differences. The performance of the decoder did not change for all 3 groups (ANOVA, $P = 0.35$). n = 7,6,4 sessions, respectively. (B) Trial-by-Trial differences. Here, the decoder was trained on the complete set of trials but its performance was evaluated separately either for failed (R^-) or successful (R^+) recall trials. Clearly, the familiarity of the stimulus could be decoded for trials with failed recall (R^-). In the 30 min delay sessions with successful recall (30 min R^+), firing during successful recall trials contained significantly more information about the familiarity of the stimulus ($P = 0.037$, paired t test, $n = 7$ sessions). All errorbars are ± SE.

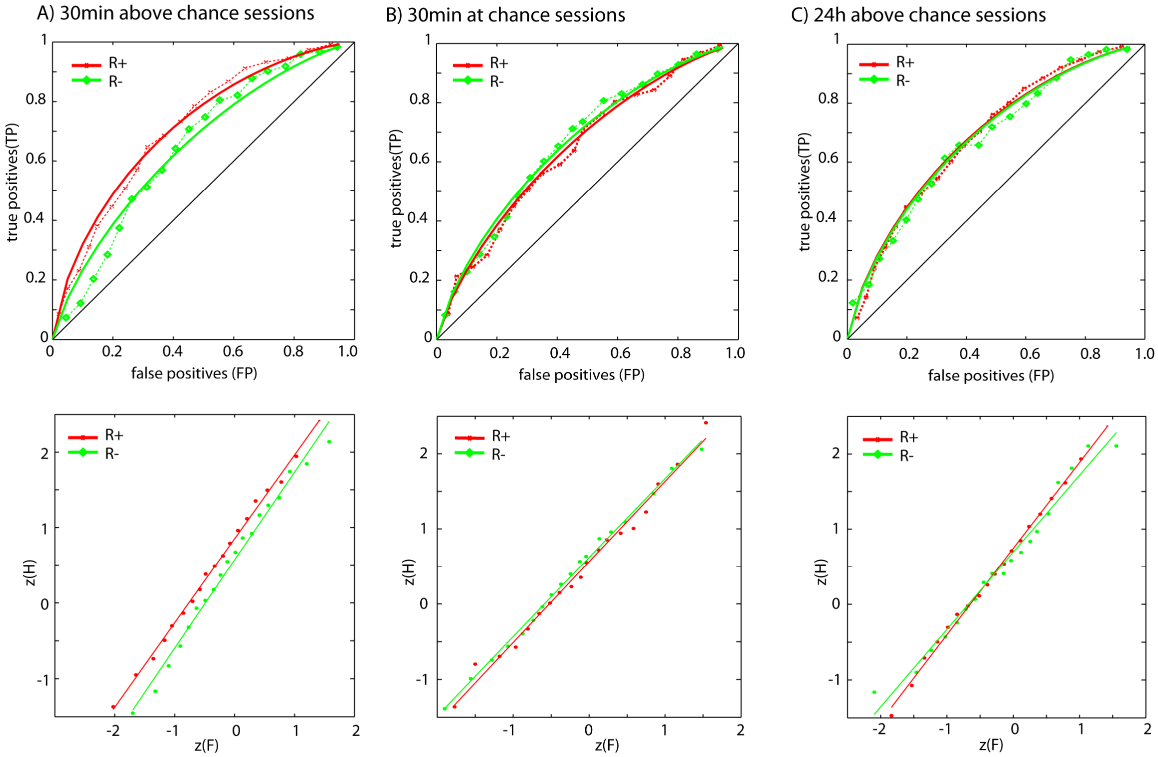


Figure 4-9. ROC analysis of the neuronal data for all 3 behavioral groups. (A: 30 min above chance, B: 30 min at chance, C: 24 h above chance). The top row shows the raw datapoints as well as fits computed from d'. The bottom row shows the same but z-transformed. R^2 is > 0.97 for all straight line fits. See the supplementary methods for how the ROC was computed. A) d' for R$^+$ and R$^-$ groups was 0.81 and 0.55, respectively. The slope (s) of the z-transformed line was 1.11 ± 0.03 and 1.16 ± 0.07, respectively. ± are 95% confidence intervals. B) d' was 0.55 and 0.61 and s was 1.07 ± 0.06 and 1.05 ± 0.04, respectively. C) d' was 0.73 and 0.69 and, was 1.14 ± 0.04 and 1.02 ± 0.08, respectively.

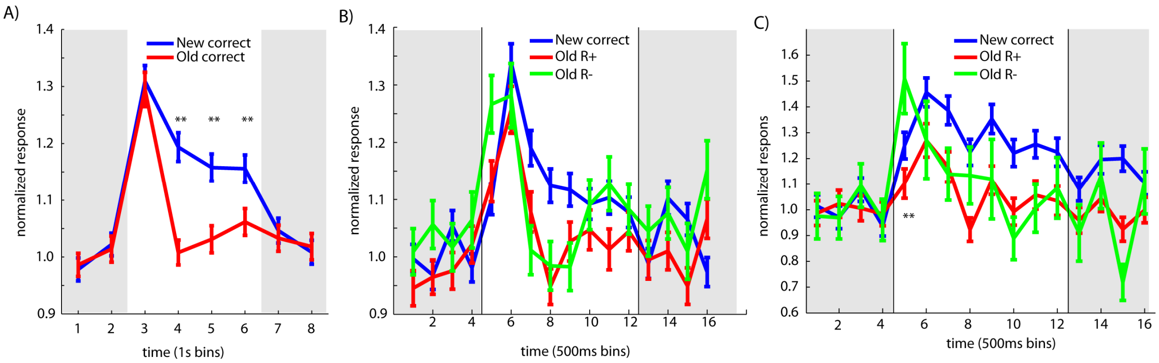


Figure 4-10. Population average of all recorded neurons.
(A) Population average of all recorded neurons that have a baseline firing rate of > 0.25 Hz (*n* = 346). While the firing of most neurons was not significantly different between new vs. old, a significant difference between new and old stimuli could still be observed in the population average. Errors are ± SEM and ** indicates significance of a one-tailed *t* test at $P \leq 0.006$ ($P \leq 0.05$ Bonferonni-corrected for 8 multiple comparisons). (B) Population average of all neurons with recollected and not recollected familiarity trials shown separately. (C) Population average of all neurons recorded in the 30 min delay sessions with above chance recollection performance. The signal for the not recollected items peaked earlier than the signal for recollected items. ** indicates a significant difference between recollect (R^+) and not recollected (R^-) items at $P \leq 0.003$ ($P \leq 0.05$ Bonferonni-corrected for 16 multiple comparisons). The only difference was for the first time bin (0–500 ms after stimulus onset). *n* = 134 neurons.

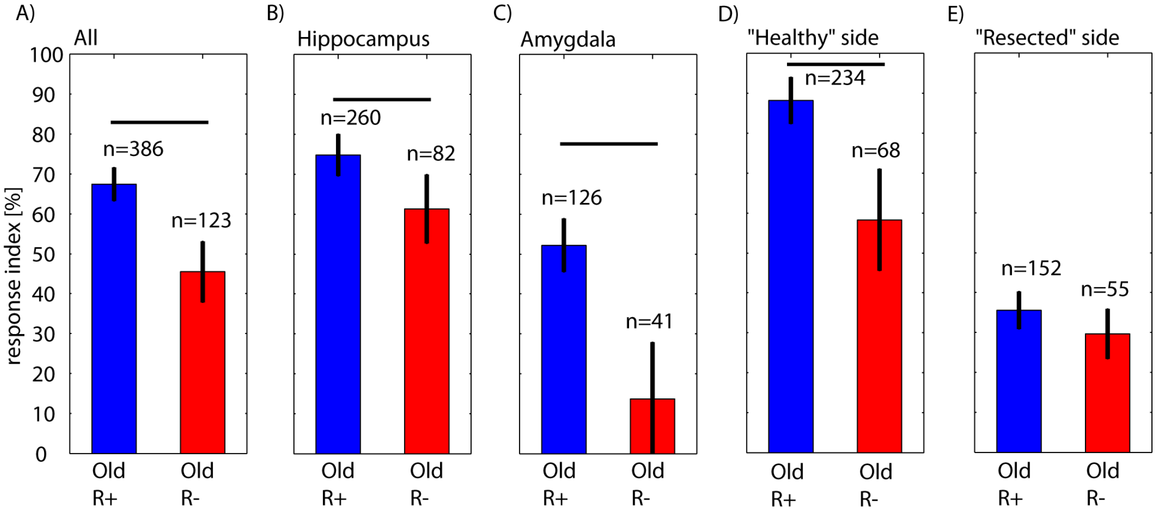


Figure 4-11. Comparison of trial-by-trial response strength for different subcategories of neurons.

In this figure, only neurons from 30 min delay with successful recollection (30 min R+) are included. **(A)** All trials from all areas (same as Figure 3B). **(B)** Only trials from hippocampal neurons. **(C)** Only trials from amygdala neurons. **(D)** Only trials from the "healthy" hemisphere. **(E)** Only trials from neurons in the eventually resected hemisphere. In **(A-D)**, the response to R+ compared to R- trials is significantly different ($P < 0.05$, two-tailed Kolmogorov-Smirnov test, compare to Figure 3B). The response in (E) is not significantly different.

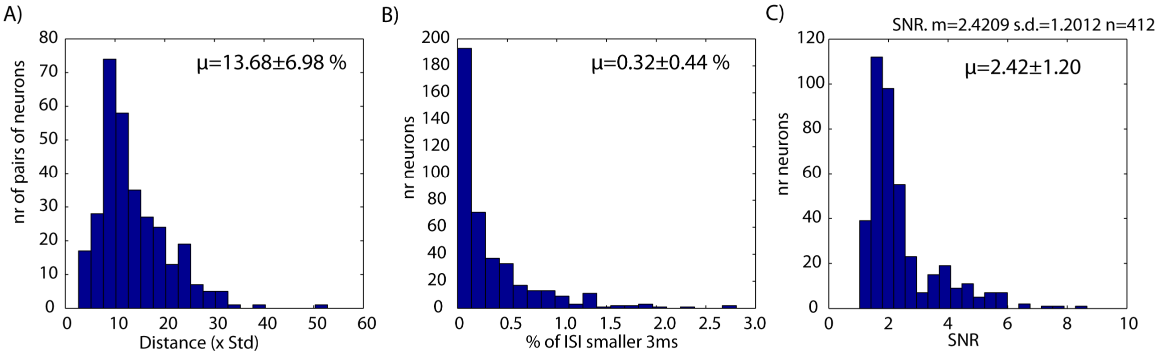


Figure 4-12. Sorting quality for the 412 recorded units.
(A) Histogram of the distance, in standard deviations, between all pairs of clusters. Only channels on which more than one unit was detected are included (315 pairs from 130 channels). The mean distance was 13.68 ± 6.98 (± s.d.) (B) Histogram of the percentage of interspike intervals (ISI) that were shorter than 3 ms. On average 0.32 ± 0.44% of all ISIs were shorter than 3 ms (n = 412). (C) Histogram of the SNR of all 412 units.

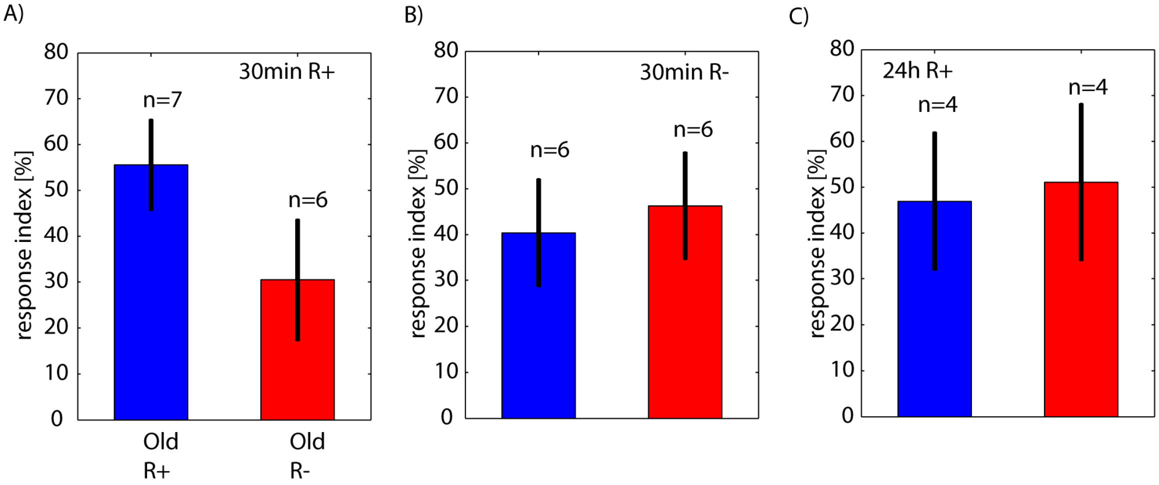


Figure 4-13. Comparison of response strength across different recording sessions (days).

The difference is only significant for the 30 min R+ sessions. The data displayed here is the same as detailed in Figure 4-3. However, here the mean response index for R+ and R- trials is compared between recording sessions. **(A)** The response index for all recording sessions that had above chance recollection. The difference approaches significance (*P* = 0.07). Number of sessions is 7 and 6, respectively (from 4 patients; one session had no R- trials). **(B)** Same as (A) but for all recording sessions with at chance recollection. Number of sessions is 6 for both groups (from 5 patients). There was no significant difference (*P* = 0.63). **(C)** Same as (A) but for all recording sessions with 24 h delay and above chance recollection. Number of sessions is 4 from 3 patients. There was no significant difference (*P* = 0.57). Errorbars are ± SEM with n as specified. p values are from a *t* test.

4.9 Supplementary tables

				WAIS-III			WMS-R					
Patient	**Age**	**Sex**	**Diagnosis**	**PIQ**	**VIQ**	**FSIQ**	**Verbal Mem**	**Mental control**	**VPA 2**	**LM 2**	**Vis Rep 1**	**Vis Rep 2**
1	28	m	left temporal	125	98	110	114	6	4	24	37	39
2	41	f	left temporal	92	91	91	91	5	8	18	37	29
3	20	f	left temporal	92	93	93	83	6	8	16	34	28
4	58	f	left temporal	85	83	83	83	6	4	10	22	7
5	23	m	left temporal & frontal pole	144	111	126	122	6	8	26	39	39
6	44	m	right temporal	76	92	84	83	6	5	10	29	14
7	51	f	left temporal	90	95	93	89	6	4	23	34	34
8	16	m	right lateral frontal	84	91	88	n/a	n/a	8	n/a	31	29
av	*35.1*	-	-	*98.5*	*94.3*	*96.0*	*95.0*	*5.9*	*6.1*	*18.1*	*32.9*	*27.5*
mean raw								5.0±1.2	7.6±0.7	21.9±9.2	32.5±5.3	29.5±7.1

Table 4-1. Neuropsychological evaluation of patients.
Intelligence was measured using the Wechsler Intelligence Scale (WAIS-III) measures of performance IQ (PIQ), verbal IQ (VIQ), and full scale IQ (FSIQ). All IQ scores have an average of 100 (by design). Memory measures are from the Wechsler Memory Scale Revised (WMS-R). Verbal memory is an WMS-R index score with a mean of 100 of the normal population (by definition). The remaining WMS-R scores are raw (unnormalized) scores. For the raw scores, the mean and standard deviation of the normal population (from WMS-R) is shown in the last row for the average age of our population. Abbreviations: Verbal paired associates 2 (VPA 2), Logical Memory 2 (LM 2), Visual Reproduction 1 (Vis Rep 1), Visual Reproduction 2 (Vis Rep 2).

	Group	**Hippocampus**			**Amygdala**			**All**		
Recorded	*30min R+*	77			103			180		
	30min R-	96			47			143		
	24h R+	45			44			89		
	all	218			194			412		
		Nov	**Fam**	**All**	**Nov**	**Fam**	**All**	**Nov**	**Fam**	**All**
New v. old	*30min R+*	25	7	32	10	5	15	35	12	47
	30min R-	11	11	22	13	3	16	24	14	38
	24h R+	11	6	17	7	5	12	18	11	29
	all			71			43	77	37	114
New v. old & baseline 1	*30min R+*	14	5	19	6	3	9	20	8	28
	30min R-	5	6	11	6	1	7	11	7	18
	24h R+	5	4	9	5	2	7	10	6	16
	all			39 (55%)			23 (53%)			62 (54%)
New v. old & baseline 2	*30min R+*	22	7	29	10	5	15	32	12	44
	30min R-	10	10	20	11	3	14	21	13	34
	24h R+	9	6	15	7	5	12	16	11	27
	all			64 (90%)			41 (95%)			105 (92%)

Table 4-2. Number of neurons recorded.
Number of neurons recorded in each area (first row) and number of neurons that responded in each behavioral group(2nd, 3rd, 4th row). The second row shows the number of neurons which had a significantly different firing rate for old vs. new trials during the post-stimulus period (6s). The last two rows show the number of neurons which are, in addition, also significantly different for two different baseline comparisons (1 and 2). The two baseline comparisons are: i) The trials associated with the type of unit are significant from baseline. (That is, if the neuron is classified as a familiarity neuron, the old trials were significantly different from baseline. The same applies for the novelty neurons, but for the new trials). ii) Either the new or the old trials are significantly different from baseline. Note that the first (i) baseline condition is the most restrictive: for example, a familiarity unit that decreases firing to novel items but remains at baseline for familiar items would not pass this test. For the second baseline condition, 92% of units (105 of 114) remain significant. Thus, almost all units fired significantly different from baseline for either the new or old condition. Note that some of the n's reported in the main analysis are slightly lower than the numbers reported in this table. This is because additional constraints were applied (for example, at least one R+ and one R- trial for each included unit).

Chapter 5. Predictors of successful memory encoding

5.1 Introduction

Whether a memory is successfully retrieved or forgotten is determined by many different factors. The first step in establishing a new memory is encoding it. The cellular, molecular, and network processes triggered during encoding set into motion a permanent change that is sufficient to later recall the memory. Many other factors influence this process, such as attention, arousal, consolidation, interference with other memories, sleep, and emotional significance (Paller and Wagner, 2002). Here we asked how much of the retrieval performance can be explained by the neural activity during initial learning. Thus, we are looking for indicators of successful memory encoding.

We recorded single units and LFP data from three areas strongly involved in memory formation: two structures in the MTL (the hippocampus and amygdala) as well as one structure of the cortex (anterior cingulate cortex). Lesions of the MTL produce severe memory deficits (see Chapter 1 for details). Also, hippocampal lesions, in particular, produce deficits in the detection of novelty (Knight, 1996). The successful detection of novelty is a prerequisite for memory formation in many instances (Rutishauser et al., 2006a). While the function of the ACC is poorly understood, it is clear that it has a prominent role in performance monitoring and attention (focus, effort), and it is thus expected that it will also contribute to memory encoding. From animal studies it is known that lesions of the ACC (particularly area 24) severely impair the acquisition of Pavlovian conditioning (Gabriel et al., 1991). Similarly, recordings from the ACC reveal prominent theta oscillations which interact with hippocampal theta, as well as single units (in the cingulate) that modulate their firing relative to hippocampal theta (Colom et al., 1988; Gabriel et al., 1991; Gabriel et al., 1987). Novelty-related responses in the ACC have also been observed. Thus, in addition to its role in attention, the ACC is likely to play an important role in learning.

Does the neural activity present during the encoding of memory (during the first stimulus presentation) predict memory success? Activity before the stimulus onset has been shown to predict successful memory recollection (Otten et al., 2002; Otten et al., 2006). This is a manifestation of the influence of the baseline state (attention, arousal, focus, motivation, or some form of task-preparation) on encoding. Otten et al. demonstrated this effect by comparing the event-related potentials (ERPs) evoked by a cue that predicts stimulus onset a fixed time later (Otten et al., 2006). The authors found that ERPs, sorted according to whether the stimuli were later recollected or not, were different. This is remarkable because it shows that not only does neural activity (measured by ERPs) before stimulus onset correlate with encoding success, but that it can change fast enough to have an effect trial-by-trial. This is difficult to reconcile with baseline states of the brain, which are thought to change on a slower timescale. Top-down attention can, however, influence processing differentially trial-by-trial (Einhauser et al., 2008; Rutishauser and Koch, 2007).

The neural activity present in the MTL shortly after the onset of a stimulus is directly and causally related to whether a memory is formed or not. A demonstration of this involves temporal disruption of neural activity in the hippocampus (of macaques) in a match-to-sample task: performance was only influenced if stimuliation onset was within 300 ms of the stimulus (Ringo, 1995). Afterwards, performance was not disrupted. In humans, intracarotid injection of amobarbital 1 min after acquiring a new memory does not disrupt memory for retrieval after recovery from anesthesia (Gleissner et al., 1997). This form of anesthesia causes extreme hyperpolarization and thus prevents spiking. This suggests that new memories become at least partially independent of electrical neural activity shortly after initial acquisition.

Mechanistically, induction of synaptic plasticity requires tightly coordinated pre- and postsynaptic activity (on the order of 10 ms). Neurons tend to fire in synchrony with others in the same circuit and thus the inputs to a particular neuron oscillate. A prominent oscillation in the hippocampus (and other areas) is the theta rhythm. In vivo, only stimulation around the peak of theta induces strong LTP (Holscher et al., 1997; Hyman et al., 2003; McCartney et al., 2004; Orr et al., 2001; Pavlides et al., 1988). Neurons are most excitable and most depolarized at the peak of theta and fire more sparsely in the presence of theta (Buzsaki et al., 1983; Fox, 1989; Wyble et al., 2000). The presence or absence of hippocampal theta also has a direct behavioral effect on learning: learning rates during conditioning are are positively affected by the presence of theta prior to training (Berger et al., 1976; Berry and Thompson, 1978).

Gamma oscillations (30–80 Hz) are very prominent in many areas of the human brain, including the hippocampus, the amygdala (Jung et al., 2006b; Oya et al., 2002), and a large number of cortical areas (see (Jensen et al., 2007) for a review). In humans, the intracranially measured power of gamma oscillations correlates with working memory load, attention, and sensory perception (Engel and Singer, 2001; Howard et al., 2003; Tallon-Baudry and Bertrand, 1999; Tallon-Baudry et al., 2005). The presentation of visual stimuli triggers gamma oscillations in many areas (Tallon-Baudry et al., 2005). The power of stimulus-triggered increases and decreases in gamma oscillations have also been shown to correlate with recall success in a free-recall task (Sederberg et al., 2007).

We recorded LFP from intracranial depth electrodes during performance of a single-trial learning task. In a similar task, we previously observed single units that indicated the novelty or familiarity of the stimulus presented (Rutishauser et al., 2006a; Rutishauser et al., 2008). Here we asked whether the LFP, recorded during learning, contained information about the success or failure of plasticity. We compared the power of oscillations (during learning) between stimuli which were later recognized and stimuli which were forgotten. Our task was a recognition memory test (new/old) with continuous confidence ratings and a long delay (> 15 min) to test for true long-term recognition memory. The stimuli that we used were all novel and had never been seen before by the patient. This is distinct from previous paradigms used by others, which used short delays, highly familiar stimuli (words), free recall of words, or subjective judgments of recollection (remember/know). We also repeated the same experiment with a longer (24 h, overnight) delay and a new set of novel stimuli. We then tested whether periods of changed oscillatory power identified from the same-day data could predict whether stimuli would be remembered after the overnight delay. We found that there are several distinct frequencies of oscillations in the hippocampus, amygdala, and anterior cingulate that are good predictors of memory success. Also, we find that the oscillatory periods that correlate with same day memory can be used to predict memory performance the next day (overnight memory).

5.2 Methods

5.2.1 Task

During each trial, the stimulus (a picture) was presented at the center of the screen. Distance to the screen was approximately 50 cm and the screen was approximately 30 by 23 degrees of visual angle. Stimuli were 9 by 9 degrees. A trial consisted of the following displays (in this order): delay (1 s), stimulus (1 s), delay (0.5 s), question (variable). During delay periods, the screen was blank. After the delay, the question (see below) was displayed until an answer was provided. The answer could only be provided when the question was on the screen to avoid motor artifacts (keys presses during stimulus presentation were ignored).

During learning trials, patients were asked to answer the question “Was there an animal in the picture?” to facilitate attention and focus. Patients answered this question almost perfectly (≥ 98%), confirming that they were looking at the images on the screen during learning.

During retrieval trials, patients were asked to indicate, for each picture, whether they had seen it before (during learning) or not (e.g. new or old). Also, patients were asked to indicate their subjective confidence of their judgment. Answers were provided on a 1-6 scale from: 1 = new, confident, 2 = new, probably, 3 = new, guess, 4 = old, guess, 5 = old, probably, 6 = old, confident).

All psychophysics was implemented using Psychophysics toolbox (Brainard, 1997; Pelli, 1997) in Matlab (Mathworks Inc).

Stimuli were photographs of natural scenes of 5 different visual categories (animals, people, cars, outdoor scenes, flowers). There were the same number of images presented for each category. Categories were balanced during retrieval to avoid any inherent bias in memory for individual subjects for certain categories. All stimuli were novel and had never been seen by the patient. Each stimulus was presented at most two times (once during learning, once during retrieval).

5.2.2 Data analysis — LFP

For details on how we analyzed LFP data (in particular wavelet decomposition, power/phase estimation), please refer to the methods chapter of this thesis. Here, only the parameter settings and techniques specific to this chapter are described.

Frequency bands were sampled logarithmically spaced: $f = 2^x$ with $x \in [2:2:52]/8$ (see appendix for details). Here, the maximal frequency examined was 90 Hz. In total 24 frequencies were examined (all in Hz): 1.68, 2.00, 2.38, 2.83, 3.36, 4.00, 4.76, 5.66, 6.73, 8.00, 9.51, 11.31, 13.45, 16.00, 19.03, 22.63, 26.90, 32.00, 38.05, 45.25, 53.81, 64.00, 76.11, 90.50.

All channels were included that contained appropriately distributed 1/f wideband signal. Channels with 60 Hz were filtered using a 4th-order Butterworth notch filter. Channels were not pre-selected for the presence of particular peaks in the spectrogram. Thus, it is expected that many of the channels have only weakly detectable energy in prominent LFP bands such as theta or gamma (due to inappropriate impedances or the location of wire). Since we could not find any good (and objective) criteria to judge what constitutes a “good” LFP channel, we opted to include all channels to avoid any biases. Also note that the LFP reported here was recorded simultaneously with spikes. Since we recorded spikes relative to a local ground (one of the other wires on the same macroelectrodes), the LFP signals reported in this chapter are also locally grounded. This implies that the signals discussed here represent the activity of a local population of neurons/synapses (maxmally a few millimeters, often much less). They are distinct from other types of recorded LFPs which are globally grounded (e.g., by an electrode in the other hemisphere or the skull). Examples of globally grounded signals include intracranial EEG and surface EEG. It is thus important to note that LFP in this thesis refers to a local signal. Due to this type of grounding, oscillations in the brain that are the same over long distances (several millimeters) can not be observed (requires global grounding). Other reports of LFP recorded from similar microwires (simultaneously with spikes) also have this caveat, although they usually neglect to mention this explicitly (Ekstrom et al., 2007; Jacobs et al., 2007; Kraskov et al., 2007; Nir et al., 2007). It is also important to keep this caveat in mind when comparing human microwire LFP to animal LFP data, which is usually not locally grounded (and similarly to intracranial EEG).

The LFP power in these 24 different frequency bands was calculated as a continuous function of time using wavelet decomposition (see appendix). We compared the mean LFP power in 250 ms bins from stimulus onset to 500 ms after stimulus offset (total duration 1500ms). We tested for differences in mean power in each bin using 5000 bootstrap samples (Efron and Tibshirani, 1993). The LFP power at a particular frequency has a heavy tail (χ^2 distributed) and it is thus inappropriate to compare these populations using parametric tests such as the t-test. The bootstrap test we used is entirely non-parametric and makes no assumptions about the distribution of the values. For each channel, there were thus 148 (6 x 24) comparisons. We corrected for multiple comparisons using false discovery rate (FDR) with a q = 0.05 across time (Benjamini and Hochberg, 1995). This thus guarantees a FDR of 5% at each frequency, regardless of the number of time bins used. Thus, it is expected that 5% of the channels will show a significant difference at each frequency due to chance. Note that FDR was thus not controlled at the level of an entire electrode (but rather at the frequency). It is thus not meaningful to state the percentage of electrodes that show a significant difference due to memory (DM) effect because the false positives are not controlled for this measure (and, in the worst case, could be very high due to 24 independent frequency bands at a 5% level each). Nevertheless, some authors have still reported % of channels significant using the same multiple comparisons approach we use here (Sederberg et al., 2003; Sederberg et al., 2007). In our opinion, these reported numbers (reported to be > 70%) are meaningless because conservative (complete independence between frequencies) chance levels are of the same magnitude.

We further confirmed that the 5% chance level enforced using FDR was appropriate. There are many reasons why the chance level could be much higher even if using $p < 0.05/q < 0.05$: i) small sample sizes (15–35 samples in each group, i.e., the stimuli that subjects remembered/forgot), ii) the heavy-tailed distribution of LFP power, iii) the imbalance between the two classes (typically more pictures are remembered then forgotten, although a high number of forgotten pictures does not indicate the absence of memory if false positives are low), or iv) the different dynamics due to the 1/f properties of the signal (faster signals can change faster, thus more noise). It was thus necessary to calculate the empirical chance (bootstrapped). To create a bootstrapped sample, we randomly re-assigned the labels "forgot" and "remembered" (sampled with replacement). This created two samples of LFP powers, which were then compared as described above (at each frequency and time bin). The same random sampling was used for all channels of one subject (since these channels were recorded simultaneously). Repeating this procedure 200 times for each subject resulted in a percentage of channels which showed significant DM effects (as a function of frequency). We found that the chance level calculated with this procedure was only marginally above 5% and our procedures are thus appropriate (see results for details). Chance levels were, however, not entirely independent of frequency (higher for higher frequencies). This further reinforces the need for empirically estimating the chance levels to assure that effects are not spurious.

5.2.3 Data analysis — LFP decoding

Decoding was performed using regularized least-square classifiers (RLSC; see appendix for details). The classifier was binary (hit or miss). Each stimulus was classified as either a hit or miss based on whether it was correctly remembered or not (regardless of confidence). Thus, the number of examples in each class was determined by the performance of the subject and varied from session to session. To avoid any biases, classes were balanced 50/50 before testing and training the classifier. This assured that the true chance performance of the classifier was 50%. Otherwise, if (for example) the subject remembered 80% of the stimuli the true chance performance would be 80% (a classifier that always says "hit" could reach this performance without even considering the input). Classifiers were always trained separately for each recording session. Data were not artificially pooled.

For the overnight sessions, we trained a classifier on all time/frequency bins that significantly differed for hit vs. miss same-day trials. The significance of bins was also determined based on the same-day trials. We then used this classifier on the trials that were used for overnight recognition to predict whether the stimuli will be remembered or not. The measure of performance was the percentage of overnight trials that the classifier predicted correctly.

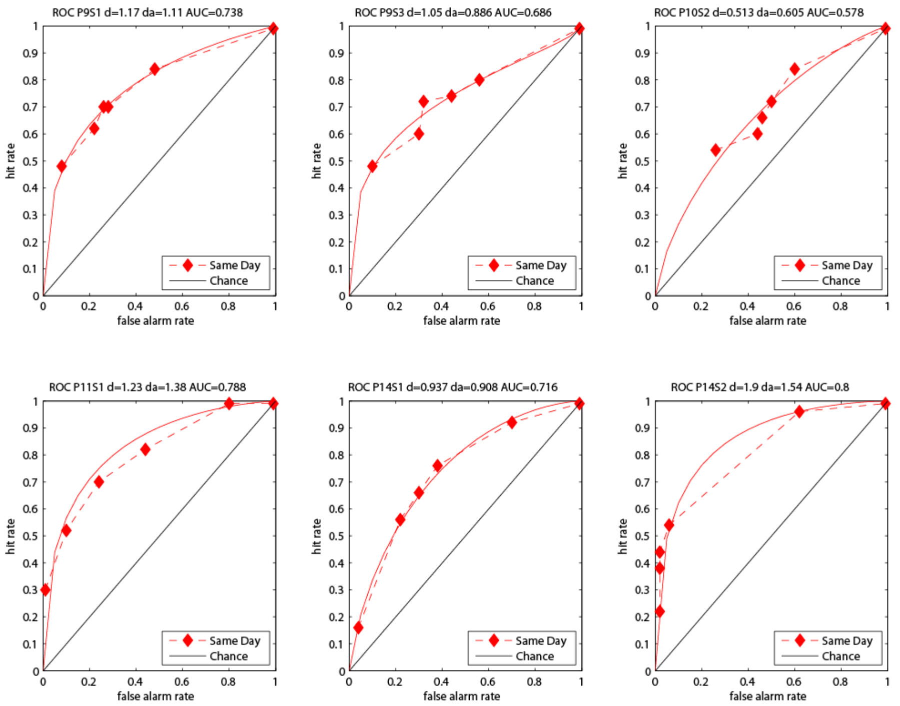


Figure 5-1. Retrieval performance (behavior) shown as a receiver operator characteristic (ROC) curve.

All subjects had above-chance performance for all confidence levels (points are above the diagonal). Also, subjects had a good sense of confidence (lower false alarms for high confidence). The summary measures d' and area under the curve (AUC) values are shown for each session (title). Each panel shows the performance for one individual retrieval session (6 are shown). The location of each data point (red dot) is determined by a pair of false alarm and true positive rates (x and y axis, respectively). Subjects rated their confidence on a 6 point scale: 1=new sure, 2=new probably, 3=new guess, 4=old guess, 5=old probably, 6=old confident. The leftmost datapoint corresponds to 6 (“old confident”) and the rightmost point is 1 (“new sure”). Also shown is the analytical fit (full line) that was used to determine the d' value.

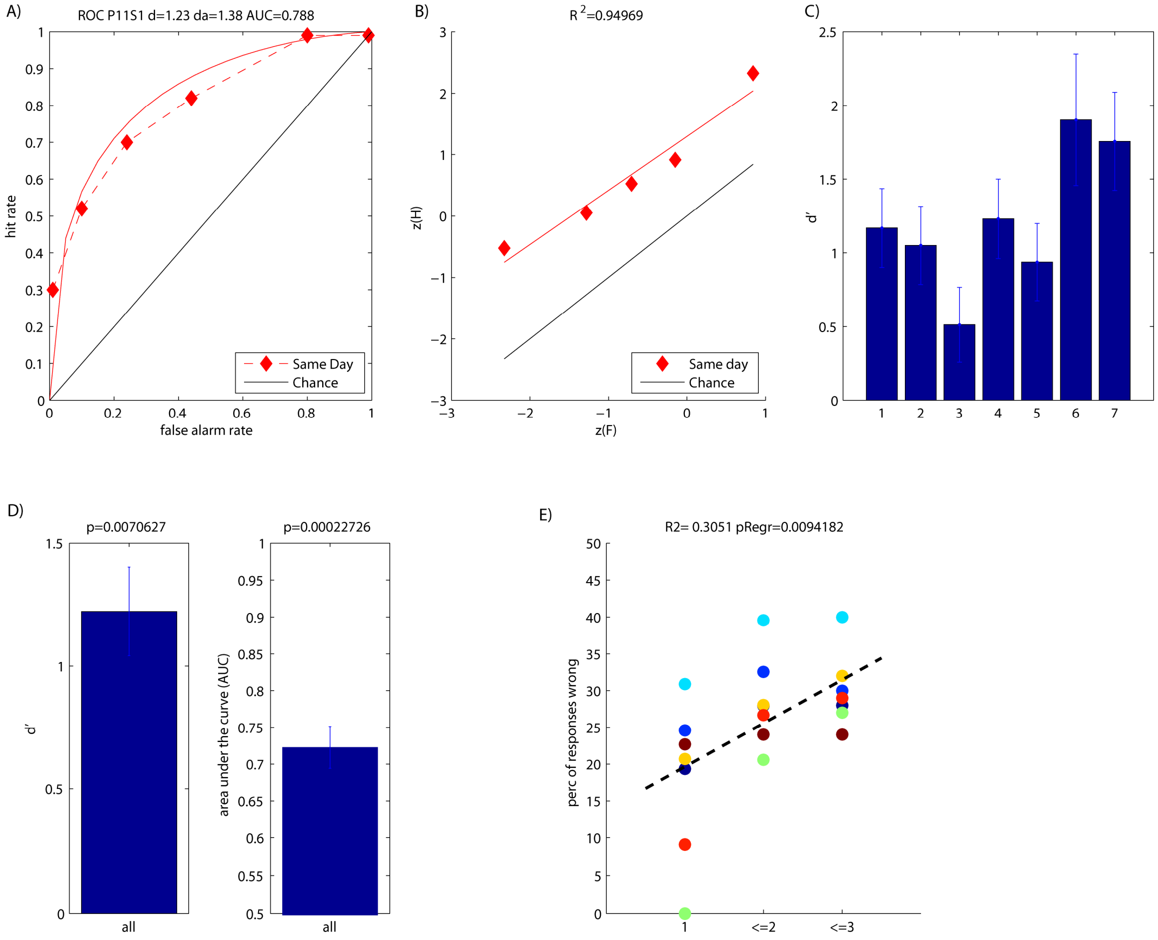


Figure 5-2. Retrieval performance for all subjects.

(a) ROC curve of one retrieval session (see Figure 5-1 for details). **(b)** The z-transformed representation of the same ROC curve as shown in (a). Each datapoints corresponds to one level of confidence. The z-transformed performance was fit well by a straight line ($R^2 = 0.95$) and thus d' is an appropriate summary measure of performance. **(c)** d' for each retrieval session. Performance was above chance (d' = 0) for all sessions. Errorbars are ±s.e. and show within-subject confidence intervals. **(d)** Average d' for all sessions (n = 7) was significantly different from chance (p = 0.007, chance is d' = 0). **(e)** Average area under the curve (AUC) for all sessions (n = 7) was significantly different from chance (p = 0.0002, chance is AUC = 0.5). AUC is a nonparametric summary measure with no assumptions and thus confirms the d' result. **(f)** Percentage of errors as a function of confidence. The lower the confidence, the higher the error rate. Each session is a different color. Subjects had a good sense of confidence: error rates decreased significantly with an increase in confidence (1 = highest confidence, 3 = lowest; $R^2 = 0.31$, p = 0.009).

5.3 Results

We administered a simple picture memorization task in two stages: learning and retrieval. Pictures were photographs of natural scenes that contained objects (see Methods). Memory was tested 10–20min after learning. A distraction task (Stroop) was administered during the delay period. During learning, 50–100 pictures (depending on the memory capacity of the patient, see Methods) were presented. Patients were instructed to remember which pictures they had seen. Each picture was shown for 1 s.

Memory was tested by asking patients to indicate whether they had seen the picture shown before as well as the confidence of their judgment (on a 1–6 scale, see Methods). Patients had both good memory for the stimuli shown as well as a good subjective sense of confidence (Figure 5-1 and Figure 5-2). We quantified retrieval performance using receiver operator characteristice (ROC) analysis and d' (Macmillan and Creelman, 2005). Example ROCs for six retrieval sessions are shown in Figure 5-1. Each data point in the ROCs illustrates one confidence level. The point in the lower left corner (lowest false as well as true positive rate) corresponds to the highest confidence level ("old confident"). As a summary measure of the entire ROC, we used d' and area under the curve (AUC) of the ROC. Using d' requires that the values underlying the ROC are normally distributed (thus, it makes assumptions about the shape of the ROC curve). For our patients this assumption was well justified: the z-transformed ROC was fit well by a straight line (an example is shown in Figure 5-2B with an $R^2 = 0.95$). The average d' ("d-Prime") for all 7 retrieval sessions (from 5 patients) was 1.22±0.18 (Figure 5-2C,D). Nevertheless we also quantified retrieval performance using the average AUC, which is the integrated area below the ROC curve. For example, the ROC shown in Figure 5-1A has an

AUC of 0.79). The AUC varies between 0.5 (chance) and 1.0 (perfect). In contrast to d', it makes no assumptions about the underlying distributions (non-parametric). Patients had an average AUC of 0.72±0.03 (Figure 5-2E). Subjects not only had good memory but they also had a good sense of subjective confidence. This is indicated by the monotonically increasing ROC curves (Figure 5-2A), as well as the increasing percentage of errors made as a function of decreasing confidence. This is illustrated in Figure 5-2F: the lower the confidence, the higher the error rate (quantified as the percentage of all responses made). Errors increased by 6% per decreased confidence level and were well fit by a linear model ($p = 0.009$, $R^2 = 0.31$).

Next, we analyzed the neural activity during learning. The general approach for this analysis was to compare learning trials for pictures that were later remembered with learning trials for pictures that were not remembered (difference due to memory (DM) effect). If the failure to retrieve the forgotten stimuli is directly attributable to a failure to evoke plasticity during learning, it is hypothesized that such differences can be observed in the LFP and/or single unit data. Obviously there could be many other reasons why retrieval failed and it is thus not expected that every retrieval failure can be attributed to a failure of plasticity during learning. Other possible factors are attention during retrieval, misattribution due to confusions with similar-looking stimuli, memory consolidation, rehearsal, incorporation into personal memories (episodic), sleep, or emotional attributes evoked by the stimuli (which differ in each patient).

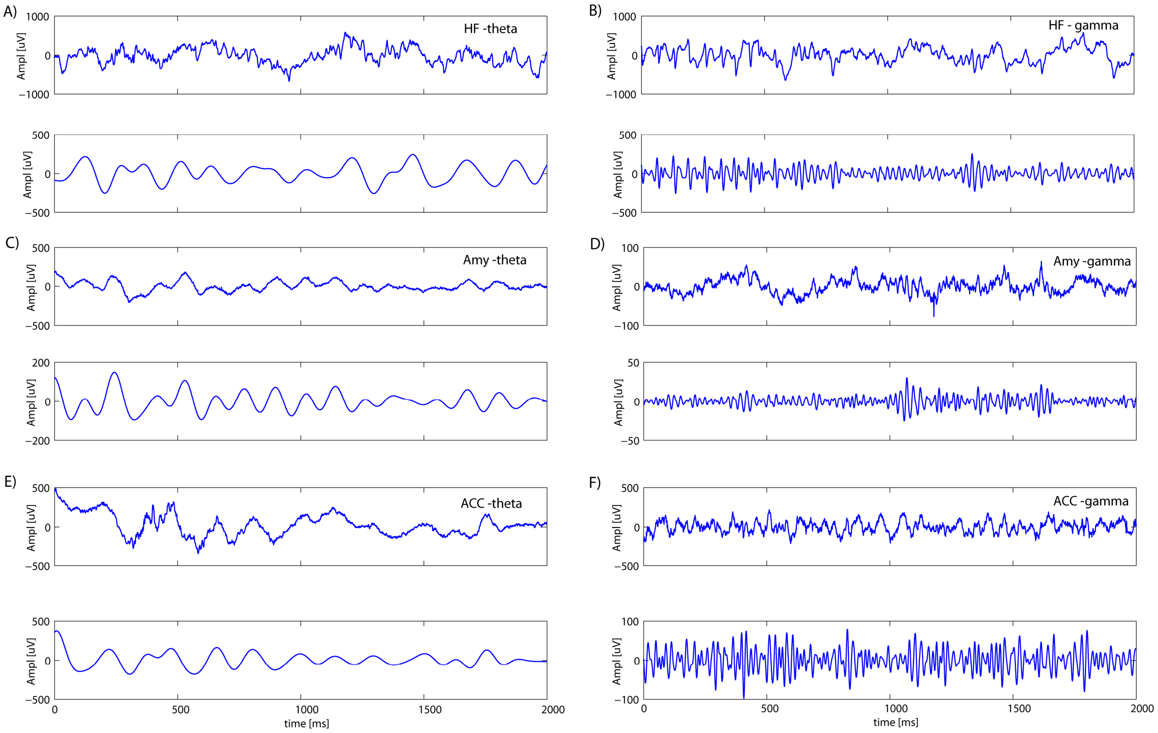


Figure 5-3. Example LFP traces (raw, theta, gamma).
Shown are 2 s of data from the hippocampus (HF; A+B), amygdala (Amy; C+D) and anterior cingulate (ACC; E+F). Traces were from data recorded during the learning part of the task (stimulus onset at 500 ms). Each panel shows the raw fullband trace (high-pass 1 Hz) and a bandpass filtered version (theta 3–10 Hz, gamma 30–80 Hz; note that these frequency bands are for illustration purposes only and were not used for analysis). Left column shows theta, right column gamma. Note the clear presence of gamma and theta oscillations in all three areas. The amplitudes of oscillations varied widely between channels.

First, we compared the power in different frequency bands of the LFP. We recorded the wideband extracellular signal from single wire electrodes in the amygdala, hippocampus, and anterior cingulate cortex bilaterally (see Methods). Many channels showed prominent activity in the gamma and theta bands, which were visible in the raw unfiltered signal (Figure 5-3). Since the traditional boundaries of which frequencies constitute a “theta” or “gamma” oscillation are somewhat arbitrary, we only use these terms here for discussion purposes. Also, there are indications that the frequency of many of the intrinsic oscillations (which are mostly defined based on recordings in small rodents) are slower in bigger mammals and particularly in humans (Buzsáki, 2006; Penttonen and Buzsaki, 2003). To avoid assumptions, all analysis was conducted independently at each frequency, regardless of which (hypothesized) band it belonged to.

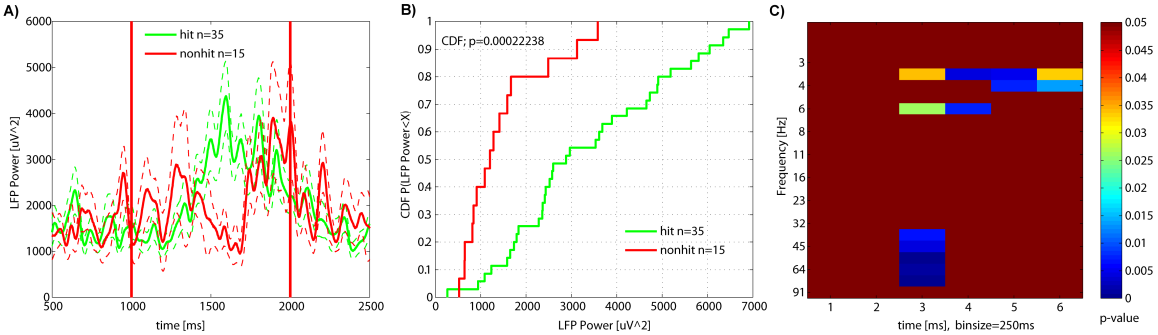


Figure 5-4. Example of LFP power difference due to memory.
All data in this figure is from a microwire in the left hippocampus. The frequency band illustrated is 53 Hz (gamma). (A) shows the LFP power (at 53 Hz) as a function of time for all learning trials. Trials for stimuli which were later remembered (green) had more gamma power compared to trials with stimuli that were not remembered (red). The stimulus is on the screen for 1 s, indicated by the vertical red lines. (B) Distribution of power for the 3rd timebin (500–750 ms) illustrated as a cdf. Notice the large shift to the right (larger values) of remembered (hit, green) trials. (C) P-Values for all timebins and all frequency bands. Each bin is 250 ms long. Only values which survived the per-frequency FDR are shown. Notice the highly significant difference for gamma-band frequencies for the 3rd timebin, an example of which is shown in A+B.

We compared, at each frequency, the power of the LFP signal between stimuli that were later remembered vs. stimuli that were forgotten (see Methods for details). We found that prominent differences exist in several distinct frequency bands. An example channel from the left hippocampus is shown in Figure 5-4A. This channel had higher power in the 53 Hz band for stimuli which were later remembered. We found similar differences due to memory in all brain areas we recorded from for a variety of frequencies (examples are shown in Figure 5-5, see below for statistics). One observation was a prominent increase in power for remembered stimuli that was seen both during (Figure 5-5F) as well as shortly after presentation of the stimulus (Figure 5-5C). Some channels also had a decrease in power that correlated with remembered stimuli (Figure 5-5B). In the anterior cingulate, some channels showed prominent overall power decreases that started shortly before stimulus onset (Figure 5-5A).

We found significant differences in several distinct frequency bands (Figure 5-6). To differentiate which frequency differences were not attributable to chance, we calculated an unbiased boostrap estimate of the chance level as a function of frequency for each brain area (Figure 5-6, blue bars; theoretical level of 5% is indicated by the black line). We found that the empirical chance level generally increased somewhat as a function of frequency. A comparison of the expected number of channels different due to chance with the observed number of channels using a goodness-of-fit χ^2 reveals a significant difference for all 3 brain areas (hippocampus χ^2 =164, amygdala χ^2 =84, cingulate χ^2 =56; all $p < 0.0001$; all df = 24). Several frequency bands with prominent DM effects become apparent (compare blue and red in Figure 5-6): < 3 Hz, 4–8 Hz, 9–12 Hz, 16–30 Hz and > 30 Hz. Differences due to very low frequency oscillations (< 3 Hz) were only apparent in the amygdala and hippocampus (Figure 5-6A,B). Gamma band differences were prominent in all brain areas (> 30 Hz). Alpha-band differences (9–12 Hz) were particularly prominent in the cingulate, present in the hippocampus and absent in the amygdala. Beta-band differences (16–30 Hz) were prominently present in the amygdala.

Are the power differences described above predictive of whether a stimulus will be remembered? So far we have only demonstrated a correlation: on some channels, power is distributed differently for stimuli which are later remembered compared to stimuli which are not. We used a decoding approach to quantify how far this activity is truly predictive. We used a regularized least square classifier (RLSC; see methods). This decoder is very simple: it takes the weighted sum of all available bins. The weights are determined based on the training samples and a regularizer term, which enforces smoothness.

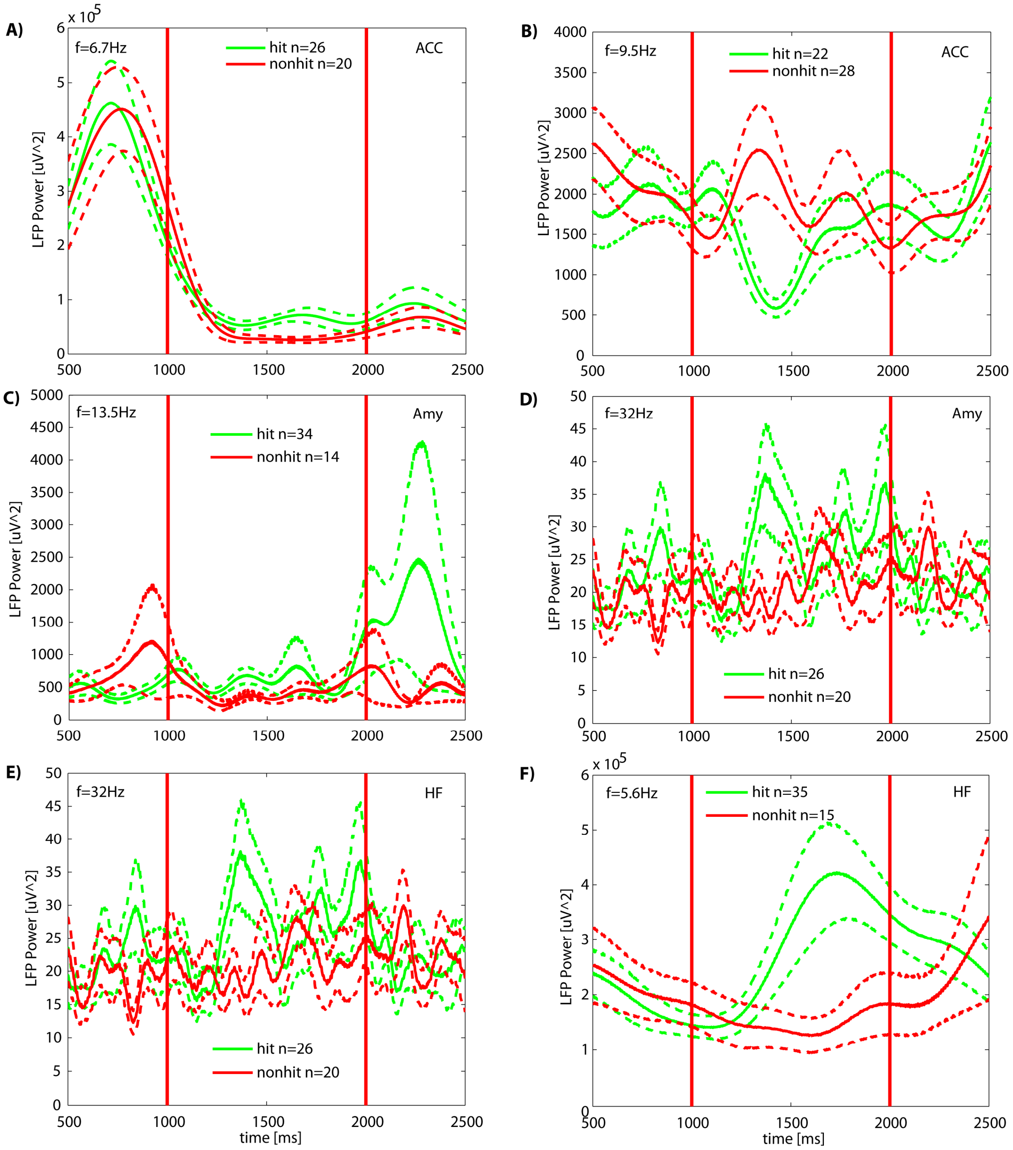


Figure 5-5. Examples of DM effects from all three brain areas as well as different frequency ranges.

Shown are two examples from each: anterior cingulate (A,B), amygdala (C,D), and hippocampus (E,F). The frequency of each is indicated in the panel (f = X Hz). Time units are in milliseconds. The stimulus is present on the screen for 1 s (between red vertical lines). Notice that for A,E the y axis is in terms of 10^5.

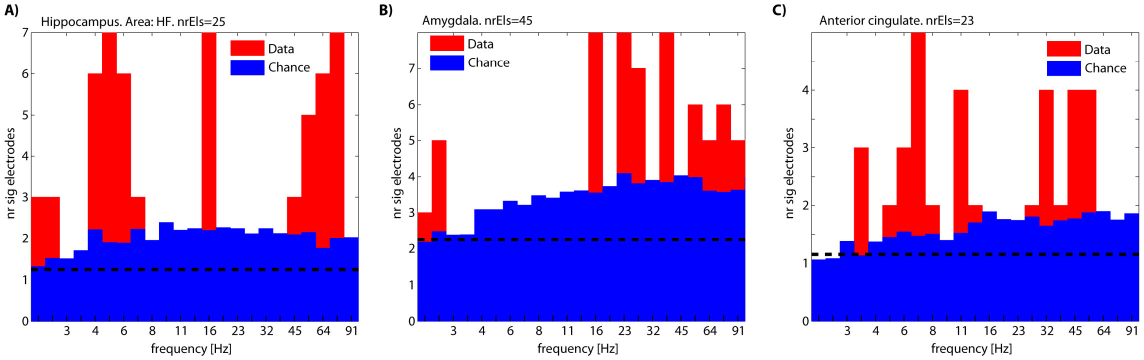


Figure 5-6. Summary of DM effects for all brain areas and frequencies. Shown are the number of electrodes as a function of frequency that have a DM effect for at least one timebin. Red bars show the real data, blue bars the bootstrapped chance level and the black line the theoretical chance level. All comparisons are multiple comparisons corrected using FDR. Note the distinct frequency bands that have significant effects: < 3 Hz, 4–8 Hz, 11–16 Hz and > 30 Hz. Data is shown separately for the hippocampus (A), amygdala (B), and anterior cingulate (C). Note the clear presence of theta-band difference in the hippocampus and cingulate, but not the amygdala (see text).

For decoding we focused on the overnight sessions. Decoding from same-day trials was possible as well (with percentage correct > 80%, compare to below), but this is not unexpected: time-frequency bins were selected such that they showed a significant difference. It is thus more meaningful to decode overnight trials, which are entirely independent. We were able to record an overnight retrieval session from a subset of our patients (2 sessions from 2 separate patients). These patients had sufficient memory capacity to learn 100 images in one session. Half (50) of these images were used for same-day retrieval (10–20 min delay) and the other half were used for retrieval 24 h later. The images the patients saw after 24 h were different from the images the patients saw after the short delay. Patients were able to remember pictures overnight: average d’ was 0.45±0.15 (excluding guess trials) and the average AUC was 0.56±0.01. Also, patients had a good sense of confidence (both FP and TP increased monotonically as a function of decreasing confidence). First, we analyzed the learning trials for the stimuli used for same-day retrieval; we identified the frequency/time bins that showed a significant difference between hits vs. non-hits (as described above). Also, we trained a classifier using this data. This analysis was based entirely on the 50 trials that were used for same-day retrieval. No data from the learning trials for overnight retrieval was used. Afterwards, we used the time/frequency bins identified by this analysis to investigate whether these had predictive power for overnight retrieval. An example of one channel is shown in Figure 5-7. Note that the distribution of the hits and non-hit trials is similar for same-day and overnight retrieval sessions. The overnight learning trials constitute a perfect out-of-sample testset. All parameters of all the analysis steps are exclusively estimated from the same-day learning trials. It is an open question as to whether overnight memory could be predicted based on firing patterns that predict same-day memory. It is conceivable that different physiological mechanisms are responsible for these different memory spans. Also, it is conceivable that the influence of the plasticity triggered during initial acquisition is less prominent the longer the time delay (due to processes such as consolidation). One indication for this is that retrieval performance is worse after the 24 h delay (average overnight AUC = 0.56 and average same-day AUC = 0.67, for the two patients that have both overnight and same day sessions). Despite this, we found that activity patterns identified from same-day activity are predictive of overnight memory: decoding overnight trials results in correct prediction (of whether the stimulus will be remembered or not) for 58.5±0.04% of all trials (Figure 5-8; significantly bigger than chance p = 0.036). Percentage correct as a summary measure of decoding performance can be misleading, and we thus also quantified performance using A' (Macmillan and Creelman, 2005). A' for overnight decoding was 0.66±0.02 (0.5 is chance). Thus, the activity patterns that predict successful same-day memory also have predictive power for long-term (overnight) memory. Decoding performance is, however, worse then for same-day retrieval (as expected, due to the factors mentioned above).

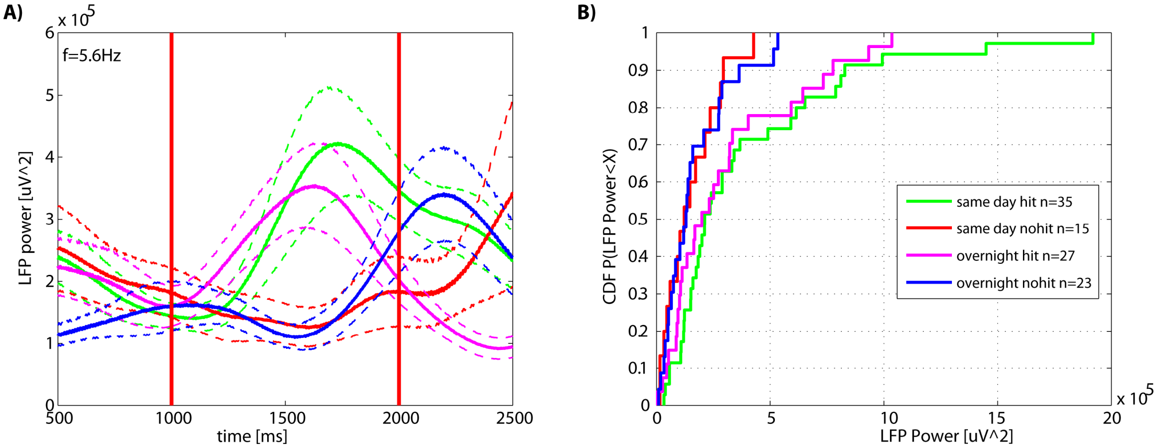


Figure 5-7 Example of LFP power difference, shown for learning trials that were retrieved on the same day (green and red) and overnight (magenta and blue). This channel was selected entirely based on the statistics for the same-day trials. (A) raw trace of LFP power in the 5.6 Hz band. Note that the units are in terms of 10^5. The stimulus was on the screen for 1 s (vertical red lines). Notation for colored lines is shown in (B). (B) Illustration of the distribution of all 4 trial types using a cdf.

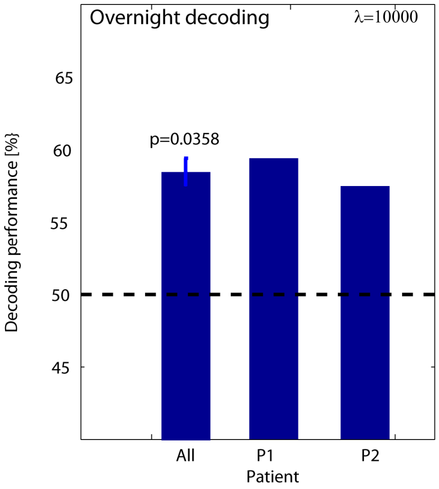


Figure 5-8. LFP power can be used to predict overnight memory.
Channels were identified that correlate with success of retrieval after the short delay (same day) and were then used to train a classifier. This classifier is able to predict overnight memory successfully if used on the learning trials for the overnight trials. Shown is the the mean performance (left) as well as the individual performance for the 2 patients that completed this task. The dashed line indicates chance performance (50%).

5.4 Discussion

We found that LFP power in different frequency bands in the hippocampus, amygdala and cingulate correlates with later retrieval success. Thus, LFP power changes (during learning) are correlates of the successful induction of plasticity and thus retrieval success. Power changes were specific to certain frequency bands (< 3 Hz, 4–8 Hz, 16–30 Hz, > 30 Hz) rather then overall increases in LFP power. We also found that the LFP power changes can be used to predict whether retrieval will be successful or not. Thus, they are not just a correlation but a valid predictor. Our findings thus represent a direct demonstration (by later behavior) that the strength of local extracellular field oscillations is a relevant factor in the induction of plasticity.

While we show that certain LFP power changes are predictive of later retrieval success it remains to be demonstrated *why* this is so. Increased power of oscillations likely indicates higher synchrony of firing between different neurons, which thus could induce plasticity more easily (Axmacher et al., 2006). It is also possible that increased LFP power enhances the effectiveness of information transmission between different areas, such as the hippocampus and the cortex. These effects could be mediated by increased phase locking due to more dominant oscillations. Phase locking to, for example, theta or gamma is a prominent feature of both hippocampal and cortical neurons (see Introduction for details). While phase locking is relatively well understood at the circuit level, its behavioral relevance is unknown. What triggers the increases in oscillatory power also remains unclear. In part these can probably be attributed to attentional processes, but there are probably also other causes of increased oscillations. Increased power can also be caused by phase resets (triggered by stimulus onset) of existing oscillations, which can be observed during memory tasks (Mormann et al., 2005; Rizzuto et al., 2006; Rizzuto et al., 2003).

Candidates for regulation of LFP oscillations are modulation by emotional factors (arousing stimuli), reward (such as reward predictors), or depth-of processing modifications. One indication that reward predictors might influence memory encoding is the correlation of retrieval success with activation (measured with BOLD) of the ventral tegmental area (VTA) (Adcock et al., 2006; Knutson et al., 2001; Wittmann et al., 2005), an area which projects dopamine releasing axons to the hippocampus (Bjorklund and Dunnett, 2007; Gasbarri et al., 1997; Gasbarri et al., 1994), amygdala (Fallon et al., 1978; Fried et al., 2001), and prefrontal areas (Bjorklund and Dunnett, 2007; Vogt et al., 1995; Williams and Goldman-Rakic, 1998). It is thus conceivable that dopamine release contributes to the increase in LFP power. Such dopamine release is also hypothesized to be triggered by novel stimuli (Lisman and Grace, 2005). In vitro, dopamine has been shown to have a strong modulatory role in the strength of plasticity (Chen et al., 1996;

Huang and Kandel, 1995; Otmakhova and Lisman, 1996; Smith et al., 2005). Of particular interest is the finding that dopamine acts as a high-pass filter at the synapse that relays direct cortical input to the hippocampus (Ito and Schuman, 2007). BOLD activity recorded in the VTA, in fact, has been shown to be activated by absolute novelty rather then emotional content, general saliency, or rarity (Bunzeck and Duzel, 2006). This indicates that a fruitful avenue for future experiments would be the modulation of reward during learning with a paradigm known to activate the VTA, while simultaneously recording LFP in the hippocampus. Similar arguments can be made for the hypothesized modulation of memory strength of emotional stimuli by the amygdala (Sharot et al., 2004). One possibility for the amygdala to achieve this is to induce or enhance oscillations in the hippocampus or other areas. Simultaneous recordings of LFP in the amygdala and hippocampus while performing a memory task comprising both emotional and non-emotional stimuli would be a useful experiment to elucidate these effects. A frequently used paradigm to change memory strength has been a modification of depth of processing, for example counting the number of characters vs. imagining a sentence describing the situation in the case of remembering words (Paller et al., 1987). Such modifications effectively modify attention. A mixture of such a paradigm and another modulator of memory strength (such as emotion) might allow one to disambiguate attentional from other effects of increased encoding success.

This study is different from others in several crucial aspects. We exclusively used novel stimuli which had never been seen by the patient. We did so to ensure that we examined the encoding of novel information rather then the judgment of recency. The time delay between learning and retrieval was substantial (> 10min). Also, a distraction task was performed immediately after completing learning. To assess memory strength, we used a recognition memory test (new/old) with confidence ratings. This allowed us to systematically assess the behavioral performance of the patients using ROC diagrams. Previous studies used lists of highly familiar words that were then freely recalled by the patient after a short (often 30 s) delay (Cameron et al., 2001; Fernandez et al., 1999; Sederberg et al., 2003; Sederberg et al., 2007). Thus, these studies report predictors of memory success for recall ("recency") of verbal memory (for words that were very familiar) after short time delays. In contrast, we report predictors of encoding success for a much more general class of novel stimuli (complex natural scenes of objects) that were learned in a single trial. Also, we show that these changes are truly predictive. One other study (using words and free recall) claims to document this too, but in fact only shows a correlation (Sederberg et al., 2007). Note also that we did not normalize the LFP power to baseline (in contrast to others). Thus, the differences that we analyzed include both stimulus-triggered as well as other differences (such as more slowly varying state changes, possibly evoked by changes in the neuromodulatory environment).

One curious aspect of our findings is the lack of specificity to a particular brain area. While there were differences in terms of the frequencies that were predictive between areas, in general all three areas investigated (amygdala, hippocampus, ACC) correlated with encoding success to a similar degree. Our recordings were locally grounded (see methods), and the LFP reported here is thus of a very local nature. This effect can thus not be explained by large-scale synchronous oscillations. Rather, it appears that all three areas contribute to encoding success to a similar degree (on average). Since our stimulus set contained a very heterogeneous set of stimuli of different categories and emotional saliency, we cannot exclude that this is an effect of averaging all stimuli. This finding is, however, in agreement with many surface EEG and MEG studies that report differences due to later memory in a widespread collection of areas (Klimesch et al., 1996; Osipova et al., 2006; Takashima et al., 2006). Our recordings, which have much higher spatial resolution, confirm that power changes can be observed very locally in all three areas we recorded from. Since the areas responsible for encoding of memories are tightly interconnected in many different ways, it is perhaps not surprising that all areas show increased activity. It is possible that one area seeds the increase in synchrony, which then quickly spreads to all the other areas such that increases in LFP power are visible in the entire network. The non-specificity of predictive oscillatory power increases also indicates that an important component of encoding success is the coordination of large-scale brain circuitry. For example, BOLD signal correlations between extrastriate visual areas (face/place selective) and prefrontal (DLPC) correlate with successful episodic memory formation (Summerfield et al., 2006). Thus, cortical-cortical correlations are important for memory success. Similarly, hippocampal-cortical interactions are crucial for memory formation (Wiltgen et al., 2004). For example, it has been demonstrated that prefrontal neurons in the rat can phase-lock to hippocampal theta (Siapas et al., 2005), and it has been proposed that this facilitates information transfer between these two structures. It is thus perhaps not surprising that power increases can be observed in both structures simultaneously.

Chapter 6. Summary and General Discussion

6.1 Novelty detection and single-trial learning

This thesis is about single neurons in the human brain that express a fundamental piece of information: whether a stimulus is novel or familiar. By definition, a stimulus is novel only once. The second time it is seen, it is familiar. Novelty- and familiarity detecting neurons follow this by rapidly changing their firing pattern. The response of novelty- and familiarity detecting neurons is not binary. Rather, the response strength (or absence, in the case of novelty neurons) is proportional to the strength of memory. Familiarity detecting neurons, by definition, increase their firing rate for stimuli which are familiar (have been seen before). Their response is strongest for stimuli that are recognized and recollected, intermediate for recognized stimuli (but not recollected), and weak (but non-zero) for forgotten stimuli. Novelty detecting neurons, by definition, only increase firing for novel items. For familiar items, however, they tend to decrease their firing. The stronger the memory, the larger the firing rate decrease (relative to the response to novel stimuli). Thus, both novelty and familiarity detecting neurons signal the strength of memory, but with opposite polarity.

The neurons discussed here are capable of the most rapid form of plasticity: single-trial learning. Most events in life occur only once. Thus it is of fundamental importance to investigate this form of learning. Responses to novel stimuli are extremely prevalent in the brain and in behavior (Sokolov, 1963). Many neurons in many areas of the brain respond differently if a stimulus is novel or otherwise salient in some way. Responses to novelty can also be observed behaviorally: animals such as rats have a natural tendency to explore novel objects. In fact, this effect is commonly used to test recognition memory for objects in rodents (Ennaceur and Delacour, 1988). Animals automatically orient towards novel stimuli (Sokolov, 1963; Vinogradova, 2001). Humans and non-human primates automatically move their eyes towards novel objects and they spend more time fixating novel objects (Althoff and Cohen, 1999; Smith et al., 2006; Yarbus, 1967). This preference exists even in infants (Fantz, 1964). Autonomic reactions such as skin conductance (Knight, 1996), heart rate (Weisbard and Graham, 1971), or pupil diameter-dilation (Hess and Polt, 1960) also show prominent novelty responses. Many of these novelty responses are severely reduced by lesions of parts of the medial temporal lobe (Honey et al., 1998; Kishiyama et al., 2004; Knight, 1996; Knight and Nakada, 1998; Yonelinas et al., 2002). This is particularly the case for hippocampal lesions. Being novel (or more generally, different) is a very effective modulator of memory strength. This is true even if the attribute that makes a stimulus novel is task irrelevant. This is the well known “von Restorff” effect (Hunt, 1995; Kinsbour and George, 1974; Kishiyama et al., 2004; Parker et al., 1998; von Restorff, 1933; Wallace, 1965). Thus, it is clear that novelty is an efficient modulator of memory strength. Some have proposed that novelty increases dopamine release, which is known to induce strong and long-lasting plasticity (Lisman and Grace, 2005). Given the persistence of this phenomenon, it seems warranted to speculate that the feature that novelty enhances memory constitutes an evolutionary advantage and is thus selected for.

While remembering something as best as possible (and thus detecting novelty as well as possible) is usually advantageous, there are also situations where strong memories are not advantageous for the individual. Examples are memories which can, despite best efforts, not be erased such as in post-traumatic stress disorder (PSTH), drug-induced place preference, or memories connected to strong emotions. What these examples have in common is that the memory was established by a single experience (a single trial). Thus, the novelty advantage conveyed to memories can also be a disadvantage.

Given the importance and prevalence of novelty-dependent effects, surprisingly little is known about the neuronal mechanisms of such rapid learning. Among the many behavioral paradigms used to study learning, most require a large number of learning trials. Examples are conditioning (both classical and instrumental), the Morris water maze (learning the escape location), and maze learning. There are sophisticated models for these types of learning (see (O'Doherty et al., 2003; Schultz, 2002; Seymour et al., 2004) for examples). Such models are typically variants of reinforcement learning (Sutton and Barto, 1998). However, in everyday life most learning tasks we face are not of this nature. Rather, they are of the more rapid kind of learning where we, at best, learn from a few trials (Exceptions are acquiring new habits). Examples of behavioral paradigms for rapid learning are those described in this thesis (for humans), conditioned taste aversion as well as some forms of fear conditioning. In these paradigms, the failure to detect a stimulus as novel impairs learning severely (Welzl et al., 2001). For such rapid learning tasks we lack the formal understanding that we have for incremental learning (such as reinforcement learning models). This also applies to learning by machines. In machine learning, learning from many examples by training classifiers is well established. There is no equivalent technique to learn from just a few trials. In tasks which require many trials to learn (such as maze navigation), it has become clear that neural activity during the first few learning trials (in a novel environment) is very different to that observed when learning is completed (see (Cheng and Frank, 2008) for an example). This stresses the importance of recording from the very initial learning trials rather then when the animal is well trained (or over-trained), as is most often done in the case of hippocampal recordings. Given that the hippocampus is thought to be most important for learning, it is likely that many important processes are never observed because they are over by the time recording starts.

6.2 The relationship between memory responses and behavior

The firing rates of single neurons in the human amygdala and hippocampus can be used to construct a simple new/old decoder that outperforms the patient. That is, if the patient made an error (either forgetting a stimulus or wrongly declaring it familiar), the neuronal responses often indicated what would have been the correct decision (which was not made by the patient). Thus, these neurons had better memory than the patient. Here, we used this fact to argue that these neurons do not represent the motor output nor the decision of the patient. This is because there is a clear dissociation between the decision (and thus motor output) and the neural response. For both types of trials (forgot and new trial), the behavioral response is the same: a press of the "New" button. The neuronal response, however, is very different. Thus, these neurons do not represent the patient's decision, rather they may represent the input to the decision-making process.

Research in decision making typically focuses on decisions about external stimuli, i.e., which cue is most likely to indicate a reward. Memory retrieval, however, involves a different kind of decision making: decisions about internal states. Deciding whether a stimulus is novel requires deciding that there is no trace or representation of the presented stimulus in the system and thus the stimulus is novel. Similarly, deciding that a stimulus is familiar requires judgment of whether there is enough evidence of previous occurence. Both types of decisions are about neuronal firing, which represents an internal state rather than an external stimulus. Little is known about how such decisions are made. The paradigms and neuronal responses presented in this thesis lend themselfs well to investigating this process. A first step would be to identify an area of the brain that contains novelty/familiarity detectors (such as the ones documented here) that follow the decision rather then the memory (and that are not trivially related to motor output). Candidates for such area(s) are frontal areas such as the anterior cingulate, medial prefrontal, or orbitofrontal areas (Badre and Wagner, 2007; Koechlin and Hyafil, 2007; Lepage et al., 2000; Wagner et al., 2001). For saccadic decisions, it is known that the frontal eye fields (FEF) represent the decision rather then the visual input (Hanes and Schall, 1996). Simultaneous recordings from the hippocampus/amygdala and this yet to be identified area (preferably in humans, which is possible for several candidate areas) would be a powerful system to investigate how decisions are made about the presence or absence of memories. Asking humans to judge their confidence would be particularly useful in this setting. While identifying the location of such neurons itself only tells us where the decision is represented, simultaneous recordings from both areas will allow detailed investigation into how the decision itself is made (e.g., by looking at their interactions in time during errors).

6.3 Novelty and familiarity responses in the amygdala

Most of the data reported in this thesis are pooled across the amygdala and the hippocampus. We also analyzed the data separately, however. Surprisingly, the differences in terms of novelty/familiarity responses are (on average) subtle. This is in agreement with previous human record data (Fried et al., 1997; Kreiman et al., 2000a). Clearly, both the amygdala and the hippocampus contain neurons which respond as described in detail in this thesis. What is remarkable, however, is that the difference between recollected and not-recollected familiar items is much more pronounced in the amygdala (Figure 4-11). The response for stimuli which are only familiar but not recollected is much smaller in the amygdala than in the hippocampus. At first, this seems surprising, as it suggests that the amygdala is more involved in recollective memory than the hippocampus is. However, an alternative interpretation is that the amygdala is proportionally more active for memories that have an emotional component. Since the emotional component is only attributed to an object if it is recollected, it seems reasonable that the response to objects which are not recollected is rather weak in the amygdala. Also note that this comparison is based on the average response to all stimuli. The stimuli we used could have an emotional value for some patients and not for others. Averaging would erase these effects. In a trial-by-trial comparison it is possible that there are stimuli which evoke a stronger amygdala familiarity/novelty response due to some stimulus property such as emotional content. This suggests a further experiment, using trial-by-trial correlations with image rankings along different dimensions (saliency, emotional content). Stimuli could either be rated by an independent subject population or a standardized dataset can be used (such as the International Affective Picture System dataset, (Lang and Cuthbert, 1993)).

The amygdala is well known to have a strong influence on memory. Emotional stimuli are remembered better than non-emotional stimuli (Heuer and Reisberg, 1990). Patients with amygdala lesions have good memory but lack the enhanced memory for emotional stimuli

(Adolphs et al., 1997; Adolphs et al., 2000; Phelps et al., 1997). Thus, the role of the amygdala in memory formation seems to be modulatory (Mcgaugh et al., 1990; Phelps, 2004). However, in some situations the amygdala is necessary for rapid (and often novelty-dependent) learning as, for example, in conditioned taste aversion (Lamprecht and Dudai, 2000) or in fear conditioning (Wilensky et al., 2006). It thus seems reasonable that neurons in the amygdala are novelty sensitive as well as plastic.

6.4 Differential response strength in epileptic tissue

All data reported in this thesis has been recorded from patients with a long history of epilepsy. Based on careful neuropsychological measures, we have argued that our patient population is not different from the normal population in their ability to learn, remember, or reason (Table 4-1). It is thus reasonable to conclude that, in the absence of seizures, their brains function comparable to normals since they achieve the same behavior. One of the primary clinical aims of intracranial electrode implantation is to determine whether seizures have a clear unilateral origin. If this is the case, unilateral resection of parts of the MTL is a possible treatment (see Introduction). Excluding all patients who did not receive a clear unilateral MTL diagnosis, we used knowledge of the laterality of the epileptic focus to compare neural responses between the epileptic and the presumably non-epileptic side. As a measure of response strength we used a response index that is equal to the absolute difference between the response to novelty and familiarity of neurons that are novelty sensitive. Using this index we find (Figure 4-11) that the response index for neurons in the epileptic hemisphere was much weaker when compared to the non-epileptic side (For recollected trials, 88% vs. 36%). Also, neurons in the epileptic side did not fire significantly differently for recollected vs. not-recollected trials (for healthy neurons there was a 20–30% difference). While we have not verified this with a predictive study, this finding nevertheless suggests potential value for this paradigm as a useful diagnostic. Due to the large difference it is imaginable that similar differences in novelty/familiarity responses also exist in multi-unit data or even LFP, which would make it easier to use this diagnostic clinically.

6.5 Predictors of successful learning

Transforming a new experience into a long-term memory is a complex process that is poorly understood. It starts with the neural activity during the initial acquisition and continues at least for hours (but probably for much longer) after initial acquisition (consolidation). At the time of retrieval, which can be many years after initial acquisition, these changes are sufficient to evoke the feeling of familiarity, sometimes together with other attributes that were part of the learning experience (an episode). While it is clear that the initial acquisition is clearly necessary for successful retrieval, it is unclear how much of the retrieval variability can be attributed directly to it rather then all the other events that contribute to a memory (Paller and Wagner, 2002). Representing a robust memory likely requires changes (plasticity) in a large number of neurons. Inducing the cellular changes thought to underlie these changes requires tightly coordinated neuronal activity. It is thus thought that one of the important contributors to successful learning is synchrony (Axmacher et al., 2006). In this thesis I show that specific components of the LFP, measured during learning, are predictive of whether a stimulus will be remembered or not. This supports the hypothesis that increased synchrony is crucial for successful learning.

6.6 The value of studying single-unit responses in humans

The observation of spike trains emitted by single neurons in the human brain while the subject is awake and engaged in a task is a tremendous opportunity. Spikes are arguably the common currency of communication of brains, and thus the appropriate units that we would like to observe and study. There are a multitude of functions that are extremely difficult (or sometimes impossible) to study in animal models (see Introduction for details). The object of study in this thesis, episodic memory, is one example of this. Recording from humans in a clinical setting has many disadvantages over animal models. For example, the experimental conditions are relatively poorly controlled. Head and eye movements can not be constrained, nor can patients be over-trained to do a task perfectly (owing to human subjects concerns). Single-unit isolation quality is often not as good as with animals (due to single-wire recordings rather then tetrodes, and to electrode movement issues). Electrode location is known only approximately and cannot be confirmed with histology. No cellular or molecular manipulations are possible. With these caveats in mind, human recordings nevertheless offer a tremendous opportunity that should be utilized as much as possible. Great care should be taken to only use this rare opportunity to address problems that are well-suited to this technique, and not better addressed in other systems. It seems of dubious value to me to simply reproduce standard experiments done in monkeys or rats to conclude that it is the same in humans. Also, there are clearly questions which are better addressed with other techniques such as surface EEG or fMRI. Examples of such experiments are questions related to which brain area responds to some particular condition. Given the restricted (and fixed) implantation sites of depth electrodes, the questions best approached with this question are distinctively different. Examples are: What subclasses of neurons respond to a given stimulus? What is the latency of the response? How does the response, trial-by-trial, relate to behavior (particularly during errors, or different confidence levels, or awareness)? How selective are the responses of the same neurons to different stimuli? What are the dynamics of interactions between neurons in the same population? These are questions which cannot be addressed using other techniques. This also stresses the importance of robust behavior. Many neuronal responses only make sense if studied in the context of an appropriately designed task where all behavioral variables are properly controlled. The true power of human recordings is combining behavior with the observation of single neurons. In the absence of behavior (such as passive viewing), many of the benefits of awake human recordings are not taken advantage of.

6.7 Note on visual tuning of MTL neurons

Many neurons in the MTL respond selectively to certain aspects of the visual input, such as its category (i.e., animal, person, house) or its identity (see Introduction for details). One curious aspect of these studies is that in almost every patient recorded, one finds such cells. This is remarkable, since in a typical recording session there are at very best several tens of neurons (and often fewer). Out of these few neurons, which are sampled entirely randomly from implanted fixed electrodes, invariably a few are tuned to the task variable (such as visual category) at hand. This indicates that the tuning of these neurons is probably not static. Rather it seems to be the case that MTL neurons are automatically tuned to all relevant attributes of a particular task. One of the earliest single-unit studies already remarked on this aspect by stating "These data suggest that MTL stimulus-specific responses represent a temporary allocation of a subset of MTL neurons to the ensemble encoding of distinct events within a given context" (Heit et al., 1988). While this aspect of sensitivity to the task has not been studied systematically, it nevertheless suggests that this aspect of MTL neuron function is distinctively different from neurons found in sensory areas, where tuning is typically thought to be static (such as receptive fields in early visual areas or even object-selective neurons in IT cortex). Alternatively, if tuning is static, this would imply that each neuron responds to many different categories (Waydo et al., 2006). However, this would make it difficult to reconcile the finding that one finds tuning to almost everything that is task relevant, whatever the task. This puzzling finding suggests further experiments such as changing the task-relevant categories or similar manipulations.

Chapter 7. Appendix A: Methods for signal-processing, analysis of spike trains, and local field potentials (LFPs)

7.1 Signal acquisition

All extracellular recordings were acquired continuously using Neuralynx Hardware (Neuralynx Inc, Tucson, AZ). We used two generations of systems: An Analog Cheetah system (25 kHz sample rate) with 32 channels, and a Digital Cheetah system (32 kHz sample rate) with 64 channels. In both systems, signals were first pre-amplified as close as possible to the source (pre-amplifiers were placed on the head of the patient). After pre-amplification, signals were fed into the acquisition system, which was located in the room of the patient (but several meters away). The acquisition system amplified the analog signals (with gain in the range of 2000–50000) and fed them into an A/D converter (analog system) or directly fed them to the A/D converter (digital system, no analog amplification). Spike times were determined offline after the recording (see spike detection and sorting chapter for details). All parts of the system that were in contact with the patient were powered by DC batteries to avoid safety problems as well as to reduce line-noise interference. The interface between the acquisition system and the recording PC (acquisition card) was optical. We used the Cheetah software to acquire all data (Neuralynx Inc, Tucson, AZ).

Each macroelectrode contained 8 microwires (see Introduction for details). One of these wires was used as ground. The choice of ground wire (based on background noise levels and impedance) occurred on the first day of recording for every patient. Special care was taken to identify a ground wire that had very low levels of electrical activity, as otherwise the activity on the groundwire would be recorded on all other wires as a signal. All our recordings were locally grounded. Thus, the measured voltage (the output of the amplifier) is the difference between the two inputs to the amplifier (differential amplification, relative): the measuring wire and the ground wire. Thus, the signal S_i (output of the amplifier) represents $S_i = V_i - V_G$, where V_i is the voltage on each microwire measured relative to a distant ground (i.e., the skull). All microwires of the same macroelectrode are located very closely together (spatially, typically < 1 mm). This kind of differential recording thus allows the measurement of very local electrical activity. All activity that is common to both wires (such as global line noise, long-range oscillations) is cancelled from the signal due to the subtraction. This has implications for the LFP signal recorded from these electrodes: It is very different from a traditional iEEG signal (see the LFP chapter for details).

Signals were acquired with the widest bandpass filter settings possible (given the level of background noise). However, emphasis was placed on recording spikes rather then LFP. Thus, if the dynamic range of the low-frequency components was too large to have appropriate amplification to see clear spikes (given the limited dynamic range), a bandpass filter was used to allow appropriate increases in gain. All gain and filter settings were determined before recording started. This limitation only applies to the first-generation system (analog) that we used. The second-generation system did not have this constraint due to the increased dynamic range of the A/D converter, which has 18 effective bits. With this system, we could always record the entire frequency band (1 Hz–9000 Hz bandpass filter).

7.2 The origin and structure of the extracellular signal

The wideband extracellular signal recorded from a microwire electrode with relatively high impedance (200 kOhm–1 MohM) and small surface area contains a mixture of electrical signals from many different sources. Electrical events in neurons occur on two fundamental timescales: i) spikes are fast events that last 0.4 – 1 ms and ii) excitatory and inhibitory post synaptic potentials (EPSPs and IPSPs), on the other hand, are slow events that last from 10–100 ms. These two timescales are reflected in the structure of the wideband extracellular signal. The high-frequency components (> 300 Hz) are dominated by spikes, whereas the low-frequency components (< 300 Hz) are dominated by synaptic events. Simulations show that spikes contribute dominantly to the 300–3000 Hz frequency band and have negligible power at lower frequencies (See Figure 15 in (Logothetis, 2002) for an insightful illustration of this fact). This is the justification for using the 300-3000Hz frequency band for extracting spikes from the extracellular signal. Simulated synaptic potentials, on the other hand, have their dominant power at frequencies lower than 150 Hz (Logothetis, 2002).

The shape of the waveform of the spikes of a particular neuron (see spike sorting chapter for examples) depends on many factors such as the location, surface shape, and impedance of the electrode, as well as neuronal morphology and type, and the expression of different ion channels (Gold, 2007; Gold et al., 2006). For this reason waveforms from different neurons recorded on the same electrode are different. We exploit this fact to attribute each waveform to a particular neuron (see spike sorting chapter). These differences can also be exploited to infer properties of the recorded neuron from the shape of the action potential. For example, inhibitory neurons have sharper waveforms than excitatory neurons (McCormick et al., 1985). This fact can be used to infer the identity of the neuron recorded from (Buzsaki and Eidelberg, 1982; Csicsvari et al., 1999; Fox and Ranck, 1981; McCormick et al., 1985; Mitchell et al., 2007; Viskontas et al., 2007). Extracellularly recorded waveforms have amplitudes on the order of 50–200 μV (peak-to-peak). Background noise levels are in the range of 5–20 μV (RMS). Given these noise levels and the fact that the amplitude decays linearly with the distance from the source, it is estimated that an extracellular electrode can record spikes from neurons within a radius of perhaps up to 140 uM (Buzsaki, 2004; Gold, 2007; Gold et al., 2006; Henze et al., 2000; Holmgren et al., 2003).

The origin of the low-frequency components of the extracellular field, the local field potential (LFP), are much less clear (Bullock, 1997). It is thought that the LFP is mostly composed of the sum of large numbers of postsynaptic discharges. It is estimated that the LFP from a single extracellular electrode is influenced by potentials within a radius of 0.5–3 mm (Juergens et al., 1999; Logothetis, 2002; Mitzdorf, 1985). Due to its (predominantly) synaptic origin, the LFP can be independent of the spiking output measured at a particular location (Logothetis, 2002). The LFP is dominated by synchronized synaptic/dendritic components of neurons that are oriented in space such that their potentials add rather then cancel. The organization of cortical pyramidal neurons yields a particularly large LFP because neurons are parallel, with dendrites in one direction and axons in the other direction. This yields an open field geometrical arrangement (Mitzdorf, 1985). Due to its dominantly synaptic origin, the LFP is thought to represent the synaptic input as well as local processing. Some have used this to argue that spikes measure the output and LFP the input to a particular area. There are cases, however, where this strict distinction does not hold. Also, the exact origin of the LFP (in general) remains unknown, and making this argument thus requires detailed knowledge about the neuronal architecture of the area under investigation.

7.3 Signal processing

7.3.1 Filtering

All filters were 4th order zero-phase-lag Butterworth filters unless otherwise noted. For spike extraction, signals were bandpass filtered between 300–3000 Hz. For LFP, signals were down-sampled to 1000 Hz sampling rate and lowpass filtered < 300 Hz. To extract specific LFP frequencies (for example 4–8 Hz), a narrow bandpass filter was applied.

7.3.2 Local field potential (LFP)

The LFP is the sum of all oscillations that influence the extracellular electrical field at the point of space where the electrode is placed. There are many different forms of oscillations of widely varying frequencies. Some of these oscillations are known to have distinct physiological mechanisms. For example, oscillations of some frequencies are only present during sleep or during motor movement. The LFP bands are traditionally (and arbitrarily) decomposed into the following frequency bands (Buzsáki, 2006; Penttonen and Buzsaki, 2003): Delta (δ, 0–4 Hz), theta (θ, 4–8 Hz) , Alpha (α, 8–12 Hz), Beta (β, 12–24 Hz), Gamma (γ, 24–100 Hz or higher for high gamma). The frequency of a particular oscillation, however, can vary substantially depending on brain state (wake, sleep, drowsy) as well as between species (Steriade et al., 1993). For example, the frequency of theta is slower in larger mammals (such as primates or cats; 3–5

Hz) compared to rodents (6–9 Hz) (Robinson, 1980). These terms should thus only be used as guidelines but not as fixed entities.

Since the recorded LFP is a mixture of many frequencies (a voltage as a function of time), it is necessary to decompose the signal into a different representation which is a function of both frequency and time W(t,f). The fundamental technique to achieve this is the Fourier transform (FT), which transforms a function of time x(t) into a function of frequency x(f) (and vice versa). While this is useful to calculate a power spectrum, all time resolution is lost. One technique to circumvent this is to split the data into small time bins and calculate the FT for each (windowed fourier transform, WFT) (Teolis, 1998). Due to the small window in time, this technique will prevent estimation of frequencies whose wavelength is less than the window size. A more sophisticated version of WFT is wavelet analysis. Wavelets (see below) are functions which are well localized in both time and frequency. Their effective window size is adapted based on the frequency and is thus always optimal. Here, wavelets or the Hilbert transform were used to compute a continuous estimate of power and phase as a function of time.

Time-frequency decomposition using wavelets: The raw signal $S(t)$ was decomposed into a function of frequency and time using the continuous wavelet transform (cwt). In the following I am using the notation developed in (Torrence and Compo, 1998). The mother wavelet used was always a complex Morlet wavelet:

$$\psi_0(\eta) = \frac{1}{\sqrt{\pi\omega}} \exp(i2\pi\eta f_0) \exp(-\frac{\eta^2}{2})$$

The two parameters are the center frequency f_0 and the number of cycles. We used $f_0 = 1$ and $\omega = 4$ cycles, unless mentioned otherwise (see below).

The cwt of the raw signal $S(t)$ is a function of both scale (frequency) and time: $W(t,s)$. It is computed by convolving the raw signal (of length N) with the wavelet function $\psi_0(\eta)$ for a number of different frequencies (scales) s.

$$W(t,s) = \sum_{t'=0}^{N-1} S(t')\psi^*\left[\frac{(t'-t)\Delta t}{s}\right]$$

$\psi^*(\eta)$ is the complex conjugate of the wavelet function $\psi(\eta)$. $\psi(\eta)$ is a normalized version of the wavelet $\psi_0(\eta)$. See (Torrence and Compo, 1998) for details.

The effective resolution of the Morlet wavelet depends on the center frequency f_0 and the scale s. If δT is the spacing between two sampled points (due to the sampling rate), the effective frequency of a Morlet wavelet at scale s is $f = \frac{f_0}{s\delta T}$. Thus, the higher the scale, the lower the frequency. The resolution is measured separately in terms of the standard deviation in time σ_t and frequency σ_f. Time resolution at scale s is $a\delta T$ and frequency resolution is $\frac{\sigma_f}{a}$. Thus, the better the resolution in time the worse it is in frequency and vice versa (uncertainty principle, a fundamental limit, dictates $\sigma_t \sigma_f \leq \frac{1}{2\pi}$). The time width of a wavelet is defined as (Najmi and Sadowsky, 1997):

$$\sigma_t^2 = \frac{\int_{-\infty}^{\infty} t^2 \psi^2(t)dt}{\int_{-\infty}^{\infty} \psi^2(t)dt}$$

Thus, $\sigma_f = \frac{1}{2\pi s \sigma_t}$. To illustrate this trade-off, Figure 7-1 shows Morlet wavelets in both time and frequency space for 3 different parameter combinations. The time and frequency resolution for the same 3 wavelets are shown in Figure 7-2. Notice the trade-off between accuracy in time and frequency clearly visible from the size of the error bars in Figure 7-2 (bottom row). Since the width in frequency space increases as a function of frequency, the frequencies at which the wavelets are calculated are logarithmically scaled. This leads to an even sampling in frequency space (Figure 7-2). Here, we sampled at frequencies of $f = 2^x$ with $x \in [2:2:52]/8$ (not all are shown in Figure 7-2).

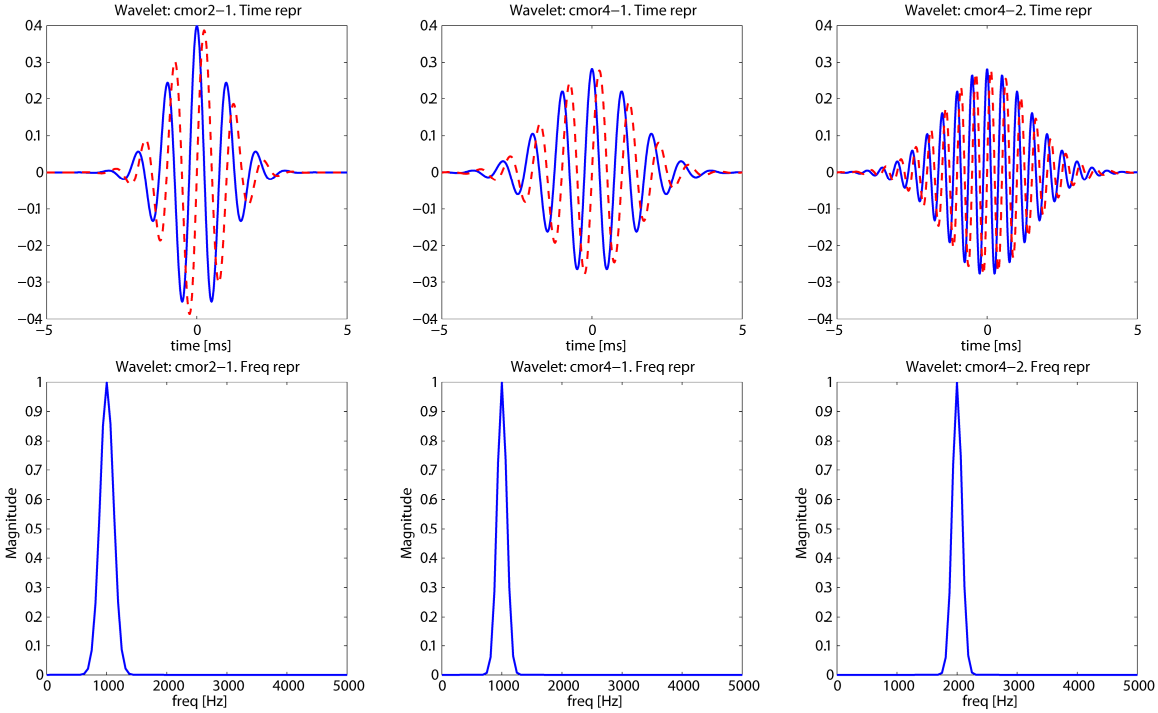


Figure 7-1. Illustration of the complex morlet wavelet.

The wavelet is illustrated for 3 different combinations of parameters of cycle number and center frequency: (4,1), (6,1), and (6,2) (from left to right). The top row shows the wavelet in time (blue: real part; red: complex part) and the bottom row in frequency (Fourier transform of the above). Notice the tight tuning in both time and frequency.

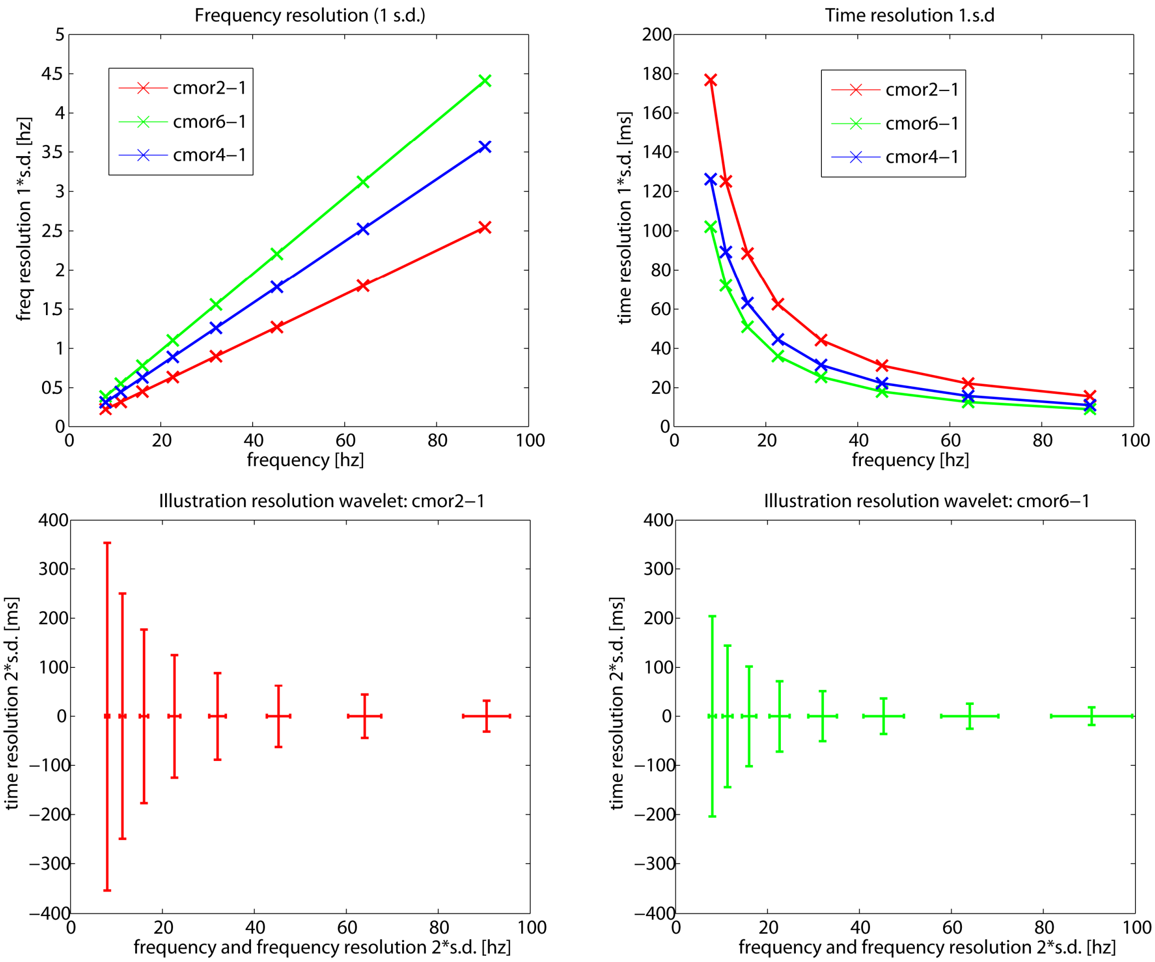


Figure 7-2. Illustration of the trade-off between specificity in time and frequency space.

Illustrated is the complex morlet wavelet for three different parameters (cycle number and center frequency): (2,1), (4,1), and (6,1). The top row shows the frequency resolution (left) and the time resolution (right). The y axis shows one standard deviation as a function of frequency. Note that whenever one wavelet has better time resolution (left, red) it has worse frequency resolution (right, red) and vica-versa. The bottom row illustrates this property by showing the 95% confidence interval (±2*s.d.) for both time (y axis) and frequency (y axis). Note how the left wavelet (2,1) has better resolution in frequency compared to the wavelet on the right (6,1). However, the left wavelet has better time resolution. Only a subset of the frequencies used for the analyis are shown (every second is shown). Note that in the bottom row, the y axis is in log units, and thus the error bars appear to be of equal length.

Computation of the analytic signal with the Hilbert transform: To estimate the phase and power of a narrowly bandpass-filtered signal without using the wavelet transform (such as in the theta band), the Hilbert transform was used to calculate an analytical signal. The analytical signal $X(t)$ is complex and can be used to calculate the phase/power with the same methods as for wavelet coefficients (see below). The Hilbert transform $S_H(t)$ is equal to the signal phase shifted by 90°. The real part of the analytic signal equals the raw signal and the complex part is the Hilbert transformed signal.

$$X(t) = S(t) + iS_H(t)$$

Estimation of instantaneous phase and power (energy): Given a complex signal as a function of time $X(t)$, the following methods were used to estimate the instantaneous phase $\phi(t)$ and power $R(t)$. $X(t)$ is either the result of a Hilbert transform or a continuous wavelet transform (see above). In the following, $\Re\{X\}$ and $\Im\{x\}$ refer to the real and imaginary part of X, respectively.

$$R(t) = \Re\{X(t)\}^2 + \Im\{X(t)\}^2$$

$$\phi(t) = \arg(X(t)) = \operatorname{atan}2(\Im\{X(t)\}, \Re\{X(t)\})$$

Wavelet power spectrum and distribution of wavelet power: The wavelet power spectrum is equivalent to $R(t)$ as defined above, as a function of frequency. The real and the imaginary parts of the wavelet coefficients $W(t,s)$ for any particular scale are normally distributed random variables with mean 0 and unknown variance. Since the wavelet power $R(t)$ is the squared sum of the real and imaginary part of $W(t,s)$, $R(t)$ is χ^2 distributed with 2 degrees of freedom. Since the variance is unknown (but not 1), however, the mean of this variable is unknown. Since the LFP is 1/f distributed, the mean of this distribution is a function of the frequency. The χ^2 distribution needs to be scaled appropriately to allow statistical tests (Caplan et al., 2003).

An alternative approach is to normalize the real and complex part of $W(t,s)$ to a variance of 1 independently before calculating the power. This removes the 1/f frequency dependency and allows easy statistical comparisons with a unscaled χ^2 distribution. I estimated the variance of $W(t,s)$ (separately for the real and imaginary part) for each scale for the entire experiment and normalized (divided) all samples by this value to assure that each are normally distributed with mean zero and variance 1. This allows the construction of a flat power spectrum (instead of 1/f) where peaks correspond to a deviation from the null hypothesis of no signal.

Statistics of phase locking: All statistics related to phases were performed using circular statistics (Batschelet, 1981; Fisher, 1993). The phase was measured (in radians) in the range $[-\pi ... \pi]$ (-180°–+180°) with 0 equal to the peak and $-\pi / \pi$ equal to the through of the oscillation. Statistics of a sample of n angles θ_i (phases) were calculated based on the mean resultant vector:

$$C = \sum_{i=1}^{n} \cos(\theta_i)\,,\ S = \sum_{i=1}^{n} \sin(\theta_i)\,,\ R^2 = C^2 + S^2\,,\ \overline{R} = \frac{R}{n}$$

The mean angle $\overline{\theta}$ is also calculated from above measures:

$$\cos(\overline{\theta}) = \frac{C}{R},\ \sin(\overline{\theta}) = \frac{S}{R}$$

The larger the length of the mean resultant $\overline{R}$ (range 0–1), the stronger the phase locking of the sample. The sample circular variance is $V = 1 - \overline{R}$. To test whether a neuron is significantly phase locked, the sample of all phase angles was compared against uniformity using a Rayleigh test. The Rayleigh test is based on the length of $\overline{R}$:

$$Z = n\overline{R}^2\,,\ P = \exp(-Z)\left[1 + \frac{2Z - Z^2}{4n} - \frac{24Z - 132Z^2 + 76Z^3 - 9Z^4}{288n^2}\right]$$

If P is sufficiently small, the null hypothesis of uniformity can be rejected. The alternative hypothesis is that the data is unimodal (one mean direction). To quantify the distribution of a sample of phase values that was significantly non-uniform, we fit a Von Mises distribution to the data using maximum likelihood. The Von Mises distribution is the normal distribution adapted for circular data. The following is its density function:

$$f(\theta) = \frac{1}{2\pi I_0(\kappa)} \exp(\kappa \cos(\theta - \mu)$$

It is fully specified by a mean direction μ and a concentration parameter κ. The concentration parameter is analogous to the standard deviation of a normal distribution, although of opposite direction: The larger κ, the more concentrated the distribution (the smaller its variance). For $\kappa = 0$, the Von Mises distribution is equivalent to the uniform distribution on the circle. $I_0()$ is the modified Bessel function of order zero. A definition of it can be found in (Fisher, 1993).

Simulated LFP: For systematic evaluation of our methods we used artificially simulated LFP which has a phase spectrum similar to real data ("red noise", i.e., "1/f"). Such LFP was simulated using sinusoidal pink noise (Cohen, 1995; Rohani et al., 2004):

$$X(t) = \sum_{i=1}^{N/2} \sin(\frac{2\pi t}{i}) + \phi_i)\sqrt{\frac{i}{N}}$$

This generates a time series of length N. The phase is sampled randomly from a uniform distribution, i.e., $\phi_i \in U[0,2\pi]$.

7.4 Spike train analysis

7.4.1 Single neurons: Spike times and the distribution of interspike intervals

In the following I describe the statistical properties of a series of spikes (a “spike train”). In probability theory, this is commonly referred to as a point process. For details and proofs refer to (Dayan and Abbott, 2001; Gabbiani and Koch, 1999; Kass et al., 2005; Koch, 1999).

For the purposes of analysis, spikes are treated as unitary events that occur at a particular point of time t_i. Spikes emitted by real neurons last 0.5–1.5 ms and are thus not restricted to a single point of time. Here, the peak (maximal deviation from baseline) of the waveform is used as the point of time the neuron spikes. The measurement accuracy of t_i is restricted by the sampling rate and uncertainty in determining the peak. Here, the accuracy is estimated to be on the order of 0.1 ms. Time is measured relative to a fixed reference point, such as the start of the experiment or trial. The unit of time is usually assumed to be milliseconds (ms), but any units can be used. Observing the N spikes emitted by a single identified neuron leads to a set of spike times $T = \{t_1, t_2, ..., t_N\}$. T is thus a list of events emitted by a point process. Observing the properties of T allows us to make inferences about the properties of this point process (which here is equal to a single neuron). The most important measure to quantify the behavior of a point process is the interspike interval (ISI). The interspike intervals are defined as the times between two neighboring spikes, i.e., $I_1 = t_2 - t_1, I_2 = t_3 - t_{2,...}$. The set of ISIs $I = \{I_1, I_2, ..., I_{N-1}\}$ is the set of all differences between neighboring spikes. The shape of the distribution of the ISIs can be used to infer a great number of properties about the neuron that emitted the spikes. Examples are:

inferences about the firing rate (the mean), the variability of the firing rate, bursting behavior, or whether the neuron fires periodically. Also, the shape of the ISI can be used to judge whether the set of spikes used for calculating it could have been emitted by a single neuron or not. This can be used to judge spike sorting quality (see the spike sorting chapter for details).

The spikes fired by a neuron are, in the great majority of cases, Poisson distributed (Dayan and Abbott, 2001; Holt et al., 1996; Softky and Koch, 1993). Due to biophysical constraints (such as the refractory period), neurons cannot fire at extremely high firing rates. Thus, it is unlikely that a neuron will fire more than one spike within approximately 3 ms (although there are cell types which have a shorter refractory period). For this reason, the firing probability at any particular point of time is a function of the time since the last spike. Such a process is modeled as a renewal process. For this reason the intervals deviate systematically from a pure Poisson distribution.

Given a homogenous Poisson process with rate r, the probability of observing n spikes within a time period T is:

$$P_T[n] = \frac{(rT)^n}{n!} e^{-rT}$$

The Poisson process is entirely defined by the rate r. Given a Poisson process, the waiting times between two spikes are exponentially distributed:

$$P_{ISI}(\tau) = re^{-r\tau}$$

The above function specifies the probability that, given a spike at $t = 0$, no spike will have occurred in the interval $t + \tau$. For a homogenous Poisson process this is the expected shape of the ISI distribution. Given a sample of ISIs, the mean ISI is $\langle ISI \rangle = \frac{1}{r}$ and the variance is $\sigma^2 = \left\langle \left[ISI - \langle ISI \rangle \right]^2 \right\rangle = \frac{1}{r^2}$. The mean and variance of the underlying Poisson process can thus be calculated from the ISI distribution.

If an absolute refractory period t_{ref} is introduced, this function is shifted (to the right on the time axis) by t_{ref}: $P_{ref}(\tau) = P_{ISI}(\tau - t_{ref})$. Another possibility for expression of the ISI distribution of a neuron with a refractory period is to use a gamma distribution:

$$P_{ISI}(\tau) = \frac{r(r\tau)^k e^{-r\tau}}{k!}$$

This representation has two parameters: the rate r and the parameter k, which is the order of the gamma distribution. When $k = 0$ an exponential distribution results. With $k > 0$ (estimated from the data using maximum likelihood), this typically provides a very good fit for ISI distributions. Also note that the mean of a gamma process with $k > 0$ is the same as for $k = 0$. The estimate of the mean rate as $\langle ISI \rangle = \frac{1}{r}$ remains thus valid, regardless of the order of the gamma process.

Computationally, a random series of spike times that are Poisson distributed can be generated by sampling randomly from an exponential distribution (and discarding the ones which are less than the refractory period). The returned numbers are wait times (interspike intervals). This is the strategy that we used whenever random spike trains were generated.

Also note that (under the Poisson assumption) the standard deviation of the ISI distribution is equal to its mean (the rate). Thus, the variance is not independent of the mean for neurons. The ratio of the standard deviation to the mean of the ISI of a perfectly homogenous Poisson process is thus equal to 1. This ratio is the coefficient of variation (CV):

$$CV_{ISI} = \frac{\sigma}{r}$$

The CV is an important measure of the regularity of firing of a neuron. A neuron that fires perfectly at a single rate has a CV of 0. A neuron that fires perfectly according to a Poisson distribution has a CV of 1.0. A neuron with highly irregular firing (for example, complex spikes, bursts) will have a CV > 1. The CV is routinely calculated for recorded neurons. There is a wide range of observed CV values. For neocortical neurons, it is often close to 1, as expected (Britten et al., 1993; Shadlen and Newsome, 1998; Softky and Koch, 1993; Tomko and Crapper, 1974). The measured relationship between mean rate and variance is approximately 1.5 (whereas the theoretical prediction is 1.0) (Shadlen and Newsome, 1998). A refractory period will impose some form of regularity and thus lowers the CV. A perfect Poisson neuron with a refractory period will thus have a CV < 1.

Chapter 8. Appendix B: Population decoding: principles and methods

8.1 Motivation and principles

A principal goal of systems neuroscience is to understand what information is present in spike trains and in which form (“neural coding”). Neurons coding for a particular variable or stimulus feature are typically identified by comparing the mean firing rate between two conditions, repeated over many trials. While this identifies neurons that differ in firing rate on average, it tells us little about how a downstream region (that receives this signal) might use this information. In reality, the brain can not average over several trials. The relevant unit of information is thus what can be decoded from a single trial and not an average. Of course a downstream region receives the simultaneous outputs of many neurons rather then just one. Thus, some form of averaging between neurons can take place for a single trial. This does not necessarily improve the information content, however. Spiking of pairs of neighboring neurons can be highly correlated, and averaging correlated signals can either increase or decrease information content based on the exact nature of the correlations (Abbott and Dayan, 1999; Mazurek and Shadlen, 2002; Seung and Sompolinsky, 1993; Sompolinsky et al., 2001). External noise that influences the spiking of several neighbouring neurons results in positive correlations, as does common input or recurrent connectivity. This can lead to a decrease in information due to correlations (Sompolinsky et al., 2001), and imposes a fundamental constraint on the amount of information that can be represented by a population of neurons. Averaging a large number of units would thus not necessarily improve information content. This effect has been demonstrated experimentally by recording from pairs of neurons in area MT of non-human primates (Zohary et al., 1994). Surprisingly, the spikes emitted by a single motion-selective MT unit allow perceptual discrimination that is as good as the psychophysical sensitivity of the animal. Thus, the sensitivity of a single neuron is indistinguishable from the sensitivity of the entire animal. The spiking of pairs of neurons was weakly positively correlated. In this particular instance this correlation resulted in no increase in information content if more than 50–100 units were averaged (Zohary et al., 1994). In this case, averaging could only improve the signal-to-noise ratio by 2–3 times (even if large numbers of units were considered).

How is information represented in single neurons as well as populations thereof? Comparing the output of neurons between different conditions requires the use of features of spike trains. The simplest example of a feature is counting the number of spikes emitted per units of time ("rate code"). However, many other features could possibly contain information as well, such as the time of occurrence of a spike ("temporal code"). Examples of more complex features are the number of spikes that are less than 5 ms apart (a burst), the correlations between different units, or the phase relationship of a spike to a particular frequency of the local field potential. Cortical as well as hippocampal neurons fire highly irregularly (Fenton and Muller, 1998; Shadlen and Newsome, 1998; Softky and Koch, 1993). The variability of spike times generally increases linearly with the mean firing rate (Poisson). Thus, increasing the firing rate alone does not reduce response variability. This inherent uncertainty imposes constraints on possible useful features, as any feature that relies on highly accurate timing is unlikely to be useful in general (with some exceptions). There are many proposals on what the fundamental unit of neural coding is (Abeles, 1991; MacKay and McCulloch, 1952; Rieke, 1997). While there is little agreement, it is acknowledged that the simplest of all codes, spike counts, is surprisingly good. In some cases, incorporating time in addition to rate improves decoding performance. There are only a few demonstrated examples where a rate code alone does not allow reading out of a substantial amount of the information available (Butts et al., 2007; Johansson and Birznieks, 2004; Laurent, 2002; MacLeod et al., 1998; Montemurro et al., 2007; Stopfer et al., 1997). This statement only refers to readout of information and does not imply that precise spike timing is not important for other processes, such as plasticity (discussed elsewhere in this thesis). However, since there are exponentially more possible combinations the more neuronal features are used, a rate code can transmit much less information per unit of time compared to a more complex code. Population decoding is a powerful technique that can address questions of coding and the feature used (see below).

An alternative method for quantifying the information present in the output of a population of neurons is decoding. This method avoids averaging entirely (both across trials and across neurons). Mathematically, a decoder is a function $y = f(x)$ that takes as input information about all available neurons (such as firing rate or spike times) and gives as output a prediction of what the input represents. For example, x could be the response of a sensory neuron in response to a stimulus y. The sensory neuron transforms the stimulus y into a neural response $x = g(y)$. The task of the decoder is to reverse this and reconstruct, from only observing the neuronal response x, the input y. Thus in this example $f = g^{-1}$. Traditional neurophysiology focuses on establishing the transformation of input to neuronal output, i.e., $x = g(y)$. However, from the perspective of the brain, estimating ("decoding") the external world from the neuronal responses is the relevant task (Bialek et al., 1991). $f(x)$ can be determined automatically by a machine learning algorithm ("classifier") from a subset of the data. Alternatively, f(x) can also be pre-defined by a model with a few parameters estimated from the data. The performance of such a decoder, on a separate test set, is an estimate of how much information about the state of the neural system can be extracted, on a trial-by-trial basis, by the decoding technique used. Note the importance of an entirely independent test set. One mistake, for example, is to pre-select neurons (or time bins for LFP) using a statistical test and then estimate decoding performance using leave-one-out cross-validation. Such a test set is not independent: the samples were already used to pre-select the input. Thus the decoder will, by definition, return above-chance performance (as the inputs passed a statistical test). While this can still be meaningful (for example to estimate how difficult it is to decode on a single trial or compare different conditions), the fact that the information is present as such is not a new finding. If the particular classifier used has a structure that could conceivably be implemented by a network of realistic neurons (such as a linear weighted sum, as used in this thesis), one can reasonably assume that a real neural network could be reading out this information in a similar fashion. If, on the other hand, the information can only be extracted by a mechanism that can hardly be implemented biologically (such as requiring sub-millisecond resolution), it is less plausible that this information could be read out by another brain area.

A decoding approach is particularly useful to quantify how much information is present about the state of the system in a population of neurons. Plotting decoding performance as a function of variables such as number of units, time relative to stimulus onset, time resolution (binning), or the neuronal code used allows quantification of such effects. Also, decoding can determine what state the neural population represented during a behavioral error (such as forgetting a picture or making the wrong perceptual decision). This allows one to clearly discern whether an area follows the decision, the motor output, or the sensory input. Another question that can be investigated using decoding is the latency of the response: By what point in time does a population of neurons distinguish best between two particular stimuli? Decoding can also be used to distinguish between different kind of neuronal codes: Does a code that distinguishes exact spike timing contain more information then a rate code? The answer to this question will depend on the exact experiment and experimental model but can easily be investigated with decoding.

There is a trade-off between readout difficulty and the robustness and richness of the code. The easiest code to read out would be if every neuron represents a particular concept exclusively ("grandmother neuron"). Thus decoding is trivial — counting the number of spikes emitted by this neuron is sufficient. The number of represented concepts is limited by the number of neurons. This representation is, however, not robust. If this single neuron dies, the representation is lost. On the other hand a fully distributed code can represent vastly more concepts (2^N), but such a representation is very difficult to read out—a decoder needs access to all neurons simultaneously. The decoding approach to neuronal spike train analysis is useful to specify the difficulty of readout, because a given decoder quantifies how much information can be read out with the given complexity of the decoder.

There are a large variety of techniques for constructing population decoders. These include simple linear-sum type decoders such as the perceptron and go all the way to highly non-linear support vector machines (SVMs). In the following section the decoding techniques used in this thesis are summarized. In our experience the performance of highly complex decoders is often only marginally better than simple linear decoders (in the context of the analysis performed here). Thus, in the interest of understanding neural coding rather then machine learning, it is (in our opinion) often advisable to use a very simple linear decoder. If it turns out that higher-order interactions between terms are important, it is still possible to include these in a linear decoder by introducing additional variables that represent the higher-order terms.

Apart from being a valuable approach for data analysis, single-trial decoding of neural activity also has practical applications. It has been demonstrated that small numbers of neurons in an appropriate area of the brain provide enough information to allow a human or monkey to remote-control a robotic device (or a computer) by thought only (Andersen et al., 2004; Carmena et al., 2003; Hochberg et al., 2006; Rizzuto et al., 2005; Serruya et al., 2002). This is a direct demonstration of the value of decoding approaches.

8.2 Definitions

The matrix Z contains the data samples. Rows correspond to training samples (n) and columns to variables (p, such as spike counts). Thus, Z is of dimensionality $n \times p$. Z contains a column of 1s to account for a linear offset relative to the mean. Each sample has one numerical label (for example -1 or 1, in the binary case), which is stored in the vector y of dimensionality $n \times 1$. The vector of weights w is of dimensionality $n \times 1$.

8.3 Multiple linear regression

With multiple linear regression (Eq B1), the weights w are determined by multiplying the inverse of data samples Z with the training labels y (Johnson and Wichern, 2002).

$$w = [Z'Z]^{-1} Z'y \qquad \text{(Eq B1)}$$

8.4 Regularized least-square regression (RLSC)

Multiple linear regression can not determine the weight vector unambiguously if the sample matrix is ill conditioned or can not be inverted for other reasons (Eq B1). Even if the matrix can be inverted, the resulting matrix can be numerically unstable. In practice, such problems often arise due to larger number of variables then training samples (i.e., 100 neurons and 50 trials). Another common source is linear dependency between variables, which leads to rank-deficiency.

To circumvent this problem, additional constraints (on the weights) such as smoothness or small numerical values need to be enforced. One way to achieve this is to add a constant term (regularizer). Here, we used regularized least squares (RLSC) to achieve this. In RLSC (Evgeniou et al., 2000; Hung et al., 2005; Rifkin et al., 2003), an additional term is added to the data samples (Eq B2). Here, I is the identity matrix and λ is a scalar parameter (the regularizer).

$$w = [Z'Z + \lambda I]^{-1} Z'y \qquad \text{(Eq B2)}$$

The value of the regularizer is arbitrary. The bigger it is, the more constraints are placed on the solution (the less the solution is determined by the data samples). A small value of the regularizer, on the other hand, makes the solution close to the multiple linear regression solution. Importantly, however, even a small value of the regularizer punishes unrealistically large weights and also guarantees full rank of the data matrix. Regularization becomes particularly important when there are a large number of input variables relative to the number of training samples.

The value of the regularizer is an important determinant of the performance of the classifier. Thus, great care must be taken to choose it appropriately. In our experience, the exact value of the regularizer is not very important (not sensitive). It is enough to get it approximately right (i.e., 10 or 100). If the only aim is to make multiple linear regression numerically stable it is sufficient to add a very small regularizer, such as $\lambda = 0.01$ (this value was used in this thesis, unless stated otherwise). Increasing the value of λ will lead to an increased training error but a decreased testing error (at least initially). It is an indicator that regularization is necessary (due to overfitting) if this pattern is observed. Thus it is possible to find a good regularizer value by plotting the testing and training error as a function of λ. The test set used to optimize the value of λ has to be different than the test set used to determine classifier performance. This is only true if λ is optimized in this way. If, on the other hand, λ is constant, no separate test set is required.

One problem with regularization is bias. For illustration, assume a binary classification problem where 80% of the samples are of one class (with label 1). A classifier trained (using RLSC) with a large λ on this dataset will predict “1” for every test sample. Due to the inherent bias in the prior distribution, the classifier will achieve 80% correct performance entirely due to chance. Thus, while this is a binary classification problem, the chance level for this regularizered classifier is 80% rather then 50%. If the bias in the test set is different than in the training set, the chance performance is unclear. One is often presented with this situation in the context of the analysis of populations of neurons, such as training on behaviorally correct trials and then decoding the error trials (which typically have different bias). To circumvent this problem, the number of samples in each class needs to be balanced artificially before training the classifier (by removing samples from the bigger class). The excluded samples can nevertheless be used by training the same classifier multiple times based on a randomly sampled subset of the training data. If excluding samples is not an option, the chance level on the test set needs to be established using a bootstrap procedure. It can not be assumed to be 50%.

References

Abbott, L.F., and Dayan, P. (1999). The effect of correlated variability on the accuracy of a population code. Neural Computation *11*, 91-101.

Abeles, M. (1991). Corticonics: neural circuits of the cerebral cortex (Cambridge: Cambridge University Press).

Abeles, M., and Goldstein, M.H. (1977). Multi-Spike Train Analysis. Proceedings of the IEEE *65*, 762-773.

Acsady, L., Kamondi, A., Sik, A., Freund, T., and Buzsaki, G. (1998). GABAergic cells are the major postsynaptic targets of mossy fibers in the rat hippocampus. J Neurosci *18*, 3386-3403.

Adcock, R.A., Thangavel, A., Whitfield-Gabrieli, S., Knutson, B., and Gabrieli, J.D. (2006). Reward-motivated learning: mesolimbic activation precedes memory formation. Neuron *50*, 507-517.

Adolphs, R., Cahill, L., Schul, R., and Babinsky, R. (1997). Impaired declarative memory for emotional material following bilateral amygdala damage in humans. Learning & Memory *4*, 291-300.

Adolphs, R., Tranel, D., and Denburg, N. (2000). Impaired emotional declarative memory following unilateral amygdala damage. Learning & Memory *7*, 180-186.

Aggleton, J.P. (2000). The Amygdala: A Functional Analysis (Oxford University Press).

Aksenova, T.I., Chibirova, O.K., Dryga, O.A., Tetko, I.V., Benabid, A.L., and Villa, A.E. (2003). An unsupervised automatic method for sorting neuronal spike waveforms in awake and freely moving animals. Methods *30*, 178-187.

Allman, J.M., Hakeem, A., Erwin, J.M., Nimchinsky, E., and Hof, P. (2001). The anterior cingulate cortex - The evolution of an interface between emotion and cognition. Unity of Knowledge: The Convergence of Natural and Human Science *935*, 107-117.

Althoff, R.R., and Cohen, N.J. (1999). Eye-movement-based memory effect: A reprocessing effect in face perception. Journal of Experimental Psychology-Learning Memory and Cognition *25*, 997-1010.

Amaral, D.G., Ishizuka, N., and Claiborne, B. (1990). Neurons, numbers and the hippocampal network. Prog Brain Res *83*, 1-11.

Amaral, D.G., and Lavenex, P. (2007). Hippocampal Neuroanatomy. In The Hippocampus Book, P. Andersen, R. Morris, D.G. Amaral, T. Bliss, and J. O'Keefe, eds. (Oxford University Press), pp. 37-114.

Andersen, P., Morris, R., Amaral, D.G., Bliss, T., and O'Keefe, J. (2007). The Hippocampus Book (Oxford University Press).

Andersen, R.A., Musallam, S., and Pesaran, B. (2004). Selecting the signals for a brain-machine interface. Curr Opin Neurobiol *14*, 720-726.

Asaad, W.F., Rainer, G., and Miller, E.K. (1998). Neural activity in the primate prefrontal cortex during associative learning. Neuron *21*, 1399-1407.

Atiya, A.F. (1992). Recognition of multiunit neural signals. IEEE Trans Biomed Eng *39*, 723-729.

Axmacher, N., Mormann, F., Fernandez, G., Elger, C.E., and Fell, J. (2006). Memory formation by neuronal synchronization. Brain Res Rev *52*, 170-182.

Badre, D., and Wagner, A.D. (2007). Left ventrolateral prefrontal cortex and the cognitive control of memory. Neuropsychologia *45*, 2883-2901.

Bancaud, J., Brunet-Bourgin, F., Chauvel, P., and Halgren, E. (1994). Anatomical origin of deja vu and vivid 'memories' in human temporal lobe epilepsy. Brain *117 (Pt 1)*, 71-90.

Bankman, I.N., Johnson, K.O., and Schneider, W. (1993). Optimal detection, classification, and superposition resolution in neural waveform recordings. IEEE Trans Biomed Eng *40*, 836-841.

Batschelet, E. (1981). Circular statistics in biology (London ; New York: Academic Press).

Bayley, P.J., and Squire, L.R. (2002). Medial temporal lobe amnesia: Gradual acquisition of factual information by nondeclarative memory. J Neurosci *22*, 5741-5748.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B-Methodological *57*, 289-300.

Berger, T.W., Alger, B., and Thompson, R.F. (1976). Neuronal substrate of classical conditioning in the hippocampus. Science *192*, 483-485.

Berry, S.D., and Thompson, R.F. (1978). Prediction of learning rate from the hippocampal electroencephalogram. Science *200*, 1298-1300.

Bi, G.Q., and Poo, M.M. (1998). Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. J Neurosci *18*, 10464-10472.

Bialek, W., Rieke, F., de Ruyter van Steveninck, R.R., and Warland, D. (1991). Reading a neural code. Science *252*, 1854-1857.

Bjorklund, A., and Dunnett, S.B. (2007). Dopamine neuron systems in the brain: an update. Trends Neurosci *30*, 194-202.

Bliss, T.V., and Lomo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. J Physiol *232*, 331-356.

Boss, B.D., Peterson, G.M., and Cowan, W.M. (1985). On the number of neurons in the dentate gyrus of the rat. Brain Res *338*, 144-150.

Botvinick, M., Nystrom, L.E., Fissell, K., Carter, C.S., and Cohen, J.D. (1999). Conflict monitoring versus selection-for-action in anterior cingulate cortex. Nature *402*, 179-181.

Brainard, D.H. (1997). The Psychophysics Toolbox. Spatial Vision *10*, 433-436.

Bremaud, P. (2002). Mathematical Principles of Signal Processing. Fourier and Wavelet Analysis. (Springer).

Britten, K.H., Newsome, W.T., Shadlen, M.N., Celebrini, S., and Movshon, J.A. (1996). A relationship between behavioral choice and the visual responses of neurons in macaque MT. Vis Neurosci *13*, 87-100.

Britten, K.H., Shadlen, M.N., Newsome, W.T., and Movshon, J.A. (1993). Responses of neurons in macaque MT to stochastic motion signals. Vis Neurosci *10*, 1157-1169.

Brown, J.W., and Braver, T.S. (2005). Learned predictions of error likelihood in the anterior cingulate cortex. Science *307*, 1118-1121.

Brown, M.W., and Aggleton, J.P. (2001). Recognition memory: what are the roles of the perirhinal cortex and hippocampus? Nat Rev Neurosci *2*, 51-61.

Buhl, E.H., Otis, T.S., and Mody, I. (1996). Zinc-induced collapse of augmented inhibition by GABA in a temporal lobe epilepsy model. Science *271*, 369-373.

Bullock, T.H. (1997). Signals and signs in the nervous system: the dynamic anatomy of electrical activity is probably information-rich. Proc Natl Acad Sci U S A *94*, 1-6.

Bunzeck, N., and Duzel, E. (2006). Absolute coding of stimulus novelty in the human substantia nigra/VTA. Neuron *51*, 369-379.

Bush, G., Whalen, P.J., Rosen, B.R., Jenike, M.A., McInerney, S.C., and Rauch, S.L. (1998). The counting Stroop: an interference task specialized for functional neuroimaging--validation study with functional MRI. Hum Brain Mapp *6*, 270-282.

Butts, D.A., Weng, C., Jin, J., Yeh, C.I., Lesica, N.A., Alonso, J.M., and Stanley, G.B. (2007). Temporal precision in the neural code and the timescales of natural vision. Nature *449*, 92-95.

Buzsaki, G. (1998). Memory consolidation during sleep: a neurophysiological perspective. J Sleep Res *7 Suppl 1*, 17-23.

Buzsaki, G. (2004). Large-scale recording of neuronal ensembles. Nat Neurosci *7*, 446-451.

Buzsáki, G. (2006). Rhythms of the brain (Oxford: Oxford University Press).

Buzsaki, G., and Eidelberg, E. (1982). Direct afferent excitation and long-term potentiation of hippocampal interneurons. J Neurophysiol *48*, 597-607.

Buzsaki, G., Leung, L.W., and Vanderwolf, C.H. (1983). Cellular bases of hippocampal EEG in the behaving rat. Brain Res *287*, 139-171.

Cameron, K.A., Yashar, S., Wilson, C.L., and Fried, I. (2001). Human hippocampal neurons predict how well word pairs will be remembered. Neuron *30*, 289-298.

Caplan, J.B., Madsen, J.R., Schulze-Bonhage, A., Aschenbrenner-Scheibe, R., Newman, E.L., and Kahana, M.J. (2003). Human theta oscillations related to sensorimotor integration and spatial learning. J Neurosci *23*, 4726-4736.

Caporale, N., and Dan, Y. (2008). Spike Timing-Dependent Plasticity: A Hebbian Learning Rule. Annu Rev Neurosci.

Carmena, J.M., Lebedev, M.A., Crist, R.E., O'Doherty, J.E., Santucci, D.M., Dimitrov, D.F., Patil, P.G., Henriquez, C.S., and Nicolelis, M.A. (2003). Learning to control a brain-machine interface for reaching and grasping by primates. PLoS Biol *1*, E42.

Carter, C.S., Braver, T.S., Barch, D.M., Botvinick, M.M., Noll, D., and Cohen, J.D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. Science *280*, 747-749.

Casey, B.J., Trainor, R.J., Orendi, J.L., Schubert, A.B., Nystrom, L.E., Giedd, J.N., Castellanos, F.X., Haxby, J.V., Noll, D.C., Cohen, J.D*., et al.* (1997). A developmental functional MRI study of prefrontal activation during performance of a Go-No-Go task. Journal of Cognitive Neuroscience *9*, 835-847.

Chandra, R., and Optican, L.M. (1997). Detection, classification, and superposition resolution of action potentials in multiunit single-channel recordings by an on-line real-time neural network. IEEE Trans Biomed Eng *44*, 403-412.

Chang, B.S., and Lowenstein, D.H. (2003). Epilepsy. N Engl J Med *349*, 1257-1266.

Chapin, J.K. (2004). Using multi-neuron population recordings for neural prosthetics. Nature Neuroscience *7*, 452-455.

Chapman, L.F., Walter, R.D., Markham, C.H., Rand, R.W., and Crandall, P.H. (1967). Memory changes induced by stimulation of hippocampus or amygdala in epilepsy patients with implanted electrodes. Trans Am Neurol Assoc *92*, 50-56.

Chapman, P.F., Kairiss, E.W., Keenan, C.L., and Brown, T.H. (1990). Long-Term Synaptic Potentiation in the Amygdala. Synapse *6*, 271-278.

Chawla, M.K., Guzowski, J.F., Ramirez-Amaya, V., Lipa, P., Hoffman, K.L., Marriott, L.K., Worley, P.F., McNaughton, B.L., and Barnes, C.A. (2005). Sparse, environmentally selective expression of Arc RNA in the upper blade of the rodent fascia dentata by brief spatial experience. Hippocampus *15*, 579-586.

Chen, Z., Ito, K., Fujii, S., Miura, M., Furuse, H., Sasaki, H., Kaneko, K., Kato, H., and Miyakawa, H. (1996). Roles of dopamine receptors in long-term depression: enhancement via D1 receptors and inhibition via D2 receptors. Receptors Channels *4*, 1-8.

Cheng, S., and Frank, L.M. (2008). New experiences enhance coordinated neural activity in the hippocampus. Neuron *57*, 303-313.

Clayton, N.S., Griffiths, D.P., Emery, N.J., and Dickinson, A. (2001). Elements of episodic-like memory in animals. Philos Trans R Soc Lond B Biol Sci *356*, 1483-1491.

Cohen, J.E. (1995). Unexpected dominance of high frequencies in chaotic nonlinear population models. Nature *378*, 610-612.

Colom, L.V., Christie, B.R., and Bland, B.H. (1988). Cingulate cell discharge patterns related to hippocampal EEG and their modulation by muscarinic and nicotinic agents. Brain Res *460*, 329-338.

Corkin, S. (2002). What's new with the amnesic patient H.M.? Nat Rev Neurosci *3*, 153-160.

Covolan, L., Ribeiro, L.T., Longo, B.M., and Mello, L.E. (2000). Cell damage and neurogenesis in the dentate granule cell layer of adult rats after pilocarpine- or kainate-induced status epilepticus. Hippocampus *10*, 169-180.

Creutzfeldt, O., Ojemann, G., and Lettich, E. (1989a). Neuronal activity in the human lateral temporal lobe. I. Responses to speech. Exp Brain Res *77*, 451-475.

Creutzfeldt, O., Ojemann, G., and Lettich, E. (1989b). Neuronal activity in the human lateral temporal lobe. II. Responses to the subjects own voice. Exp Brain Res *77*, 476-489.

Critchley, H.D., Mathias, C.J., and Dolan, R.J. (2001). Neural activity in the human brain relating to uncertainty and arousal during anticipation. Neuron *29*, 537-545.

Csicsvari, J., Hirase, H., Czurko, A., and Buzsaki, G. (1998). Reliability and state dependence of pyramidal cell-interneuron synapses in the hippocampus: an ensemble approach in the behaving rat. Neuron *21*, 179-189.

Csicsvari, J., Hirase, H., Czurko, A., Mamiya, A., and Buzsaki, G. (1999). Oscillatory coupling of hippocampal pyramidal cells and interneurons in the behaving Rat. J Neurosci *19*, 274-287.

Davis, C.D., Jones, F.L., and Derrick, B.E. (2004). Novel environments enhance the induction and maintenance of long-term potentiation in the dentate gyrus. J Neurosci *24*, 6497-6506.

Davis, H.P., and Squire, L.R. (1984). Protein synthesis and memory: a review. Psychol Bull *96*, 518-559.

Dayan, P., and Abbott, L.F. (2001). Theoretical Neuroscience. Computational and Mathematical Modeling of Neural Systems (MIT Press).

Diba, K., and Buzsaki, G. (2007). Forward and reverse hippocampal place-cell sequences during ripples. Nat Neurosci *10*, 1241-1242.

Doclo, S., and Moonen, M. (2002). GSVD-based optimal filtering for single and multimicrophone speech enhancement. IEEE Transactions on Signal Processing *50*, 2230-2244.

Donaldson, W. (1996). The role of decision processes in remembering and knowing. Memory & Cognition *24*, 523-533.

Duvernoy, H.M. (2005). The Human Hippocampus (Springer).

Efron, B., and Tibshirani, R.J. (1993). An Introduction to the Bootstrap (London: Chapman&Hall).

Einhauser, W., Rutishauser, U., and Koch, C. (2008). Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. J Vis *8*, 2 1-19.

Ekstrom, A., Viskontas, I., Kahana, M., Jacobs, J., Upchurch, K., Bookheimer, S., and Fried, I. (2007). Contrasting roles of neural firing rate and local field potentials in human memory. Hippocampus *17*, 606-617.

Eldridge, L.L., Knowlton, B.J., Furmanski, C.S., Bookheimer, S.Y., and Engel, S.A. (2000). Remembering episodes: a selective role for the hippocampus during retrieval. Nat Neurosci *3*, 1149-1152.

Engel, A.K., Moll, C.K., Fried, I., and Ojemann, G.A. (2005). Invasive recordings from the human brain: clinical insights and beyond. Nat Rev Neurosci *6*, 35-47.

Engel, A.K., and Singer, W. (2001). Temporal binding and the neural correlates of sensory awareness. Trends Cogn Sci *5*, 16-25.

Engel, J., Jr. (2001). Mesial temporal lobe epilepsy: what have we learned? Neuroscientist *7*, 340-352.

Ennaceur, A., and Delacour, J. (1988). A new one-trial test for neurobiological studies of memory in rats. 1: Behavioral data. Behav Brain Res *31*, 47-59.

Evgeniou, T., Pontil, M., and Poggio, T. (2000). Regularization networks and support vector machines. Advances in Computational Mathematics *13*, 1-50.

Fahy, F.L., Riches, I.P., and Brown, M.W. (1993). Neuronal activity related to visual recognition memory: long-term memory and the encoding of recency and familiarity information in the primate anterior and medial inferior temporal and rhinal cortex. Exp Brain Res *96*, 457-472.

Fallon, J.H., Koziell, D.A., and Moore, R.Y. (1978). Catecholamine innervation of the basal forebrain. II. Amygdala, suprarhinal cortex and entorhinal cortex. J Comp Neurol *180*, 509-532.

Fanselow, M.S., and LeDoux, J.E. (1999). Why we think plasticity underlying pavlovian fear conditioning occurs in the basolateral amygdala. Neuron *23*, 229-232.

Fantz, R.L. (1964). Visual Experience in Infants - Decreased Attention to Familiar Patterns Relative to Novel Ones. Science *146*, 668-670.

Fee, M.S., Mitra, P.P., and Kleinfeld, D. (1996a). Automatic sorting of multiple unit neuronal signals in the presence of anisotropic and non-Gaussian variability. J Neurosci Methods *69*, 175-188.

Fee, M.S., Mitra, P.P., and Kleinfeld, D. (1996b). Variability of extracellular spike waveforms of cortical neurons. J Neurophysiol *76*, 3823-3833.

Felleman, D.J., and Van Essen, D.C. (1991). Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex *1*, 1-47.

Fellows, L.K., and Farah, M.J. (2005). Is anterior cingulate cortex necessary for cognitive control? Brain *128*, 788-796.

Fenton, A.A., and Muller, R.U. (1998). Place cell discharge is extremely variable during individual passes of the rat through the firing field. Proc Natl Acad Sci U S A *95*, 3182-3187.

Fernandez, G., Effern, A., Grunwald, T., Pezer, N., Lehnertz, K., Dumpelmann, M., Roost, D.V., and Elger, C.E. (1999). Real-Time Tracking of Memory Formation in the Human Rhinal Cortex and Hippocampus. Science *285*, 1582-1585.

Fisher, N.I. (1993). Statistical analysis of circular data (Cambridge [England] ; New York, NY, USA: Cambridge University Press).

Fletcher, P.C., Shallice, T., and Dolan, R.J. (1998). The functional roles of prefrontal cortex in episodic memory - I. Encoding. Brain *121*, 1239-1248.

Flexner, J.B., Flexner, L.B., and Stellar, E. (1963). Memory in Mice as Affected by Intracerebral Puromycin. Science *141*, 57-&.

Fonseca, R., Nagerl, U.V., and Bonhoeffer, T. (2006). Neuronal activity determines the protein synthesis dependence of long-term potentiation. Nat Neurosci *9*, 478-480.

Foster, D.J., and Wilson, M.A. (2006). Reverse replay of behavioural sequences in hippocampal place cells during the awake state. Nature *440*, 680-683.

Fox, S.E. (1989). Membrane potential and impedance changes in hippocampal pyramidal cells during theta rhythm. Exp Brain Res *77*, 283-294.

Fox, S.E., and Ranck, J.B., Jr. (1981). Electrophysiological characteristics of hippocampal complex-spike cells and theta cells. Exp Brain Res *41*, 399-410.

Frankland, P.W., and Bontempi, B. (2005). The organization of recent and remote memories. Nat Rev Neurosci *6*, 119-130.

Frankland, P.W., Bontempi, B., Talton, L.E., Kaczmarek, L., and Silva, A.J. (2004). The involvement of the anterior cingulate cortex in remote contextual fear memory. Science *304*, 881-883.

Franklin, J., and Bair, W. (1995). The Effect of a Refractory Period on the Power Spectrum of Neuronal Discharge. Siam Journal on Applied Mathematics *55*, 1074-1093.

Frey, U., Schroeder, H., and Matthies, H. (1990). Dopaminergic antagonists prevent long-term maintenance of posttetanic LTP in the CA1 region of rat hippocampal slices. Brain Res *522*, 69-75.

Fried, I., Cameron, K.A., Yashar, S., Fong, R., and Morrow, J.W. (2002). Inhibitory and excitatory responses of single neurons in the human medial temporal lobe during recognition of faces and objects. Cereb Cortex *12*, 575-584.

Fried, I., MacDonald, K.A., and Wilson, C.L. (1997). Single neuron activity in human hippocampus and amygdala during recognition of faces and objects. Neuron *18*, 753-765.

Fried, I., Wilson, C.L., Maidment, N.T., Engel, J., Behnke, E., Fields, T.A., MacDonald, K.A., Morrow, J.W., and Ackerson, L. (1999). Cerebral microdialysis combined with single-neuron and electroencephalographic recording in neurosurgical patients - Technical note. Journal of Neurosurgery *91*, 697-705.

Fried, I., Wilson, C.L., Morrow, J.W., Cameron, K.A., Behnke, E.D., Ackerson, L.C., and Maidment, N.T. (2001). Increased dopamine release in the human amygdala during performance of cognitive tasks. Nat Neurosci *4*, 201-206.

Gabbiani, F., and Koch, C. (1999). Principles of Spike Train Analysis. In Methods in Neuronal Modeling: From Synapses to Networks, C. Koch, and I. Segev, eds. (MIT Press), pp. 313-360.

Gabriel, M., Kubota, Y., Sparenborg, S., Straube, K., and Vogt, B.A. (1991). Effects of cingulate cortical lesions on avoidance learning and training-induced unit activity in rabbits. Exp Brain Res *86*, 585-600.

Gabriel, M., Sparenborg, S.P., and Stolar, N. (1987). Hippocampal control of cingulate cortical and anterior thalamic information processing during learning in rabbits. Exp Brain Res *67*, 131-152.

Gallant, J.L., Connor, C.E., Rakshit, S., Lewis, J.W., and Van Essen, D.C. (1996). Neural responses to polar, hyperbolic, and Cartesian gratings in area V4 of the macaque monkey. J Neurophysiol *76*, 2718-2739.

Gallant, J.L., Shoup, R.E., and Mazer, J.A. (2000). A human extrastriate area functionally homologous to macaque V4. Neuron *27*, 227-235.

Gallo, M., and Candido, A. (1995). Reversible inactivation of dorsal hippocampus by tetrodotoxin impairs blocking of taste aversion selectively during the acquisition but not the retrieval in rats. Neurosci Lett *186*, 1-4.

Garcia, A.D., Doan, N.B., Imura, T., Bush, T.G., and Sofroniew, M.V. (2004). GFAP-expressing progenitors are the principal source of constitutive neurogenesis in adult mouse forebrain. Nat Neurosci *7*, 1233-1241.

Gasbarri, A., Sulli, A., and Packard, M.G. (1997). The dopaminergic mesencephalic projections to the hippocampal formation in the rat. Prog Neuropsychopharmacol Biol Psychiatry *21*, 1-22.

Gasbarri, A., Verney, C., Innocenzi, R., Campana, E., and Pacitti, C. (1994). Mesolimbic dopaminergic neurons innervating the hippocampal formation in the rat: a combined retrograde tracing and immunohistochemical study. Brain Res *668*, 71-79.

Gaspar, P., Berger, B., Febvret, A., Vigny, A., and Henry, J.P. (1989). Catecholamine Innervation of the Human Cerebral-Cortex as Revealed by Comparative Immunohistochemistry of Tyrosine-Hydroxylase and Dopamine-Beta-Hydroxylase. Journal of Comparative Neurology *279*, 249-271.

Gilbert, P.E., Kesner, R.P., and Lee, I. (2001). Dissociating hippocampal subregions: double dissociation between dentate gyrus and CA1. Hippocampus *11*, 626-636.

Gleissner, U., Helmstaedter, C., Kurthen, M., and Elger, C.E. (1997). Evidence of very fast memory consolidation: an intracarotid amytal study. Neuroreport *8*, 2893-2896.

Gold, C. (2007). Biophysics of extracellular action potentials. (California Institute of Technology).

Gold, C., Henze, D.A., Koch, C., and Buzsaki, G. (2006). On the origin of the extracellular action potential waveform: A modeling study. J Neurophysiol *95*, 3113-3128.

Gould, E., Reeves, A.J., Graziano, M.S., and Gross, C.G. (1999). Neurogenesis in the neocortex of adult primates. Science *286*, 548-552.

Green, D., and Swets, J. (1966). Signal Detection Theory and Psychophysics (Wiley).

Grunwald, T., Lehnertz, K., Heinze, H.J., Helmstaedter, C., and Elger, C.E. (1998). Verbal novelty detection within the human hippocampus proper. Proc Natl Acad Sci U S A *95*, 3193-3197.

Halgren, E., Babb, T.L., and Crandall, P.H. (1978a). Activity of human hippocampal formation and amygdala neurons during memory testing. Electroencephalogr Clin Neurophysiol *45*, 585-601.

Halgren, E., Walter, R.D., Cherlow, D.G., and Crandall, P.H. (1978b). Mental phenomena evoked by electrical stimulation of the human hippocampal formation and amygdala. Brain *101*, 83-117.

Halgren, E., and Wilson, C.L. (1985). Recall deficits produced by afterdischarges in the human hippocampal formation and amygdala. Electroencephalogr Clin Neurophysiol *61*, 375-380.

Halgren, E., Wilson, C.L., and Stapleton, J.M. (1985). Human medial temporal-lobe stimulation disrupts both formation and retrieval of recent memories. Brain Cogn *4*, 287-295.

Hampton, R.R. (2001). Rhesus monkeys know when they remember. Proc Natl Acad Sci USA *98*, 5359-5362.

Han, C.J., O'Tuathaigh, C.M., van Trigt, L., Quinn, J.J., Fanselow, M.S., Mongeau, R., Koch, C., and Anderson, D.J. (2003). Trace but not delay fear conditioning requires attention and the anterior cingulate cortex. Proc Natl Acad Sci U S A *100*, 13087-13092.

Hanes, D.P., and Schall, J.D. (1996). Neural control of voluntary movement initiation. Science *274*, 427-430.

Hansen, P.C. (1998). Rank-deficient prewhitening with quotient SVD and ULV decompositions. Bit *38*, 34-43.

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. J Neurophysiol *84*, 401-414.

Hasselmo, M.E., Schnell, E., and Barkai, E. (1995). Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hippocampal region CA3. J Neurosci *15*, 5249-5262.

Hebb, D.O. (1949). The Organization of Behavior; A Neuropsychological Theory. (Wiley).

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.

Heit, G., Smith, M.E., and Halgren, E. (1990). Neuronal activity in the human medial temporal lobe during recognition memory. Brain *113 (Pt 4)*, 1093-1112.

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.

Henze, D.A., Wittner, L., and Buzsaki, G. (2002). Single granule cells reliably discharge targets in the hippocampal CA3 network in vivo. Nat Neurosci *5*, 790-795.

Hess, E.H., and Polt, J.M. (1960). Pupil Size as Related to Interest Value of Visual Stimuli. Science *132*, 349-350.

Heuer, F., and Reisberg, D. (1990). Vivid Memories of Emotional Events - the Accuracy of Remembered Minutiae. Memory & Cognition *18*, 496-506.

Hochberg, L.R., Serruya, M.D., Friehs, G.M., Mukand, J.A., Saleh, M., Caplan, A.H., Branner, A., Chen, D., Penn, R.D., and Donoghue, J.P. (2006). Neuronal ensemble control of prosthetic devices by a human with tetraplegia. Nature *442*, 164-171.

Holmgren, C., Harkany, T., Svennenfors, B., and Zilberter, Y. (2003). Pyramidal cell communication within local networks in layer 2/3 of rat neocortex. J Physiol *551*, 139-153.

Holscher, C., Anwyl, R., and Rowan, M.J. (1997). Stimulation on the positive phase of hippocampal theta rhythm induces long-term potentiation that can Be depotentiated by stimulation on the negative phase in area CA1 in vivo. J Neurosci *17*, 6470-6477.

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.

Honey, R.C., Watt, A., and Good, M. (1998). Hippocampal lesions disrupt an associative mismatch process. J Neurosci *18*, 2226-2230.

Hopfield, J.J. (1982). Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci U S A *79*, 2554-2558.

Howard, M.W., Rizzuto, D.S., Caplan, J.B., Madsen, J.R., Lisman, J., Aschenbrenner-Scheibe, R., Schulze-Bonhage, A., and Kahana, M.J. (2003). Gamma oscillations correlate with working memory load in humans. Cereb Cortex *13*, 1369-1374.

Huang, Y.Y., and Kandel, E.R. (1995). D1/D5 receptor agonists induce a protein synthesis-dependent late potentiation in the CA1 region of the hippocampus. Proc Natl Acad Sci U S A *92*, 2446-2450.

Huang, Y.Y., Nguyen, P.V., Abel, T., and Kandel, E.R. (1996). Long-lasting forms of synaptic potentiation in the mammalian hippocampus. Learn Mem *3*, 74-85.

Hung, C.P., Kreiman, G., Poggio, T., and DiCarlo, J.J. (2005). Fast readout of object identity from macaque inferior temporal cortex. Science *310*, 863-866.

Hunt, R.R. (1995). The Subtlety of Distinctiveness - What Vonrestorff Really Did. Psychonomic Bulletin & Review *2*, 105-112.

Hyman, J.M., Wyble, B.P., Goyal, V., Rossi, C.A., and Hasselmo, M.E. (2003). Stimulation in hippocampal region CA1 in behaving rats yields long-term potentiation when delivered to the peak of theta and long-term depression when delivered to the trough. J Neurosci *23*, 11725-11731.

Ito, H.I., and Schuman, E.M. (2007). Frequency-dependent gating of synaptic transmission and plasticity by dopamine. Front. Neural Circuits *1*.

Jacobs, J., Kahana, M.J., Ekstrom, A.D., and Fried, I. (2007). Brain oscillations control timing of single-neuron activity in humans. J Neurosci *27*, 3839-3844.

Jensen, O., Kaiser, J., and Lachaux, J.P. (2007). Human gamma-frequency oscillations associated with attention and memory. Trends Neurosci *30*, 317-324.

Johansson, R.S., and Birznieks, I. (2004). First spikes in ensembles of human tactile afferents code complex spatial fingertip events. Nat Neurosci *7*, 170-177.

Johnson, R.A., and Wichern, D.W. (2002). Applied multivariate statistical analysis (New York: Prentice Hall).

Jolliffe, I.T. (2002). Principal component analysis (New York: Springer).

Juergens, E., Guettler, A., and Eckhorn, R. (1999). Visual stimulation elicits locked and induced gamma oscillations in monkey intracortical- and EEG-potentials, but not in human EEG. Exp Brain Res *129*, 247-259.

Jung, H.H., Kim, C.H., Chang, J.H., Park, Y.G., Chung, S.S., and Chang, J.W. (2006a). Bilateral anterior cingulotomy for refractory obsessive-compulsive disorder: Long-term follow-up results. Stereotact Funct Neurosurg *84*, 184-189.

Jung, J., Hudry, J., Ryvlin, P., Royet, J.P., Bertrand, O., and Lachaux, J.P. (2006b). Functional significance of olfactory-induced oscillations in the human amygdala. Cereb Cortex *16*, 1-8.

Jung, M.W., and McNaughton, B.L. (1993). Spatial selectivity of unit activity in the hippocampal granular layer. Hippocampus *3*, 165-182.

Kalivas, P.W., and Volkow, N.D. (2005). The neural basis of addiction: a pathology of motivation and choice. Am J Psychiatry *162*, 1403-1413.

Kanerva, P. (1988). Sparse distributed memory (Cambridge: MIT Press).

Kang, H., and Schuman, E.M. (1996). A requirement for local protein synthesis in neurotrophin-induced hippocampal synaptic plasticity. Science *273*, 1402-1406.

Kang, H.J., and Schuman, E.M. (1995). Neurotrophin-induced modulation of synaptic transmission in the adult hippocampus. J Physiol Paris *89*, 11-22.

Kass, R.E., Ventura, V., and Brown, E.N. (2005). Statistical issues in the analysis of neuronal data. J Neurophysiol *94*, 8-25.

Kay, S.M. (1993). Fundamentals of statistical signal processing (Englewood Cliffs, N.J.: PTR Prentice-Hall).

Kelleher, R.J., 3rd, Govindarajan, A., and Tonegawa, S. (2004). Translational regulatory mechanisms in persistent forms of synaptic plasticity. Neuron *44*, 59-73.

Kelley, W.M., Miezin, F.M., McDermott, K.B., Buckner, R.L., Raichle, M.E., Cohen, N.J., Ollinger, J.M., Akbudak, E., Conturo, T.E., Snyder, A.Z., and Petersen, S.E. (1998). Hemispheric specialization in human dorsal frontal cortex and medial temporal lobe for verbal and nonverbal memory encoding. Neuron *20*, 927-936.

Kennerley, S.W., Walton, M.E., Behrens, T.E., Buckley, M.J., and Rushworth, M.F. (2006). Optimal decision making and the anterior cingulate cortex. Nat Neurosci *9*, 940-947.

Kerns, J.G., Cohen, J.D., MacDonald, A.W., 3rd, Cho, R.Y., Stenger, V.A., and Carter, C.S. (2004). Anterior cingulate conflict monitoring and adjustments in control. Science *303*, 1023-1026.

Kim, K.H., and Kim, S.J. (2003). A wavelet-based method for action potential detection from extracellular neural signal recording with low signal-to-noise ratio. IEEE Trans Biomed Eng *50*, 999-1011.

Kinsbour, M., and George, J. (1974). Mechanism of Word-Frequency Effect on Recognition Memory. Journal of Verbal Learning and Verbal Behavior *13*, 63-69.

Kirwan, C.B., Bayley, P.J., Galvan, V.V., and Squire, L.R. (2008). Detailed recollection of remote autobiographical memory after damage to the medial temporal lobe. Proc Natl Acad Sci U S A *105*, 2676-2680.

Kishiyama, M.M., Yonelinas, A.P., and Lazzara, M.M. (2004). The von Restorff effect in amnesia: the contribution of the hippocampal system to novelty-related memory enhancements. J Cogn Neurosci *16*, 15-23.

Klimesch, W., Doppelmayr, M., Russegger, H., and Pachinger, T. (1996). Theta band power in the human scalp EEG and the encoding of new information. Neuroreport *7*, 1235-1240.

Knight, R. (1996). Contribution of human hippocampal region to novelty detection. Nature *383*, 256-259.

Knight, R.T., and Nakada, T. (1998). Cortico-limbic circuits and novelty: a review of EEG and blood flow data. Rev Neurosci *9*, 57-70.

Knutson, B., Adams, C.M., Fong, G.W., and Hommer, D. (2001). Anticipation of increasing monetary reward selectively recruits nucleus accumbens. J Neurosci *21*, RC159.

Koch, C. (1999). Biophysics of computation : information processing in single neurons (New York: Oxford University Press).

Koechlin, E., and Hyafil, A. (2007). Anterior prefrontal function and the limits of human decision-making. Science *318*, 594-598.

Kohonen, T., and Lehtio, P. (1981). Storage and Processing of Information in Distributed Associative Memory Systems. In Parallel Models of Associative Memory, G. Hinton, and J. Anderson, eds. (Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.), pp. 105-143.

Kraskov, A., Quiroga, R.Q., Reddy, L., Fried, I., and Koch, C. (2007). Local field potentials and spikes in the human medial temporal lobe are selective to image category. J Cogn Neurosci *19*, 479-492.

Kreiman, G. (2007). Single unit approaches to human vision and memory. Curr Opin Neurobiol *17*, 471-475.

Kreiman, G., Fried, I., and Koch, C. (2002). Single-neuron correlates of subjective vision in the human medial temporal lobe. Proc Natl Acad Sci U S A *99*, 8378-8383.

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

Kreiman, G., Koch, C., and Fried, I. (2000b). Imagery neurons in the human brain. Nature *408*, 357-361.

Lamprecht, R., and Dudai, Y. (2000). The amygdala in conditioned taste aversion: it's there, but where. In The Amygdala, J.P. Aggleton, ed. (NY: Oxford Press), pp. 311–329.

Lang, P.J., and Cuthbert, B.N. (1993). International affective picture system standardization procedure and initial group results for affective judgements. (Gainesville, FL, University of Florida).

Laurent, G. (2002). Olfactory network dynamics and the coding of multidimensional signals. Nat Rev Neurosci *3*, 884-895.

Lemaire, V., Aurousseau, C., Le Moal, M., and Abrous, D.N. (1999). Behavioural trait of reactivity to novelty is related to hippocampal neurogenesis. Eur J Neurosci *11*, 4006-4014.

Lepage, M., Ghaffar, O., Nyberg, L., and Tulving, E. (2000). Prefrontal cortex and episodic memory retrieval mode. Proceedings of the National Academy of Sciences of the United States of America *97*, 506-511.

Leutgeb, J.K., Leutgeb, S., Moser, M.B., and Moser, E.I. (2007). Pattern separation in the dentate gyrus and CA3 of the hippocampus. Science *315*, 961-966.

Levine, E.S., Dreyfus, C.F., Black, I.B., and Plummer, M.R. (1995). Brain-derived neurotrophic factor rapidly enhances synaptic transmission in hippocampal neurons via postsynaptic tyrosine kinase receptors. Proc Natl Acad Sci U S A *92*, 8074-8077.

Lewicki, M.S. (1998). A review of methods for spike sorting: the detection and classification of neural action potentials. Network *9*, R53-78.

Li, L., Miller, E.K., and Desimone, R. (1993). The representation of stimulus familiarity in anterior inferior temporal cortex. J Neurophysiol *69*, 1918-1929.

Li, S., Cullen, W.K., Anwyl, R., and Rowan, M.J. (2003). Dopamine-dependent facilitation of LTP induction in hippocampal CA1 by exposure to spatial novelty. Nat Neurosci *6*, 526-531.

Lisman, J.E., and Grace, A.A. (2005). The hippocampal-VTA loop: controlling the entry of information into long-term memory. Neuron *46*, 703-713.

Lisman, J.E., and Otmakhova, N.A. (2001). Storage, recall, and novelty detection of sequences by the hippocampus: elaborating on the SOCRATIC model to account for normal and aberrant effects of dopamine. Hippocampus *11*, 551-568.

Lledo, P.M., Alonso, M., and Grubb, M.S. (2006). Adult neurogenesis and functional plasticity in neuronal circuits. Nat Rev Neurosci *7*, 179-193.

Logothetis, N.K. (2002). The neural basis of the blood-oxygen-level-dependent functional magnetic resonance imaging signal. Philos Trans R Soc Lond B Biol Sci *357*, 1003-1037.

Logothetis, N.K., Pauls, J., Augath, M., Trinath, T., and Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. Nature *412*, 150-157.

Lohof, A.M., Ip, N.Y., and Poo, M.M. (1993). Potentiation of developing neuromuscular synapses by the neurotrophins NT-3 and BDNF. Nature *363*, 350-353.

Maass, W., and Markram, H. (2004). On the computational power of circuits of spiking neurons. Journal of Computer and System Sciences *69*, 593-616.

Maass, W., Natschlager, T., and Markram, H. (2002). Real-time computing without stable states: a new framework for neural computation based on perturbations. Neural Comput *14*, 2531-2560.

MacDonald, A.W., 3rd, Cohen, J.D., Stenger, V.A., and Carter, C.S. (2000). Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. Science *288*, 1835-1838.

MacKay, D., and McCulloch, W. (1952). The limiting information capacity of a neuronal link. Bulletin of Mathematical Biology *14*, 127-135.

MacLeod, K., Backer, A., and Laurent, G. (1998). Who reads temporal information contained across synchronized and oscillatory spike trains? Nature *395*, 693-698.

Macmillan, N.A., and Creelman, C.D. (2005). Detection theory, 2nd edn (Mahwah, NJ: Lawrence Associates).

Mandler, G. (1980). Recognizing - the Judgment of Previous Occurrence. Psychological Review *87*, 252-271.

Manns, J.R., Hopkins, R.O., Reed, J.M., Kitchener, E.G., and Squire, L.R. (2003). Recognition memory and the human hippocampus. Neuron *37*, 171-180.

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.

Marr, D. (1970). A Theory for Cerebral Neocortex. Proceedings of the Royal Society of London Series B-Biological Sciences *176*, 161-234.

Marr, D. (1971). Simple memory: a theory for archicortex. Philos Trans R Soc Lond B Biol Sci *262*, 23-81.

Martin, S.J., Grimwood, P.D., and Morris, R.G. (2000). Synaptic plasticity and memory: an evaluation of the hypothesis. Annu Rev Neurosci *23*, 649-711.

Mazurek, M.E., and Shadlen, M.N. (2002). Limits to the temporal fidelity of cortical spike rate signals. Nat Neurosci *5*, 463-471.

McCarthy, G., Wood, C.C., Williamson, P.D., and Spencer, D.D. (1989). Task-dependent field potentials in human hippocampal formation. J Neurosci *9*, 4253-4268.

McCartney, H., Johnson, A.D., Weil, Z.M., and Givens, B. (2004). Theta reset produces optimal conditions for long-term potentiation. Hippocampus *14*, 684-687.

McCormick, D.A., Connors, B.W., Lighthall, J.W., and Prince, D.A. (1985). Comparative electrophysiology of pyramidal and sparsely spiny stellate neurons of the neocortex. J Neurophysiol *54*, 782-806.

Mcgaugh, J.L., Introinicollison, I.B., Nagahara, A.H., Cahill, L., Brioni, J.D., and Castellano, C. (1990). Involvement of the Amygdaloid Complex in Neuromodulatory Influences on Memory Storage. Neuroscience and Biobehavioral Reviews *14*, 425-431.

McHugh, T.J., Jones, M.W., Quinn, J.J., Balthasar, N., Coppari, R., Elmquist, J.K., Lowell, B.B., Fanselow, M.S., Wilson, M.A., and Tonegawa, S. (2007). Dentate gyrus NMDA receptors mediate rapid pattern separation in the hippocampal network. Science *317*, 94-99.

Messinger, A., Squire, L.R., Zola, S.M., and Albright, T.D. (2005). Neural correlates of knowledge: stable representation of stimulus associations across variations in behavioral performance. Neuron *48*, 359-371.

Milner, B., Corkin, S., and Teuber, H.L. (1968). Further Analysis of Hippocampal Amnesic Syndrome - 14-Year Follow-up Study of HM. Neuropsychologia *6*, 215-234.

Mitchell, J.F., Sundberg, K.A., and Reynolds, J.H. (2007). Differential attention-dependent response modulation across cell classes in macaque visual area V4. Neuron *55*, 131-141.

Mitzdorf, U. (1985). Current source-density method and application in cat cerebral cortex: investigation of evoked potentials and EEG phenomena. Physiol Rev *65*, 37-100.

Montemurro, M.A., Panzeri, S., Maravall, M., Alenda, A., Bale, M.R., Brambilla, M., and Petersen, R.S. (2007). Role of precise spike timing in coding of dynamic vibrissa stimuli in somatosensory thalamus. J Neurophysiol *98*, 1871-1882.

Mormann, F., Fell, J., Axmacher, N., Weber, B., Lehnertz, K., Elger, C.E., and Fernandez, G. (2005). Phase/amplitude reset and theta-gamma interaction in the human medial temporal lobe during a continuous word recognition memory task. Hippocampus *15*, 890-900.

Musallam, S., Corneil, B.D., Greger, B., Scherberger, H., and Andersen, R.A. (2004). Cognitive control signals for neural prosthetics. Science *305*, 258-262.

Mussa-Ivaldi, F.A., and Miller, L.E. (2003). Brain-machine interfaces: computational demands and clinical needs meet basic neuroscience. Trends Neurosci *26*, 329-334.

Najmi, A.H., and Sadowsky, J. (1997). The continuous wavelet transform and variable resolution time-frequency analysis. Johns Hopkins Apl Technical Digest *18*, 134-140.

Nakazawa, K., Quirk, M.C., Chitwood, R.A., Watanabe, M., Yeckel, M.F., Sun, L.D., Kato, A., Carr, C.A., Johnston, D., Wilson, M.A., and Tonegawa, S. (2002). Requirement for hippocampal CA3 NMDA receptors in associative memory recall. Science *297*, 211-218.

Nakazawa, K., Sun, L.D., Quirk, M.C., Rondi-Reig, L., Wilson, M.A., and Tonegawa, S. (2003). Hippocampal CA3 NMDA receptors are crucial for memory acquisition of one-time experience. Neuron *38*, 305-315.

Neuman, R.S., and Harley, C.W. (1983). Long-lasting potentiation of the dentate gyrus population spike by norepinephrine. Brain Res *273*, 162-165.

Neves, G., Cooke, S.F., and Bliss, T.V.P. (2008). Synaptic plasticity, memory and the hippocampus: a neural network approach to causality. Nature Reviews Neuroscience *9*, 65-75.

Nicolelis, M.A., Ghazanfar, A.A., Faggin, B.M., Votaw, S., and Oliveira, L.M. (1997). Reconstructing the engram: simultaneous, multisite, many single neuron recordings. Neuron *18*, 529-537.

Nimchinsky, E.A., Gilissen, E., Allman, J.M., Perl, D.P., Erwin, J.M., and Hof, P.R. (1999). A neuronal morphologic type unique to humans and great apes. Proc Natl Acad Sci U S A *96*, 5268-5273.

Nir, Y., Fisch, L., Mukamel, R., Gelbard-Sagiv, H., Arieli, A., Fried, I., and Malach, R. (2007). Coupling between neuronal firing rate, gamma LFP, and BOLD fMRI is related to interneuronal correlations. Curr Biol *17*, 1275-1285.

Noton, D., and Stark, L. (1971). Scanpaths in eye movements during pattern perception. Science *171*, 308-311.

O'Doherty, J.P., Dayan, P., Friston, K., Critchley, H., and Dolan, R.J. (2003). Temporal difference models and reward-related learning in the human brain. Neuron *38*, 329-337.

O'Reilly, R.C., and McClelland, J.L. (1994). Hippocampal conjunctive encoding, storage, and recall: avoiding a trade-off. Hippocampus *4*, 661-682.

Ojemann, G., and Fedio, P. (1968). Effect of stimulation of the human thalamus and parietal and temporal white matter on short-term memory. J Neurosurg *29*, 51-59.

Ojemann, G.A. (1997). Treatment of temporal lobe epilepsy. Annu Rev Med *48*, 317-328.

Ojemann, G.A., Creutzfeldt, O., Lettich, E., and Haglund, M.M. (1988). Neuronal activity in human lateral temporal cortex related to short-term verbal memory, naming and reading. Brain *111 (Pt 6)*, 1383-1403.

Ojemann, G.A., and Dodrill, C.B. (1985). Verbal memory deficits after left temporal lobectomy for epilepsy. Mechanism and intraoperative prediction. J Neurosurg *62*, 101-107.

Ojemann, G.A., Schoenfield-McNeill, J., and Corina, D.P. (2002). Anatomic subdivisions in human temporal cortical neuronal activity related to recent verbal memory. Nat Neurosci *5*, 64-71.

Orr, G., Rao, G., Houston, F.P., McNaughton, B.L., and Barnes, C.A. (2001). Hippocampal synaptic plasticity is modulated by theta rhythm in the fascia dentata of adult and aged freely behaving rats. Hippocampus *11*, 647-654.

Osipova, D., Takashima, A., Oostenveld, R., Fernandez, G., Maris, E., and Jensen, O. (2006). Theta and gamma oscillations predict encoding and retrieval of declarative memory. J Neurosci *26*, 7523-7531.

Otmakhova, N.A., and Lisman, J.E. (1996). D1/D5 dopamine receptor activation increases the magnitude of early long-term potentiation at CA1 hippocampal synapses. J Neurosci *16*, 7478-7486.

Otten, L.J., Henson, R.N., and Rugg, M.D. (2002). State-related and item-related neural correlates of successful memory encoding. Nat Neurosci *5*, 1339-1344.

Otten, L.J., Quayle, A.H., Akram, S., Ditewig, T.A., and Rugg, M.D. (2006). Brain activity before an event predicts later recollection. Nat Neurosci *9*, 489-491.

Oya, H., Kawasaki, H., Howard, M.A., and Adolphs, R. (2002). Electrophysiological responses in the human amygdala discriminate emotion categories of complex visual stimuli. Journal of Neuroscience *22*, 9502-9512.

Paller, K.A., Kutas, M., and Mayes, A.R. (1987). Neural correlates of encoding in an incidental learning paradigm. Electroencephalogr Clin Neurophysiol *67*, 360-371.

Paller, K.A., and Wagner, A.D. (2002). Observing the transformation of experience into memory. Trends in Cognitive Sciences *6*, 93-102.

Parent, J.M. (2007). Adult neurogenesis in the intact and epileptic dentate gyrus. Dentate Gyrus: A Comphrehensive Guide to Structure, Function, and Clinical Implications *163*, 529-+.

Parent, J.M., Yu, T.W., Leibowitz, R.T., Geschwind, D.H., Sloviter, R.S., and Lowenstein, D.H. (1997). Dentate granule cell neurogenesis is increased by seizures and contributes to aberrant network reorganization in the adult rat hippocampus. J Neurosci *17*, 3727-3738.

Parker, A., Wilding, E., and Akerman, C. (1998). The von Restorff effect in visual object recognition memory in humans and monkeys: The role of frontal/perirhinal interaction. Journal of Cognitive Neuroscience *10*, 691-703.

Paus, T. (2001). Primate anterior cingulate cortex: where motor control, drive and cognition interface. Nat Rev Neurosci *2*, 417-424.

Paus, T., Koski, L., Caramanos, Z., and Westbury, C. (1998). Regional differences in the effects of task difficulty and motor output on blood flow response in the human anterior cingulate cortex: a review of 107 PET activation studies. Neuroreport *9*, R37-47.

Pavlides, C., Greenstein, Y.J., Grudman, M., and Winson, J. (1988). Long-term potentiation in the dentate gyrus is induced preferentially on the positive phase of theta-rhythm. Brain Res *439*, 383-387.

Pavlov, L.P. (1927). Conditioned Reflexes. An Investigation of the Physiological Activity of the Cerebral Cortex. (Toronto: Oxford University Press).

Pelli, D.G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. Spatial Vision *10*, 437-442.

Penfield, W. (1958). Functional localization in temporal and deep sylvian areas. Res Publ Assoc Res Nerv Ment Dis *36*, 210-226.

Penfield, W., and Perot, P. (1963). The Brain's Record of Auditory and Visual Experience. a Final Summary and Discussion. Brain *86*, 595-696.

Penttonen, M., and Buzsaki, G. (2003). Natural logarithmic relationship between brain oscillators. Thalamus & Related Systems *2*, 145-152.

Perrine, K., Devinsky, O., Uysal, S., Luciano, D.J., and Dogali, M. (1994). Left temporal neocortex mediation of verbal memory: evidence from functional mapping with cortical stimulation. Neurology *44*, 1845-1850.

Phelps, E.A. (2004). Human emotion and memory: interactions of the amygdala and hippocampal complex. Current Opinion in Neurobiology *14*, 198-202.

Phelps, E.A., LaBar, K.S., and Spencer, D.D. (1997). Memory for emotional words following unilateral temporal lobectomy. Brain and Cognition *35*, 85-109.

Pouget, A., Dayan, P., and Zemel, R.S. (2003). Inference and computation with population codes. Annu Rev Neurosci *26*, 381-410.

Pouzat, C., Delescluse, M., Viot, P., and Diebolt, J. (2004). Improved spike-sorting by modeling firing statistics and burst-dependent spike amplitude attenuation: a Markov chain Monte Carlo approach. J Neurophysiol *91*, 2910-2928.

Pouzat, C., Mazor, O., and Laurent, G. (2002). Using noise signature to optimize spike-sorting and to assess neuronal classification quality. Journal of Neuroscience Methods *122*, 43-57.

Prinz, A.A., Abbott, L.F., and Marder, E. (2004). The dynamic clamp comes of age. Trends Neurosci *27*, 218-224.

Quirk, M.C., and Wilson, M.A. (1999). Interaction between spike waveform classification and temporal sequence detection. J Neurosci Methods *94*, 41-52.

Quiroga, R.Q., Nadasdy, Z., and Ben-Shaul, Y. (2004). Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. Neural Computation *16*, 1661-1687.

Quiroga, R.Q., Reddy, L., Koch, C., and Fried, I. (2007). Decoding visual inputs from multiple neurons in the human temporal lobe. J Neurophysiol *98*, 1997-2007.

Quiroga, R.Q., Reddy, L., Kreiman, G., Koch, C., and Fried, I. (2005). Invariant visual representation by single neurons in the human brain. Nature *435*, 1102-1107.

Rainer, G., and Miller, E.K. (2000). Effects of visual experience on the representation of objects in the prefrontal cortex. Neuron *27*, 179-189.

Ratcliff, R., Gronlund, S.D., and Sheu, C.F. (1992). Testing Global Memory Models Using Roc Curves. Psychological Review *99*, 518-535.

Reddy, L., Quiroga, R.Q., Wilken, P., Koch, C., and Fried, I. (2006). A single-neuron correlate of change detection and change blindness in the human medial temporal lobe. Curr Biol *16*, 2066-2072.

Redish, A.D. (2003). MClust-3.3 (software).

Rempel-Clower, N.L., Zola, S.M., Squire, L.R., and Amaral, D.G. (1996). Three cases of enduring memory impairment after bilateral damage limited to the hippocampal formation. J Neurosci *16*, 5233-5255.

Richardson, M.P., Strange, B.A., and Dolan, R.J. (2004). Encoding of emotional memories depends on amygdala and hippocampus and their interactions. Nature Neuroscience *7*, 278-285.

Riches, I.P., Wilson, F.A., and Brown, M.W. (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. J Neurosci *11*, 1763-1779.

Rieke, F. (1997). Spikes: exploring the neural code (Cambridge: MIT Press).

Rifkin, R., Yeo, G., and Poggio, T. (2003). Regularized Least Squares Classification. In Advances in Learning Theory: Methods, Model and Applications, J.A.K. Suykens, ed. (Amsterdam: IOS Press), pp. 131-146.

Rinberg, D., Bialek, W., Davidowitz, H., and Tishby, N. (2003). Spike sorting in the frequency domain with overlap detection. In ArXiv Physics e-prints, p. 0306056.

Ringo, J.L. (1995). Brevity of processing in a mnemonic task. J Neurophysiol *73*, 1712-1715.

Rizzuto, D.S., Madsen, J.R., Bromfield, E.B., Schulze-Bonhage, A., and Kahana, M.J. (2006). Human neocortical oscillations exhibit theta phase differences between encoding and retrieval. Neuroimage *31*, 1352-1358.

Rizzuto, D.S., Madsen, J.R., Bromfield, E.B., Schulze-Bonhage, A., Seelig, D., Aschenbrenner-Scheibe, R., and Kahana, M.J. (2003). Reset of human neocortical oscillations during a working memory task. Proc Natl Acad Sci U S A *100*, 7931-7936.

Rizzuto, D.S., Mamelak, A.N., Sutherling, W.W., Fineman, I., and Andersen, R.A. (2005). Spatial selectivity in human ventrolateral prefrontal cortex. Nat Neurosci *8*, 415-417.

Robinson, T.E. (1980). Hippocampal rhythmic slow activity (RSA; theta): a critical analysis of selected studies and discussion of possible species-differences. Brain Res *203*, 69-101.

Rogan, M.T., Staubli, U.V., and LeDoux, J.E. (1997). Fear conditioning induces associative long-term potentiation in the amygdala. Nature *390*, 604-607.

Rohani, P., Miramontes, O., and Keeling, M.J. (2004). The colour of noise in short ecological time series data. Math Med Biol *21*, 63-72.

Rolls, E.T. (1996). A theory of hippocampal function in memory. Hippocampus *6*, 601-620.

Rolls, E.T. (1999). Spatial view cells and the representation of place in the primate hippocampus. Hippocampus *9*, 467-480.

Rolls, E.T. (2007). An attractor network in the hippocampus: theory and neurophysiology. Learn Mem *14*, 714-731.

Rolls, E.T., Cahusac, P.M., Feigenbaum, J.D., and Miyashita, Y. (1993). Responses of single neurons in the hippocampus of the macaque related to recognition memory. Exp Brain Res *93*, 299-306.

Romo, R., Brody, C.D., Hernandez, A., and Lemus, L. (1999). Neuronal correlates of parametric working memory in the prefrontal cortex. Nature *399*, 470-473.

Rugg, M.D., Mark, R.E., Walla, P., Schloerscheidt, A.M., Birch, C.S., and Allan, K. (1998). Dissociation of the neural correlates of implicit and explicit memory. Nature *392*, 595-598.

Rutishauser, U., and Koch, C. (2007). Probabilistic modeling of eye movement data during conjunction search via feature-based attention. Journal of Vision *7*, 5.

Rutishauser, U., Mamelak, A.N., and Schuman, E.M. (2006a). Single-trial learning of novel stimuli by individual neurons of the human hippocampus-amygdala complex. Neuron *49*, 805-813.

Rutishauser, U., Schuman, E.M., and Mamelak, A.N. (2006b). Online detection and sorting of extracellularly recorded action potentials in human medial temporal lobe recordings, in vivo. J Neurosci Methods *154*, 204-224.

Rutishauser, U., Schuman, E.M., and Mamelak, A.N. (2008). Activity of human hippocampal and amygdala neurons during retrieval of declarative memories. Proc Natl Acad Sci U S A *105*, 329-334.

Sahani, S., Pezaris, J.S., and Andersen, R.A. (1998). On the Separation of Signals from Neighboring Cells in Tetrode Recordings. In Advances in Neural Information Processing Systems 10, J.I. Jordan, M.J. Kearns, and S.A. Solla, eds. (MIT Press), pp. 222-228.

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

Schacter, D.L., and Dodson, C.S. (2001). Misattribution, false recognition and the sins of memory. Philos Trans R Soc Lond B *356*, 1385-1393.

Schacter, D.L., Reiman, E., Curran, T., Yun, L.S., Bandy, D., McDermott, K.B., and Roediger, H.L. (1996). Neuroanatomical correlates of veridical and illusory recognition memory: Evidence from positron emission tomography. Neuron *17*, 267-274.

Schmidt-Hieber, C., Jonas, P., and Bischofberger, J. (2004). Enhanced synaptic plasticity in newly generated granule cells of the adult hippocampus. Nature *429*, 184-187.

Schultz, W. (2000). Multiple reward signals in the brain. Nat Rev Neurosci *1*, 199-207.

Schultz, W. (2002). Getting formal with dopamine and reward. Neuron *36*, 241-263.

Schultz, W., and Dickinson, A. (2000). Neuronal coding of prediction errors. Annu Rev Neurosci *23*, 473-500.

Schuman, E.M. (1999). mRNA trafficking and local protein synthesis at the synapse. Neuron *23*, 645-648.

Schwartz, A.B. (2004). Cortical neural prosthetics. Annu Rev Neurosci *27*, 487-507.

Scoville, W.B., and Milner, B. (1957). Loss of Recent Memory after Bilateral Hippocampal Lesions. Journal of Neurology Neurosurgery and Psychiatry *20*, 11-21.

Sederberg, P.B., Kahana, M.J., Howard, M.W., Donner, E.J., and Madsen, J.R. (2003). Theta and gamma oscillations during encoding predict subsequent recall. J Neurosci *23*, 10809-10814.

Sederberg, P.B., Schulze-Bonhage, A., Madsen, J.R., Bromfield, E.B., McCarthy, D.C., Brandt, A., Tully, M.S., and Kahana, M.J. (2007). Hippocampal and neocortical gamma oscillations predict memory formation in humans. Cerebral Cortex *17*, 1190-1196.

Serruya, M.D., Hatsopoulos, N.G., Paninski, L., Fellows, M.R., and Donoghue, J.P. (2002). Instant neural control of a movement signal. Nature *416*, 141-142.

Seung, H.S., and Sompolinsky, H. (1993). Simple models for reading neuronal population codes. Proc Natl Acad Sci U S A *90*, 10749-10753.

Seymour, B., O'Doherty, J.P., Dayan, P., Koltzenburg, M., Jones, A.K., Dolan, R.J., Friston, K.J., and Frackowiak, R.S. (2004). Temporal difference models describe higher-order learning in humans. Nature *429*, 664-667.

Shadlen, M.N., and Newsome, W.T. (1998). The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. J Neurosci *18*, 3870-3896.

Shapiro, M.L., and Olton, D.S. (1994). Hippocampal function and interference. In Memory Systems, D.L. Schacter, and E. Tulving, eds. (London: MIT Press), pp. 141-146.

Sharot, T., Delgado, M.R., and Phelps, E.A. (2004). How emotion enhances the feeling of remembering. Nature Neuroscience *7*, 1376-1380.

Shidara, M., and Richmond, B.J. (2002). Anterior cingulate: single neuronal signals related to degree of reward expectancy. Science *296*, 1709-1711.

Shima, K., and Tanji, J. (1998). Role for cingulate motor area cells in voluntary movement selection based on reward. Science *282*, 1335-1338.

Shoham, S., Fellows, M.R., and Normann, R.A. (2003). Robust, automatic spike sorting using mixtures of multivariate t-distributions. J Neurosci Methods *127*, 111-122.

Shors, T.J., and Matzel, L.D. (1997). Long-term potentiation: what's learning got to do with it? Behav Brain Sci *20*, 597-614.

Siapas, A.G., Lubenov, E.V., and Wilson, M.A. (2005). Prefrontal phase locking to hippocampal theta oscillations. Neuron *46*, 141-151.

Smith, C.N., Hopkins, R.O., and Squire, L.R. (2006). Experience-dependent eye movements, awareness, and hippocampus-dependent memory. Journal of Neuroscience *26*, 11304-11312.

Smith, M.E., Stapleton, J.M., and Halgren, E. (1986). Human medial temporal lobe potentials evoked in memory and language tasks. Electroencephalogr Clin Neurophysiol *63*, 145-159.

Smith, W.B., Starck, S.R., Roberts, R.W., and Schuman, E.M. (2005). Dopaminergic stimulation of local protein synthesis enhances surface expression of GluR1 and synaptic transmission in hippocampal neurons. Neuron *45*, 765-779.

Softky, W.R., and Koch, C. (1993). The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. J Neurosci *13*, 334-350.

Sokolov, E.N. (1963). Higher nervous functions; the orienting reflex. Annu Rev Physiol *25*, 545-580.

Sompolinsky, H., Yoon, H., Kang, K.J., and Shamir, M. (2001). Population coding in neuronal systems with correlated noise. Physical Review E *6405*, -.

Spencer, S.S., Sperling, M.R., Shewmon, D.A., and Kahane, P. (2007). Intracranial electrodes. In Epilepsy. A comprehensive Textbook., J. Engel, Jr., and T.A. Pedley, eds. (Lippincott), pp. 1791-1815.

Squire, L.R. (1992). Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. Psychol Rev *99*, 195-231.

Squire, L.R., and Alvarez, P. (1995). Retrograde amnesia and memory consolidation: a neurobiological perspective. Curr Opin Neurobiol *5*, 169-177.

Squire, L.R., and Bayley, P.J. (2007). The neuroscience of remote memory. Curr Opin Neurobiol *17*, 185-196.

Squire, L.R., Stark, C.E., and Clark, R.E. (2004). The medial temporal lobe. Annu Rev Neurosci *27*, 279-306.

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

Standing, L., Conezio, J., and Haber, R.N. (1970). Perception and Memory for Pictures - Single-Trial Learning of 2500 Visual Stimuli. Psychonomic Science *19*, 73-74.

Stark, C.E., and Squire, L.R. (2000). Functional magnetic resonance imaging (fMRI) activity in the hippocampal region during recognition memory. J Neurosci *20*, 7776-7781.

Stark, C.E.L., Bayley, P.J., and Squire, L.R. (2002). Recognition memory for single items and for associations is similarly impaired following damage to the hippocampal region. Learn Mem *9*, 238-242.

Stark, C.E.L., and Squire, L.R. (2001). When zero is not zero: The problem of ambiguous baseline conditions in fMRI. Proceedings of the National Academy of Sciences of the United States of America *98*, 12760-12765.

Steinlein, O.K. (2004). Genetic mechanisms that underlie epilepsy. Nature Reviews Neuroscience *5*, 400-408.

Steriade, M., McCormick, D.A., and Sejnowski, T.J. (1993). Thalamocortical oscillations in the sleeping and aroused brain. Science *262*, 679-685.

Stern, C.E., Corkin, S., Gonzalez, R.G., Guimaraes, A.R., Baker, J.R., Jennings, P.J., Carr, C.A., Sugiura, R.M., Vedantham, V., and Rosen, B.R. (1996). The hippocampal formation participates in novel picture encoding: evidence from functional magnetic resonance imaging. Proc Natl Acad Sci U S A *93*, 8660-8665.

Stevens, C.F. (1998). A million dollar question: does LTP = memory? Neuron *20*, 1-2.

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.

Stuss, D.T., Floden, D., Alexander, M.P., Levine, B., and Katz, D. (2001). Stroop performance in focal lesion patients: dissociation of processes and frontal lobe lesion location. Neuropsychologia *39*, 771-786.

Summerfield, C., Greene, M., Wager, T., Egner, T., Hirsch, J., and Mangels, J. (2006). Neocortical connectivity during episodic memory formation. PLoS Biol *4*, e128.

Sutton, M.A., and Schuman, E.M. (2006). Dendritic protein synthesis, synaptic plasticity, and memory. Cell *127*, 49-58.

Sutton, R.S., and Barto, A.G. (1998). Reinforcement learning : an introduction (Cambridge, Mass.: MIT Press).

Sutton, S., Braren, M., Zubin, J., and John, E.R. (1965). Evoked-potential correlates of stimulus uncertainty. Science *150*, 1187-1188.

Suzuki, W.A., and Amaral, D.G. (2004). Functional neuroanatomy of the medial temporal lobe memory system. Cortex *40*, 220-222.

Takahashi, S., Anzai, Y., and Sakurai, Y. (2003). Automatic sorting for multi-neuronal activity recorded with tetrodes in the presence of overlapping spikes. J Neurophysiol *89*, 2245-2258.

Takashima, A., Jensen, O., Oostenveld, R., Maris, E., van de Coevering, M., and Fernandez, G. (2006). Successful declarative memory formation is associated with ongoing activity during encoding in a distributed neocortical network related to working memory: a magnetoencephalography study. Neuroscience *139*, 291-297.

Tallon-Baudry, C., and Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. Trends Cogn Sci *3*, 151-162.

Tallon-Baudry, C., Bertrand, O., Henaff, M.A., Isnard, J., and Fischer, C. (2005). Attention modulates gamma-band oscillations differently in the human lateral occipital cortex and fusiform gyrus. Cereb Cortex *15*, 654-662.

Teolis, A. (1998). Computational Signal Processing with Wavelets (Birkhaeuser).

Tomko, G.J., and Crapper, D.R. (1974). Neuronal variability: non-stationary responses to identical visual stimuli. Brain Res *79*, 405-418.

Torrence, C., and Compo, G.P. (1998). A practical guide to wavelet analysis. Bulletin of the American Meteorological Society *79*, 61-78.

Treves, A., and Rolls, E.T. (1994). Computational analysis of the role of the hippocampus in memory. Hippocampus *4*, 374-391.

Tulving, E., Markowitsch, H.J., Craik, F.E., Habib, R., and Houle, S. (1996). Novelty and familiarity activations in PET studies of memory encoding and retrieval. Cereb Cortex *6*, 71-79.

Uylings, H.B.M., Groenewegen, H.J., and Kolb, B. (2003). Do rats have a prefrontal cortex? Behavioural Brain Research *146*, 3-17.

van Praag, H., Kempermann, G., and Gage, F.H. (1999). Running increases cell proliferation and neurogenesis in the adult mouse dentate gyrus. Nature Neuroscience *2*, 266-270.

van Praag, H., Schinder, A.F., Christie, B.R., Toni, N., Palmer, T.D., and Gage, F.H. (2002). Functional neurogenesis in the adult hippocampus. Nature *415*, 1030-1034.

Vendrell, P., Junque, C., Pujol, J., Jurado, M.A., Molet, J., and Grafman, J. (1995). The Role of Prefrontal Regions in the Stroop Task. Neuropsychologia *33*, 341-352.

Verzeano, M., Crandall, P.H., and Dymond, A. (1971). Neuronal Activity of Amygdala in Patients with Psychomotor Epilepsy. Neuropsychologia *9*, 331-344.

Vinogradova, O.S. (2001). Hippocampus as comparator: role of the two input and two output systems of the hippocampus in selection and registration of information. Hippocampus *11*, 578-598.

Viskontas, I.V., Ekstrom, A.D., Wilson, C.L., and Fried, I. (2007). Characterizing interneuron and pyramidal cells in the human medial temporal lobe in vivo using extracellular recordings. Hippocampus *17*, 49-57.

Viskontas, I.V., Knowlton, B.J., Steinmetz, P.N., and Fried, I. (2006). Differences in mnemonic processing by neurons in the human hippocampus and parahippocampal regions. Journal of Cognitive Neuroscience *18*, 1654-1662.

Vives, K., Lee, G., Doyle, W., and Spencer, D.D. (2007). Anterior Temporal Resection. In Epilepsy. A comprehensive Textbook., J. Engel, Jr., and T.A. Pedley, eds. (Lippincott), pp. 1859-1867.

Vogt, B.A., Nimchinsky, E.A., Vogt, L.J., and Hof, P.R. (1995). Human cingulate cortex: surface features, flat maps, and cytoarchitecture. J Comp Neurol *359*, 490-506.

von Restorff, H. (1933). Uber die Wirkung von Bereichsbildungen im Spurenfeld. Pyschologische Forschung *18*, 299-342.

Wagner, A.D., Pare-Blagoev, E.J., Clark, J., and Poldrack, R.A. (2001). Recovering meaning: Left prefrontal cortex guides controlled semantic retrieval. Neuron *31*, 329-338.

Wais, P.E., Wixted, J.T., Hopkins, R.O., and Squire, L.R. (2006). The hippocampus supports both the recollection and the familiarity components of recognition memory. Neuron *49*, 459-466.

Wallace, R.H., Wang, D.W., Singh, R., Scheffer, I.E., George, A.L., Jr., Phillips, H.A., Saar, K., Reis, A., Johnson, E.W., Sutherland, G.R*., et al.* (1998). Febrile seizures and generalized epilepsy associated with a mutation in the Na+-channel beta1 subunit gene SCN1B. Nat Genet *19*, 366-370.

Wallace, W.P. (1965). Review of the Historical, Empirical, and Theoretical Status of the Von Restorff Phenomenon. Psychological Bulletin *63*, 410-424.

Wang, C., Ulbert, I., Schomer, D.L., Marinkovic, K., and Halgren, E. (2005). Responses of human anterior cingulate cortex microdomains to error detection, conflict monitoring, stimulus-response mapping, familiarity, and orienting. J Neurosci *25*, 604-613.

Wang, X.J. (2001). Synaptic reverberation underlying mnemonic persistent activity. Trends Neurosci *24*, 455-463.

Ward, A.A., and Thomas, L.B. (1955). The Electrical Activity of Single Units in the Cerebral Cortex of Man. Electroencephalography and Clinical Neurophysiology *7*, 135-136.

Waydo, S., Kraskov, A., Quian Quiroga, R., Fried, I., and Koch, C. (2006). Sparse representation in the human medial temporal lobe. J Neurosci *26*, 10232-10234.

Weisbard, C., and Graham, F.K. (1971). Heart-Rate Change as a Component of Orienting Response in Monkeys. Journal of Comparative and Physiological Psychology *76*, 74-83.

Welzl, H., D'Adamo, P., and Lipp, H.P. (2001). Conditioned taste aversion as a learning and memory paradigm. Behav Brain Res *125*, 205-213.

West, M.J., and Slomianka, L. (1998). Total number of neurons in the layers of the human entorhinal cortex. Hippocampus *8*, 69-82.

Whitlock, J.R., Heynen, A.J., Shuler, M.G., and Bear, M.F. (2006). Learning induces long-term potentiation in the hippocampus. Science *313*, 1093-1097.

Wilensky, A.E., Schafe, G.E., Kristensen, M.P., and LeDoux, J.E. (2006). Rethinking the fear circuit: The central nucleus of the amygdala is required for the acquisition, consolidation, and expression of pavlovian fear conditioning. Journal of Neuroscience *26*, 12387-12396.

Williams, S., and Johnston, D. (1988). Muscarinic depression of long-term potentiation in CA3 hippocampal neurons. Science *242*, 84-87.

Williams, S.M., and Goldman-Rakic, P.S. (1998). Widespread origin of the primate mesofrontal dopamine system. Cereb Cortex *8*, 321-345.

Williams, Z.M., Bush, G., Rauch, S.L., Cosgrove, G.R., and Eskandar, E.N. (2004). Human anterior cingulate neurons and the integration of monetary reward with motor responses. Nat Neurosci *7*, 1370-1375.

Wilson, C.L. (2004). Intracranial electrophysiological investigation of the human brain in patients with epilepsy: contributions to basic and clinical research. Exp Neurol *187*, 240-245.

Wilson, F.A., and Rolls, E.T. (1993). The effects of stimulus novelty and familiarity on neuronal activity in the amygdala of monkeys performing recognition memory tasks. Exp Brain Res *93*, 367-382.

Wilson, M.A., and McNaughton, B.L. (1994). Reactivation of hippocampal ensemble memories during sleep. Science *265*, 676-679.

Wiltgen, B.J., Brown, R.A., Talton, L.E., and Silva, A.J. (2004). New circuits for old memories: the role of the neocortex in consolidation. Neuron *44*, 101-108.

Wirth, S., Yanike, M., Frank, L.M., Smith, A.C., Brown, E.N., and Suzuki, W.A. (2003). Single neurons in the monkey hippocampus and learning of new associations. Science *300*, 1578-1581.

Witter, M.P. (1993). Organization of the entorhinal-hippocampal system: a review of current anatomical data. Hippocampus *3 Spec No*, 33-44.

Wittmann, B.C., Schott, B.H., Guderian, S., Frey, J.U., Heinze, H.J., and Duzel, E. (2005). Reward-related FMRI activation of dopaminergic midbrain is associated with enhanced hippocampus-dependent long-term memory formation. Neuron *45*, 459-467.

Wixted, J.T. (2007). Dual-process theory and signal-detection theory of recognition memory. Psychol Rev *114*, 152-176.

Wyble, B.P., Linster, C., and Hasselmo, M.E. (2000). Size of CA1-evoked synaptic potentials is related to theta rhythm phase in rat hippocampus. J Neurophysiol *83*, 2138-2144.

Xiang, J.Z., and Brown, M.W. (1998). Differential neuronal encoding of novelty, familiarity and recency in regions of the anterior temporal lobe. Neuropharmacology *37*, 657-676.

Yamaguchi, S., Hale, L.A., D'Esposito, M., and Knight, R.T. (2004). Rapid prefrontal-hippocampal habituation to novel events. J Neurosci *24*, 5356-5363.

Yanike, M., Wirth, S., and Suzuki, W.A. (2004). Representation of well-learned information in the monkey hippocampus. Neuron *42*, 477-487.

Yarbus, A.L. (1967). Eye movements and vision (New York: Plenum).

Yonelinas, A.P. (2001). Components of episodic memory: the contribution of recollection and familiarity. Philos Trans R Soc Lond B *356*, 1363-1374.

Yonelinas, A.P., Kroll, N.E., Quamme, J.R., Lazzara, M.M., Sauve, M.J., Widaman, K.F., and Knight, R.T. (2002). Effects of extensive temporal lobe damage or mild hypoxia on recollection and familiarity. Nat Neurosci *5*, 1236-1241.

Yonelinas, A.P., Otten, L.J., Shaw, K.N., and Rugg, M.D. (2005). Separating the Brain Regions Involved in Recollection and Familiarity in Recognition Memory. J Neurosci *25*, 3002-3008.

Zohary, E., Shadlen, M.N., and Newsome, W.T. (1994). Correlated neuronal discharge rate and its implications for psychophysical performance. Nature *370*, 140-143.