Neural Correlates of Memory Consolidation during Waking State and Sleep

Inaugural-Dissertation
zur Erlangung der Doktorwürde
der
Philosophischen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität
zu Bonn

vorgelegt von
Lorena Deuker
aus
Gießen

Bonn 2014

Gedruckt mit der Genehmigung der Philosophischen Fakultät
der Rheinischen Friedrich-Wilhelms-Universität Bonn

Zusammensetzung der Prüfungskommission:

PD Dr. Christian Montag
(Vorsitzender)

Prof. Dr. Martin Reuter
(Betreuer und Gutachter)

PD Dr. med. Nikolai Axmacher
(zweiter Gutachter)

Prof. Dr. Ulrich Ettinger
(weiteres prüfungsberechtigtes Mitglied)

Tag der mündlichen Prüfung: 20.12.2013

Acknowledgements - Danksagung

An dieser Stelle möchte ich mich bei allen bedanken, ohne deren Unterstützung die vorliegende Arbeit nicht möglich gewesen wäre. Allen voran sind das natürlich Professor Martin Reuter als Betreuer meiner Doktorarbeit, der mir über die gesamte Dauer meiner Disseration hinweg beratend und motivierend zur Seite stand, sowie Nikolai Axmacher und Jürgen Fell, welche mir in ihren Arbeitsgruppen ein harmonisches Umfeld boten, in dem ich selbständig und kreativ wissenschaftlich tätig sein konnte. Die beiden hatten immer ein offenes Ohr für meine Fragen und Probleme und eröffneten mir neue Perspektiven, wenn ich einmal nicht weiterkam. Ich bedanke mich außerdem bei Professor Christian Elger für die Möglichkeit, intrakranielle Daten in seiner Klinik zu erheben sowie bei Professor Bernd Weber für den umkomplizierten Zugang zum MRT-Scanner, ebenso bedanke ich mich bei Peter Trautner, der bei Problemen mit dem Scanner stets zügig und kompetent geholfen hat.

Mein Dank gilt auch allen aktuellen und ehemaligen Mitgliedern der AG Cortical Oscillations für die vielen Diskussionen und Tee-Küchen-Gespräche. Besonders bei Thorsten Kranz möchte ich mich bedanken, der mir beim Erlernen der Programmiersprache Python immer geduldig geholfen hat. Anne DoLam möchte ich danken, weil sie mir den Umgang mit Patientenmessungen beigebracht hat. Jan Olligs gilt mein Dank für die Hilfestellung beim Erlernen der Mustererkennungsanalyse-Toolbox PyMVPA. Ebenso Yilmaz Sagik, der mich bei den Patientenmessungen unterstützt hat. Alejandro Suarez, Linda Wilbers und Anna Müller für ihre Unterstützung beim Erheben der MRT-Daten. Marcin Leszczynski, Amir Jahanbekam, Leila Chaieb und Paul Jacks gebührt mein Dank für Korrekturlesen und hilfreiches Feedback.

Ich möchte mich außerdem bei meinen Eltern bedanken, die mich während meines Studiums nicht nur finanziell, sondern auch moralisch unterstützt haben und mir auch während der Doktorarbeitszeit eine große Hilfe waren. Zuletzt bedanke ich mich auch bei Claudio Bierig, der immer für mich da war und sich geduldig Erfolge und Probleme meiner Arbeit angehört hat.

Contents

1 Introduction

The ability to remember the past is probably one of the most valuable assets of any organism. It enables us to benefit from successful experience and prevents us from repeating the same mistakes over and over again. More philosophically, memory can be viewed as the glue that keeps our life together, provides a red line which links disparate events, places and people into unified experiences and ultimately enables us to form a sense of personhood and identity.

The complexity of memory functions ranges from more basic forms such as modifying behavior with operant conditioning, to motoric skills and procedural memory and finally to hallmarks of human experience such as “mental time travel” and autobiographic memory.

Every day, a person encounters an overwhelming amount of information. Some of it will be kept in their memory for a long time, while most of it will soon be forgotten. According to two-step models of memory formation, certain neuronal processes have to take place in order to transform initially labile memories into permanent memory traces. This process of strengthening, or stabilization, has been called consolidation.

The neuronal mechanisms underlying this stabilization remain to be completely understood. However, there is mounting evidence that sleep is important for successful memory consolidation, maybe even essential. The study of sleep as a potential memory enhancer merits special attention in a society in which, on the one hand, sleep disorders are on the rise and which, on the other hand, faces the challenge of an aging population with the associated increase in mild cognitive impairment and dementia.

In recent years, it has been suggested that the beneficial effect of sleep on memory consolidation may be due to specific neuronal processes which happen exclusively or most efficiently during sleep, a state in which the brain is mostly insulated from external influences. One of these processes is a reactivation of the same neuronal activity that was involved in the original learning process: In rodents, it has been demonstrated that sequences of neuronal activity, which are associated with specific content during a learning task, are spontaneously “replayed” during sleep after learning.

In this thesis, three studies are presented which used a novel method of analysis to investigate whether a reactivation of stimulus-specific activity can also be identified in humans. In these studies, stimulus-specific neuronal activity patterns were extracted from functional imaging and electrophysiological data that was recorded during a learning task and the reoccurrence of these patterns was tracked in subsequent waking state and sleep.

In the first part of this thesis, the theoretical concepts related to memory consolidation, sleep and reactivation are discussed. In addition, important methods are introduced. In the second part, three empirical studies are presented that investigated stimulus-specific neuronal replay with different methods and different paradigms. Lastly, the results of the three studies are compared and discussed.

Part I

Theoretical Part

2 Memory systems

Memory is not a unitary concept, but can be divided into several subcomponents. One of the most widely accepted models of memory proposes two main memory systems – declarative and non-declarative memory (Squire and Zola, 1996). An overview of this model is given in Figure 2.1. Declarative memory refers to consciously accessible memory, such as semantic knowledge of facts or the memory about one's own past. Non-declarative memory subsumes several functions which are not consciously accessible, such as procedural memory (knowing how to ride a bike) or implicit knowledge, such as having learned statistical regularities without being consciously aware of it (Knowlton et al., 1994).

The main evidence for the proposed distinction between declarative and non-declarative memory comes from patients with brain damage, in which a dissociation has been observed: Depending on which regions are affected, patients may be unable to form declarative memories, but they can still acquire new motor skills (Squire and Zola, 1996).

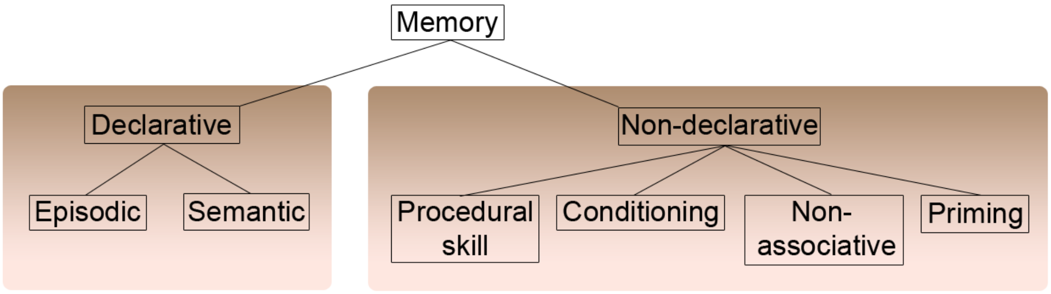

Figure 2.1: Different types of memory can be grouped into declarative and non-declarative memory. Declarative memory includes all forms that are consciously accessible and is divided into episodic memory (e.g. "memory of the first day of school") and semantic memory, which is memory about facts (e.g. "Paris is the capital of France"). Non-declarative memory usually cannot be verbalized and is often implicit, i.e. a person knows something without having conscious access to the knowledge. For example, procedural memory includes skills such as riding a bike, which most of us can do but have difficulty explaining how. Figure adapted from Walker and Stickgold (Walker and Stickgold, 2004), based on Squire and Zola (Squire and Zola, 1996).

From the study of neuropsychological patients, it has become apparent that declarative memory depends on the integrity of the hippocampus (Scoville and Milner, 1957) and other medial temporal lobe structures, while non-declarative memory seems to be independent of these brain areas. The role of the hippocampus is discussed in more detail below.

This thesis focuses on episodic memory, which is a subcomponent of declarative memory and which refers to the memory for events in a person's past (Walker and Stickgold, 2004).

3 Memory consolidation

3.1 Origins and development

What do we really talk about when we talk about memory consolidation? When the term was first coined by Mueller and Pilzecker (Mueller and Pilzecker, 1900), they used it to describe resistance against interference: In nonsense-syllable learning paradigms similar to those employed by Ebbinghaus (Ebbinghaus, 1885), they found that memory recall for syllable pairs was worse when additional material had to be learned between encoding and retrieval of the original material, an effect they termed "retroactive interference". This effect was less pronounced when the time increased between encoding of the original material and exposure to the interfering material. This led them to conclude that soon after encoding of new material, a process takes place which stabilizes or "consolidates" the memory and renders it more resistant against interference.

A second, complementary avenue of research began nearly at the same time with the study of amnesic patients. The loss of memory in global amnesics was often found to display a prominent temporal gradient that can be summed up as "last in, first out": Recent memories are more likely to be lost than remote memories, an observation that had been described some years earlier by a French psychologist (Ribot, 1882) and is epitomized today as "Ribot's law". What is striking in these amnesic patients is that the gradient of memory loss can span years – in some cases all memory is lost except for episodes from early childhood. But equally puzzling is the fact that memory is not lost completely. Thus, the underlying problem in these patients is not simply disrupted retrieval – they can still access some of their memories. Over the course of months and years, the older memories must have undergone a transition which makes them less vulnerable to forgetting,

Thus, in both of these two approaches, memory becomes more resistant with longer time periods since encoding. However, the stabilization process which was observed by Mueller and Pilzecker (Mueller and Pilzecker, 1900) took place within minutes, not days, weeks or even years as in the neuropsychological patients. Is this at all the same process then? Today, it is commonly understood that the effects which can be observed at the scale of seconds and minutes reflect synaptic consolidation, while processes spanning months and years are connected to the somewhat harder to grasp concept of system consolidation (Dudai, 2004; Frankland and Bontempi, 2005), both of which will be discussed in more detail below. But even if the underlying neuronal substrate is probably quite different, both phenomena have in common that the passage of time seems to change something about the memory traces that leaves them more resistant against disturbance.

3.2 Interference and forgetting

The two approaches have another thing in common: They describe consolidation in terms of forgetting. This highlights that consolidation is a theoretical construct that can only be inferred from behavior, either from post-learning improvement (e.g. in procedural memory, see below in the chapter "Sleep") or from reduced or increased forgetting relative to a control condition. As such, much of the genuinely psychological evidence that consolidation exists and that it constitutes an important mechanism is intricately linked to research into forgetting.

Thus, when ideas regarding the causes of forgetting changed, the concept of memory consolidation virtually disappeared from psychological research. In 1957, an influential article was published, in which retroactive interference was largely dismissed as a cause of forgetting (Underwood, 1957). The author reviewed studies in which a word list had to be learned and recall of this list was tested 24 hours later. The percentage of correctly recalled items varied greatly from one study to the other. Underwood discovered that almost all of the variance in forgetting could be explained by the number of word lists that participants in these studies had learned *prior* to the learning of the target word list, including preceding practice runs. For this observation, the term "proactive" interference was coined. This refers to an interference which acts in a forward manner and describes a disturbance of subsequently learned material by previously learned material. This new view on forgetting had a profound impact on consolidation research – or lack thereof: If memory success was indeed determined already at the time of learning (by what was learned before), there is no room for a process like consolidation, which takes place after learning.

Today, this change in doctrine clearly seems like a "wrong turn" (Wixted, 2004, p. 240) in the study of the causes of forgetting. Proactive interference could not explain many of the phenomena that had already been observed, for example studies demonstrating that subsequent periods of sleep as compared to wakefulness led to better memory performance (Jenkins and Dallenbach, 1924; Ekstrand, 1967). It also fails to explain why recall of a list that is learned first (i.e. for which there is no preceding list) is worse when a second list is interjected between learning and recall than when no list intervenes, and why the performance decline is not as strong when the intervening list is presented later rather than earlier after initial learning (Mueller and Pilzecker, 1900). For many results, proactive interference theory has to be bent and twisted in order to fit.

Wixted (Wixted, 2004) offers an excellent account of the history of this avenue of research and describes fatigue among psychologists following years of debate and research on the causes of forgetting with ever more complicated theories and experimental designs without gaining any progress. This fatigue might have been the reason why psychological research into consolidation lay dormant for most of the second part of the 20th century.

There might have been another, more fundamental reason for psychologists' departure from this area of research: Trying to elucidate an abstract concept such as consolidation solely with behavioral paradigms often poses a classical problem of reverse inference: A result – such as diminished forgetting relative to a control condition – is observed, and any number of underlying constellations could be the cause for this result: retroactive interference, proactive interference, simple decay or a mixture of all of these processes, not to mention processes related to attention, encoding and retrieval success, which also affect memory performance.

Biologists and neuroscientists seem to be at an advantage here, because they can test theories that take into account different sub-components, anatomical regions and their respective contributions. In animals, they can manipulate brain function with lesions or by the administration of drugs. They can also invasively measure activity in brain regions of interest.

Today, consolidation research seems to be firmly in the hands of neuroscience. While the existence of consolidation is an accepted fact among neuroscientists, it is barely even discussed by psychologists (Wixted, 2004). In fact, even in a recent psychological textbook on memory (Baddeley et al., 2009), the concept of consolidation is only mentioned twice in brief paragraphs.

In the next chapter, memory consolidation will be described from a neuroscientific perspective.

4 The neural correlates of memory consolidation

4.1 Synaptic consolidation

The neuronal foundations of learning are the forming of new synapses and changes in synaptic strength at existing synapses. These changes happen as a consequence of neuronal activity during a learning experience.

According to Hebb's rule, "cells that fire together, wire together" (Hebb, 1949). What does this rule express exactly? Consider two neurons that are connected via a synapse. Neuron A is pre-synaptic and neuron B is post-synaptic. If neuron B repeatedly fires shortly after neuron A, then processes will take place at their synapse which make it more likely in the future that an action potential in neuron A induces an action potential in neuron B. If, however, neuron A repeatedly fires shortly after neuron B, processes will take place at their synapse which make it less likely that, in the future, neuron A induces an action potential in neuron B. This restructuring of synapses is also referred to as spike-timing dependent plasticity and is achieved by long-term potentiation and long-term depression, respectively. Long-term potentiation can be observed especially well in the human hippocampus (Birbaumer and Schmidt, 2010, p. 66), which plays an important role in memory formation, as will be discussed below.

Strengthening of synapses depends on a cascade of molecular changes, which may be disrupted by behavioral interference, drugs, seizures or anatomical lesions (Dudai, 2004). For example, a tonic-clonic seizure often induces an amnesic gap of several minutes up to a couple of hours. During that amnesic gap, the patient often appears already fully oriented, reacts appropriately to questions and can keep up a conversation; however, when asked later, he or she has no memory of these episodes.

One common aspect of interventions that disrupt synaptic consolidation is that they cause amnesia or memory loss when applied within a certain time window, but do not impair memory when applied later (Frankland and Bontempi, 2005). The length of this time window ranges from seconds and minutes to several hours (Dudai, 2004). However, temporal gradients of amnesia in neuropsychological patients imply that memory stabilization happens over time periods as long as years and decades. This long-term stabilization is conceptualized by system consolidation.

4.2 System consolidation

Whereas synaptic consolidation happens relatively short-term and at the level of synapses and neurons, system consolidation refers to long-term changes at the level of the whole brain.

Knowledge in this field was first gained by systematic research, instead of anecdotal reports, about the memory loss in amnesic patients, using standardized questionnaires (Sanders and Warrington, 1971). In a review of the literature up until then, Squire and Alvarez (Squire and Alvarez, 1995) concluded that in most cases, amnesia is temporally graded, confirming earlier studies (Ribot, 1882; Burnham, 1903).

It should be noted that some of the studies with globally amnesic patients have not found temporally graded amnesia (Sanders and Warrington, 1971), but instead described extensive memory loss with no apparent discrimination between recent and remote memories. However, this might be due to differences in patient populations: It appears that patients with damage restricted to the medial temporal lobe (MTL), i.e. the hippocampus and adjacent entorhinal and perirhinal cortices, display temporally graded amnesia while those patients who have extensive and “flat” memory loss often have broader lesions beyond the hippocampus, including neocortical regions in the lateral and anterior temporal lobe (Squire et al., 2001).

This last finding already points to the pivotal role of the hippocampus in long-term memory consolidation.

4.2.1 Patient H.M.

The starting point for hippocampus research arguably was the unfortunate case of Henry Molaison (commonly referred to as H.M.), who underwent bilateral resection of his medial temporal lobes after having suffered from severe, pharmaco-resistant epilepsy for years (Scoville and Milner, 1957). While the surgery was successful in reducing the frequency of his epileptic seizures, it also quickly became apparent that H.M was no longer able to form new declarative long-term memory; he suffered from a severe case of anterograde amnesia. His memory for events preceding the surgery was largely unaffected, even though there is evidence for temporally graded retrograde amnesia spanning a period somewhere between 3 (Scoville and Milner,

1957) and 11 (Corkin, 2002) years. Curiously, procedural learning (i.e. motor skills) and formation of implicit memory were still possible. Also, his working memory remained intact.

This case has been cited and reviewed so often that it could be considered trivial. Still, some of the conclusions that can be drawn shall be briefly discussed here to the extent in which they relate directly to the study of consolidation. One of the most compelling conclusions from the case H.M. certainly is that the neuronal substrate which stores memory and the substrate which promotes or achieves such storing have to be different: The patient still possessed remote declarative memories while the storing of new declarative content had become impossible.

The fact that the patient could hold items in working memory proves that it was not initial encoding of new material that prevented him from forming new memories. Likewise, the deficit was not solely in retrieval of items from memory, as he was still able to retrieve episodes preceding his surgery. What apparently had become dysfunctional was a process taking place between initial encoding and later recall – the stabilization, or consolidation, of newly learned material.

The second conclusion is that this deficient process is most likely linked to the anatomical regions which were removed in the patient. The loss of the hippocampus was identified to be responsible for H.M.'s memory dysfunction – even though it should be noted that the resection in H.M. was massive and included the entorhinal cortices and amygdalae as well as more of the surrounding tissue. However, it has since been replicated that it is damage to the hippocampus which leads to the type of amnesia observed in H.M (Penfield and Milner, 1958; Corkin, 2002).

4.2.2 Lesion studies in rodents

Conclusions drawn solely from neurospsychological patients easily inspire doubt in their validity. For example, H.M. suffered from severe epilepsy prior to his surgery; his memory might have been affected as a consequence of his illness prior to removal of the hippocampi without anyone noticing it. Likewise, patients suffering from damage in the medial temporal lobe due to stroke or cerebral injury very often have wide-spread lesions far beyond the hippocampus or parahippocampal cortex.

Lesion studies in rodents can complement insights derived from neuropsychological patients. Even though it is difficult to compare memory systems between rodents and humans, there are some approaches that convincingly model human learning.

In one study (Kim and Fanselow, 1992), rats were fear conditioned to either an environment (context) or to a sound by the application of electrical foot shocks. Memory of the fear conditioning was later assessed by the degree of freezing that the rat displayed when confronted again with the context or the sound. At a timepoint that was either 1, 7, 14 or 28 days after fear conditioning, the hippocampus was selectively lesioned. The rats displayed temporary graded amnesia for the fear conditioning, that is, they displayed less or no freezing for recently acquired fear conditioning (indicating lack of memory for the conditioning), but normal freezing for remote memories. Importantly, this was only true for fear conditioning that involved a context, but not a sound. As the context-related conditioning is more similar to episodic memory in humans, and the graded amnesia is specific for this condition, this provides good support for the role of the hippocampus in consolidation of episodic-like memory.

In another study, social transmission of food preference was investigated (Clark et al., 2002). In this experimental setup, rats display a preference for food they have previously smelled on the breath of their peers without actually having tasted the food. The preference is likely an adaptive behavior, as prior sampling by peers signals the food is safe to eat. Such acquired food preference lasts up to several months and thus constitutes a good model for non-spatial memory. Again, electrolytic lesions of the hippocampus caused anterograde amnesia and temporally graded retrograde amnesia for socially transmitted food preference.

These are only two examples which were selected from a large body of lesion studies in animals (reviewed in Frankland and Bontempi, 2005) which support the idea that lesions to the hippocampus induce temporally graded retrograde amnesia.

4.2.3 Beyond the medial temporal lobe

For a long time, the hippocampus was the sole center of attention in research of declarative memory consolidation, but other brain areas have increasingly gained more consideration.

The amygdala is a small almond shaped structure in the immediate vicinity of the hippocampus. When it is damaged, the processing of emotional material is adversely affected (Adolphs et al., 1997; Adolphs et al., 1999), especially with regard to negative emotions such as fear. On the other hand, it has been shown that memory for emotionally arousing material in healthy humans is superior to that of non-arousing material (Hamann, 2001; Kensinger and Corkin, 2003; Kensinger, 2004). It therefore seems plausible to assume that a structure which is associated with the processing of emotions is also involved in memory consolidation.

There are several studies which confirm the role of the amygdala in memory. For example, in a PET study, amygdala activation during encoding was associated with performance at recall across healthy participants (Cahill et al., 1996). In an fMRI study, event-related amygdala activation during the viewing of emotionally negative and neutral scenes was predictive of performance in an unexpected memory test three weeks later – but only for scenes that were rated by the subjects as the most emotionally intense (Canli et al., 2000). It has been suggested that the amygdala induces this enhancement of memory performance for emotional content by modulating hippocampal activity (Cahill and McGaugh, 1998; McGaugh, 2004).

There is good evidence that the effects of emotional arousal on memory are not only due to enhanced attention or saliency at encoding, but that memories for emotional content are consolidated differently (Hamann, 2001). In fact, studies have shown convincingly that emotional memories benefit more from sleep than neutral memories, especially during REM sleep (Wagner, 2001; Hu et al., 2006; Payne et al., 2008; Nishida et al., 2009). The role of sleep in memory consolidation will be further discussed below.

Another important brain area in long-term memory is prefrontal cortex. One could call it the remote-memory counterpart of the hippocampus. While disruption of the hippocampus affects recent memories, lesions in prefrontal cortex lead to a loss of remote memories (Takehara et al., 2003; Beeman et al., 2013).

The exact role of prefrontal cortex in the recall of remote memory is not clear at the moment. It has been suggested that prefrontal cortex, over time, takes over the role of the hippocampus in combining different parts of an episode into one memory trace and that it is necessary for strategic retrieval of memory content (Frankland and Bontempi, 2005).

In humans, the prefrontal cortex has been found to be involved in memory retrieval already in the earliest of imaging studies (Rugg et al., 1996; Henson et al., 1999), but these studies probed memory for recent memories. One study found that activity in prefrontal cortex was modulated by the age of a memory (Maguire et al., 2001), but BOLD activity actually *decreased* with increasing age of the memory.

Investigation of remote memories in humans is probably impeded by the methodological difficulty of assessing very old memory content, especially with regard to the accuracy of recall. For example, when probing memory for early childhood events, researchers usually have to rely on information given by the participants. But clearly, this area of research merits further attention and the role of prefrontal cortex in memory consolidation should be considered more carefully.

4.3 The standard model of memory formation

Even though different brain areas should be considered in memory consolidation research, the hippocampus clearly plays a very important role in declarative memory. Its exact contribution has been discussed and described in models of two-step memory formation (Marr, 1970; Marr, 1971) and is an integral part of what is now considered the standard model of memory formation (Squire and Alvarez, 1995; McClelland et al., 1995).

This model postulates that new information is initially represented by neuronal activity in disparate cortical modules, such as visual or somatosensory areas. The hippocampus binds features from these disparate modules into a coherent memory trace, or episode, and stores it in a rapid and temporary manner. Here, synaptic consolidation is assumed to achieve this initial storing.

In a second step, information related to this memory trace is transferred from the hippocampus to cortical areas in which they are then stored as long-term memory (Squire and Alvarez, 1995; McClelland et al., 1995; Squire et al., 2004; Hasselmo, 2005). Figure 4.1 provides an overview of the model.

In this model, the hippocampus is often conceptualized as a “fast learner” while the cortex is deemed a “slow learner” (McClelland et al., 1995; Frankland and Bontempi, 2005). It is not clear why new information is not stored directly within cortical areas. It has been suggested that only the hippocampus is capable of performing the necessary rapid synaptic changes (Lisman and Morris, 2001; Frankland and Bontempi, 2005).

The process by which memory traces become independent of the hippocampus and can be accessed even if the hippocampus is removed, is proposed to be a gradual

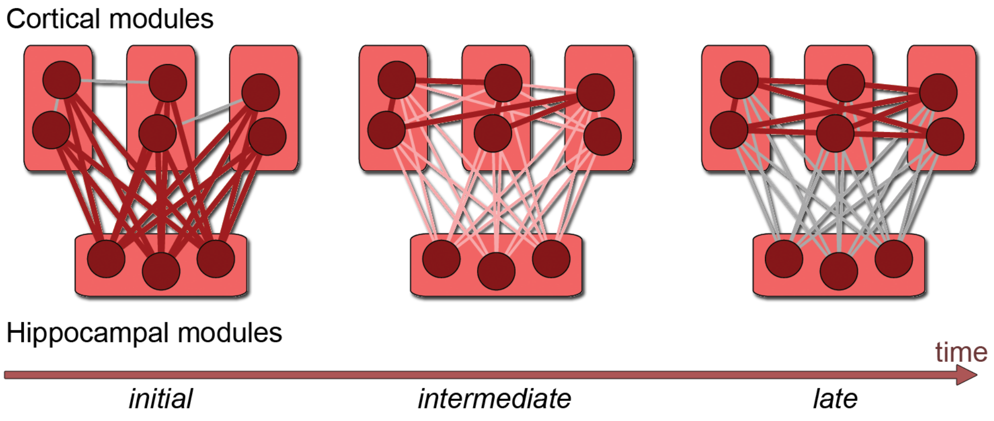

Figure 4.1: The standard model of memory formation as it has been described in the literature (Squire and Alvarez, 1995; McClelland et al., 1995). While initial encoding of new material relies on different cortical areas (such as sensory or motor cortices), it is the hippocampus that binds these representations together into a coherent memory trace in a fast manner. Through repeated reactivation of this hippocampo-cortical network, the intrinsic connections between the cortical modules become stronger. Finally, the memory trace is fully represented by strong connections within cortical modules and becomes independent of the hippocampus. Figure adapted from Frankland and Bontempi (Frankland and Bontempi, 2005).

information transfer from hippocampus to the cortex. This transfer is assumed to happen by a repeated reactivation of the neuronal activity that was associated with encoding (Marr, 1971; Káli and Dayan, 2004), probably driven by the hippocampus. This reactivation is assumed to induce a gradual reshaping of cortical connections so that the new information is, ultimately, represented by cortical modules alone.

4.4 Why do memories have to be consolidated at all?

There are different ideas on why such a two-step process as described above might have evolved. One aspect could be limits of the neuronal substrate (Dudai, 2004). Assuming that the human brain can only store a finite number of representations, simply adding every new episode indiscriminately is not efficient. Integrating new information into existing networks, building on similarities and altogether getting rid of information which proves to be unimportant would, in contrast, save capacity.

Another reason might be that immediate integration of new memories into existing memory networks might lead to catastrophic loss or distortion of older memories (Frankland and Bontempi, 2005). This in turn is linked to the implicit understanding that during the integration process, old memories become temporarily vulnerable again. Studies investigating retrieval induced forgetting indirectly support this notion. They show that simply by retrieving material which is in some way related to previously studied items (e.g. by sharing the same category), memory for the items is impaired during later recall (Anderson et al., 1994; Ciranni and Shimamura, 1999).

Thus, one prevalent view is that memory consolidation takes place most efficiently during periods of cortical “silence”. The best known period of cortical silence is sleep, which will be discussed in the next section. After this, the focus of this introduction will return to the process of reactivation and how it has been studied in rodents and humans.

5 Sleep

5.1 Physiology of sleep

In recent years a major research effort has been devoted to investigate the role of sleep in system consolidation. Apart from many empirical findings, this is also based on theoretical considerations: Sleep, in which humans are largely insulated from stimuli in their environment, might provide an optimal environment for the reactivation, or replay, of neuronal activity which is necessary for the gradual information transfer from hippocampus to neocortex.

Sleep is a physiological state that can be observed in many species, and in virtually all mammals. It is characterized by “a rapidly reversible state of reduced responsiveness, reduced motor activity and reduced metabolism” (Siegel, 2009, p. 747).

Human sleep has been extensively studied with electroencephalography (EEG). It has been found that human sleep can be divided into different stages, which can be visually identified from EEG, using additional data from electromyography (EMG) and electrooculography (EOG). The most common system of classifying sleep stages (Rechtschaffen et al., 1968) distinguishes five sleep stages: sleep stages 1-4 with increasing depth of unconsciousness and rapid eye movement (REM) sleep, the appearance of which in EEG is similar to waking state or the lightest stage of sleep but in which a distinct flat muscle tone combined with rapid, large eye movements can be observed.

Healthy sleep is usually organized in cycles: A person progresses from waking state consecutively through sleep stages 1, 2, 3 and 4. After remaining some time in slow-wave sleep (SWS), i.e. stages 3 and 4, the person cycles back through stages 2 and 1. Sometimes, the person shortly wakes up after sleep stage 1, which is usually not remembered. He or she may also have an episode of REM sleep or go back directly to the deeper stages of sleep for a second cycle. Figure 5.1 shows an example. As a rule of thumb, the first two or three cycles (the first half of the night) have extended periods of slow-wave sleep and relatively little REM sleep. In the second half of the night, REM sleep periods become increasingly longer and often the person does not reach slow-wave sleep anymore.

It is not clear what happens when humans do not sleep at all. There have been

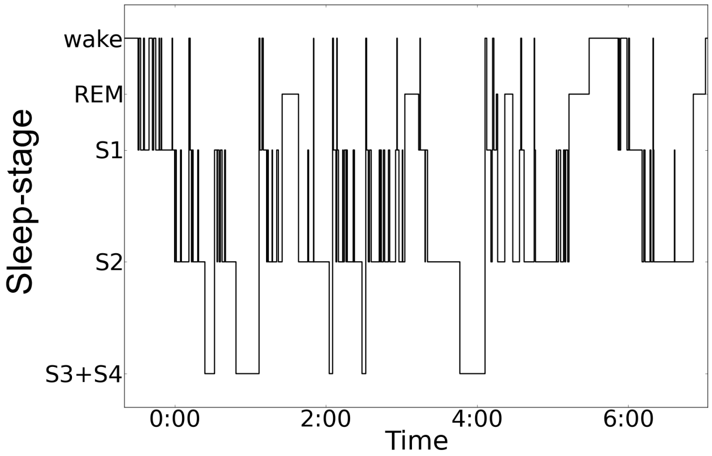

Figure 5.1: Examplary hypnogram. The subject cycles from waking state through sleep-stages 1 and 2 (S1 and S2), and reaches slow-wave sleep (S3+S4) for the first time around 0:30. Note that slow-wave sleep occurs before 4:00, after which the subjects does not go back again into slow-wave sleep. Rapid eye movement (REM) sleep is plotted between waking state and S1, which is often done in order to illustrate that REM sleep is neither wake nor deep sleep. In this subject, REM sleep occurs predominantly after 2:00, i.e. during the second half of the night. The subject also has short returns to the waking state throughout the night, which is not unusual. Very often, these short periods of wakefulness are not remembered in the morning.

studies in which rats died subsequent to massive sleep deprivation (Rechtschaffen and Bergmann, 2002), but it is difficult to assess whether lack of sleep or the associated severe stress is responsible for this: The "Disc-Over-Water" (DOW) method entails pushing the animal into water whenever their electrophysiological recording indicates the first signs of sleep. A study with pigeons did not find lethal effects of sleep deprivation (Newman et al., 2008). In humans, the longest scientifically documented case of voluntary sleep deprivation of 11 days did not result in death or even severe adverse health effects (Ross, 1965). However, systematic research with enforced total sleep deprivation in humans cannot be done for obvious ethical reasons.

Still, it has become more obvious in recent years that insufficient or disrupted sleep is associated with a number of detrimental effects. Even though the causal direction has not been established, abnormal sleep patterns are associated with a number of psychiatric illnesses (Benca, 1992; Roth et al., 2010; Kyung Lee and Douglass, 2010). Investigation of the exact role of sleep in humans may one day provide therapeutic potential for these disorders.

5.2 Effect of sleep on memory consolidation

5.2.1 Behavioral findings

In recent years, it has been suggested that sleep plays an important role in memory consolidation (Walker and Stickgold, 2004; Stickgold and Walker, 2005; Ellenbogen et al., 2006b; Diekelmann and Born, 2010). Empirical evidence that declarative memory is retained better after a period of sleep than after a period of wakefulness is quite old (Jenkins and Dallenbach, 1924). However, many issues have to be considered before a definite statement can be made. For example, in a typical experimental setup, one group studies a list in the morning, stays awake for 12 hours during the day and performs a memory test in the evening. The sleep group studies the list in the evening, sleeps and is tested in the morning, again, 12 hours after studying the list. Obviously, other factors than sleep could explain a memory benefit for the sleep group – such as time of day effects, which might well be related to varying hormonal levels that influence memory formation. In another experimental setup, sleep deprivation is used: both groups study a list in the evening, one group sleeps and the other is kept awake during the night. Group differences at memory recall in the morning could then easily explained by mere tiredness of the wake group or the stress that is associated with being kept awake.

Ellenbogen and colleagues (Ellenbogen et al., 2006b) have grouped the controversial views on sleep and memory consolidation into four main categories of sentiments: First, sleep has no impact whatsoever on memory formation. This view is mainly supported by the existence of patients who have virtually no REM sleep, either due to certain anti-depressant drugs or brain stem damage (Siegel, 2001; Vertes, 2004). However, systematic research has apparently not been performed in these patients and even if they were found to be cognitively unimpaired, this would only challenge the role of REM sleep. This first view is also easily refuted by studies which show an advantage of a sleep group over a wake group, especially when time of day effects are excluded, e.g. by experiments that use an afternoon nap with identical timings for the sleep and wake group (Tucker et al., 2006).

The second view (which is not easily separable from the third view) holds that the only beneficial influence of sleep consists in reducing interference from normal, everyday activities that participants might carry out if they were awake. In this view, there is no critical time window for consolidation, and memory content thus becomes equally vulnerable again as soon as participants wake up. This view, along with a number of other concerns, has been elegantly refuted by a study that experimentally manipulated the timing of interference as well as the time of learning and testing: Subjects had to learn an A-B list first, then spent 12 hours either awake or asleep (during the day and during the night, respectively). Half of the participants of both groups had to learn an A-C list 12 minutes prior to testing of the original A-B list (interference groups) while the other half was tested without a preceding interfering list (no-interference groups). Subjects who had slept performed slightly better than those who had stayed awake in the no-interference groups. However, participants who had slept performed considerably and significantly better than those who had been awake in the interference groups: correct recall of the A-B list was 76% in the group who had slept and 32% in the group who had stayed awake. Of note is a fifth group of participants in this study who learned material in the evening and were tested 24 hours later, also in the evening and also with interference prior to testing. This group still performed better (correct recall: 71%) than the subjects who had stayed awake for 12 hours and had been given an interfering list. This fifth group addresses concerns about time of day effects as they were tested at the same time of day as the awake group (i.e., in the evening). Also, this group spent the same amount of time awake as the awake group (12 hours during the day), thus being exhibited to the same possibly detrimental effects of normal day-time interference. Also, the time between learning and testing was much longer in this fifth group which should have led to worse performance. But instead, this group performed better than the awake-interference group – apparently a night of sleep after learning preserves memory even during longer intervals between learning and testing.

The third view identified by Ellenbogen and colleagues (Ellenbogen et al., 2006b) posits that sleep has a positive influence on memory recall by providing a time window during which memory consolidation can happen effectively. This view suggests a passive or permissive influence of sleep. It acknowledges that sleep provides a good, maybe even unique environment for the stabilization of memories. However, there is no active mechanism that causes memory consolidation.

This is what the fourth view suggests: That there are specific characteristics and neuronal processes during sleep that actively promote consolidation.

It is very difficult if not impossible to assess the claims of either the third or the fourth view without using neuroscientific methods, and they will be discussed in more detail below. First, I will consider some of the behavioral evidence for an active role of sleep.

As outlined earlier in this introduction, there are different types of memory. Interestingly, sleep seems to have a differential impact upon them. Tasks involving procedural memory (such as texture discrimination or motor sequence tasks) typically not only show less deterioration after sleep as compared to wakefulness, but the skill is even *enhanced* by sleep (Fischer et al., 2002; Walker et al., 2003; Stickgold et al., 2000). This in itself might provide a first tentative argument against a passive mechanism – obviously, something happens after procedural training that leads to an even better performance. However, moderate improvement in these tasks can usually also be observed after a period of waking state (Walker et al., 2003).

In declarative memory, sleep shows the more historically established stabilization against forgetting, that is, one can observe less forgetting in participants who slept compared to those who stayed awake. Some studies have shown that sleep's beneficial effect on declarative memory is more pronounced when the learning was difficult (Drosopoulos et al., 2007). All of this begs the question: If sleep simply permitted passive consolidation to take place, why would the impact of sleep be different depending on the task?

Perhaps the most intriguing behavior-based evidence for an active role of sleep stems from studies that show that sleep not only preserves memory, but also leads to a restructuring of learned elements, that is, it has not only a quantitative, but also a qualitative influence on memory.

For example, Ellenbogen and colleagues let subjects study individual relation pairs such as $A > B$, $B > C$, $C > D$, $D > E$ and $E > F$ (Ellenbogen et al., 2007). After learning, subjects were tested on these studied pairs but were also asked to give relational judgements about new pairs (e.g. $B?D$). Unbeknownst to the subjects, there was a transitive relation between the items (i.e. $A > B > C > D > E > F$). Immediately after learning, subjects had no insight into this relation as revealed by testing of the novel pairs. One group of participants was retested after 20 minutes, and two groups were retested after 12 hours of either sleep or waking state. While the group with 20 minutes test-retest interval gained no insight into the transitive relational structure of the pairs, both 12 hour groups were better at making first-order transitive inferences (such as $B > D$). However, the subjects who had slept had significantly better performance at second-order transitive inference ($B > E$). These results can be interpreted in such a way that the mere passage of time leads to an adaptive reconfiguration of memory content, but that sleep might provide a better environment for this process to take place or might actively enhance it.

Such qualitative changes in memory, in some cases referred to as “insight”, have been reported in numerous studies (Wagner et al., 2004; Fischer et al., 2006; Ellenbogen et al., 2007) and they can be taken as behavioral evidence that sleep actively reconfigures memory traces.

5.2.2 Neuroscientific findings

Neuroscience can contribute to the study of sleep’s role in memory consolidation by identifying neuronal processes or characteristics that are unique to sleep, or to individual sleep stages, and then relating the frequency or strength of these characteristics to memory performance. In addition, the mechanisms may be experimentally manipulated, with resulting changes providing a strong argument for a causal role of the mechanisms.

In a first approach, different sleep stages can be considered. Neither memory nor sleep are simple, unitary concepts and one of the major challenges has been to disentangle which type of memory benefits from which type of sleep and how. This is typically investigated by comparing the effect of the first half of a night (predominantly slow-wave sleep) to the second half of the night (predominantly REM sleep) on different types of memory tasks.

Early on, it had been suggested that episodic memories benefit especially from slow-wave sleep (Barrett and Ekstrand, 1972; Plihal and Born, 1997; Plihal and Born, 1999), while improvements in procedural tasks and memory for emotional events are more dependent on REM sleep (Plihal and Born, 1997; Plihal and Born, 1999; Wagner, 2001). However, the picture is not quite so clear anymore (Diekelmann and Born, 2010), with some studies reporting improved procedural memory after slow-wave sleep (Gais et al., 2000; Huber et al., 2004; Aeschbach et al., 2008) and others improved declarative memory after REM sleep (Rauchs et al., 2004; Fogel et al., 2007). Also, procedural memory often has a declarative component. Sleep has even been shown to transform non-declarative memory in a serial reaction time task into explicit knowledge, i.e. declarative memory (Fischer et al., 2006). Ironically, sleep-induced gains in motor speed could no longer be observed in participants who had gained explicit knowledge.

The lack of consistent results with regard to the effect of different sleep stages on different memory types may be due to differences in paradigms or experimental setups. Alternatively, it might be indicative of a more complex relationship. It is not unlikely that the different sleep stages have evolved and take place in their specific, multi-cycle fashion in order to provide optimal environments for different aspects of memory consolidation (Ficca and Salzarulo, 2004; Stickgold and Walker, 2005; Diekelmann and Born, 2010). This has also been called the “sequential hypothesis” (Giuditta et al., 1995).

In a recent study, Rolls and colleagues used optogenetics in mice to target hypocretin/orexin neurons which play an important role in arousal (Rolls et al., 2011). With this method they fragmented sleep into shorter intervals than they occur in normal sleep, while at the same time avoiding the usually necessary waking by touch or through aversive events (which might cause stress and in itself impair memory). In addition, the fragmented sleep had the same total duration and composition as undisturbed sleep, unlike in many sleep deprivation studies. The authors found significant memory impairment if the duration of the sleep segments was reduced by more than 62% of normal segment length. This supports the notion that sleep as a whole, without breaks or distortions, benefits memory best.

Individual sleep characteristics and their relationship to consolidation have also been investigated. Sleep spindles are transient oscillatory patterns of 10-16 Hertz which can be observed in EEG during sleep stage 2, of which they are also a defining characteristic. In a less discrete form, they also appear in sleep stages 3 and 4 (Gennaro and Ferrara, 2003). However, spindles are a unique EEG characteristic only observed during sleep. The amount of sleep spindles was found to be associated with lexical integration of novel spoken words (Tamminen et al., 2010). Another study found longer duration sleep stage 2 and increased spindle density following massed motor learning (Fogel and Smith, 2006). After a face-scene learning task, category-related neuronal activity was higher during spindle events and was modulated by the amplitude of the spindle events (Bergmann et al., 2012). In addition, across subjects spindle-coupled hippocampal activity was stronger when memory performance in the preceding task had been better.

Neurochemically, slow-wave sleep is characterized by minimal cholinergic activity. Artificially increasing cholinergic tone during slow-wave sleep by administering physostigmine blocked declarative memory consolidation (Gais and Born, 2004). Reducing cholinergic tone during wakefulness lead to improved consolidation of previously learned material but impaired acquisition of subsequent similar material (Rasch et al., 2006). It has been suggested that cholinergic tone acts as a “switch” in brain modes between encoding and consolidation (Hasselmo, 1999; Diekelmann and Born, 2010). Cortisol is also low during slow-wave sleep. Again, an experimentally induced increase during slow-wave sleep inhibits memory consolidation (Kloet et al., 1998; Wagner and Born, 2008). Interestingly, cortisol infusion impaired retention of temporal order information when administered during a nap, but increased retention when administered during a waking period (Wilhelm et al., 2011). These results relating to neurotransmitter levels again support the notion that there are specific mechanisms and characteristics in sleep that promote memory consolidation, rather than simply permitting it.

In summary, there is currently little doubt that sleep is, if not crucial, then at least beneficial for memory consolidation. Sleep’s influence on memory can be either seen in diminished forgetting (stabilization) or even improved performance (enhancement). The exact relationship between sleep stages and different types of memory is not clear at the moment. Improved consolidation probably relies on the overall integrity of sleep. Also, there is mounting evidence that sleep not only promotes memory success because it passively provides protection from interference, but that in addition to that, unique mechanisms during sleep actively enhance consolidation. One of them might be replay, which is discussed in the next section.

6 Neuronal replay as a mechanism for consolidation

6.1 Evidence from rodent studies

In the last 20 years, supporting evidence for both two-step models of memory formation and the notion that sleep promotes memory consolidation has been found in spatial memory studies in rodents.

In the rodent hippocampus, some cells reliably increase their firing rate whenever the animal is at a specific location in an environment (O'Keefe and Dostrovsky, 1971). These cells have been called “place cells”. They are especially interesting because they provide a relatively simple neuronal code for behavior.

If a rat runs along a track or a maze repeatedly, the same sequence of increased firing rate across the cells will be observed during each run. Interestingly, the same sequence has been found to spontaneously reoccur more often than would be expected by chance in task-subsequent sleep (Skaggs and McNaughton, 1996; Louie and Wilson, 2001; Lee and Wilson, 2002) and also quiet resting state (Foster and Wilson, 2006; Karlsson and Frank, 2009; Carr et al., 2011; Jadhav et al., 2012). As a mechanism, this kind of replay would correspond well to the reactivation that has been proposed to be necessary for consolidation in two-step models. This notion is supported by the finding that replay of place cells is behaviorally relevant (Dupret et al., 2010).

6.2 Studies in humans

In humans, a simple neural code such as the firing of place cells is not available. Still, inspired by the results in rodents, several studies have found evidence that reactivation can be observed in humans as well. In a series of elegantly designed experiments, Rasch and colleagues (Rasch et al., 2007) presented their participants with a declarative object-place association task as well as a procedural motor skill task. During both tasks, subjects were exposed to either rose odor or to no odor. Presentation of the rose odor during slow-wave sleep, but not during either waking state or REM sleep lead to improvement in the declarative memory task, but not in the procedural task. Also, rose odor presentation did not lead to improvement in those participants that had not experienced it during learning, precluding a simple odor-related memory enhancement effect. In fact, some of the participants were scanned with fMRI after being exposed to the rose odor during learning and tried to sleep inside the fMRI scanner. During presentation of the odor cues, hippocampus activity was greater than during phases without presentation of odor cues, but only if rose odor was presented during slow-wave sleep and not if it was presented during waking state. This is a strong indicator that activation of memory related neuronal structures can be triggered by an associated cue.

A similar design was employed in another study, in which individual object-place associations that participants had to learn were presented together with an auditory cue (Rudoy et al., 2009). For half of these object-place associations, the related auditory cue was subliminally played back to participants during a nap (masked by white-noise to avoid arousal or detection). The pairs for which the related sound had been played were remembered better at recall. This neatly shows a specific benefit for cued items rather than a general improvement.

Some studies in humans also find evidence for reactivation-like neuronal activity during waking state. In a recent fMRI study (Tambini et al., 2010), increased hippocampal-neocortical correlation was found in resting state after a memory task as compared to a resting state prior to the task and the increase in correlation was associated with better memory performance across participants. Also, persistent task-specific brain activation was found in resting state after either a declarative or procedural memory task (Peigneux et al., 2006).

Thus, there is first evidence that the neuronal correlate of reactivation as a mechanism for memory consolidation can be detected in humans. However, the question remains as to just how specific the neuronal correlates that have been found so far really are. In rodents, specific neuronal firing sequences that correspond to specific experiences have been found to be replayed. In humans, it has only been shown that learning related brain areas are reactivated.

The goal of this thesis, as outlined below, is to find a way to demonstrate *stimulus-specific* replay in humans. For this, one first has to find a way to identify the neuronal signatures of individual stimuli. One possibility for this is the application of multi-variate pattern analysis (MVPA), which will be discussed below. First, a short description of neuroscientific methods will be given.

7 Neuroimaging methods

7.1 Functional magnetic resonance imaging

7.1.1 Background

In the last 30 years, magnetic resonance imaging (MRI) has been applied in neuroscience for visualizing brain structures as well as brain function. The basis of this signal is briefly recounted here. The description is based on a standard textbook on functional magnetic resonance imaging (Huettel et al., 2008).

In short, a strong magnetic field in the center of an MRI scanner aligns the spin axes of hydrogen atomic nuclei in the human body so that a net magnetization can be measured. In a process called excitation, a radio frequency pulse, which is adjusted to the resonance frequency (Larmor frequency) of the hydrogen nuclei and magnetic field strength, is then used to flip the net magnetization 90 (i.e. to the transverse plane). During relaxation, the spins return back to their previous state and thereby emit a signal that can be detected with receiver coils.

Using various methods such as application of temporary magnetic gradients and Fourier analysis, the recorded signal can be decomposed to reflect signal strength at different locations inside the receiver coil. Typically, MRI images of the brain are scanned as multiple 2D slices which are reconstructed to form a 3D image of the brain. A 3D image of the brain, especially in functional imaging (see below), is referred to as a volume. One unit in such a 3D image is called voxel.

The strength and time-course of the emitted signal depends on the type of tissue. Thus, bones emit a different signal than blood, spinal fluid or lung tissue and cancerous tissue emits a different signal than healthy tissue. Settings in the sequence of the radio pulses can be used to maximize the difference between tissue types and this property has long been used for medical purposes, e.g. in detecting cancerous tissue, subdural hematoma or bone fractures.

7.1.2 The BOLD response

At first, MRI does not seem helpful for neuroscientists who want to image brain activation rather than brain structure. Luckily, oxygenated and desoxygenated blood emit different MR signals. The strength of blood flow and blood oxygenation level, in turn, are related to brain activity. When neurons increase their firing rate because they are part of circuits that are involved in a task, they have an increased need for glucose. Blood flow into this region is subsequently increased. As oxygen is not needed at the same rate as glucose, this leads to a temporary relative increase in oxygenated hemoglobin in the vicinity of active neuron populations. This so-called hemodynamic response is very slow: The peak is reached 4-6 *seconds* after the initial neuronal reaction. This has to be considered both in design of studies and in analysis of the data.

Fluctuations in the level of desoxygenated blood are called the blood oxygenation level dependent (BOLD) signal, which is measured with functional magnetic resonance imaging (fMRI). In fMRI, 3D images of the brain are usually scanned rapidly (e.g. every 3 seconds) for a certain interval of time (e.g. 20 minutes). This results in a time-series for each voxel (400 datapoints over 20 minutes in the example). This time series can then be related to psychological states that were induced during that time. For example, if the BOLD signal in voxel 1 increases every time a visual stimulus is presented but does not change when an acoustic stimulus is presented, it is likely that the voxel is involved in some form of visual processing.

7.1.3 Traditional analysis methods

Traditionally, time-series of individual voxels have been investigated independently from one another, i.e. one voxel at a time. Usually, a general linear model is set up that includes the different factors (or conditions) that were present in an experiment. Then, it is estimated for every factor, how much it contributes to explaining the signal observed in the single voxel. The degree to which a certain factor is associated with BOLD changes in a voxel is captured by a beta estimate.

Often, an experimental factor is compared to a control factor by calculating the difference between the two beta estimates. If the difference between experimental and control factor is consistent across participants for a given voxel, then the voxel is considered to be involved in the experimental condition. This is done for every voxel. Often, clusters of voxels are found which show the same response and regions in which such a cluster is located are then thought to be involved in the neuronal processing of the experimental condition.

This procedure has been called a “mass-univariate approach” (Bonnici et al., 2012) because it performs uni-variate statistical analyses independently on a large number of voxels.

7.1.4 Multi-variate approaches

Recently, the pattern of BOLD signal across voxels has been taken into account by methods such as representational similarity analyses (Kriegeskorte et al., 2008) or pattern classification (Norman et al., 2006). The idea with these new approaches is that there might be differences between two conditions even though they are not apparent (or statistically significant enough) in any one voxel. This idea is further illustrated in the section about multi-variate pattern analysis, specifically in Figure 8.2.

7.2 Electrophysiological methods

7.2.1 Electroencephalography (EEG)

Electrophysiological methods record changes in electrical potential which are generated by synchronized activity across populations of thousands of neurons. In contrast to fMRI, which records a substitute marker for neuronal activity, EEG records the potentials induced by neuronal activity directly and at a high temporal resolution (e.g. at 5000 Hertz) with electrodes placed on the scalp.

The most common approach to analysis of EEG signals is to present different types of trials repeatedly, “cut-out” the time-series signal around the onset of each trial and then average across trials. Assuming that the underlying neuronal activity is the same across repeated trials, the random parts of the time-series signal (“noise”) cancel each other out with increasing number of trials, and the average represents the “real” part of the signal (Luck et al., 2000).

Using this method, various typical event-related potential (ERP) components have been identified which can reliably be observed in certain types of tasks. Also, comparing the average across all trials in Condition1 with the average across all trials in Condition2 permits conclusions as to whether they are significantly different at certain time-points. This is another example for a univariate approach: various time-points are compared separately between Condition1 and Condition2, even though it is usually desired that contiguous parts of the signal exhibit a significant difference.

One of the major drawbacks of EEG is that the electrical potential is generated in the brain and is recorded with electrodes at the scalp. Depending on the distance between the source of the signal and the recording site, considerable distortions may occur making it difficult to localize the origin of any signal detected at the scalp.

Elaborate algorithms are employed for source localization, i.e. for determining where the EEG signal originates from (Pascual-Marqui, 1999; Michel and Murray, 2012). These algorithms work better with increasing number of electrodes (Michel et al., 2004). Modern high density EEG systems record from up to 256 electrodes narrowly spread across the head surface. The origin of the electrical potentials can then be narrowed down to a matter of centimeters (Lantz et al., 2003). However, the spatial resolution is not as high or as reliable as in fMRI recordings.

7.2.2 Intracranial EEG

In intracranial EEG, electrophysiological activity is recorded either from the cortex surface with strips and grids or from within the brain using depth electrodes. Position of the electrodes can be confirmed with MRI recordings and localized with a precision of millimeters. Thus, intracranial EEG combines excellent temporal resolution with good spatial resolution and is thus a valuable recording method for scientists, especially for structures deep within the brain such as the medial temporal lobe.

However, due to the invasive nature of this method, it is obviously only employed for medical purposes. Most often, intracranial EEG recordings are performed in patients with severe pharmaco-resistant epilepsy who might undergo surgical treatment. In these patients, intracranial EEG is used to localize or narrow down the brain region which causes the epilepsy, the epileptic focus. Surgical removal of the epileptogenic tissue is a drastic step, but often reduces the frequency of seizures or leads to complete remission in patients who had not responded to anti-epileptic medication (Kohrman, 2007).

In some patients, the clinical manifestations of their seizures or structural MRI scans are sufficient to reliably determine the epileptic focus; these patients may undergo surgery without prior recording from intracranial electrodes. However, if the epileptic focus is not entirely clear, the implantation of intracranial electrodes is performed as a means to localize the focus and to avoid unnecessary resection of tissue.

Electrodes are implanted according to the specific diagnostic question for every patient. Depth electrodes, electrode strips and grids are employed. After surgery, patients are taken off their usual medication and seizures are recorded until a diagnosis can be made. After clinical diagnostics are concluded, patients often stay at the ward for a couple of days until electrodes are explanted. During this time, the patients may be asked whether they would be willing to participate in scientific studies. Obviously, strict ethical guidelines apply.

Of course, there are some theoretical and methodological issues with intracranial EEG. First, the data is obviously recorded in patients who suffer from a severe form of epilepsy. This might influence the recorded results in any number of ways. Second, the patients have per definition a history of severe epilepsy and many have developed cognitive deficits as a result. Experimental paradigms have to be adjusted so they are not too difficult or too exhausting. Third, some of the recorded data may contain epileptic activity. Most of this activity can be avoided if all electrodes are excluded which are located in or near an epileptic focus. Careful artifact correction in the remaining data usually leads to good data quality.

7.2.3 Simultaneous EEG and fMRI

From a theoretical point of view, combining EEG recordings with functional MRI scanning is an excellent idea: EEG has a high temporal resolution and low spatial resolution while fMRI has a high spatial resolution and a low temporal resolution. Thus, each method could compensate the deficit of the other.

Naturally, from a practical point of view, combining the two methods is a very bad idea: EEG recordings rely on metal electrodes, cables for transmission and amplifiers which contain metal and need electrical power – none of these elements should normally be introduced into the strong magnetic field of an MR scanner.

In the last two decades, however, EEG systems have become commercially available which have been specifically designed for use in an MR scanner. Electrodes are made of material with minimal magnetic properties and the amplifiers are equipped with batteries and shielded in a way that they are not influenced by the magnetic field.

Even with these new systems, the EEG system and electrodes will cause artifacts in MRI images and MRI scanning will cause massive artifacts on EEG recording (Ritter and Villringer, 2006). Artifacts in MRI scans caused by the electrodes can be neglected as the resulting extinction often only affects small areas at the scalp.

Two major types of EEG artifacts are caused by fMRI:

1. Gradient artifacts are caused by the switching of the magnetic gradients which is necessary for scanning. As this is a technical, highly consistent artifact and its onset can be reliably recorded by the EEG via triggers, it is easily filtered out by subtracting a "template" artifact, which is determined by averaging across all scanner artifacts.

2. Cardio-ballistic artifacts are caused by the heartbeat and are seen strongest around the R-peak of the electrocardiography (ECG) signal. On a much smaller scale, this artifact may also be seen in regular scalp EEG, but it rarely poses a problem. Within the magnetic field of an fMRI scanner, the artifact is much more pronounced. It is currently not entirely clear whether this is caused by the rapid flow of blood (which has weak magnetic properties and may thus induce electrical field changes in nearby conductors) or due to small movements of the electrodes associated with blood pumping through the vessels. In any case, it is a biological and thus variable artifact. Therefore, it is much harder to filter out than the gradient artifact. Usually, a basic electrocardiogram is recorded together with EEG. The R-peak of the ECG can then be detected manually or with automated algorithms. Based on these R-peaks, a template artifact can again be computed and subtracted around each R-peak.

Additional artifacts may be caused by the helium pump which circulates the helium around the scanner's electromagnetic coil to keep it cool and superconducting. This artifact is also highly variable and almost impossible to filter out. Often, the pump is simply switched off during simultaneous EEG/fMRI, which is safe to do for up to 1.5 hours.

Even with all the sophisticated artifact correction tools that have been made available by commercial software (e.g. Brain Vision Analyzer 2.0, Brain Products, Munich, Germany), it should be kept in mind that the EEG signal will never be as good as if it had been recorded outside of the scanner in a shielded room. Using simultaneous EEG/fMRI should thus always be motivated by an important question that can only be answered by using the two methods together. One such case is scanning participants when they are sleeping. In this case, EEG is the only way to determine whether participants are asleep and which sleep-stage they are in.

7.2.4 Time-frequency analysis

As outlined above, the classical approach to EEG analysis has been to average EEG amplitude across trials and look at the resulting ERP components. However, taking into account the power of different frequency bands at specific points in time or the interactions between oscillations may provide a more complete picture of neuronal activity (Makeig et al., 2004).

A now widely popular method of analyzing electrophysiological signals is to break down the signal into time and/or frequency bands, using methods such as fast fourier transform, wavelet decomposition or Hilbert transform (Wacker and Witte, 2013).

It has been shown that increases or decreases in the power of specific frequency bands are related to cognitive functions (Klimesch, 1999; Doppelmayr et al., 2002; Jensen et al., 2012). Also, interactions between different frequency bands are increasingly considered (Jensen and Colgin, 2007; Canolty and Knight, 2010). The pattern of time-frequency components has also been used to decode different brain states (van Gerven et al., 2013).

8 Multi-variate pattern analysis

8.1 Development of multi-variate approaches

For a long time, analysis of fMRI data was restricted to a so-called mass-univariate approach (Bonnici et al., 2012), which looks at activity in individual voxels separately without considering the pattern of responses across voxels. The same is true for electrophysiological data, in which in most cases simple differences in event-related potentials (ERPs) were compared between conditions. With methods like these, differences between conditions have to be massive in the single units (e.g. voxels, time-points) that are looked at to survive correction for multiple comparisons. Small differences will be discarded, even if, across multiple units, they offer enough information to distinguish between conditions.

This unsatisfactory use of the rich neuroimaging data changed with the turn of the century, when the first groups started to look at information content across voxels rather than at mere voxel-wise differences (Haxby et al., 2001; Cox and Savoy, 2003). It quickly became apparent that patterns of voxel activity contained more information about mental states than activity differences in single voxels.

Based on these pattern differences, it also became possible to predict which condition was present during a given fMRI scan, an approach for which the term “decoding” has been coined. In the last ten years, an astounding variety of subtle mental states has been decoded with pattern classification approaches, such as the orientation of subliminally presented lines (Haynes and Rees, 2005a), the alternating conscious perception in a binocular rivalry task (Haynes and Rees, 2005b), hidden intentions (Haynes et al., 2007) and which category a participant was thinking of during memory search (Polyn et al., 2005).

Of course, these approaches are far from perfect or infallible in decoding psychological experience from brain states, but they manage accuracy far above chance level and, given enough training material, they can be applied to differentiate quite fine-grained percepts or thoughts. As such, this method is well suited to detect the neuronal signature of individual items that have to be learned in a declarative memory task and might thus also be capable of detecting a reoccurrence, or replay, of these signatures during task-subsequent resting state or sleep.

8.2 What is Pattern Classification?

Imagine you have stranded on an unknown island and you encounter an unfamiliar kind of fruit. Actually, there are two different kinds of fruit, but it is very difficult to tell them apart. Differentiating between them soon becomes vitally important as you realize that the fruits are the only means of sustenance on the island and one kind of them makes you sick. You begin to notice that, on average, the poisonous kind of fruit is a bit larger, slightly less red, the peel is rougher and the smell is sweeter. For each of these features, however, the overlap between the two kinds of fruit is large and judging an exemplar by only one of these properties (univariate approach) has a high likelihood of false classification. The solution obviously is to take into account as many predictive features as possible (multivariate approach) to maximize the probability of correct classification.

Humans implicitly use this method of differentiating every day. We are ourselves very good pattern classifiers. Based on the information we have in a given situation and experience gathered throughout our lives, we constantly make classifications: Is someone trustworthy? Is someone writhing in pain or shaking with laughter? Is this a male or a female? Of course, humans do not use exact solutions to multivariate problems, but base their decisions on “intuitive”, or heuristic judgment. However, human decision making is often flawed by systematic biases and distortions (Tversky and Kahneman, 1974).

Computers, however, are not subject to such distortions. When all available information is presented to a computer in an appropriate format, it can take into account all features simultaneously and return an exact or optimized solution. With the arrival of powerful computers, solutions to high-dimensional classification problems have become easily attainable. Applications range from face recognition and text-to-speech software to weather prediction and autopilots in cars.

In neuroscience, pattern classification approaches have gained influence in the last decade. They have been employed primarily in functional magnetic resonance imaging (Norman et al., 2006; Haynes and Rees, 2006), supplementing classical analysis methods that focus on activity in each individual voxel separately such as general linear models. In fact, with pattern classification it was possible to identify different brain states for stimuli so similar that a classical general linear model (GLM) approach would not have detected the differences (Haynes and Rees, 2005a).

The parsimonious idea behind the trend towards pattern classification is that if cognition is supported by patterns of brain activation, analysis methods should also take patterns into account.

In the following, an introduction into pattern classification is given from the point of view of a psychologist, focusing on general ideas and practical application.

8.3 Terminology

With a method as complex as pattern classification, it is important to be precise with the words that are being used, therefore the most important expressions will be introduced here briefly.

A **classifier** generally is any algorithm that is employed to differentiate between two or more distinct classes. The classes are characterized by **features**. Features are properties that can be used to describe classes. They can be everyday properties like “height, weight, color” or more abstract like “activation in voxel 1, voxel 2, voxel 3”. Features can be thought of as dimensions along which exemplars of different classes can be qualified or quantified. Refer to Figure 8.1 for illustration.

A classifier has to “learn” before being able to perform classifications. Learning can be supervised or unsupervised. Unsupervised classifiers are not employed in this thesis and will thus not be further discussed. Supervised learning is often called **training**. During training, the classification algorithm is fed with already classified (or labeled) data and, based on this training data, it will arrive at a classification boundary – a set of rules which the classifier will use when differentiating between the classes.

The training data are made up of **samples** and **labels**. Usually, samples are observations of real instances of the classes. Let us consider our fruit example from the introduction. Suppose you have had the opportunity to observe the consequences of eating 20 exemplars of fruit, so you know whether they were fruit A or fruit B. For each of the exemplars, you have made note of 5 physical properties or features. The dataset will then consist of 20 samples, and each of the samples will consist of a vector of length 5 providing information on the five features. During training, each sample is accompanied by a label which tells the classifier which class the given sample belongs to.

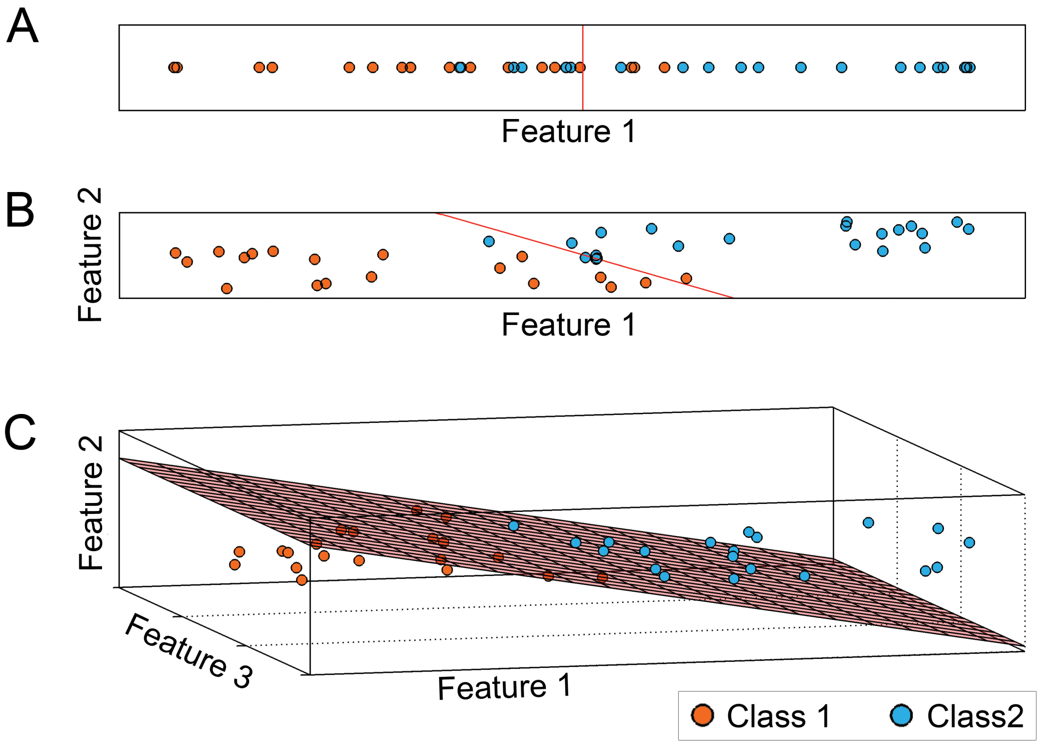

Figure 8.1: Classification problem with different numbers of features. Every dot represents a sample that was observed for one of two classes (orange and blue). The position of the dot indicates the parameter value with regard to a quantitative feature. **A**: With only one feature, the distinction is made by setting a vertical boundary somewhere between the centers of the two distributions. **B**: With two features, a linear function can be used to draw a distinction in a two dimensional space. **C**: When three features are considered, the different observations can be visualized in a three-dimensional space and a two-dimensional plane can be used to separate the classes. Any classification problem with more than three features is difficult to visualize or even imagine; however, the basic principles are the same.

.

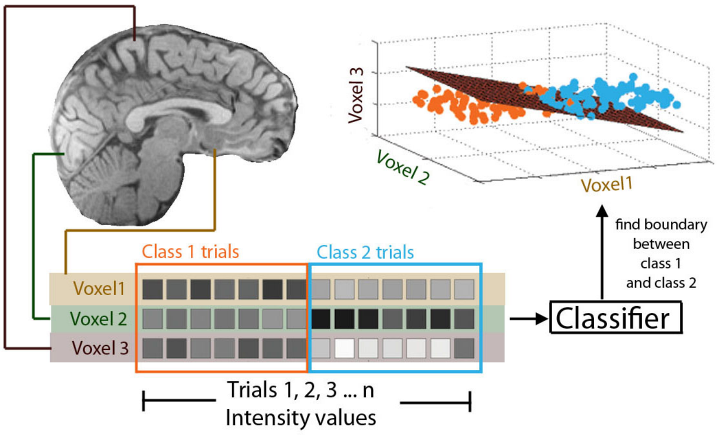

Figure 8.2: Basic principles of pattern classification in fMRI data. The raw signal for selected voxels is extracted from the fMRI data as a time-series: The gray values correspond to blood level oxygenation dependent signal intensity during different trials in which stimuli belonging to either class 1 or class 2 were presented. Voxels are useful as features when they show consistent difference in activation between the two classes. For example, Voxel 1 displays on average less activity during trials of class 1 than class 2. The gray values form selected voxels are used to train the classifier, which tries to find a good rule for differentiating between the samples of the two classes in a high-dimensional space.

.

It should be noted that with sufficiently many features the classifier can almost always find a perfect solution for differentiating the training dataset. What is more interesting, of course, is whether the classifier generalizes, that is, whether the algorithm can make correct classification on data it has not been trained on. This is done during **cross-validation**. The classifier makes **predictions** on new samples and these predictions are then compared to the actual label of the sample (**target**). Calculating the percentage of matches between predictions and targets yields a measure of **classifier accuracy**, or classifier performance, i.e. how well the classifier recognizes classes in unknown data.

How do these terms relate to the practical application of pattern classification to neuroimaging data? In the classification of fMRI data, samples usually consist of 3D brain scans (volumes) that were scanned at a particular point in time, for example when a specific experimental manipulation took place. Mostly, voxels are then taken as features. Voxels can be selected from the whole brain, or taken from regionally restricted areas, depending on the hypothesis. Refer to Figure 8.2 for an example.

In electrophysiological data, the samples typically are comprised of time series data, again during a time window when a given psychological event took place. Features can be raw amplitude values, power of specific frequency bands, wavelet coefficients or virtually any other characteristic than can be identified in time series data.

8.4 Basic steps of pattern classification

The basic steps in pattern classification analyses as they are currently employed in neuroscientific research are similar across experiments and imaging modalities. The steps listed below have been adapted from a review on the application of MVPA on fMRI data (Norman et al., 2006) in a way that they also encompass electrophysiological data.

1. **Feature selection:** Not every feature that can be or has been measured is contributing to good classification results. On the contrary, including features which have no discriminative value for the classification task might even be detrimental to the classifier performance. This is especially true for datasets with a very high number of potential features, e.g. fMRI data with 30.000 to 50.000 voxels that can serve as features.

   Thus, only features with good discriminability should be selected. There are numerous methods for determining good features. One of the simplest is to perform an ANOVA for each feature. In such an ANOVA, the different classes would be the group factors and the different samples would be observations of the independent variable. By doing this, features can be identified that display high variance between and little variance within classes. This is the method mainly used in this thesis.

   If the classifier's ability to generalize to new data is to be assessed with cross-validation, the feature selection should be performed only on the respective training dataset. Performing feature selection on the complete dataset (including the later test dataset) leads to artificially high classification performance. Generally, this is not seen as a correct use of the method.

2. **Pattern assembly:** Based on the extracted features, a dataset of "brain patterns" is assembled which relate to specific psychological states during a

recording session. In fMRI data, activity levels from the selected voxels (features) at the timepoints of interest (e.g. whenever a picture has been presented) are extracted and organized into samples. Every sample is given a label for its respective class. In EEG data, the same is done with amplitude levels or frequency power values.

3. **Classifier training:** Training of the algorithm is, as mentioned above, necessary in supervised learning. The algorithm is confronted with a training dataset and an accompanying list of labels for each sample. The training or "learning" process per se is different for each classifier and shall be briefly discussed below for the two algorithms that are used in this thesis. Importantly, at the end of classifier training, the algorithm arrives at a set of rules that are then used to make predictions on unknown data. The quality of these rules should then be assessed in the next step, cross-validation. In general, classifier training will be more successful with increasing number of samples that are available for training, because the probability is higher that the classifier arrives at more generalized classification boundaries.

4. **Generalization testing:** To assess the classifier's ability to generalize, some form of cross-validation is typically done. As mentioned above, classifiers very often find a perfect solution for separating the training dataset. That does not necessarily imply that they will perform well on new data that have not been included in the training process. In certain cases, the rules for separating the training dataset are too strict, or too specialized – this is called "over-fitting". Therefore, the complete datatset is usually split; one part is then used during training (training dataset), the other part is used as a validation run for the classifier (test dataset). When the classifier makes more correct predictions than would be expected by chance on this test dataset, it is an indication that the classifier is not too specialized and generalizes to new data.

   The methods for splitting up the dataset into training and test data are manifold. A popular method is the "odd/even method", in which odd trials are taken as training data and even trials as test data (and vice versa). Note that this reduces the number of training samples by half, which might by itself affect classifier performance. Another popular method is the "leave-one-out cross-validation", in which the dataset is split into any number of different blocks, or chunks. All of the chunks except one are then used for training the classifier and the left-out chunk serves as the test dataset. This is repeated with leaving every chunk out as test dataset once and thus this method provides a good estimate of classifier performance on the complete dataset.

8.5 Pattern classification algorithms

8.5.1 Linear Support Vector Machines

The linear support vector machine (SVM) is probably the most widely used algorithm in neuroscience today. The following very basic explanation is based on two standard textbooks (Duda et al., 2001; Bishop, 2009).

In the SVM framework, n features span up an n-dimensional space. Samples from two classes can then be represented as points in this space based on the quantity of their features (as is demonstrated for 3D-space in Figure 8.2). During classifier training, an $n-1$-dimensional hyperplane is drawn in this feature space which separates the points/samples of the two classes. New samples are then mapped into this feature space and classified as either of two classes depending which side of the boundary they are on.

Many different hyperplanes may achieve a separation between the two classes in a training dataset. However, the linear SVM has the constraint that the hyperplane must have a wide margin, that is distance, between the nearest points/samples of the two classes and the decision hyperplane. Effectively, this constraint maximizes the distance between the hyperplane and those samples from the two classes which are most difficult to separate (because they are closest to the boundary). These "difficult" samples are called support vectors. The wide margin between support vectors and decision hyperplane is enforced based on the assumption that the resulting decision boundary will be more general, i.e. that new samples can be classified more accurately.

Sometimes, the algorithm will even permit misclassification of some samples during training if the margin to the other support-vectors becomes greater as a result. The degree to which the classifier accepts misclassifications in favor of a wider margin can be controlled with parameter C, which can be manipulated in most available implementations of the algorithm.

As noted above, the linear SVM is essentially a two-class classifier. It can be applied to a multi-class scenario by training binary classifiers on each pair of classes separately. Every binary classifier then makes predictions. The class that "wins" out most often in these binary predictions is the final label the classifier returns as prediction. This feature is implemented in the available libraries (such as libSVM) so seamlessly that the classifier appears to be a multi-class classifier. However, one should keep in mind the underlying binary nature of the linear SVM.

8.5.2 Sparse multi-nomial logistic regression

This classifier is a true multi-class classifier that has only recently been developed (Krishnapuram et al., 2005). It performs, as the name implies, a regression between the features (predictors) and the discrete class-label (target).

The important development with this algorithm is that it automatically finds a sparse solution, that is, regression weights are either very small or very large, which is important for large number of features.

The algorithm has been tested on a number of well known, freely available classification datasets (Krishnapuram et al., 2005) and performs favorably in comparison with other established algorithms in terms of classification accuracy.

9 Summary and goal of this thesis

The idea of consolidation, or stabilization over time, is an integral part of many two-step models of memory formation, which posit that memory traces are initially stored mainly by the hippocampus (the "fast learner") in a labile state and that by repeated coordinated information transfer between hippocampus and neocortex (the "slow learner") become represented by increasingly strong neocortical connections until the hippocampus is no longer necessary for retrieval of these memory traces (Frankland and Bontempi, 2005).

A vast amount of research has been conducted in the last two decades to show that sleep enhances memory consolidation, arguably by providing an interference-free window of time for an active information transfer between hippocampus and neocortex. The same may be true for quiet resting state.

In rodents, a likely neural correlate of reactivation has been identified both during sleep and quiet resting state in the form of a coordinated replay of the same place-cell sequences which had been also observed during prior learning (Louie and Wilson, 2001; Foster and Wilson, 2006). Can a similar correlate of reactivation also be identified in humans?

Although some studies in humans have found first evidence for reactivation (Rasch et al., 2007; Rudoy et al., 2009; Tambini et al., 2010), none has done so with the same specificity as has been demonstrated in the rodent studies.

The goal of this thesis was to investigate *stimulus-specific* reactivation in humans. But how can this be done? Place-cell recording as it can be done in rodents is not possible in humans due to obvious ethical reasons. A more indirect route has to be taken.

Neuroscientific methods, as introduced above, can be used to record event-related neuronal activity in humans. Multi-variate pattern analysis can then be applied to these recordings to reliably decode a "neuronal signature" for individual stimuli that were encountered by participants during recording. A classifier trained on these specific stimuli may then be able to track their neuronal signatures during phases of resting state and sleep.

In this thesis, three empirical studies will be presented that attempted exactly this approach for detecting replay. Despite focusing on slightly different aspects, all three studies have the following in common: First, they presented participants with a declarative, associative memory task in which individual objects were shown repeatedly. Second, the neural signatures related to encoding of these individual object-place pairs were extracted with pattern classification algorithms. And third, the neural signatures were tracked during resting state or sleep recordings that followed the learning task.

Each of the three studies investigated a variation of the following assumptions:

1. The neuronal signature of individual items presented during a learning task can be reliably decoded from the recorded data with multi-variate pattern analysis.

2. A pattern classifier that has been trained on data from the learning task can be applied to periods of subsequent resting state and make predictions about them, thereby tracking possible reoccurrence of the original learning related activity patterns.

3. Compared to a baseline condition, there is significant reoccurrence of stimulus-specific neuronal activity.

4. The frequency of this reoccurrence of individual items is associated with subsequent memory performance for these items.

This is, to my knowledge, the first attempt to use multi-variate pattern analysis to directly detect replay events in resting state and sleep in humans.

Part II

Empirical Part

Three studies are presented here that investigate different aspects of reactivation during resting state and sleep. The first study employed simultaneous EEG/fMRI and tracked reactivation of regular object-place associations during quiet resting state and sleep. The second study investigated reactivation of emotionally negative as compared to neutral stimuli during resting state with fMRI. The third study was recorded using intracranial EEG in patients suffering from pharmaco-resistant epilepsy and allowed us to take a closer look at the temporal and frequency specific dynamics of reactivation during sleep.

All three studies presented here use pattern classification as the main method of analysis and are therefore specifically designed to accommodate the requirements of the method.

The most important restraint when using a pattern classification approach on neuroimaging data is that every class/stimulus one wants to decode should be presented multiple times to guarantee good classifier performance and generalization. Drawing from previous MVPA studies and our own extensive piloting, it was concluded that every stimulus should be presented between 20 and 30 times.

This precludes the use of simple recognition tasks in which stimuli are presented and subsequently probed with a forced choice "old/new" task, because either the task would be too easy or the experiment would last too long. If one presented 20 stimuli 20 times each, this would very likely result in performance at ceiling. As one of the goals of this thesis was to show that replay has an impact on memory performance, a memory task with a broader range of performance was desirable. If, on the other hand, one presented 100 stimuli for 20 times, this would lead to excessive task length, especially as the optimal trial length was determined to be two MRI volumes (five seconds).

For these reasons and because the hippocampus was an anatomical region of interest because of its role in two-step models of memory formation, an associative, hippocampus-dependent memory task was used, similar to tasks employed in previous studies (Rasch et al., 2007; Rudoy et al., 2009). In this task, a stimulus was always paired with a location on the screen. Over repeated presentation of the object-place pairs, participants were supposed to memorize which stimuli is associated with which location.

This memory task yields a continuous measure for memory performance: During testing, the stimulus is shown and participants mark the position they believe the stimulus was associated with. Memory performance can then be operationalized as the distance between the correct position and the position given by the participant. If this distance is small, memory performance for the tested object is high.

In addition, the task is more difficult than a forced-choice “old/new” task. Participants not only have to remember which object was presented, but also which location it was associated with. And even if participants perform very well on this task in general, looking at the error distance during memory recall allows one to identify nuanced performance for individual object-place pairs which would be lost in a task in which an item is either remembered or forgotten.

Thus, this task was well suited for our MVPA approach and, in slightly different variations, was used in all three studies.

10 Replay of stimulus-specific neuronal activity during resting state and sleep

10.1 Introduction

The goal of the first study was to investigate the model developed in the Theoretical Part of this thesis. In this model, memory consolidation is thought to depend on reactivation of the same neuronal activity patterns that were present during initial learning.

In this study, a declarative, associative memory task was performed by participants in an fMRI scanner while simultaneous EEG was recorded. After completing the task, participants tried to fall asleep inside the scanner for an afternoon nap. After this resting period, participants performed the same memory task again, but with different stimuli.

A pattern classification algorithm was trained on stimuli from the tasks preceding and following the nap and then made predictions on the resting period (see Figure 10.2). Predictions were expected to be more frequent for stimuli from the first memory task, for which replay was actually possible. The frequency of predictions of individual stimuli from the first memory task was further expected to be related to memory performance in a memory test that was completed after the second memory task.

The study was designed to be as simple as possible. Stimuli were normal and not particularly exciting. The task was a straight-forward object-place association task that has in a similar form been employed in other reactivation studies in humans (Rasch et al., 2007; Rudoy et al., 2009). In many ways, this first study served as reference and starting point for the other two studies.

Many months of piloting were invested for this first study and the results of this pilot phase, though not further described here for reasons of space, had great impact on the design of the paradigm. The second and third studies have designs very similar to this first study, precisely because it has been piloted so carefully and found to be efficient. Several of the design considerations and methodological details (mostly pertaining to multi-variate pattern analysis) will be provided for this first study in the methods section that will not be mentioned again in the other two

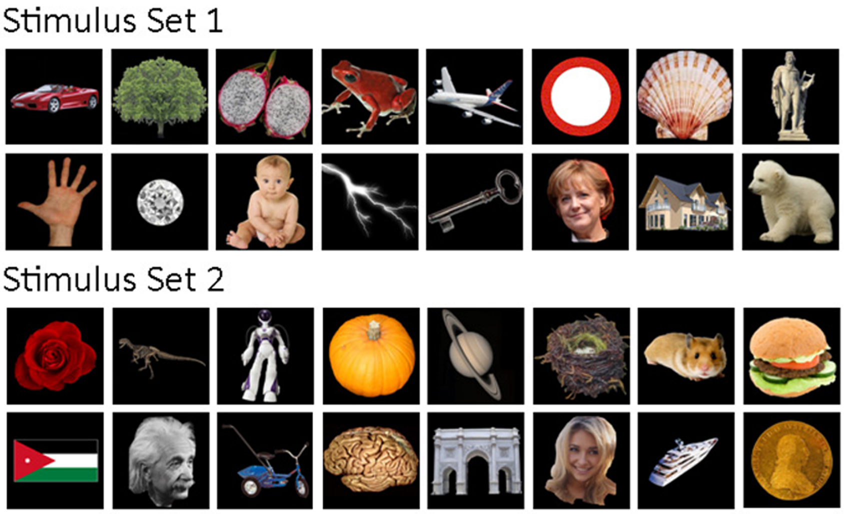

Figure 10.1: Overview of the stimuli.

.

studies. However, it should be assumed that these considerations are valid for the second and third study as well unless stated otherwise.

10.2 Methods

10.2.1 Participants

Seventeen healthy right-handed participants (10 female, age 24.1 ± 2.6 years) with no history of a neurological or psychiatric disease participated in this study. The study was approved by the local ethics committee, and all participants provided written informed consent. Participants were reimbursed for their time.

One participant aborted the experiment due to the need of a restroom break, one participant had to be excluded because of excessive movement inside the MR scanner and five participants were not analyzed further because of low general classifier performance (see below), resulting in a final dataset of 10 participants (6 female, age 23.7 ± 2.8 years).

10.2.2 Stimuli

Bitmap pictures of 32 real-life objects from the internet that were cut out and presented on a black background were used in this study. These 32 objects were grouped into 2 sets for use in the two different tasks. For every object, 6 different

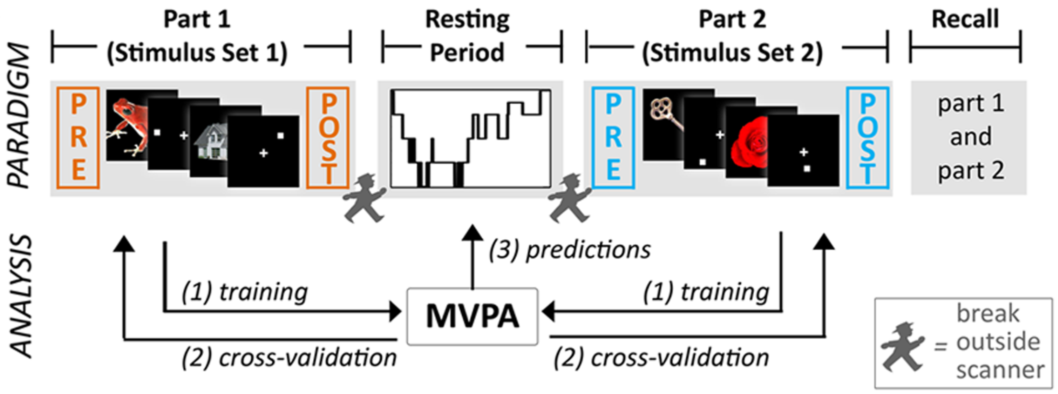

Figure 10.2: Overview of the paradigm: Subjects learned associations between 32 different stimuli (e.g., a red frog) and spatial locations that were indicated by a white square. Every object was presented 30 times followed by the corresponding location. Half of the object-location associations had to be learned in the first part of the experiment, the other half in the second part. During the main resting period between the two learning sessions, subjects slept inside an MR scanner with simultaneous EEG. Both memory tasks were flanked by 5 minutes of resting state scanning ("task-adjacent resting periods"). In a memory test subsequent to the second learning task, each of the 32 objects was presented again and subjects had to indicate the position of the associated white square.

.

exemplars were used, e.g. six different pictures of a red frog, six different pictures of German chancellor Angela Merkel and so on (an overview of all image categories is shown in Figure 10.1).

The use of different exemplars was intended to make sure that processing of the stimuli was not solely based on low-level visual features. Thus, the classifier was actually trained on a generalized version of each stimulus, which should facilitate the recognition of slightly altered activity patterns during the resting periods. In summary, two sets of 16 objects represented by 6 different exemplars each were used, resulting in $2x16x6 = 192$ pictures.

The 16 objects in each set cannot be readily grouped into obvious categories but were carefully selected to differ on dimensions such as large/small real life size, rare/common, living/nonliving, natural/man-made. The two sets were balanced with regard to luminance and spatial extent of the objects.

10.2.3 Design

In order to increase the likelihood of falling asleep, participants were instructed to sleep 2 hours less than usual in the night before the experiment and to refrain from drinking alcohol or going out late. They were also told to refrain from consuming caffeine, smoking cigarettes and taking any medication on the day of the experiment.

Participants arrived between 12.30pm and 1.30pm, and after familiarizing them with the surroundings and procedures, giving instructions and applying the EEG cap, the MRI scanning and actual experiment started between 2pm and 3pm. The total duration of the experiment was 7 to 8 hours but with several breaks in between.

A general overview of the experiment is given in Figure 10.2. Participants had to learn 16 object-place associations each during two separate sessions. Between these sessions, they attempted to take an afternoon nap inside the MR scanner ("main resting period"). Because of the long duration of the experiment (7 to 8 hours, including $\tilde{4}$ hours MRI scanning), subjects were allowed breaks outside the scanner immediately before and after the main resting period. The second, post-resting memory task was included because it served as a control condition for reactivation of Set1 stimuli. Also, it introduced interference with the stimulus-position associations learned in the first task and consolidation should predominantly stabilize memories against such interference (Mueller and Pilzecker, 1900; McGaugh, 2000; Ellenbogen et al., 2006a).

During each of the two associative memory tasks, 16 objects were paired with 16 locations on the screen as marked by a white square. Every object-place association was presented 30 times. Five minutes of scanning preceding and following each of the two tasks were included, resulting in four short resting periods ("task-adjacent resting periods").

Each trial consisted of presentation of the object for 1000ms, followed by presentation of the corresponding location for 1000ms and a fixation cross for 3000ms before the next trial started. The delay between presentation of the item and the associated spatial position was introduced because the learning paradigm was supposed to be hippocampus-dependent, and previous studies had shown that the hippocampus is particularly relevant for the formation of memory associations across a temporal distance (Staresina and Davachi, 2009). Each trial lasted exactly as long as the acquisition of two fMRI volumes and the next trial would only start with the beginning of a new volume. This was designed to always capture the same part of the cognitive task in an MRI volume. Using trial lengths at a constant multiple of the repetition time is non-optimal for general linear models, but consistent with previous pattern classification studies which are not based on a GLM (Kay et al., 2008; Harrison and Tong, 2009; Bode and Haynes, 2009). Each of the two experiments was divided into 5 blocks, separated by a one minute break. In each block, every object was presented 6 times by showing each of the 6 different exemplars once.

Within each block, stimuli were presented in randomized order. One stimulus set ("Set 1") was presented in the first memory task (before the nap), the other in the second task (after the nap, "Set 2"). The order of stimulus sets was counterbalanced across 17 participants, and in the 10 subjects who met the inclusion criteria, 6 participants saw stimulus set 1 first.

Participants were instructed to memorize the location of the white square for every object, and they were told that after finishing the second memory task, they would be shown every object again and would be required to indicate the position of the white square. They were not told that there would also be a free recall (naming every object they had seen). In addition, they were asked to give a subjective "like/dont like" evaluation of the object presented in every trial, captured by pressing a button with the left or right thumb while the image was presented. Out of the 10 participants who met the inclusion criteria, 7 pressed the right thumb to indicate a "like" decision and the left thumb for a "dislike" decision, in the remaining 3 participants the contingency was opposite. The "like/dont like" evaluation was asked of the participants to make sure they were attending the task and to induce a deeper level of processing.

After the first memory task, which lasted about 50 minutes, participants left the scanner for a 5 minute break period, then returned inside the scanner and attempted to fall asleep. Participants were told to take their time trying to fall asleep and to notify experimenters if either they felt they would not manage to fall asleep anymore, or if they had woken up and felt they would not fall asleep again. If they did not notify the experimenters, the main resting period ended after 120 minutes. A a variable duration of the main resting period was permitted, even though it was clear that time since encoding is a major factor for retrieval success and would influence performance in the final memory test. However, such an instruction was supposed to permit participants to feel more relaxed and in control during this period, making it easier for them to actually fall asleep. The main objective of this study was not to determine the general effects of sleep, sleep duration and time since encoding on memory performance, but to investigate neuronal correlates of spontaneous replay during rest and sleep. Whenever replay was related to memory performance, it was done intra-individually, thus preventing bias resulting from interindividual differences in sleep length and depth.

Importantly, the time between the first memory task and the main resting period was matched with the time between the main resting period and the second memory task, so that the temporal distance between the two tasks and the main resting period was always symmetrical. This is a necessary prerequisite for several of the following analyses.

After the main resting period, participants were again allowed to spend some time outside the scanner and then returned inside the scanner to perform the second memory task, which also lasted 50 minutes. After finishing the second task, participants left the scanner.

Outside the scanner, memory was tested first for stimuli from the first task and then for stimuli from the second task. Participants were first asked to name all objects they could remember from either task (free recall), then were shown one exemplar of each object and were asked to indicate with a mouse cursor the position of the corresponding white square (cued recall) in a similar way as was done before (Rudoy et al., 2009). While the free recall task resulted in a binary remembered/forgotten-measure of memory performance, the cued recall task allowed us to evaluate memory performance with a continuous metric, i.e. the closer the indicated position was to the actual position, the better recall was conceptualized to be.

The entire experimental paradigm was presented using Presentation software (http://www.neurobs.com). Images were transmitted inside the scanner via MR-compatible video-goggles (Nordic Neuro Lab, Bergen, Norway) with a resolution of $800x600dpi$.

10.2.4 Functional magnetic resonance imaging

MR scanning was performed with a 3 Tesla scanner (TRIO, Siemens, Erlangen, Germany) using echoplanar imaging. For each volume, 37 slices covering the whole brain were measured with a thickness of $2.5mm$ at $2500ms$ repetition time and $35ms$ echo time, a field of view of $210mm$ and a distance factor of 25%. In addition, a high-resolution structural T1-weighted image of the whole brain was collected for coregistration purposes with 160 slices with a thickness of $1mm$ at $1570ms$ repetition time and $3.42ms$ echo time, a field of view of $256mm$ and a distance factor of 50%.

Functional images were transformed from DICOM to NIfTI format using MRIcron (http://www.cabiatl.com/mricro/mricron/dcm2nii.html). Preprocessing was done with FSL (Smith et al., 2004; Woolrich et al., 2009), the steps including motion correction, $5mm$ Gaussian spatial smoothing and a linear detrending. Participants who exceeded a mean relative movement of $0.2mm$ as estimated by FSL were excluded from further analysis (one participant out of the original group of 17 participants). A z-transformation was then performed in order to have the same mean activity in each of the three scanning sessions. Images from the two memory task sessions were then spatially aligned to the images from the sleep session. Note that both task sessions were thus not in their original space but symmetrically mapped onto a third space.

10.2.5 Electroencephalographic recording and sleep staging

A 14-channel EEG was simultaneously recorded with fMRI for sleep staging during the resting period. An Easycap (EASYCAP, Herrsching, Germany) MR-compatible cap was used with 10 cortical electrodes, two of which also served to record eye muscle activity, 3 EMG electrodes at the chin and one ECG electrode at the back. This layout was according to the American Academy of Sleep Medicine (AASM) guidelines (Iber et al., 2007). All electrodes were sintered AG/AG-Cl electrodes suitable for use in a 3 Tesla scanner. The BrainProducts MR Plus amplifier (Brain Products, Munich, Germany) was also suitable for use in a 3 Tesla scanner. Data were sampled at $5000Hz$.

Offline processing of the data included scanner artifact removal, cardio-ballistic artifact removal, notch filtering at $50Hz$ and high-pass filtering at $0.01Hz$, using the available modules from Brain Vision Analyzer 2.0 (Brain Products, Munich, Germany). Data were then segmented into 20s epochs and scored for sleep stages according to Rechtschaffen and Kales (Rechtschaffen et al., 1968). Sleepstages 3 and 4 were combined for the rest of the analysis.

10.2.6 Multi-variate pattern classification

All pattern classification analysis of the fMRI data was carried out using the PyMVPA toolbox (Hanke et al., 2009a; Hanke et al., 2009b) for Python. For all classification tasks, linear support vector machines with a coefficient of C=0.1 were used (this is the default value). Classifiers were always trained within participants, never across participants. The third MRI volume after stimulus onset was used for training in order to account for the latency in the peak of the hemodynamic response. At a TR of 2500ms, this volume encompassed the time window of 5000 to 7500ms after stimulus onset.

Classification was not based on all fMRI voxels but on a subset of voxels (features) that were most discriminative: For each of the roughly 50000 voxels, a one-way ANOVA was conducted prior to classification with the 32 different objects as independent or group variable and the BOLD signal during the presentations as dependent variable (based on the respective training dataset only to avoid circularity). 1000 voxels with the highest F-values in these ANOVAs were then selected for classification, a number consistent with previous studies (Johnson et al., 2009; Ethofer et al., 2009). The F-value in this case represents a measure of general variability of a given voxel with regard to the 32 different objects. After a voxel was selected as a feature, the size of the F-value did not matter anymore. All voxels were treated the same by the classifier.

10.2.7 Classifier accuracy

To assess the classifiers ability to distinguish between the neural representations of individual objects, a cross-validation procedure was used. A linear support vector machine was trained on four of the five blocks from the paradigm (training dataset) and made predictions on the remaining block (testing dataset). This was done five times, so that every block served as testing dataset once. Comparing the classifiers output (prediction) for a given trial with the actually presented object (target) across all 960 trials in all testing datasets yields an estimate of classifier accuracy. It should be noted that there was a one-minute break between every block in this paradigm; the classifier was thus trained on data that was temporally separated from the testing data. Any confounds artificially increasing accuracy due to hemodynamic similarity of neighboring trials were thus avoided.

Excellent classifier accuracy was an important prerequisite in this study. Therefore, participants with insufficient classifier accuracy were excluded. In order to determine a suitable cut-off, a linear support vector machine was trained in the same way as was done with the real experimental data, except that the data were shuffled with regard to the contingency between samples and labels. In effect, classifiers were thus trained on nonsense data. Data were shuffled within the two experimental blocks only in order to preserve the overall structure of the data. The shuffling was done 50 times during each of the five cross-validation runs for each participant. The nonsense-trained classifier was then applied to the respective testing dataset and accuracy was determined as it was determined in the real data. Thus, 250 surrogate accuracy values were obtained for each of the 17 participants.

The resulting distribution of accuracy values was used to determine a cut-off value and all participants were excluded in whom classifier accuracy for either Experiment 1 or Experiment 2 objects was worse than the maximal value of the surrogate distribution plus three standard deviations.

10.2.8 Evaluating classifier predictions for objects from Set1 and Set2

During the paradigm-free periods of the experiment, there is no direct way to assess the external validity of the classifier predictions. Classifiers were trained on all data from the two memory tasks and returned one vote per MRI volume of the resting state. This vote reflected which stimulus from the training data the given resting state MRI volume was most similar to and either referred to a stimulus from Set1 (before the main resting period) or from the stimulus Set2, which served as a control. The ratio of classifier votes for Set1 objects to all classifier votes in a given period will be termed “Set1 ratio”. If the classifier were not able to detect any valid information in the main resting period fMRI, then the Set1 ratio should be at 0.5 (the classifier making random guesses, evenly distributed across all 32 stimuli).

One major problem for this analysis is that data from the two memory tasks and the main resting period were recorded in three different sessions. Despite careful preprocessing and coregistering, subtle differences between sessions are likely to remain. When training the classifier on the objects of the different sessions, it is hard to determine whether it is picking up on differences in the data that are merely session-related. For example, including a voxel that is completely inside the brain in one session, and only half inside the brain in the other session will allow the classifier to distinguish Set1 objects from Set2 objects. In addition, slow, long-term changes in brain activity over the sessions may contribute to a classifier bias. To elucidate these session-specific and temporal effects, a surrogate approach was used again: Linear SVMs were trained data in which labels were shuffled trial-wise, but independently within memory task 1 and memory task 2. The structure of the experiment was thus conserved and allowed us to determine the potential bias introduced by temporal and spatial proximity to the resting periods. Data was shuffled 100 times for each participant, linSVMs were trained on the shuffled data and votes for the different resting periods were derived analogous to the approach with real data, resulting in a surrogate distribution of Set1 ratios. The median of this distribution for each participant was taken as comparison value for pair-wise t-tests.

10.2.9 Relating classifier predictions to memory performance

In addition to analyzing the ratio of votes for Set1 object to all votes in the resting periods, classification frequency of individual items from Set1 and Set2 was correlated to subsequent memory success in the cued recall task. Spatio-temporal bias due to different sessions does not play a role here as Set1 objects and Set2 objects are analyzed separately, and objects from the same set were always presented in the same encoding session, evenly distributed across the five blocks of the task. To maximize power, all votes during the four different resting states following the first memory task (i.e., all phases during which replay is possible) were analyzed.

Classification frequency values were obtained for each of the 16 objects of a set and for each participant a Spearman correlation was calculated between these frequencies and the respective memory error values at later recall. Correlation coefficients were then tested against zero with a one-sided t-test. As more replay was expected to be associated with less memory error, correlation coefficients were expected to be below zero, resulting in negative T-values. This analysis was further validated using surrogate statistics. Classification frequency for individual objects found in the resting periods was randomly shuffled with respect to the item-specific memory performance within each participant and again calculated a correlation coefficient based on these shuffled data. Then, a T-test was calculated against zero with the resulting Fisher-z-transformed correlation coefficients across participants in the same way as was done for the real data. 10000 permutations were computed in this fashion. Then it was tested whether the T-value from the empirical data was below (more negative than) the $5th$ percentile of this surrogate distribution.

Replay may not only correlate with the continuous measure of associative retrieval during cued recall, but also with memory for the individual items. Therefore,a logistic regression was also calculated between the number of classifier votes for individual stimuli from the first memory task (again during the combined resting periods after presentation of the first memory task) and the remembered/forgotten dichotomous values from the free recall memory task, during which participants either did or did not freely remember each object that had been presented. For every participant, a logistic regression was calculated with “number of classifier votes” as predictor and “remembered/not-remembered during free recall” as criterion. The beta coefficients were again tested against zero with a one-sided t-test.

10.3 Results

10.3.1 Sleepstaging

All 10 participants considered reached at least sleep stage 2. The mean time spent inside the scanner during the resting period was 88.8 ± 30.2 (mean±std) minutes. Subjects spent 27.4 ± 25.6 (mean±std) minutes awake, 26.5 ± 23.5 minutes in sleep stage 1 and 25.8 ± 19.4 in stage 2. Five subjects reached sleep stages 3 and 4 for 13.1 ± 6.1 minutes. Four subjects reached REM sleep for 6.8 ± 5 minutes.

10.3.2 Behavioral results

In the free recall condition, subjects had to name every object which they remembered from the memory tasks. Participants remembered 5.5 ± 2.1 (mean±std.) objects from the first memory task and 10.5 ± 3.3 objects from the second memory task.

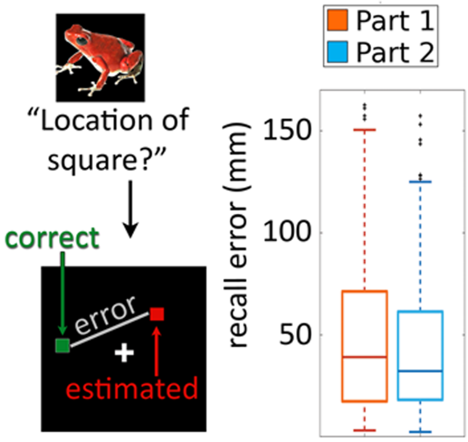

Figure 10.3: Behavioral results. Memory performance was measured as the distance between the correct and the indicated spatial position of the square which was associated with an item during the encoding phase. The box plots showing median and variance of memory performance across all recall trials and participants demonstrate relatively high intra- and inter-individual variability.

.

This increase in memory performance was highly significant ($t_9 = -4.4$, $p = 0.0017$) and is probably due to the relative recency of objects from memory task 2. Results from cued-recall show the same direction, but the difference is not significant. Memory performance in the cued-recall task was operationalized as the distance in mm from the correct position of the white square (“correct position”) to the position indicated by the participant (“estimated position”) (see Figure 10.3). Thus, larger values indicate worse memory performance. This distance was $50.6 \pm 28.1mm$ for objects from the first memory task and $45.1 \pm 28.0mm$ for objects from the second memory task ($t_9 = 0.84$, $p = 0.42$; Fig. 10.3).

10.3.3 Pattern classification accuracy

In the 16 participants who completed the study, classification accuracy for the 32 different objects from Set1 and Set2 varied between 12% and 59% (mean±std.: $33\% \pm 15.3\%$), which was highly above chance level ($100\%/32 = 3.125\%$; $t_{15} = 8.27$; $p < 0.0001$). As excellent classifier performance was a prerequisite for the identification of possible stimulus-specific reactivation during the resting periods, participants with insufficient classifier accuracy were excluded (cut-off determined by a surrogate approach: 15.12%), resulting in a final sample of 10 participants (Fig. 10.4A). Classification of the experimental stimuli was mainly based on voxels

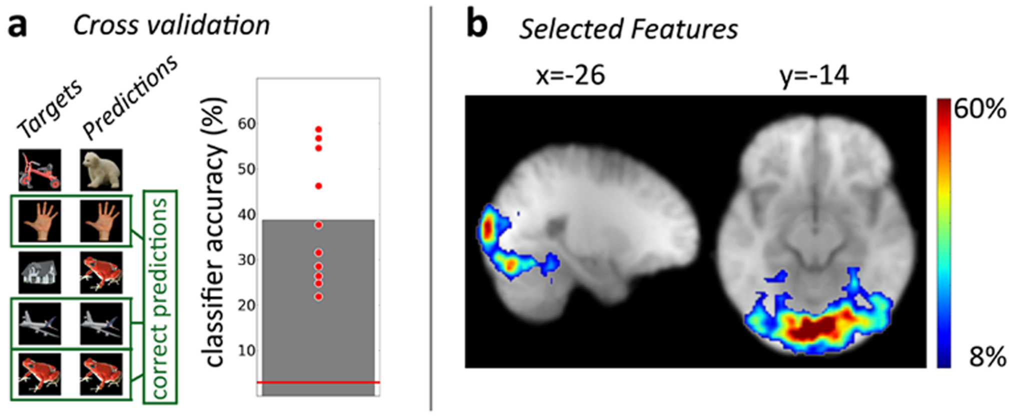

Figure 10.4: **A:** Pattern classification accuracy as assessed by a cross-validation approach. Each red point indicates results from one participant, the red line indicates chance performance (3.125%). **B:** The classifier was trained on the 1000 most discriminative features (i.e. voxels) from each subject. The figure shows the regional distribution of features that were selected most often across participants, which were most abundant in the occipital lobe but reached into inferior temporal cortex.

.

from the visual cortex, which extended into the ventral visual stream and even the posterior parahippocampal gyrus (Fig. 10.4B).

10.3.4 Pattern classifier predictions for Set1 versus Set2 objects

Figure 10.5 provides an overview of classifier predictions during all resting periods when trained on empirical data and on trial-shuffled shuffled surrogate data. The main resting period can be further divided into waking state and the five different sleep stages. A repeated measures ANOVA revealed significant differences of classifier votes (ratio of Set1 votes to all votes) during the different resting periods (Pre1, Post1, complete main resting period, wake, S1, S2, Pre2 and Post2; $F_{7,63} = 4.66$, $p < 0.001$). The ratio of Set1 votes to all votes was significantly above 0.5 in all periods except Post2 ($t_9 = 1.477$, $p = 0.088$). This result is in accordance with prior hypotheses for the main resting period. Surprisingly, however, it was also found that classifier predictions favor objects from Set1 already during phase Pre1 ($t_9 = 2.64$, $p = 0.013$), during which no replay is possible.

To better understand this apparent bias, votes from a surrogate classifier trained on trial-shuffled data were investigated. Again, an ANOVA revealed significantly

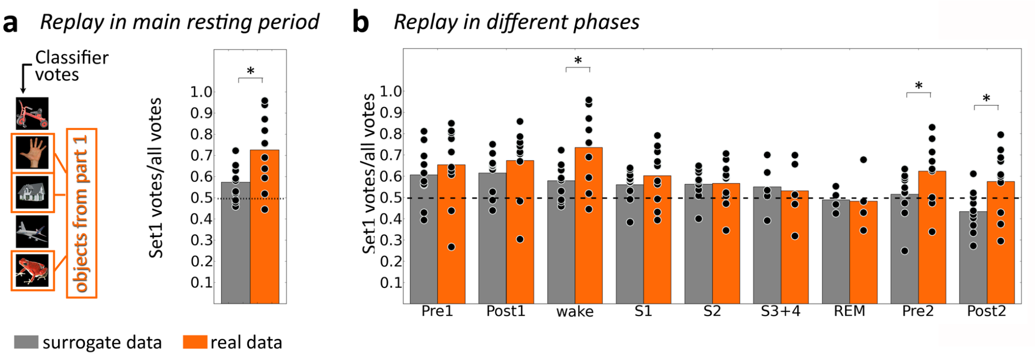

Figure 10.5: **A:** Results from the main resting period between the two experiments. The frequency with which objects from the first memory task were voted for by the classifier compared to the total amount of votes. Gray bars indicate results derived from a surrogate approach, orange bars refer to results in the empirical data. Objects from experiment 1 are voted for significantly more often than would be expected by chance in both empirical and surrogate data, but ratios for experiment 1 votes to all votes are significantly higher in the empirical than in the surrogate data. **B:** Frequency of votes for objects from the first memory task in the task-adjacent resting periods (Pre1, Post1, Pre2, Post2) and in the different stages of the main resting period in the empirical and surrogate data. The ratio of votes for objects from the first memory task to all votes was higher in the empirical vs. the surrogate classifier during the waking period, as well as during Pre2 and Post2.

.

different votes during the different stages ($F_{7,63} = 7.29$, $p < 0.001$). During the entire experiment (Pre1, Post1, main resting period, Pre2, Post2), the ratio of Set1 votes to all votes decreased monotonically for the surrogate classifier, as indicated by a significant linear trend ($F_{1,9} = 14.56$, $p = 0.004$).

Next, the results from the empirical and the surrogate classifier were compared during the different stages. It was found that the empirical classifier generated a significantly higher ratio of Set1 votes to all votes than the surrogate classifier during the main resting period ($t_9 = 3.14$, $p = 0.006$), as well as during Pre2 ($t_9 = 2.93$, $p = 0.008$) and Post2 ($t_9 = 3.48$, $p = 0.003$). In contrast, there was no significant difference during Pre1 ($t_9 = 1.64$, $p = 0.067$ [note that this test, as all others, is one-sided even though there was no one-sided hypothesis for Pre1, making this test conservative]) or Post1 ($t_9 = 1.49$, $p = 0.085$). This result strongly suggest that the apparent bias during the Pre1 period, but not the effect during the main resting period, is attributable to the temporal proximity of the presentation of Set1 items. When the different stages of alertness were analyzed, a significantly higher ratio of Set1 votes to all votes generated by the empirical vs. the surrogate classifier was

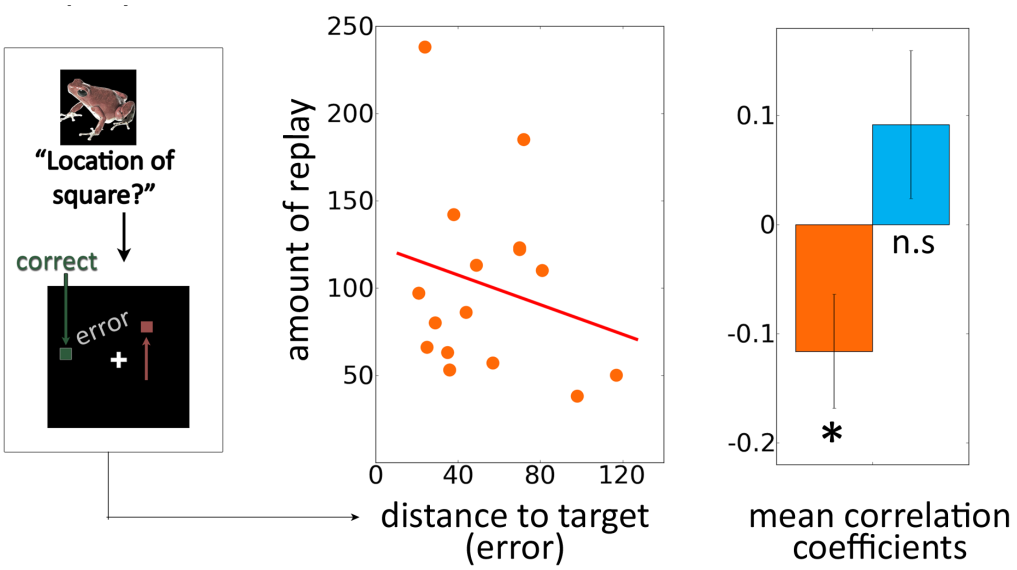

Figure 10.6: Illustrative scatter plot for one participant of the relationship between the number of classifier votes for a given stimulus and the distance to target during memory recall for the respective stimulus. Right: Fisher-z-transformed correlation coefficients between stimulus-wise error during behavioral recall and stimulus-wise number of classifier votes for objects from the first memory task (orange) and the second memory task (blue).

.

found only during the waking state ($t_9 = 3.87$, $p = 0.002$). Moreover, the difference in the ratio of Set1 votes to all votes between empirical data and surrogate data was significantly greater during waking state than during Pre1 ($t_9 = 3.07$, $p = 0.007$).

10.3.5 Association of classifier votes with memory performance

Next, it was analyzed whether reactivation of individual Set1 stimuli was related to subsequent memory of the positions associated with these stimuli (Fig. 10.6). Importantly, this analysis is independent from the analysis of the ratio of Set1 votes to all votes reported above. For example, there can be a high correlation with behavioral accuracy for objects from the first memory task, even when the total ratio of all votes for objects from the first memory task is low and vice versa.

Memory was tested by presenting each stimulus and asking the participant to indicate the associated position. Recall error (the distance between the correct and the indicated position) is then an inverse measure of memory accuracy. For each participant, a Spearman correlation was calculated between the number of classifier votes for an individual stimulus (classification frequency) and recall error.

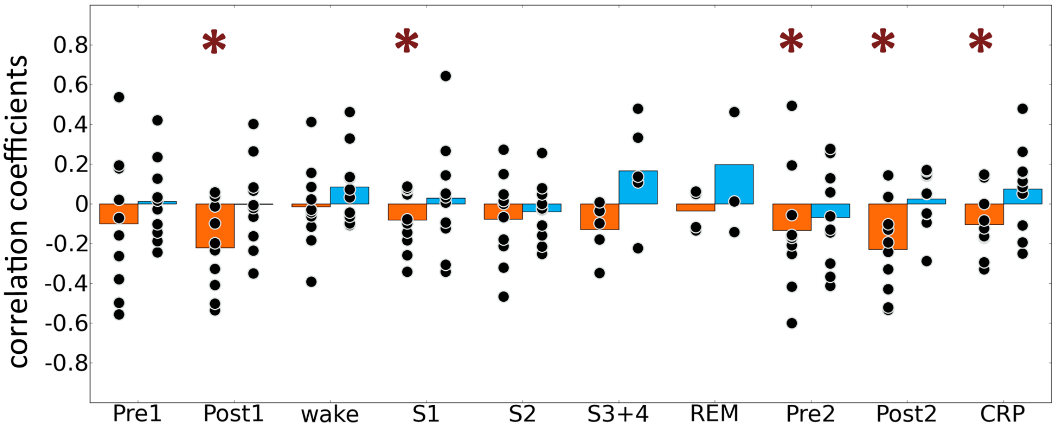

Figure 10.7: Fisher-z-transformed Spearmans correlation coefficients for Set1 objects (orange) and Set2 objects (blue) across different phases of the experiment, including waking state, sleep-stage 1 (S1), sleep-stage 2 (S2), slow-wave sleep (S3+4) and REM. Combined resting period (CRP) includes all resting periods following presentation of the first memory task. Stars indicate phases with significant consistent negativity (one-sided t-test against zero). There was no consistent negativity in any phase for correlations involving Set2 objects.

.

For stimuli from the first task, a significant negative correlation between the amount of replay and recall error during all resting periods after presentation of the first task was expected. In contrast, there should be no correlation with memory performance during Pre1. For stimuli from the second task, there should be a significant negative correlation during the Post2 period, but not during the other resting periods.

It was found that the (Fisher-z-transformed) correlation coefficients across 10 participants were significantly smaller than zero (one-sided t-test: $t_9 = -2.20$; $p = 0.027$; Fig. 10.6). This replay cannot be solely related to covert rehearsal by subjects, because the consistently negative correlation is also evident during sleep stage 1 ($t_9 = -2.81$; $p = 0.02$), and shows a trend during sleep stage 2 ($t_9 = -1.98$; $p = 0.08$). Importantly, no such consistently negative correlation was observed in the resting period preceding the first memory task ("Pre1"; $t_9 = -1.82$; $p > 0.1$) and none for Set2 stimuli (highest T-value for any of the different phases including Post2: $t_9 = 1.35$; $p > 0.1$). Data from all individual phases of the experiment are presented in Figure 10.7.

These results were confirmed by a boot-strapping approach (during which the number of classifier votes was randomly permutated with respect to the item-specific memory performance): For objects from the first memory task, the T-value for correlation coefficients was above the 5th percentile for the resting state before the first memory task (Pre1, $percentile = 11.32$), indicating lack of a significant effect, and below the 5^{th} percentile for the combined resting period after the first task (all four resting periods after task 1, $percentile = 2.83$). When surrogate data were generated in the same fashion for classifier votes for objects from the second memory task, T-values were never below the 5th percentile for either resting period (smallest percentile=89.53).

There was no relationship between number of classifier votes and behavioral performance in the free recall test: For the combined resting period (see above), the beta-values were not significantly different from zero across participants for stimuli from the first memory task ($t_9 = -1.40$; $p = 0.19$) or for stimuli from the second memory task ($t_9 = -0.77$; $p = 0.46$).

10.4 Discussion

Taken together, MVPA was used on fMRI data to decode stimulus-specific activity patterns and to investigate spontaneous replay of these patterns during awake resting state and sleep. Most importantly, it was shown for the first time that the amount of replay for a specific stimulus correlates with memory performance for this stimulus.

First, methodological considerations will be discussed and then the present findings will be related to previous work on memory reactivation. MVPA training required repeated presentations of each stimulus in a sufficiently slow event-related design, which limits the number of different stimuli that could be presented. The use of a second, post-resting memory task induced interference to the items learned before sleep and allowed us to directly explore the effect of reactivation on the stabilization of associations against interference, the major function of memory consolidation (Mueller and Pilzecker, 1900; McGaugh, 2000; Ellenbogen et al., 2006a). The retrieval task of Rudoy and colleagues (Rudoy et al., 2009) was adopted and item-specific memory performance was measured by the distance between the actual and the remembered position of each stimulus. This continuous metric allowed us to detect differences in memory even at relatively high levels of performance during cued recall. The relatively bad memory performance during free recall (mean: 5.5 items) might be explained by the transfer-appropriate processing theory (Morris et al., 1977; Stein, 1978), since participants were prepared for cued recall, but not for free recall.

In order to facilitate subjects falling asleep inside the scanner, participants were told to sleep two hours less in the night preceding the experiment. In principle, using even this mild form of sleep deprivation may not only affect attention due to increased sleepiness, but might also disturb sleep structure in a subsequent nap. However, such a procedure is difficult to avoid in fMRI sleep studies – we even used a milder sleep deprivation scheme than in previous studies (Rasch et al., 2007; Bergmann et al., 2012). Furthermore, sleep structure was found to be relatively typical for an afternoon nap.

As a result of pattern classification analyses, a vote was obtained for every single fMRI volume during all phases of the main resting period and the task-adjacent resting periods. Every vote referred to one of the 32 different objects that were studied during the two memory tasks and reflected that the fMRI data of the given volume was most similar to the data of one particular stimulus during the encoding phase. Surprisingly, more replay of Set1 than Set2 stimuli was found not only during the main resting period, but even during the task-adjacent resting period Pre1, i.e. before any stimuli had been presented. This apparent bias may be due to two main factors.

First, the experiment was split up into three fMRI sessions due to its long duration, and participants left the scanner between the sessions. Even though coregistering the sessions to one another is capable of aligning the MRI images reasonably well, images within one session will be more similar to one another than images between sessions. However, no such bias would be expected during the main resting period, which is in a different session between presentation of both Set1 and Set2 stimuli. Second, a bias would be expected due to MRI-related temporal autocorrelations, which may arise from slow metabolic processes or even circadian rhythms. As the main resting period was at an equal temporal distance to both stimulus sets, this should not affect the voting behavior for the main resting period, but might be relevant for the short task-adjacent resting periods. These problems were addressed by using a surrogate approach in which data were shuffled but in which the temporal structure of the experiment was conserved. Even though classifiers were trained on nonsense data, classifier output still showed a bias during the task-adjacent resting periods. By showing that Set1 ratios in the empirical data are higher than Set1 ratios in the surrogate data in the waking state of the main resting period, reactivation could be demonstrated over and above any bias that may be caused by the temporal structure of the experiment.

10.4.1 Relationship to previous studies on replay

The novelty in the presented study was the use of a multivariate technique for detecting and tracking neuronal activity related to specific stimuli, an approach which has also been recently suggested in a review article (Rissman and Wagner, 2012). While the motivation for this study study was derived from studies in rodents that show replay of hippocampal place cells after the learning of a spatial task (Skaggs and McNaughton, 1996; Ji and Wilson, 2007; Foster and Wilson, 2006; Karlsson and Frank, 2009; Carr et al., 2011; Jadhav et al., 2012), several technical and theoretical differences between these approaches have to be mentioned. First, electrophysiological recordings in rodents allow one to directly measure neuronal activity of individual cells. Second, the increase of neuronal firing rates on specific spatial locations represents a simple spatial rate code, which simplifies the subsequent detection of replay. On the other hand, as the number of recording sites and thus the spatial coverage is inherently limited in these studies, fMRI recordings allow one to indirectly measure neuronal activity patterns in the entire brain and to explore their potential contribution to replay. Of course, the nature of replay activity traceable with fMRI differs drastically from single-cell recordings. However, several previous studies have already indicated that content-specific reactivation can be detected in fMRI data as well. Tambini and colleagues (Tambini et al., 2010) investigated BOLD correlations between hippocampus and lateral occipital cortex in resting state fMRI preceding and following a hippocampus-dependent memory task and found a significant increase in correlations which was inter-individually related to memory performance during recall. In addition, there is evidence that category-specific fMRI activity can be detected prior to free recall of learned stimuli from that category using multivariate pattern classification (Polyn et al., 2005). Finally, Bergmann et al. (Bergmann et al., 2012) found a reactivation of BOLD responses in category-specific regions during sleep which was triggered by spindle events in simultaneously recorded EEG.

Investigating spontaneous replay in the absence of stimulation requires one to detect stimulus patterns very reliably (with high classification accuracy). Accuracy in this experiment was relatively high as compared to previous studies due to the selection of very diverse items in this study (see Figure 10.1). Indeed, during piloting with a version of the paradigm using only face stimuli, much lower accuracy values were obtained (n=8, 16 faces, mean accuracy 13.1%). Additionally, only participants with high classification accuracy were included in the main analysis (this procedure is not circular, because classifier cross-validation is done on data during encoding, while reactivation is tested during resting periods that have not been assessed during cross-validation at all).

Interestingly, reactivation was most evident during awake resting state. These results are in apparent discrepancy to a recent study by Diekelmann and colleagues (Diekelmann et al., 2011) who showed that presentation of an odor cue that was previously paired with an associative learning paradigm only improved memory stability if it occurred during slow-wave sleep, but not if it was presented during awake resting state. Similarly, previous behavioral studies on reconsolidation indicate that presentation of a learning-related context during subsequent waking state destabilizes memory traces (Hupbach et al., 2007). Several differences between these studies and the current study might explain the apparent discrepancy to the results in this study. Most importantly, reactivation was cued in the studies by Hupbach et al. (Hupbach et al., 2007) and Diekelmann and colleagues (Diekelmann et al., 2011), whereas it occurred spontaneously in this study. It could be speculated that sensory stimulation triggers bottom-up information flow into the hippocampus (Hasselmo, 2005; Takeuchi et al., 2011) which might affect reactivation differently than if it occurs spontaneously. Indeed, several electrophysiological studies in rodents (Foster and Wilson, 2006; Karlsson and Frank, 2009; Carr et al., 2011; Jadhav et al., 2012) as well as fMRI (Peigneux et al., 2006; Tambini et al., 2010) and iEEG results from humans (Axmacher et al., 2008) are consistent with the hypothesis that reactivation and memory consolidation may occur also during awake resting state. Alternatively, this discrepancy may be due to the choice of the sleeping phase, namely an afternoon nap instead of a night sleep period. However, previous studies have provided evidence that an afternoon nap affects memory consolidation similar to night sleep (Takashima et al., 2006; Lau et al., 2010), even if it lasts only a few minutes (Lahl et al., 2008).

Most importantly, a consistently negative correlation between the amount of reactivation and later memory error for these stimuli was found. While replay was not observed during phases of slow-wave sleep, this might well be due to the relatively small number of subjects who actually reached slow-wave sleep in this study (N=5). Alternatively, the pattern classification algorithm might have been unable to detect the activity patterns from waking state during slow-wave sleep, which shows significantly altered BOLD activation (Dang-Vu et al., 2008).

The role of sleep, especially slow-wave sleep, in the reactivation processes reported here might be investigated better by using an electrophysiological method such as intracranial EEG. Even though the appearance of EEG is also drastically altered during sleep, as was outlined in the introduction, it might not be affected as much by metabolic changes. Also, if the signal is decomposed into different frequency bands, replay might be detected by the relative pattern of frequency band power regardless of the overall makeup of the signal. In the third study of this thesis, a very similar paradigm to the one that was used in this study was applied to investigate reactivation in intracranial EEG.

In the future, more insight might also be gained by performing a similar study with high resolution hippocampal fMRI imaging. Then, hippocampal patterns might be decoded and they might be detected more reliably in sleeping periods. In addition, longer periods of slow-wave sleep in a greater sample of participants might help clarifying the usefulness of the MVPA method for detecting replay during deep stages of sleep.

11 Replay of neuronal activity associated with emotional stimuli

11.1 Introduction

In the first study, it was investigated whether any signs for reactivation of stimulus-specific neuronal activity patterns associated with normal stimuli could be detected. Evidence was found that multi-voxel pattern classification can identify such reactivation and that the frequency of this reactivation is associated with later memory strength. The second study investigates the influence of emotional arousal on this kind of reactivation process.

It would be naive to assume that all memory content is treated equally by the brain. It is adaptive for an organism to remember especially those events which will promote survival and reproductive success. Based on introspective experience alone, it is an intuitive assumption that episodes and facts of special importance or saliency are remembered better.

Accordingly, it has often been demonstrated that memory for emotionally arousing material is superior to that of non-arousing material (Hamann, 2001; Kensinger and Corkin, 2003; Kensinger, 2004).

Of special importance in neuronal models of the consolidation of emotional material is the amygdala. Patients with damage to the amygdala have been found to lack enhanced memory for emotional content (Cahill et al., 1995; Adolphs et al., 1997) and functional imaging studies have confirmed the involvement of the amygdala in memory formation for emotional content, as has been described above in section 4.2.3. In addition, emotional memory seems to be not only enhanced by increased attention or saliency at encoding, but the memory might indeed be consolidated differently (again, see section 4.2.3).

This raises the question investigated in this study: If enhanced memory for emotionally arousing material is indeed associated with improved consolidation, this preferential treatment should be reflected in increased neuronal reactivation during resting state or sleep.

Thus, instead of observing reactivation frequencies and relating them to subsequent memory performance as was done in the first study, here, the differential memory effects for emotionally neutral as compared to emotionally negative stimuli are taken as a starting point and reactivation frequencies are investigated separately for these two stimulus-classes.

Investigating reactivation of neuronal activity associated with emotional stimuli as compared to emotionally neutral stimuli is not only an obvious follow-up, but it also addresses an issue at the heart of any two-step model of memory formation: The question why there should even be a need for consolidation. Why not "chisle every memory into stone" the minute it has been encoded? The synaptic processes necessary for establishing a memory trace happen on a much shorter time-scale than system consolidation (McGaugh, 2000), so one can conclude that the long time-span is not immediately due to a biological limitation of the neuronal substrate. So why take so much time for consolidation?

According to McGaugh (McGaugh, 2000), one reason for ongoing consolidation and re-consolidation might be to avoid an overload of the memory system and to allow enough time for the most salient memories to take superiority in strength at the expense of less important memory traces. An equally important issue could be to carefully integrate new memories in the appropriate existing networks. In any case, the study of emotionally charged stimuli is an important step to understanding real-life memory, which is usually affected by emotional influences.

In this study, better memory performance for emotionally negative as compared to emotionally neutral items was expected. This, in turn, was hypothesized to be related to increased reoccurrence of neuronal activity patterns associated with negative items during a resting state following the learning task.

11.1.1 Changes in experimental design

In principle, the experiment was kept as similar as possible to the first study, both because the design was found to be effective and in order to be able to compare the results.

One caveat when designing this study was that the literature is actually not quite consistent when it comes to better memory performance for emotionally negative images. It seems that it depends on the exact task and experimental conditions. It has been argued that good memory for individual items does not necessarily imply good memory for the association between these items (Mather, 2007). The effect of emotion on memory has usually been tested in terms of item memory, often with recognition tasks such as “Have you seen this picture before?”. However, there are some studies that suggest that association binding might be impaired if highly arousing stimuli are involved (Mather et al., 2006; Onoda et al., 2009; Okada et al., 2011). But improved recall of locations associated with emotional pictures has also been reported (Mather and Nesmith, 2008). Mather suggests that if associated details are an intrinsic part of the emotional stimulus (such as the color of an emotional word), they are remembered better than those associated with neutral stimuli, but in cases of associations between different items (e.g. emotional word and neutral face), memory performance might be impaired (Mather, 2007).

Simple recognition tasks could not be used in this study, because pictures had to be presented multiple times to ensure good classifier performance. On the other hand, between-item associative memory tasks as used in the first study might have resulted in worse memory for negative pictures (or, at least, their associated items).

The paradigm was adapted accordingly: In the first study, there were two parts of each trial: Presentation of the picture followed by the presentation of a white square that marked the corresponding position. In this study, each trial had only one part: The stimulus was shown directly at the associated position. This was supposed to make the position more of an inherent part of the stimulus and was considered a good compromise to accommodate the necessities of a pattern classification approach on the one hand and findings related to emotional processing on the other hand.

For this study, emotionally neutral and emotionally negative pictures were taken from the International Affective Picture System (IAPS). This is a commonly used database of pictures, which have been rated by a large sample of participants with regard to their emotional valence as well as their arousal (Lang et al., 1999).

Another change concerned the resting state. As the first study showed that reactivation was tracked best during waking state, participants were not asked to nap in this study. Instead, only quiet resting state was recorded. Because of this, sleep staging was no longer necessary and simultaneous EEG was not conducted.

11.1.2 Investigation of regions of interest

The processing, encoding and retrieving of emotional stimuli is associated with several brain regions. Foremost, the amygdala has been implicated in emotional processing (Hamann, 2001; LaBar and Cabeza, 2006). Other important brain regions are the insula, anterior cingulate and medial prefrontal cortex (Phan et al., 2002; Phan et al., 2004).

To account for this, regional analyses were also performed in this study. In this approach, the classifier was only trained on voxels from specific brain regions that were determined with anatomical atlases. The differential contribution of individual brain regions to encoding and replaying emotional content was expected to provide additional insight.

11.2 Methods

11.2.1 Participants

Twenty-one young healthy participants took part in this experiment, 11 of whom were female. The mean age and standard deviation were 24.2+2.86 years within a range of 19-30 years. All proceedings were approved by the ethics committee of the University of Bonn. Participants were recruited via the job exchange at the University of Bonn. They gave written informed consent and were compensated for their time. They were informed prior to the experiment that stimulus material would be presented which might be upsetting or aversive and examples were sent via email on demand. Some potential participants opted not to take part in the experiment after viewing example pictures.

Before they were given the written instruction, participants were reminded again that highly aversive pictures would be presented and they were confronted with some examples if they had not already seen them. This was done to ensure that participants knew what to expect and were able to make an informed decision about participating. However, no participant decided against participation at this point. None of the example pictures were used in the subsequent study. One participant had to be excluded due to erroneous settings in the program that ran the experiment.

11.2.2 Paradigm and Stimulus Material

While the memory task itself was similar to the paradigm used in the first study, the overall structure of this experiment differed. Instead of two memory tasks and one nap break in between, the experiment now consisted of two periods of resting

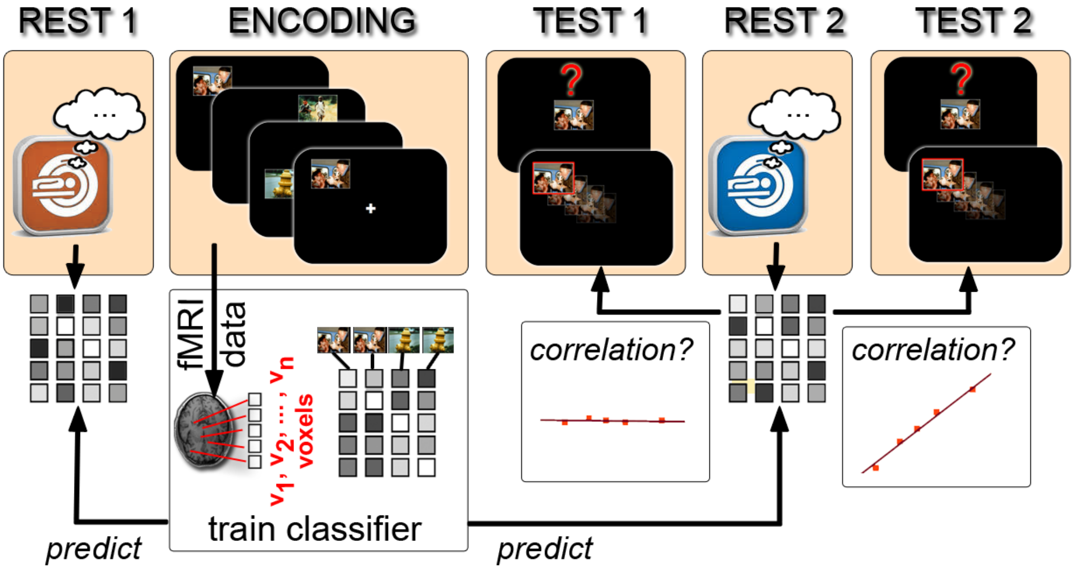

Figure 11.1: Overview of the paradigm: Two resting periods were recorded before and after a declarative associative memory task. In this task, 12 emotionally negative and 12 emotionally neutral pictures were presented at specific positions on the screen, 24 times each. Participants were instructed to learn the position for each of the pictures. One memory test was conducted immediately after the memory task and a second memory test took place after the second resting period. A linear SVM was trained on the fMRI data of the learning task to discriminate between the 24 different pictures and then made predictions on fMRI volumes during the resting states. More evidence for reactivation of negative as compared to neutral stimuli was expected. In addition, correlation between reactivation frequency for single items and subsequent memory performance was expected for predictions in the resting state following but not preceding the task.

.

state and a memory task in between. In addition, the two periods of resting state were now waking state only. Figure 11.1 provides an overview of the structure of the experiment.

Two resting periods of 30 minutes each preceded and followed the memory task (named “Rest 1” and “Rest 2”, respectively). During these periods, participants were instructed to lie still and relax. To ensure that they stayed awake during that time, a simple button press task was introduced: Every 40-80s a large red dot appeared on the screen and participants were instructed to press a button when this occurred, after which the red dot disappeared. After 5s without button press, the dot began blinking to encourage reaction. Participants were told that the sole purpose of this task was to ensure that they stayed awake and that the reaction to the appearance of the red dot was not at all about speed. This explanation was meant to reduce stress for participants and to minimize the perceived task character of this resting period.

The main task, scheduled between the two resting periods, was a declarative, associative memory task again during which pictures were associated with a specific position. 24 different stimuli were used: 12 neutral and 12 negative pictures from the IAPS collection (Lang et al., 1999). The dimensions most interesting to us were arousal (ranging from 1=“low arousal” to 9=“high arousal”) and valence (ranging from 1=“low pleasure” to 9=“high pleasure”). The neutral pictures were selected so that they had low arousal and neutral valence (arousal 2.62±0.23 mean±std, valence 4.96±0.24 mean±std), while the negative pictures were selected to have high arousal and negative valence (arousal 6.69±0.36 mean±std, valence 2.03±0.38 mean±std).

In each trial, one of the 24 pictures was shown at a specific position on the screen. The picture size was 150*100 pixels, which was large enough to get the gist of the scene. The picture was on the screen for 4s, then a fixation cross was presented until the beginning of the next trial. The inter trial interval was 5s, which again corresponded to the time needed for collection of two fMRI volumes. In total, each of the 24 pictures was presented for 24 times. The experiment was divided into twelve blocks which were separated by 1 minute of resting state. During each block, every picture was shown twice. The sequence of pictures was randomized within blocks. In total, the memory task lasted about 55 minutes.

Participants were instructed to judge for each picture in each trial whether they thought the scene was outdoors or indoors and indicate their answer via button press. Note that for some pictures this was an easy judgment while for some pictures it was completely arbitrary. Participants were instructed that, when in doubt, they should choose whichever they thought was more likely. Again, this was a task that was supposed to induce a deeper level of processing and was not analyzed further.

The instruction further stated that participants were supposed to look closely at the pictures in each trial and to memorize the position of each scene on the screen. Participants were aware that two memory tests would follow the experiment in which they would be presented with the scenes and would have to indicate the position that they were shown at during the learning task.

One test immediately followed the learning task and the second test was administered after Rest 2. Each memory test consisted of 24 trials. In each trial, one of the previously presented 24 IAPS pictures was shown in the center of the screen, a position at which none of the pictures had actually been presented during learning. Participants then moved the picture with four buttons up, down, left or right until it was at the position that participants thought it had been at during the learning task. The duration of trials was self-paced as every trial automatically concluded eight seconds after the last button press was entered by the participant. Five seconds after the last button press, a bright red frame appeared around the stimulus picture to warn participants that the trial would end in three seconds if they did not press any button. If participants pressed any button when the red frame was present, the frame disappeared and the time limit of $8s$ was set to zero again. The order of the 24 stimuli was randomized across each memory test. In addition to distance error, reaction time was also calculated in these memory tests (an improvement over the first study, in which this had not been possible). Reaction time was defined as the time from presentation of the stimulus until the last button press.

The paradigm was presented using the software Presentation (http://www.neurobs.com) inside the scanner with video goggles (NordicNeuroLab) with a resolution of 800x600dpi.

The memory task combined with the two resting periods and the memory tests lasted close to two hours in total. Since this is a duration which is too long to be scanned in one session (at least for most participants), a break was introduced after half of the memory task, which corresponded roughly to half of the experiment.

During this break the scanner was stopped and participants were taken out of the scanner to allow them to drink something, relax their muscles or take a bathroom break. After five minutes, they were taken inside the scanner again and continued with the rest of the experiment. This break was meant to prevent high drop-out due to excessive scanning length.

The break was placed so that a symmetric structure of the paradigm was preserved. In the previous experiment it had become apparent that separate scanning sessions can have considerable impact on classifier output. With the chosen setup, half of the memory task was in the same scanning session as Rest1 and the other half of the memory task was in the same scanning session as Rest2.

11.2.3 fMRI scanning and preprocessing

MR scanning was performed with a 3 Tesla scanner (TRIO, Siemens, Erlangen, Germany) using echoplanar imaging with the same settings as in the first study: For each volume, 37 slices covering the whole brain were acquired with a thickness of $2.5mm$ at $2500ms$ repetition time and $35ms$ echo time, a field of view of $210mm$ and a distance factor of 25%. Again, a high-resolution structural T1-weighted image of the whole brain was collected for coregistration purposes with 160 slices with a thickness of $1mm$ at $1570ms$ repetition time and $3.42ms$ echo time, a field of view of $256mm$ and a distance factor of 50%.

As was done in the first study, functional images were transformed from DICOM to NIfTI format using MRIcron (http://www.cabiatl.com/mricro/mricron/dcm2nii.html). Preprocessing was done with FSL (Smith et al., 2004; Woolrich et al., 2009) in the same way as in the first study, the steps including motion correction, $5mm$ Gaussian spatial smoothing and a linear detrending. No participant had to be excluded to due excessive movement (same cut-off as in the first study: mean relative movement $0.2mm$). There were four scanning sessions that were used for the pattern classification analysis: the two resting states ("Rest1" and "Rest2") and the two parts of the memory task, as participants were taken out of the scanner in the middle of the task (see above). All of these functional recordings were coregistered to "Rest1". A z-transformation was again performed in order to have the same mean activity in each of the four scanning sessions.

11.2.4 Pattern classification

Pattern classification analyses were again performed with a linear support vector machine (linSVM) using the PyMVPA toolbox available for Python (Hanke et al., 2009a; Hanke et al., 2009b). To account for the sluggish hemodynamic response which peaks at roughly 5 seconds after stimulus onset, again only the second MR volume after an event of interest was included in the dataset. This corresponded to the time window of 5000-7500ms after stimulus presentation. The classifier was trained two distinguish every individual stimulus from the memory task (24 different classes).

Feature selection: The classifier was trained on the z-scored raw BOLD signal of 1000 voxels across the trials of the encoding task. Voxels were selected as features based on the F-values that were derived from one-way ANOVAs which were performed on each voxel separately. In every ANOVA, the groups were the different classes that the classifier was trained on (the 24 individual stimuli in the first approach, negative and neutral in the second approach) and the dependent variable was the z-scored raw value of the given voxel in the second volume after stimulus presentation.

Cross-validation: In a first step, classifier accuracy was assessed with a cross-validation approach that tested classifier performance on the fMRI data of the memory task. The memory task was divided into 12 blocks which in turn were divided into two sessions (see above). The cross-validation was done sixfold: The classifier was trained on 10 of the 12 blocks (training dataset), which included 5 blocks from the first half of the memory task and five blocks from the second half, and made predictions on the remaining two blocks (test dataset). This was done six times so that every block in each part of the memory task was left out once. The classifier predictions on the samples of the test dataset were compared with the actual labels for these samples. The number of accurate predictions divided by all predictions yielded a percentage of correct predictions which served as an assessment of classifier accuracy.

11.2.5 Classifier predictions on resting state

After the classifier was trained on all samples from the memory task, it was set up to make predictions on each volume of the two resting states.

Predictions could be evaluated by counting the prediction frequencies of individual stimuli or of the overarching categories, negative and neutral. If the classifier was making random guesses on the resting state, no clear majority of either negative or neutral items should be evident: the ratio of predictions for negative items to all predictions should be 0.5. If, however, the classifier detected more reoccurrence of negative items, the ratio should be significantly larger than 0.5. During Rest1, when no stimulus has yet been presented, classifier predictions should not reflect increased neuronal activity related to negative item and the ratio should not be different from 0.5. In Rest2, during which negative items are supposedly consolidated preferentially, classifier predictions should be in favor of negative items. Thus, the ratio of predictions for negative items to all predictions in Rest2 should be both higher than 0.5 and higher than the ratio during Rest1.

11.2.6 Relationship between classifier predictions and memory performance

Again, prediction frequencies for individual items were related to later memory performance. In this study, free recall was not tested (due to practical reasons and because this measure of memory had not yielded any interesting result in the first study). Memory in this study was tested twice: immediately after the learning task (Test1) and half an hour later, after Rest2 (Rest2).

When relating classifier predictions to memory, there are multiple possible combinations: There are prediction frequencies from Rest1 and Rest2 that can be related to memory performance in Test1 and Test2. The correlation can be done across all 24 items or separately for the 12 negative and the 12 neutral items. In addition, prediction frequencies could be related to memory performance as defined by the distance error, or it could be related to reaction time.

The main assumption in these analyses was that prediction frequencies during Rest1 should not be correlated to memory performance at either Test1 or Test2, because no consolidation is possible during that time. Prediction frequencies during Rest2 were assumed to correlate stronger with memory performance during Test2 (as this test was performed after consolidation had a chance to take place) than during Test1. Also, if negative items are remembered better, it is a reasonable assumption that correlation between Rest2 prediction frequencies and memory performance at Test2 should be stronger for negative items.

As was done in the first study, Spearman's correlation coefficient was calculated between item-wise classifier prediction frequencies and item-wise memory performance for every participant. Then, correlation coefficients were fisher-z-transformed and tested against zero across participants to determine consistent negativity with a one-sample one-sided t-test against zero.

The vast possibilities for comparison combinations mentioned above raise the question of how one should correct for multiple comparisons. The strictest way would be to correct for all 24 tests (2 performance measures x 2 resting states x 2 memory tests x 3 valence categories [negative, neutral, both]). However, it seems more sensible to consider the 2 performance measures and the 3 valence categories separately. Then, there would be 4 comparisons to correct for (2 resting states x 2 memory tests). As it may be difficult to find an optimal solution here, uncorrected p-values will be given and marked as such; it will be noted if they survive Bonferroni correction for 4 comparisons.

11.2.7 Regional analysis

The paradigm used stimuli that were either negative or neutral. As mentioned in the introduction, there are several regions of interest that have been identified as relevant for the processing of emotional stimuli (Phan et al., 2002; Phan et al., 2004). Based on these meta-analyses, the following regions were investigated: amygdala, insula, anterior cingulate and medial prefrontal cortex. In all four cases, regions from both hemispheres were collapsed. To consider wide-spread activity in an emotional memory network, the four regions were combined to form a fifth mask, in which voxels of all the four regions were included. An overview of the regions of interest is given in Figure 11.2.

For the regional analysis, four anatomical masks with the bilateral regions of interest were extracted based on the Automated Anatomical Labeling (AAL) atlas (Tzourio-Mazoyer et al., 2002). They were mapped to the functional image space of every participant with FSL (Smith et al., 2004; Woolrich et al., 2009) and subsequently served as a means of feature selection: If a voxel was included in a given anatomical mask, it was also selected for classifier training, which was then done

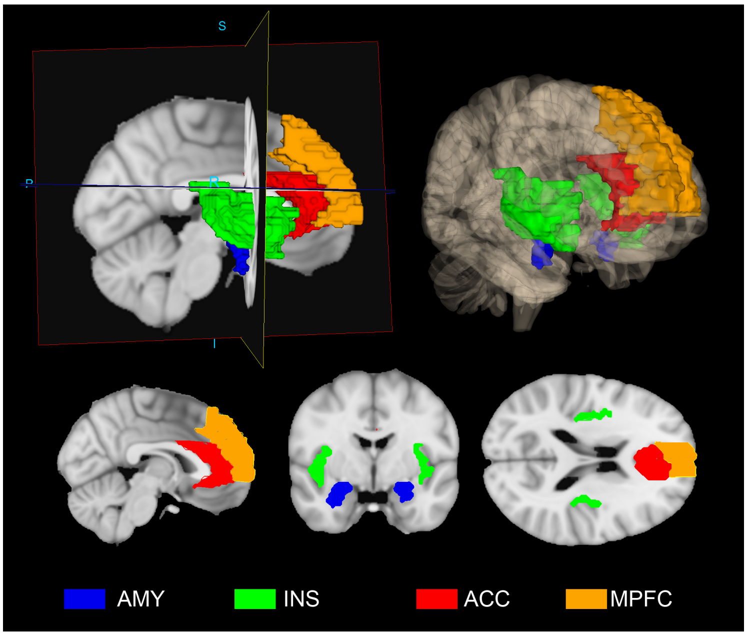

Figure 11.2: Regions of interest for the regional analysis. AMY=amygdala, INS=insula, ACC=anterior cingulate cortex, MPFC=medial prefrontal cortex.

.

based solely on the values from these voxels. No additional feature selection was performed. Classifier training was performed separately with voxels from each of the four anatomical mask, and then with a combination of all four masks, which resulted in five different classifiers for every participant.

Since the anatomical masks only influenced which voxels were included in the classifier training while the rest stayed the same, cross-validation and predictions on the two resting states were performed in the same way as was done with the regular classifier. In order not to multiply the results section five-fold, only selected aspects were investigated with the regional approach.

11.3 Results

11.3.1 Behavioral results

Two memory tests were performed by the participants (in the following referred to as “Test1” and “Test2”), one immediately after the end of the encoding task and the second after 30 minutes of resting state.

For each memory test, the mean error (distance between the actual position of the picture during the experiment and the position indicated by the participants) was

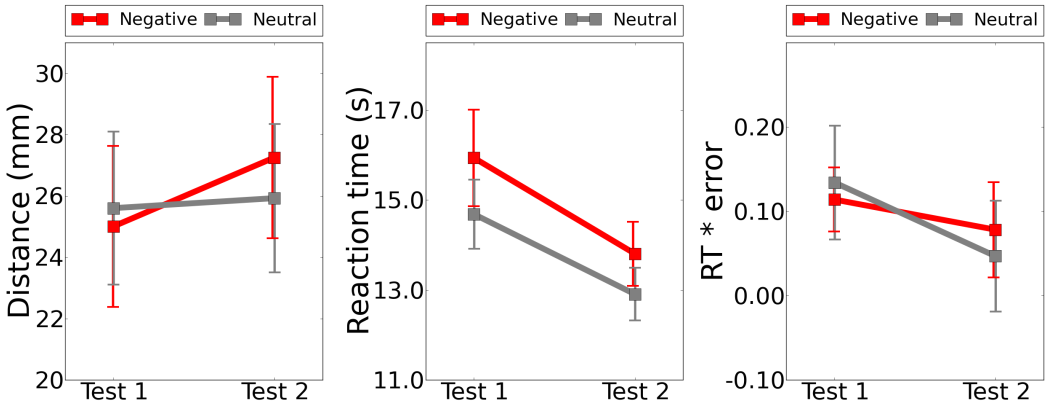

Figure 11.3: Behavioral results. A: Distance error between the actual location during encoding and location given by participants for negative and neutral items during Test1 and Test2. B: Time between presentation of picture and the last button press which was used to adjust position of the picture. C: Composite measure of reaction time and distance error.

.

calculated for negative and neutral pictures. In the first memory test, the mean error distance for negative items was $25.00mm \pm 11.45mm$ (mean±std), for neutral items it was $25.60mm \pm 10.88mm$. In the second memory test, mean error for negative items was $27.25mm \pm 11.50mm$ and for neutral it was $25.93mm \pm 10.53mm$. The results are shown in Figure 11.3A. A two-factor repeated-measures ANOVA with "Test 1 vs. Test 2" as first factor and "Neutral vs. Negative" as second factor and distance error as dependent variable revealed neither significant main effects nor an interaction (factor 1: $F_{1,19} = 4.184$, $p = 0.55$; factor 2: $F_{1,19} = 0.58$, $p = 0.813$; interaction: $F_{1,19} = 3.729$, $p = 0.069$). Looking at post-hoc contrasts with dependent t-tests, only the deterioration of memory recall for negative items between Test 1 and Test 2 was significant ($t_{19} = 2.432$, $p_{uncorr} = 0.025$), but this obviously does not survive Bonferroni correction for multiple comparisons.

Another measure of memory performance is reaction time. Even though participants were not specifically instructed to give their response as fast as possible, a speedy and correct reaction is probably an indicator of good and readily available memory. The reaction time considered here was not the time from stimulus onset until the first button press but the time until the last button press, because moving of the stimulus picture was concluded only then. Such defined mean reaction times during the first memory test were $15.93s \pm 4.69s$ (mean±std) for negative pictures and $14.68s \pm 3.33s$ for neutral pictures. In the second memory task, the mean reaction time for negative pictures was $13.80s \pm 3.11s$ and for neutral pictures $12.90s \pm 2.56s$. A repeated-measures two-way ANOVA with "Test 1 vs. Test 2" as one factor and "Neutral vs. Negative" as second factor and reaction time as dependent variable revealed significant main effects but no interaction (factor 1: $F_{1,19} = 9.238$, $p = 0.007$; factor 2: $F_{1,19} = 8.88$, $p = 0.008$; interaction: $F_{1,19} = 0.419$, $p = 0.525$). Looking at the data in Figure 11.3, this indicates that reactions times are slower at the first memory test and they are slower for negative items.

It should be noted that correct responses for pictures which were presented in the periphery of the screen during encoding necessarily take slightly longer because the starting point during recall is always in the center of the screen. Thus, moving the stimulus to the outer parts of the screen requires more time. However, as the position of each stimulus during encoding was randomly assigned, there should be no systematic difference in "distance from center" between negative and neutral pictures. This was confirmed when a t-test did not reveal any significant differences between mean distance from center for negative and neutral items across participants ($t_{19} = 1.382$, $p = 0.182$).

Of course, sheer reaction time might not be the best indicator for good memory. It is very likely that in case a participant does not remember a stimulus, he or she will simply not move the picture at all from its starting position. This would also lead to fast reaction time. Therefore, precise distance recall combined with fast reaction time is what would reflect a good memory performance best. In an approximation, z-transformed reaction times were multiplied with z-transformed distance errors for each trial, resulting in a composite measure of reaction time and distance error ($RT * error$). A two-way repeated measures ANOVA with "Test 1 vs. Test 2" as one factor and "Neutral vs. Negative" as second factor and reaction time as dependent variable revealed no significant differences (factor 1: $F_{1,19} = 1.061$, $p = 0.316$; factor 2: $F_{1,19} = 0.008$, $p = 0.93$; interaction: $F_{1,19} = 0.218$, $p = 0.646$).

Taken together, the behavioral results provide no evidence at all that emotionally negative items are remembered better than emotionally neutral items. On the contrary, reaction times were even slower for negative items than for neutral items. Reasons for this unexpected result will be discussed below.

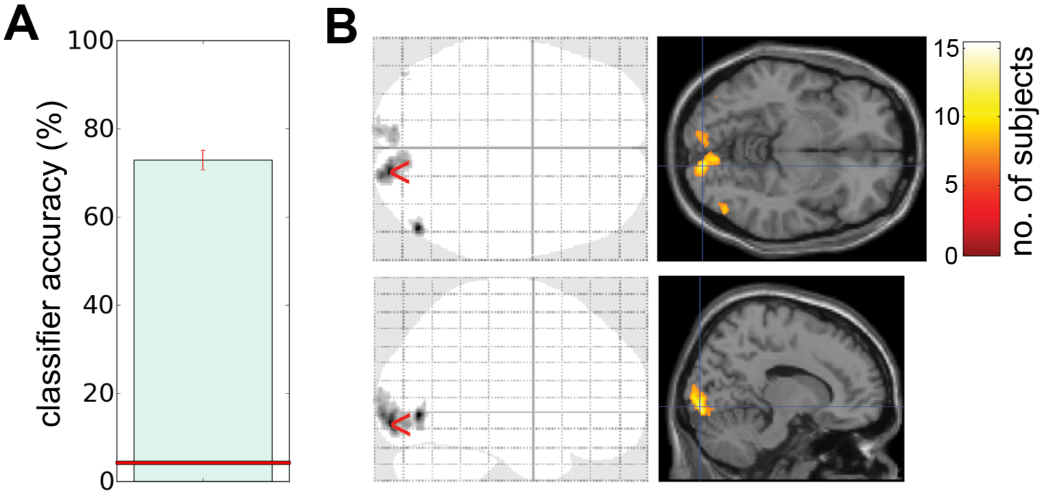

Figure 11.4: Classifier properties. **A**: Classifier performance in 20 subjects (mean and standard error of the mean). Red line indicates chance level. **B**: Features selected during classifier training. The color indicates in how many participants a given voxel was selected. Voxels selected most consistently across subject were again located in the occipital lobe.

.

11.3.2 Classifier accuracy

The ability of the pattern classification algorithm to reliably distinguish between the different stimuli presented during the memory task was assessed with a cross-validation approach as described in the methods section. Overall classifier accuracy was very good with 72.95% ± 9.77% (mean±std) in a range of 44.62-85.42. This was significantly better than the chance level of 4.17 ($t_{19} = 30.696$, $p < 0.0001$). No participant had to be excluded due to low classifier performance. The results are shown in Figure 11.4A.

Voxels selected for the classification of the 24 stimuli are shown in Figure 11.4B.

11.3.3 Ratio of negative items during the resting state

After establishing that classification accuracy was very good, predictions of the classifier on the two resting states were investigated. The ratio of predictions for negative items to all predictions was 0.46 ± 0.06 (mean±std) during Rest1 and 0.45 ± 0.05 (mean±std) during Rest2. The ratios were transformed with Daniel's arcsin transform and tested against 0.5 with a two-sided one-sample t-test (Rest 1: $t_{19} = 3.08$, $p_{uncorr} = 0.006$; Rest 2: $t_{19} = 4.33$, $p_{uncorr} = 0.0004$). Thus, in both

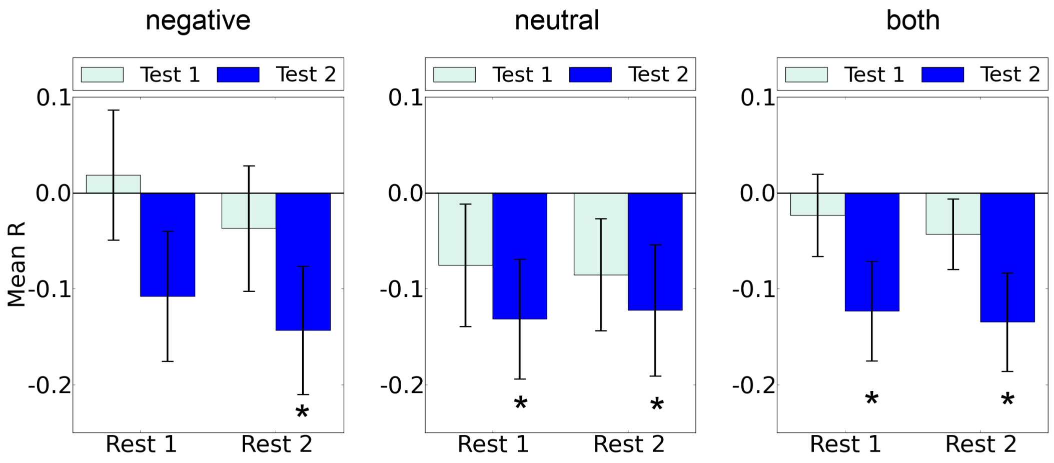

Figure 11.5: Association between classifier prediction frequencies during the two resting periods and memory performance (distance error) during Test1 and Test2 for negative, neutral or both items. For every participant, a correlation coefficient was calculated between item-wise classifier prediction frequency and item-wise memory performance. Mean correlation coefficients are shown for the different conditions. Error bars indicate standard error of the mean (sem). Across participants, Fisher-z-transformed correlation coefficients were tested against zero with a one-sided t-test. Stars mark consistently negative correlation coefficients ($p_{uncorr} < 0.05$).

.

resting states, the ratios were significantly smaller than 0.5. However, the ratio was not higher in Rest2 as compared to Rest1 (one-sided paired t-test: $t_{19} = 1.327$, $p = 0.100$). As with the behavioral results, this result is in contrast to prior expectations.

11.3.4 Relationship between classification frequency and memory performance

Next, the relationship between the frequency of classifier predictions during the two resting states and memory performance at Test1 and Test2 were investigated. For every participant, a correlation coefficient was calculated between item-wise classifier prediction frequency during the resting states and item-wise memory performance, then the fisher-z-transformed coefficients across participants were tested against zero with one-sided, one-sample t-test. Results are shown in Figure 11.5. Negative correlation coefficients indicate increased replay for items that were subsequently remembered better (more replay, less distance error). In a first step, negative and neutral items were considered separately.

For negative items, correlation coefficients were consistently negative when clas-

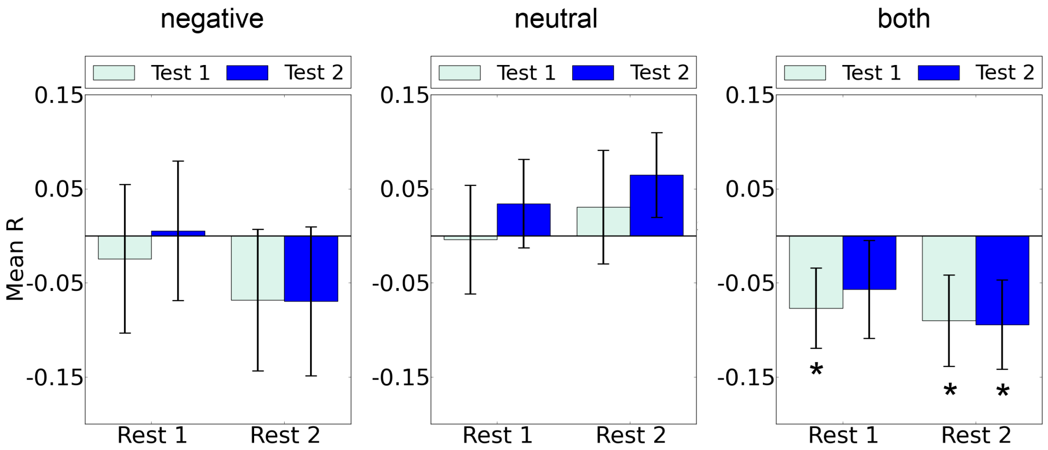

Figure 11.6: Association between classifier prediction frequencies during the two resting periods and reaction time during Test1 and Test2 for negative, neutral or both items. For every participant, a correlation coefficient was calculated between item-wise classifier prediction frequency and item-wise reaction time. Mean correlation coefficients are shown for the different conditions. Error bars indicate standard error of the mean (SEM). Across participants, Fisher-z-transformed correlation coefficients were tested against zero with a one-sided, one-sample t-test. Stars mark consistently negative correlation coefficients ($p_{uncorr} < 0.05$).

.

sifier predictions frequencies during Rest2 were correlated with memory performance during Test2 ($t_{19} = -2.126$, $p_{uncorr} = 0.023$). This is consistent with what was observed in the first study: reoccurrence during task-subsequent rest is associated with memory performance in a later test. No such association was observed for memory performance in Test1 ($t_{19} = -0.535$, $p_{uncorr} = 0.299$). There was also no association between classifier predictions during Rest1 and memory performance (Test1: $t_{19} = -0.178$, $p_{uncorr} = 0.430$; Test2: $t_{19} = -1.600$, $p_{uncorr} = 0.063$).

For neutral items, correlation coefficients were also consistently negative when classifier prediction frequencies during Rest2 were correlated with memory performance during Test2 ($t_{19} = -1.785$, $p_{uncorr} = 0.045$). Again, there was no association for memory performance in Test1 ($t_{19} = -1.501$, $p_{uncorr} = 0.075$). However, for neutral items there was an association between memory performance at Test2 and classifier predictions during Rest1, in which no replay is possible ($t_{19} = -2.109$, $p_{uncorr} = 0.024$). This is a puzzling finding that cannot be readily explained. There was no association between classifier prediction frequencies in Rest1 and memory performance at Test1 ($t_{19} = -1.258$, $p_{uncorr} = 0.112$).

When negative and neutral items were collapsed, a similar picture emerged: There was a consistently negative association between classifier prediction frequencies during Rest2 and memory performance at Test2 ($t_{19} = -2.627$, $p_{uncorr} = 0.008$; this survives Bonferroni correction for 4 comparisons) but not Test1 ($t_{19} = -1.187$, $p_{uncorr} = 0.125$). However, there was again a consistently negative association between classifier prediction frequencies during for Rest1 and memory performance at Test2 ($t_{19} = -2.368$, $p_{uncorr} = 0.014$), even though this does not survive the Bonferroni correction for 4 comparisons. There was no association between classifier prediction frequencies in Rest1 and memory performance at Test1 ($t_{19} = -0.578$, $p_{uncorr} = 0.285$).

When relating classifier predictions frequencies to reaction times, there was no significant association for either negative or neutral items (see Figure 11.6). Only when negative an neutral items were collapsed, there was a consistently negative association between classifier prediction frequencies during Rest2 and reaction time at Test2 ($t_{19} = -1.984$, $p_{uncorr} = 0.031$), as well as with reaction times at Test1 ($t_{19} = -1.861$, $p_{uncorr} = 0.039$). Surprisingly, there was also a consistently negative association between classifier prediction frequencies during Rest1 and reaction times at Test1 ($t_{19} = -1.810$, $p_{uncorr} = 0.043$), but not Test2 ($t_{19} = -1.077$, $p_{uncorr} = 0.147$).

11.3.5 Regional analyses

For the regional analysis, voxels from five different anatomical masks (amygdala, insula, anterior cingulate, medial prefrontal cortex and a combination of the four) were used for classifier training, cross-validation and prediction on the resting state.

Classifier accuracy: As can be seen in Figure 11.7A, classifier accuracy generally was quite low for all anatomical regions. However, across participants, accuracy was significantly greater than the chance level of $100/24 = 4.17$ in all regions except for bilateral amygdalae in a two-sided one-sample t-test (insula: $t_{19} = 6.252$, $p_{uncorr} < 0.0001$; anterior cingulate: $t_{19} = 4.083$, $p_{uncorr} = 0.0006$; medial prefrontal cortex: $t_{19} = 5.649$, $p_{uncorr} < 0.0001$; combined regions: $t_{19} = 5.911$, $p_{uncorr} < 0.0001$).

Ratio of negative items: Figure 11.7B shows the percentage of classifier predictions for emotionally negative items of all classifier predictions in the two resting states, Rest1 and Rest2. The percentage was never significantly larger than 50%,

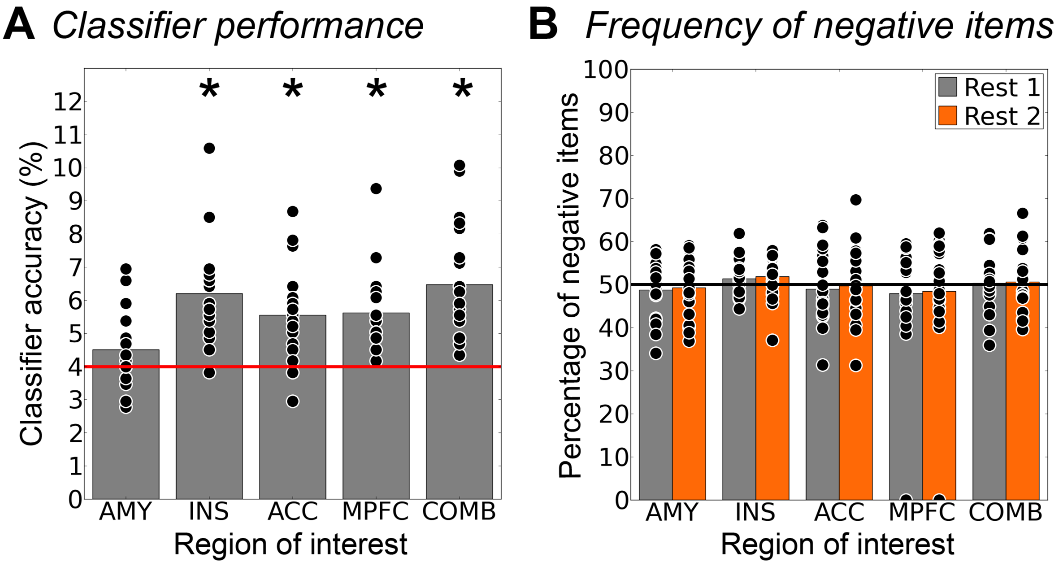

Figure 11.7: **A:** Classifier accuracy when classifiers were trained on voxels from five different anatomical masks. Stars mark accuracy levels which are better than chance across subjects ($p_{uncorr} < 0.05$). **B:** Percentage of classifier predictions for negative items across regions of interest. The percentage is never larger than what would be expected if the classifier made predictions randomly. Abbreviations: AMY=amygdala, INS=insula, ACC=anterior cingulate cortex, MPFC=medial prefrontal cortex.

.

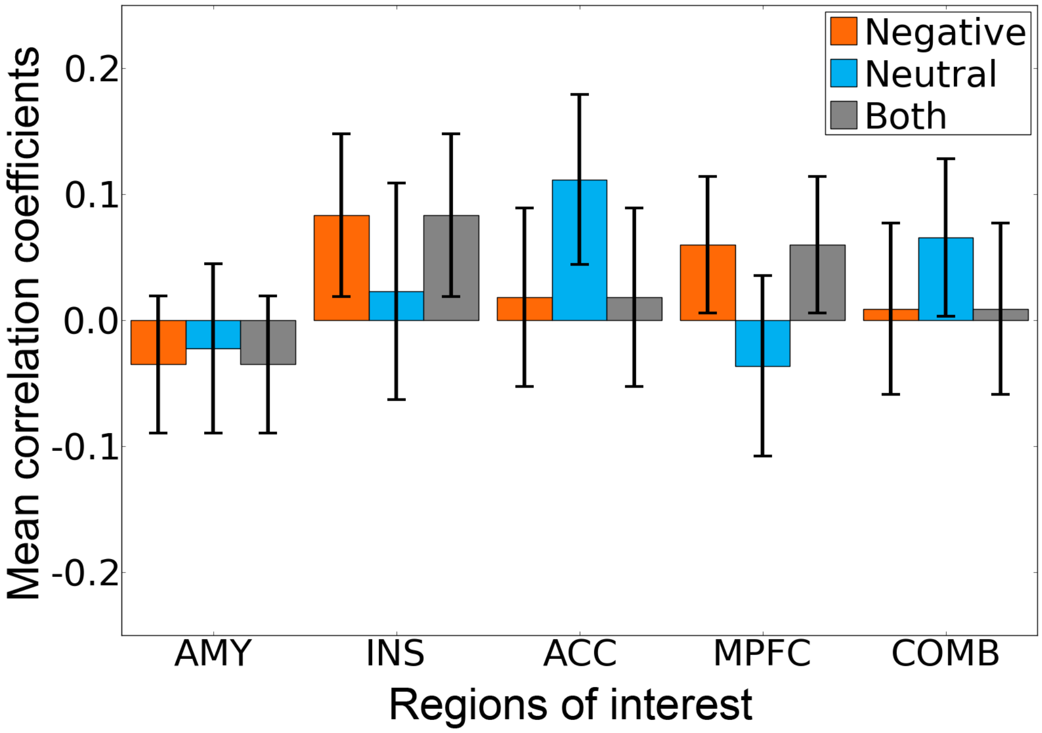

Figure 11.8: **A:** Mean correlation coefficients and SEM for the association between classifier prediction frequencies during Rest2 and memory performance (distance error) at Test2 for negative items, neutral items and both types combined when classifiers were trained on voxels from selected ROIs. Correlation coefficients were not consistently negative in any condition. Abbreviations: AMY=amygdala, INS=insula, ACC=anterior cingulate cortex, MPFC=medial prefrontal cortex.

.

which would be expected under random classifier predictions (tested with one-sided, one-sample t-tests). Also, the percentage never differed significantly between Rest1 and Rest2.

Relationship with memory performance: The association between classifier prediction frequencies in Rest2 and memory performance (distance error) at Test2 is depicted in Figure 11.8. This particular combination (Rest2, Test2) was selected because it is the condition with the strongest prior hypothesis and because it was the most consistently significant association in the regular classifier approach. Again, a correlation coefficient was calculated between item-wise classifier prediction frequency and item-wise memory performance. Negative correlation coefficients would indicate increased replay for items that were subsequently remembered better. Thus, correlation coefficients were tested against zero across participants in one-sided one-sample t-tests. This was done for negative items, neutral items and both item types combined as well as for the five different anatomical masks. None of these statistical tests yielded significant results, not even uncorrected.

11.3.6 Cross-participant analyses

One possible explanation for the mixed results so far – especially with regard to the lack of behavioral findings and missing evidence for increased reactivation of emotionally negative stimuli – would be that any effects might have been masked by inter-individual differences. For example, only some of the participants might have had better memory for negative items, while the others showed an effect in the opposite direction. In this case, even if there was more reactivation for the stimulus-class which was remembered better, the effect would not be detected with the analyses performed so far.

Therefore, an exploratory cross-participants analysis was carried out. A Spearman correlation was calculated across participants between the average memory performance for negative items and the ratio of classifier predictions for negative items to all predictions. Again, four combinations are possible between the two memory tests, Test1 and Test2, and the two resting states, Rest1 and Rest2. However, none of the four correlations across participants proved to be significantly different from zero ($p_{uncorr} < 0.05$). The same was done for the average memory of neutral items. Again, this analysis did not yield significant results.

11.4 Discussion

In this study, the reactivation of neuronal patterns for emotionally neutral versus emotionally negative stimuli during two resting states (before and after a memory task) was investigated and related to memory performance at two different time points, one immediately after encoding, the other after 30 minutes of resting state.

The analysis was guided by the following three questions, which were also investigated in the first study presented in this thesis: First, can a pattern classification algorithm reliably distinguish between different stimuli? Second, during a resting state after encoding, does the trained algorithm detect more of some neuronal patterns than of others? Third, is the item-wise frequency of classifier predictions during a resting state associated with memory performance at a subsequent memory test?

11.4.1 Classifier performance

The first question can be answered affirmatively: Classification accuracy was very high – even higher than in the first study – and was sufficient for use of the classifier in the resting state in all participants that were scanned.

11.4.2 Ratio of predictions for negative items

The other two questions are more difficult to answer. One of the main problems in this study was that the behavioral effects were in contrast to what was expected based on the literature. There is no evidence that negative and neutral items are remembered more or less accurately than the other. There is a slight effect in reaction times, but even this is in the opposite direction (negative items being responded to more slowly). Reasons for this discrepancy will be discussed below.

The initial idea in this study was that a difference in memory performance for two sets of stimuli (i.e. negative and neutral stimuli) should be reflected in a differential reactivation of associated patterns during a resting state after learning, but not prior to learning. As there was no reliable difference in memory performance between the two stimulus sets, the hypotheses for classifier predictions during the resting states became inappropriate.

The ratio of predictions for negative items to all items was significantly *below* 0.5 in both resting states, so there were more predictions for neutral items. Given that behavioral data show a trend for neutral items to be retrieved faster, this might be consistent with the idea that the set with superior memory will have more classifier predictions. However, this bias for neutral items is already there in the resting state before the encoding and it does not increase from Rest1 to Rest2. Taken together, the ratio of predictions for negative items to all predictions is inconclusive and flawed by the lack of a clear behavioral effect.

11.4.3 Relationship between classifier predictions and behavior

Independent of memory performance for a set of negative versus a set of neutral stimuli, memory performance for individual items can still be related to classifier prediction frequencies during the two resting states. Again, an association during Rest1 was not expected, because stimuli had not been encoded in that phase, and neither was an association with memory performance in the first test, at which time items had not been consolidated.

This pattern was found for negative items only, for which correlation coefficients were consistently negative only for the combination of Test2 and Rest2 (even though this does not survive correction for multiple comparisons). For neutral items, there was also an association between memory performance at Test2 and classifier prediction frequencies during Rest2, but the same was true for Rest1, which is a puzzling result and will be discussed below. When both negative and neutral items were collapsed, there was again a significant association between memory performance at Test2 and classifier prediction frequencies at Rest2 (the only significant result which survives correction for four comparisons). Here, too, there was an association between Rest1 frequencies and Test2 performance.

In part, these results replicate findings from the first study in the sense that “replay frequency” is associated with later memory performance. However, the fact that this pattern or a trend towards it can also be observed in resting states during which replay per definition is not possible begs the question whether this finding can solely be attributed to replay.

11.4.4 Regional analyses

There is a rich body of research concerned with the brain areas that are involved in the processing and retention of emotional stimuli. Here, four of the most prominent were looked at by using anatomical maps: amygdala, insula, anterior cingulate and medial prefrontal cortex, as well as a combination of the four. For each map, a different classifier was trained only on voxels included in the map. Even though this approach was more exploratory than the others, one idea was that the processing and replay of emotionally negative stimuli might be more pronounced in these areas.

Notably, classifiers performed better than chance for all anatomical maps except bilateral amygdalae during cross-validation. Even though classifier accuracy was much lower than what was achieved with the conventional approach, this finding is still interesting. With the conventional approach, which selects those voxels with good discriminability between classes, it is for the major part voxels from the occipital lobe which are almost exclusively used for classifier training. With the regional approach it could be shown that regions outside of visual areas contain information about the stimuli used. The amygdala was the only region in which classifier accuracy was not better than chance level. This might be due to the low number of voxels that were selected from each participant for this classification analysis: 40.15 ± 4.82 mean±std voxels were selected for this anatomical mask, which compares to 703.3 ± 51.9 mean±std voxels for the medial prefrontal cortex, the largest region.

However, classifier results with regard to reactivation proved to be inconclusive for the regional approach. Neither a preference for negative items during Rest2 nor a consistently negative association between item-wise memory performance and replay frequency during Rest2 was observed in any of the regions.

One obvious explanation for this lack of findings is that classifier accuracy simply was not sufficient to detect anything during the resting state with any degree of certainty. In fact, with classifier accuracies this low, participants would have been excluded in the regular approach. Another explanation might be that even if the regions are involved in the encoding and retrieval of emotional memory traces, they might simply not be involved in the *consolidation* of these traces.

The hippocampus was not investigated during this regional analysis on purpose, even though it is a region clearly implicated in consolidation of memories (McGaugh, 2000). Besides not being mentioned in the literature as one of the main regions of interest (Phan et al., 2002; Phan et al., 2004), the hippocampus is not specifically involved in consolidation of emotional memories. It has been suggested that emotion-related hippocampal activity is modulated by the amygdala (McGaugh, 2004). This would have made any effects detected with a hippocampal mask very hard to interpret.

11.4.5 Lack of behavioral effects

One of the most unexpected results in this study was that there was no memory advantage for negative stimuli. If there was an effect, it was in the opposite direction: there was a trend for increased forgetting of negative item position and response time was slower for negative items, even though these effects were weak.

There may be several reasons for this result. First, as mentioned in the introduction, good memory for individual items does not necessarily imply good memory for the association between these items (Mather, 2007). Even though efforts were made to include the associated position of the picture into the stimulus by showing the picture at the position, this experimental manipulation might not have been enough.

Another reason might be found in the short time between learning and testing. Some studies have found that memory for arousing stimuli is low in immediate recall and improved in later recall across days and weeks (Kleinsmith and Kaplan, 1963; Sharot and Phelps, 2004). A retention interval of half an hour might not have been enough for the memory benefit for emotionally negative items to develop.

Lastly, the actual effect of images on participants was not assessed. While it is reasonable to assume that pictures of mutilated bodies are experienced as aversive by almost everyone, different pictures might cause different degrees of negative affect across participants. Assessing individually experienced emotional arousal, for example with skin conductance recording or via self-report in a subsequent questionnaire, and including this metric in statistical analyses as a covariate could help to understand behavioral effects better in future studies.

11.4.6 Reasons for apparent preplay

Another finding that is difficult to interpret is that there was not only an association between memory performance and replay frequency of individual items in a resting state *after* the memory task, but that such an association, albeit not as strong, could also be observed in the resting state *prior* to the memory task.

A similar phenomenon has also been described as “preplay” in rodents (Diba and Buzsáki, 2007; Dragoi and Tonegawa, 2010). Sometimes, a place-cell sequence spontaneously occurs in resting state prior to the task in which this sequence happens. While it should be considered that the animals in these studies are highly trained to perform the kind of maze task which is required and have potentially been exposed to similar tasks before, it might also hint at a different mechanism. Resting state might not only consolidate previous experience but it might also “set the stage” for future experience. Maybe better memory results are not only a sign of enhanced consolidation, but also reflect how well a given stimulus fits into pre-configured neuronal layouts.

Another explanation might be connected to the kind of activity which is usually recorded during resting state. An intriguing suggestion has been that resting state activity ("default mode network (DMN)" activity) displays regional activity overlaps with and is closely related to self-referential processes and that a sense of self results from rest-stimulus interactions (Northoff, 2011; Qin and Northoff, 2011). It seems quite plausible that participants, during the 30 minutes resting state, were involved in thoughts about themselves, e.g. things that happened prior to their arrival at the scanning facility or plans for the evening or next day – in essence, that they had self-related thoughts. It is also plausible to assume that some of the stimuli that participants saw between the two resting states had more relevance to their self than others. These might very well have been the stimuli that were subsequently remembered best. Stimuli with more relevance to participants' selves would also recruit more of the self-referential processing related brain activity that might have been present during the resting states. Pattern classification algorithms could have picked up on that connection during the first resting state. Also, such connections to self-related processing might be particularly relevant for emotional stimuli.

Even if a connection between resting-state activity and self-referential processing cannot be substantiated or discarded with the present dataset, one should consider that brain activity during resting state is not random and that stimuli that are presented to participants are not presented to "blank slates". Participants perceive everything as the person they are, which is the same person that is scanned during a resting state. For future studies, especially if they employ emotional stimuli, it would be interesting to use questionnaires to assess the degree of self-relatedness for individual stimuli and to determine whether such a self-relatedness might even have been actively constructed in order to memorize a stimulus better.

11.4.7 Outlook

Several points have already been mentioned which likely have contributed to the mixed results in this study. The overarching mistake might have been the strict adherence to the design of the first study. While, as explained above, this design is the result of extensive piloting and works very well for pattern classification issues, the design might have been less than adequate for investigating emotional memory.

If one has to adhere to repeated stimulus presentation in an emotional memory paradigm (in order to get enough training data for the pattern classifier), it might be a good idea to make the emotional stimuli as complex as possible. Short video clips with emotionally neutral as compared to emotionally negative story lines or key elements might be an exciting alternative to mere picture stimuli. Depending on the context, the same video might even be considered neutral or negative. Also, clips could be filmed and cut in a way that certain scenes are present both in an emotionally negative and an emotionally neutral clip (e.g., the same start, different ending). Even across repeated presentation, the clips might remain engaging. Various elements of the environment, storyline and sequence of events could be tested afterwards, so that a “memory quotient” for each clip could be assessed.

Regardless of the stimuli that might be used in a follow-up study, skin conductance recording would be useful to assess arousal during encoding. Memory testing should definitely take place not only on the day of learning, but several days after to more fully assess memory consolidation for emotionally negative stimuli

12 Replay of stimulus-specific activity in intracranial EEG

12.1 Introduction

The last of the three studies in this thesis again uses a very similar paradigm to the first two studies. In fact, it is nearly identical to the first study, and the general theoretical motivations are the same. Therefore, the reader kindly refer to section 9 and 10.1 for more information on the goals of this study.

The major difference to the first study is of a methodological nature, as this last study was not performed with fMRI in healthy participants but with intracranial EEG in epilepsy patients at the Clinic for Epileptology in Bonn, who received presurgical implantation of intracranial electrodes for the diagnostic purpose of clarifying their epileptic foci.

This third study complements the previous two studies because the data allow analysis of processes that happen at timescales much faster than those that can be investigated with fMRI. This permits the investigation of the role of different frequencies in memory consolidation, which might have a differential contribution to encoding. Changes in alpha and theta frequency power have been found to be related to memory performance (Klimesch, 1999). In a different study, theta and gamma increases (Osipova et al., 2006) were associated with memory performance. Power in these frequency bands might be a valuable feature for accurate classifier predictions and might provide more information than mere amplitude. Therefore, classifiers were not only trained on raw amplitude values, but also on time-frequency-decomposed values.

12.2 Material and Methods

12.2.1 Participants

12 patients were included in this study (4 female, age 34.5 ± 10.6 mean±std years, range 20-57 years). All patients suffered from pharmaco-resistant focal epilepsy and were considered for surgical treatment. Presurgical intracranial recording was medically indicated and recording sites were selected by the attending epileptologist.

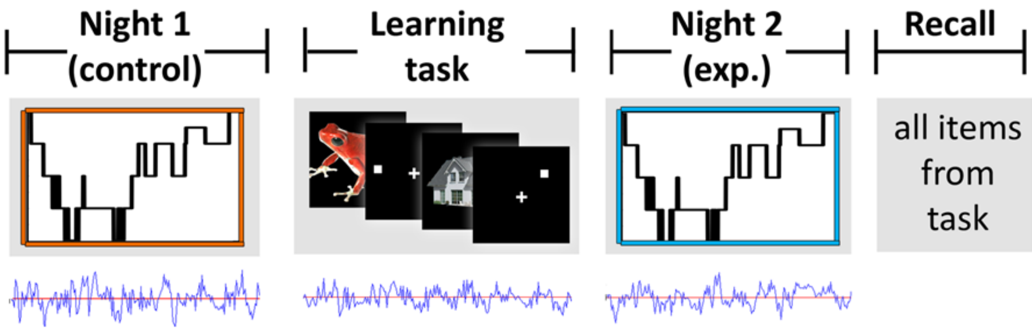

Figure 12.1: Overview of the study. Intracranial EEG was recorded during two complete nights of sleep, one preceding and one following the memory task. The memory task again was a declarative assocative memory task in which 16 object-place associations had to be learned in the course of repeated presentation of the pairs (30 times). The memory task was performed in the evening before the second night. On the morning after, memory for all 16 object-place pairs was tested.

Thus, the position were not chosen according to hypotheses, and electrode positions were different in all patients.

When patients were asked if they wanted to take part in a scientific study, they were given information of the structure of the experiment and the task that they would be performing. It was made clear to them that participation was voluntary and that there was no medical gain for them if they participated. Also, they were informed that they could decline or abort participation without any disadvantage to them and that their stay at the ward would in no case be prolonged by their participation. Also, they were made aware that the data recorded in the studies would be kept and that demographic information from their patient files such as age and gender as well as location of their electrodes would be used in a pseudonomized fashion. Written informed consent was obtained from all of the patients in accordance with guidelines of the local ethics committee.

Patients were recruited from 2010 until 2013. This long time window of data collection is due to the circumstance that the opportunity to record in these patients is quite rare.

12.2.2 Paradigm and Stimuli

An overview of the paradigm is given in Figure 12.1. Participants underwent a control night, during which brain activity was recorded for the entire night. This served as a baseline condition for later classification results. Then, the patient performed the experimental paradigm in the evening of the following day (except for one patient who, due to medical complications, performed the experiment one week after the control night). The experimental paradigm was performed between 8pm and 9.30pm depending on availability of the patient. After the experiment, another complete night of sleep was recorded in the patient. A memory test was performed on the morning after this second night.

As in the previous studies, the paradigm was an object-place association memory task. Sixteen different objects were presented, which were identical to the 16 objects from Set1 of the first study. Every object was associated with a specific location on the screen that was marked by a white square. Every trial consisted of the presentation of the object for 1s, then presentation of the white square for 1s, then a fixation cross for 3s until the next trials started. Patients were asked to indicate via button-press whether they liked or did not like the presented object. Again, this was done only to encourage deeper level of processing and was not analysed further.

Each of the 16 object-place pairings was presented 30 times, resulting in 480 trials that were evenly distributed across five blocks. After each block, there was a countdown of 60s, after which the patient could press a button to proceed with the experiment whenever he or she felt ready. The experiment lasted approximately 50 minutes. During the memory test that was conducted on the morning after the learning task, each of the 16 different objects was presented and patients had to indicate the location of the white square that was associated with the stimulus during learning.

In the first 7 of the 12 patients, a slightly different version of the experiment was performed. The 16 stimuli belonged to four categories: There were 4 houses, 4 faces, 4 landscapes and 4 tools. In addition, the position of the white square was discrete instead of continuous: The center of the screen was divided into a 4x4 grid that was visible for the patient. The white square was in the center of one of the 16 tiles of this grid. Accordingly, the grid was shown during recall and only one of the 16 discrete grid-tiles could be selected as belonging to a stimulus.

This earlier version of the paradigm corresponded to a version of the paradigm that was used during piloting for the first study described in this thesis. When it became apparent during this fMRI piloting that the use of different stimuli and continuous positions was superior to the previous version, this improved version was adapted for the intracranial EEG study as well. As noted above, the possibility to record in epileptic patients is rare and it takes long periods of time to collect the data. Thus, the existing recordings were deemed too valuable to discard altogether. Also, the changes in the paradigm are relatively minor. The earlier version of the paradigm will be referred to as the "category" version, while the later version will be called the "individual" version because the two versions are mostly different with regard to the type of their stimuli.

The paradigm was presented to patients using the software Presentation (http://www.neurobs.com) on a notebook computer with 15.4 inches diagonal screen size and a resolution of 800x600dpi. Responses were logged with the mousepad of the notebook computer.

12.2.3 Recording and initial filtering of intracranial EEG data

Intracranial EEG recordings were referenced to linked mastoids, recorded at a sampling rate of 1000 Hz, and band-pass filtered (0.5305164 Hz [12 dB/octave] to 125 Hz [12 dB/octave], including a notch filter at 50Hz). In addition to the intracranial electrodes, which were implanted according to medical necessity, regular EEG was recorded from the following electrode positions: T5, T6, C3, C4, Cz and Oz – according to the International 10-20 system. In addition, two ECG, two EOG and two two EMG electrodes were recorded from. Apart from sleep-staging, data from these external electrodes were not used in the analyses.

12.2.4 Automated artifact correction

One of the most prominent concerns with intracranial EEG data recorded from epileptic patients is that the data could contain epileptic activity. Obviously, testing would have been aborted if patients had had a seizure during the experiment, which did, however, not happen in any of the patients reported here. But even between seizures (inter-ictal), the EEG of epileptic patients may contain epileptic forms and other abnormal EEG characteristics.

Usually, visual inspection is employed to find artifacts in the episodes of interest. However, doing manual artifact rejection sometimes was not a viable option in this study due to the long duration of recording (i.e., during the two nights). Thus, a computer algorithm was employed to automatically find episodes with artifacts.

The algorithm was developed by Thorsten Kranz and is available online as part of a Python package for analyzing EEG data (https://github.com/thorstenkranz/eegpy). In short, the algorithm is set up to detect parts of the signal in which either the amplitude was too high or the gradient too steep as compared to "normal" parts of the data. Thus, the standard deviation of amplitude across all episodes of interest is calculated as well as the standard deviation of the first derivative of the signal. The first derivative reflects steepness of the slope of a tangent at each point of the signal. In this context, it is simply calculated as the difference between one time-point and the preceding time-point.

To account for the fact that EEG differs between persons, the two standard deviation values are calculated for each participant individually. They are also calculated separately for each electrode, or channel. This is sensible because some channels are more noisy than others and calculating the standard deviation across all, possibly very dissimilar channels, would increase false detection of artifacts in the most noisy channels. Based on the standard deviation, a cut-off is then determined for each participant and each channel: It is derived by multiplying the standard deviation of amplitude and first derivative each with a certain factor. For example, if the standard deviation for EEG amplitude in channel TL09 in participant A was found to be $15\mu V$, the cut-off for this channel could be set to three times the standard deviation ($45\mu V$) or six times the standard deviation ($90\mu V$), depending on how strict or liberal the cut-off is meant to be.

If the signal in any channel at any given point in the epoch of interest exceeds this individualized cut-off, the episode is rejected for containing an artifact. Note that this is quite a conservative approach – if there is an artifact in only one of the channels, the complete epoch will still be rejected. Figure 12.2 shows the idea of the artifact correction in more detail.

In summary, epochs were rejected if either amplitude or slope exceeded a cut-off value based on what was found in "normal data" (which was reflected by the standard deviation of both amplitude and slope) and a mulitplication factor that could be freely chosen.

After systematically varying the factor with which the standard deviation was

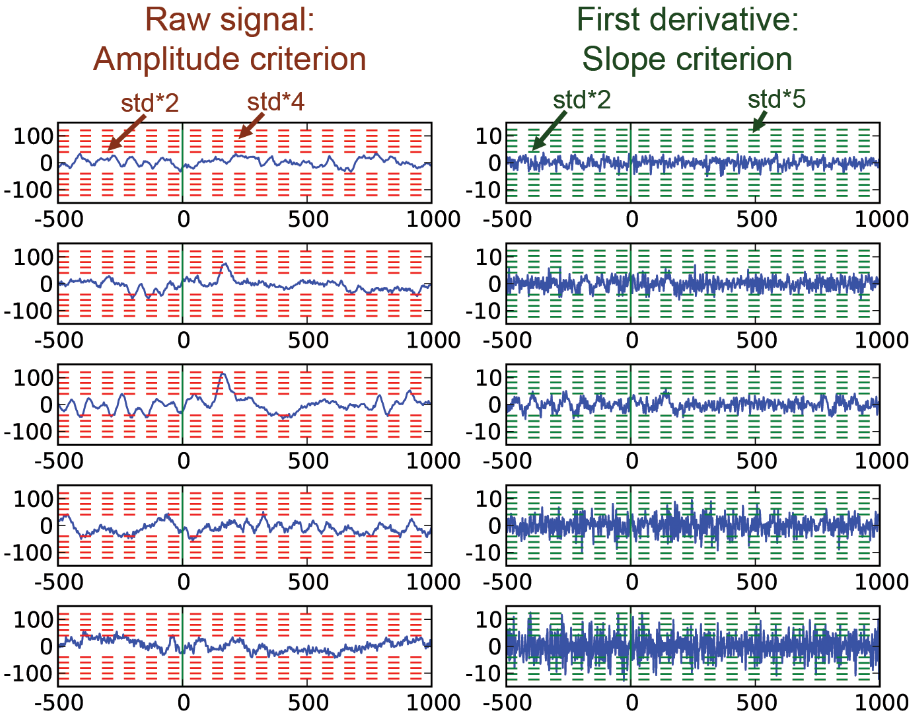

Figure 12.2: Overview of the automated artifact rejection procedure. On the left side, five epochs are presented which correspond to EEG recordings of individual trials in one electrode. The blue line is the raw signal. The red dashed lines mark different cut-off values corresponding to multiples of the amplitude standard deviation (std), which was calculated in the same channel across all trials. It provides an indication which amplitude values can normally be expected in this particular electrode. The first epoch does not exceed any of these standard deviation based thresholds. The second epoch would be rejected if three times the standard deviation was the criterion. The third epoch exceeds even a threshold based on 7 times the standard deviation. On the right side of the figure, the first derivative of every left-side epoch is shown as blue line, indicating the steepness of the slope in the raw signal. The dashed green line marks the standard deviation of the first derivative that was computed for each channel across all trials. While the fourth epoch would have passed a liberal *amplitude* criterion, its first derivative exceeds five times the standard deviation of the *slope* and would have been rejected.

.

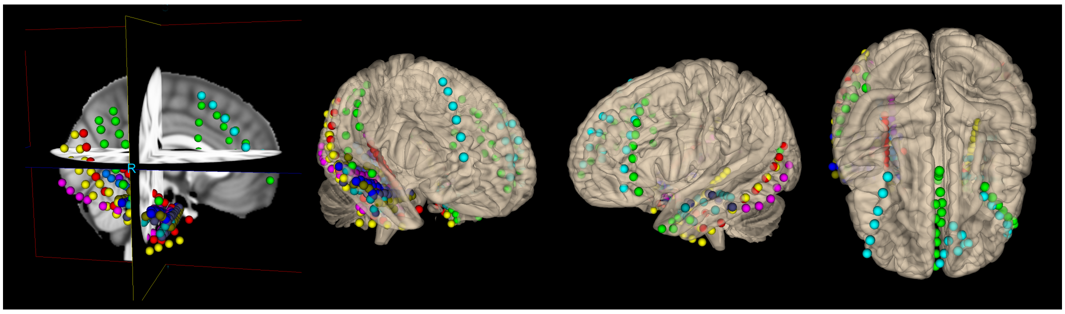

Figure 12.3: Overview of all electrodes which were used for further analyses. Every color denotes a different patient. Electrodes were excluded if they contained an epileptic focus, appeared noisy in visual inspection or if more than 25% of epochs contained artifacts.

.

multiplied, it was determined that a factor of 8 for amplitude and 8.5 for the first derivative was best suited for the data recorded in this study. This is, of course, a somewhat arbitrary decision. In the current dataset, an average of 86.86% ± 8.27% mean±std of all trials during the memory task were retained with this cut-off criterion. As such, it satisfied both the need for good data quality and the necessity of retaining enough trials. The automated artifact rejection algorithm was performed with custom code for Python which is implemented in the eegpy-package by Thorsten Kranz (https://github.com/thorstenkranz/eegpy).

12.2.5 Selection of electrodes

In principle, all intracranial electrodes were eligible for use in the classification approach. However, electrodes were excluded in the following three steps to guarantee good data quality:

1. Electrodes were excluded if the patients' medical report stated that they were located in an epileptic focus or were involved early in seizure onset. This medical report is written by the Clinic for Epileptology's medical doctors as a final report on the diagnostic results of the intracranial recording. It informs any surgical tissue removal that might take place subsequent to the implantation of the intracranial electrodes and can therefore be considered a very trustworthy source of information.

2. Electrodes were excluded when visual inspection of their activity during the

Band no.	frequency range	referred to as...
Band 1	4-8 Hertz	theta
Band 2	8-12 Hertz	alpha
Band 3	12-20 Hertz	low beta
Band 4	20-30 Hertz	high beta
Band 5	30-60 Hertz	low gamma
Band 6	60-90 Hertz	gamma
Band 7	90-125 Hertz	high gamma

Table 1: Overview of different frequency bands

paradigm revealed abnormalities such as excessive spiking, high amplitudes, flat signal or noise.

3. Electrodes were rejected if an automated artifact detection algorithm (see above) found artifacts in a significant number of trials ($> 25\%$ of all trials) during the paradigm.

In the remaining channels (see Figure 12.3), a second artifact detection run was performed and all trials with artifacts were not included in further analyses.

12.2.6 Frequency band decomposition

In the first approach of this study, the feature selection and classifier training was done on raw EEG amplitude values in the time window $0 - 1000ms$ after stimulus onset (i.e., a $480x1000x10$ dataset resulted in a patient with 10 electrodes, 480 being the number of trials). This is an intuitive approach as it involves little change to the data and will be termed “unfiltered” from now on, even though it should be noted that preprocessing included band-pass filtering for very low and very high frequencies and a notch filter at 50 Hertz. “Unfiltered” is supposed to highlight the contrast to the other approach that was taken in this study.

As stated in the introduction, one of the most exciting possibilities in an electrophysiological dataset is to look at the contributions of different frequency bands. Therefore, in the second approach, the feature selection and classifier training was performed on data that had been broken down into different frequency bands.

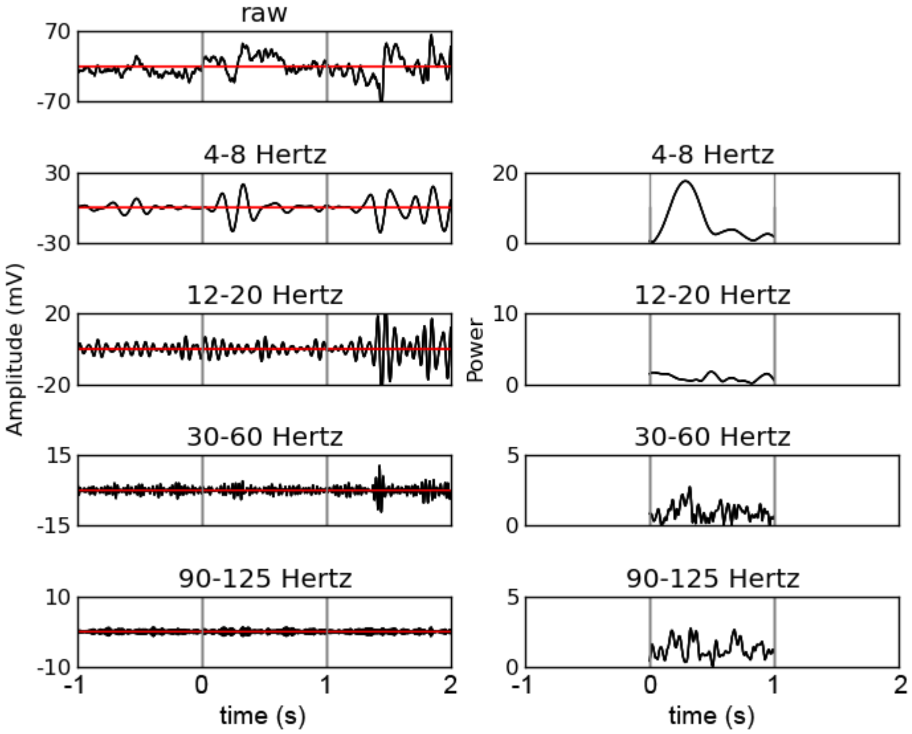

Figure 12.4: Example for frequency decomposition method. Top left shows the original epoch (inside the two vertical lines) that has been buffered on both sides with the $1000ms$ preceding and following the epoch. On the left below is the same epoch after it has been filtered into specific frequency bands. On the right side, for each frequency band, the power values from the hilbert transform are plotted. Note that the transform was performed on the extended $3000ms$ epoch, then the inner $1000ms$ are cut out to avoid edge artifacts.

.

For this, every epoch was first band-pass filtered into seven different frequency bands with a butterworth filter. See Table 1 for detailed information on the bands. This step led to a dataset with shape $480x1000x10x7$ for a patient with 10 electrodes. Then, a Hilbert transformation was applied to each epoch in each filtered band and electrode, and the absolute of the resulting complex number was taken. This corresponds to the power that a specific frequency band has at a specific point in time. After this step, the dataset again had the shape $480x1000x10x7$.

Edge artifacts are to be expected in this filtering regime. The hilbert transform returns bad results for both ends of a time series. Therefore, a buffering approach was used: for each $1000ms$ epoch, an additional $1000ms$ was included before and after the epoch (i.e. from $-1000ms$ until $2000ms$ with regard to stimulus onset, see Figure

12.4). The filtering was then performed on the extended 3000ms epoch. After the filtering was completed, the additional 1000ms on both sides were excluded again, removing any edge artifacts from the inner 1000ms epoch. Figure 12.4 shows an example of this frequency decomposition. The filtering and frequency decomposition was again performed using Thorsten Kranz's eegpy-package (https://github.com/thorstenkranz/eegpy).

12.2.7 Feature selection

In the first two studies, pattern classifiers were trained on fMRI data. Feature selection for these datasets was relatively easy. Voxels served as features and they were selected based on voxel-wise ANOVAs.

With electrophysiological data, the feature selection process is more complex. The following problems have to be considered.

1. The number of possible features is higher: There are different electrodes, different time-points during the course of a trial and – in the second approach – different frequency bands. With 1000 timepoints, 7 frequency bands and 10 electrodes, this already leads to 70000 features, and many patients have a lot more than 10 electrodes.

2. Neighboring time points are not independent of one another: If one time-point got selected based on any criterion, the neighboring time-point would likely get selected as well (to a lesser degree, this is also true of fMRI voxels). This leads to cluster-like feature selection.

3. Even after artifact rejection, there will still be some peculiarities in the signal that might lead to artificially high F-values in an ANOVA-based approach.

4. Some electrodes might systematically contain more time points with high F-values due to differences in signal quality even though they do not differentiate well between the classes. If a uniform cut-off criterion is taken for all electrodes, there might be a disproportional amount of selected features from a few "bad" electrodes.

Taken together, there is a high probability that clusters of features might be selected which contain little valuable information for a pattern classification algorithm with regard to the different classes.

Which parts of the signal should then be selected? How can one reliably distinguish between “real” clusters of features and those that are caused solely by signal disturbances? One approach in dealing with time-frequency data has been described by Maris and Oostenveldt (Maris and Oostenveld, 2007). The basic idea is to find clusters of significant signal differences and compare the cluster size in the real data to cluster sizes found in shuffled data. Only those clusters that exceed cluster sizes found in shuffled data are then retained.

Accordingly, in every electrode and every frequency band, an one-way ANOVA was performed on every time-point with the 16 different classes as group variable. For every electrode and frequency band, this resulted in 1000 F-values (one for each time-point). Clusters were then defined starting from the first F-value that exceeded 1.67 and ending with the last F-value that was still above this threshold.

The F-value was taken as cut-off instead of the p-value because in different patients, different numbers of trials were analyzed, for example 160 trials in patient A, 320 trials in patient B – depending on how many epochs with artifacts were removed. As a result, the F-value would have to be much higher in patient A than in patient B for the same p-value. F-values can be better compared across subjects with varying numbers of trials. As this F-value cut-off is only the first step in selecting clusters, a liberal F-value threshold of 1.67 can be well justified.

Cluster size for each cluster was determined as the sum of all F-values of the timepoints that were included in the cluster. The data on which the original cluster search was performed was shuffled with regard to the labels and ANOVAs were again performed, this time on nonsense classes. This was done 20 times for each channel and each frequency band. While sufficient for determining a cut-off (see below), this is a relatively low number of surrogates. However, increasing this number would have resulted in excessive length of computation (analyses during cross-validation lasted more than a week for *one* regular patient even with 20 surrogates). Only the maximum of all cluster sizes for each of these surrogate runs was included in a distribution of surrogate cluster sizes, further making the comparison conservative. In each channel and each frequency band, clusters found in the real data were only retained if they exceeded the 95th percentile of the surrogate cluster size distribution (with 20 repetitions, this amounted to exceeding the highest surrogate cluster size).

During initial analysis of the data with MVPA (carried out by Thorsten Kranz), it was found that the pattern classification algorithm could not reliably distinguish between classes when clusters were not restricted with this conservative surrogate-cluster approach. Thus, throwing out clusters that do not exceed cluster-size in shuffled data is a necessary step in selecting valuable features.

In summary, the following should be pointed out:

1. The cluster approach attenuates the effects of disturbances and noise in the real data: The original data are shuffled with regard to their label only while retaining the temporal structure of individual epochs. Thus, single-trial oddities in the real data are preserved in the surrogate data. Because the real data has to measure up to and exceed the surrogate data with regard to cluster size, this leads to selection of more reliably relevant clusters.

2. The surrogate cluster approach is performed separately on each electrode. Comparing the real clusters in every electrode to surrogate clusters in the same electrode prevents selection of overly many clusters in an electrode if they are caused solely by electrode-specific quirks.

3. In the second venue of analysis, the surrogate cluster approach is applied separately to every frequency band. Especially in the lower frequency bands, large clusters are easily found because the power values follow a slower drift (see Figure 12.4). This could lead to a dominance of low-frequency clusters in the dataset. With the surrogate cluster approach, the large clusters in the low frequency bands are compared to large clusters found in shuffled data in the low frequency bands and, if they do not exceed the surrogate clusters, are not included.

4. During cross-validation, the feature selection and the surrogate cluster approach were performed on the training data only (to ensure independence of the test data). That means that for every fold of the cross-validation, new features were selected with the surrogate cluster approach.

The data-points in the clusters that were identified in this surrogate approach were then used for classifier training. To keep the number of features small, the clusters were down-sampled by a factor 10, i.e. 10 subsequent data-points were averaged, starting from the first data-point in a cluster and including the mean of the remainder of the division by ten (e.g. in a cluster of 35 data-points, the mean of the last five constituted the fourth data-point after down-sampling).

12.2.8 Classifier training

A sparse multi-nomial logistic regression (Krishnapuram et al., 2005) implemented in the PyMVPA package (Hanke et al., 2009a; Hanke et al., 2009b) was used in this study. It was trained on all features from the surrogate cluster approach. The punishment term lambda was 0.1 (the default in PyMVPA). The choice of this classifier is explained in the discussion.

12.2.9 Cross-validation

For the cross-validation, the dataset was split into five parts and then balanced so that the same number of labels for every class was present in the training dataset and in the test dataset. Feature selection was performed on the training dataset. Features were selected according to the surviving clusters. For example, if electrode 2 in frequency band 3 had a surviving cluster from $202 - 355ms$, these datapoints were collected in every trial from electrode 2 in frequency band 3, amounting to 153 datapoints. Datapoints were then averaged in blocks of $10ms$ to avoid too many features, leaving the example cluster with 16 features. The same was done with the test dataset. The classifier was trained on the training dataset and made predictions on the test dataset, analogous to what was done in the previous two studies. By comparing the classifier prediction on samples of the test dataset with the actual target labels of these samples, one can calculate a measure of classifier accuracy.

12.2.10 Classifier predictions on the nights

After ascertaining classifier performance with the cross-validation approach, a classifier was trained on selected features of data from all five blocks of the memory task and made predictions on the iEEG recording during the two nights.

For this, a sliding window approach was used: A $1000ms$ window was cut-out every $100ms$ throughout the two nights. In the second ("filtered") approach, every epoch was split up into the seven frequency bands with the buffering approach described above. The same features that were used for classifier training were extracted from the the epochs of the two nights.

Obviously, the sliding window approach led to a massive amount of data that the classifier made predictions on. On average, $461,796 \pm 75,978$ (mean±std) epochs were extracted from Night1 and $392,369 \pm 27,387$ (mean±std) from Night2. It is fair to assume that very little of this data actually contains traces of neuronal replay. Careful consideration of classifier output is necessary to sort real replay from mere noise. In a first step, the automated artifact rejection described above was applied to the epochs of the two nights to exclude bad signal.

Then, the classifier made predictions on every artifact-free epoch of the two nights. The label of the prediction was obtained as well as the probability level that was associated with the prediction.

The probability level served as the main distinguishing factor between the two nights. The classifier returns a prediction on every epoch it is presented with. Thus, for both nights, a label will be returned for every epoch that was presented to the classifier. How could an increase in replay then be determined? The solution is to use the probability level of every prediction. Real replay of neuronal activity should be associated with a higher classifier confidence for the prediction. If neuronal patterns reoccur more often during the experimental than during the control night, more predictions with high confidence should be observed.

Thus, classifier probability was used as a cut-off: only predictions whose probability value exceeded a given level were retained. For systematically increasing confidence levels, the ratio of remaining predictions to all predictions becomes smaller. The drop in this ratio should be more pronounced for predictions in Night1 (during which no replay is possible) as compared to Night2 (in which some epochs are expected to contain replay). The main variable of interest is the difference in the ratios of "surviving" predictions to all predictions in Night2 as compared to Night1: $Ratio_{night2} - Ratio_{night1}$, which will be called probability difference. If this difference is positive, it means that more high confidence votes are present in Night2 as compared to Night1.

12.2.11 Surrogate classifiers

Influences from many different factors other than real replay might influence classifier confidence (e.g. the memory task is closer in time to Night2 than to Night1). This, combined with the huge amount of data, poses a statistical challenge. Even if more high confidence votes can be detected in Night2 as compared to Night1, how can one be sure that this is not due to oddities or drifts in the data?

To lend the results more credibility, another surrogate approach was chosen. Again, 20 surrogate classifiers were trained on data that was shuffled with regard to the labels, i.e. on nonsense data. The surrogate classifiers also returned predictions and probability values for every epoch of the two nights. Predictions were also thresholded based on the probability level and the ratio of surviving predictions to all predictions in Night2 was compared to that of Night1. If the probability difference found in the real data exceeds the largest probability difference found in the 20 surrogate runs, this is a strong indication that the detected replay is not merely due to distortions in the signal as these would also be picked up upon by the surrogate classifiers.

12.2.12 Sleep staging

Sleep-staging was performed for both nights according to the guidelines by Rechtschaffen and Kales (Rechtschaffen et al., 1968) by an experienced member of the Cortical Oscillations lab. Classifier predictions were then considered separately for every sleep stage.

12.2.13 Relationship with behavior

Similar to the proceedings of the first two studies, it was attempted to relate classifier predictions for the two nights with memory performance. For this, the probability difference was calculated for each of the 16 different stimuli shown. Then, the probability difference for later remembered stimuli was compared to the probability to later forgotten stimuli. In the five patients who performed the second ("individual") version of the memory test (using a continuous instead of a discrete memory measure), the memory results were dichotomized. If the distance error (the distance between the actual associated position of an item and the location given by the participant) exceeded $19mm$ (50 pixels), the stimulus was counted as forgotten. This limit corresponds to the size of the fields of the grid in the first version (one field in the grid was $100x100$ pixels, thus a maximum of 50 in each direction from the center of the field).

As the probability difference in this study is considered to operationalize the amount of replay, it was expected to be higher for subsequently remembered items than for subsequently forgotten items.

12.3 Results

12.3.1 Behavioral results

Behavioral performance could be assessed in 10 patients only. In one patient, data from the memory test were lost, and in the other, the memory test could not be performed because the patient did not feel up to it.

For the patients who performed the first ("category") version of the paradigm, a binary remembered/forgotten result was obtained for every object-place association. In six patients, 38.5 ± 33.3 mean±std percent of items were remembered (range from 6.25 to 93.7), the variance being quite large. In the patients who performed the second ("individual") version of the paradigm, in which an error distance like in the other two studies was obtained, the error distance in four patients was $36.34mm \pm 18.58mm$ mean±std (range $8.9mm$ to $54.4mm$). When memory performance was binarized (counting error distances exceeding $19mm$ as forgotten), 46.9 ± 31.1 mean±std percent of items were remembered (range 18.8 to 93.7).

12.3.2 Sleepstaging

Figure 12.5 provides an overview of the time spent in different sleep-stages in the two nights. All patients reached all sleep stages in both nights. As can be seen in Figure 12.5, the total duration of sleep recording was significantly longer in Night1 than in Night2 as confirmed by a two-sided paired t-test ($t_{11} = 3.019$, $p = 0.011$).

Looking at the individual sleep-phases, a two-way repeated-measures ANOVA with "Night1 vs. Night2" as first factor and "sleep-stage" as second factor revealed significant main effects for both factors but no interaction (factor 1: $F_{1,11} = 9.420$, $p = 0.011$; factor 2: $F_{4,44} = 36.119$, $p < 0.001$; interaction: $F_{4,44} = 1.137$, $p = 0.351$). This confirms that the duration of sleep-stages was longer in Night1 than in Night2 and that sleep-phase duration was generally different between stages. Post-hoc two-sided paired t-tests between Night1 and Night2 sleep-stage duration for individual sleep-stages revealed that only sleep-stage 1 was significantly longer in Night1 than

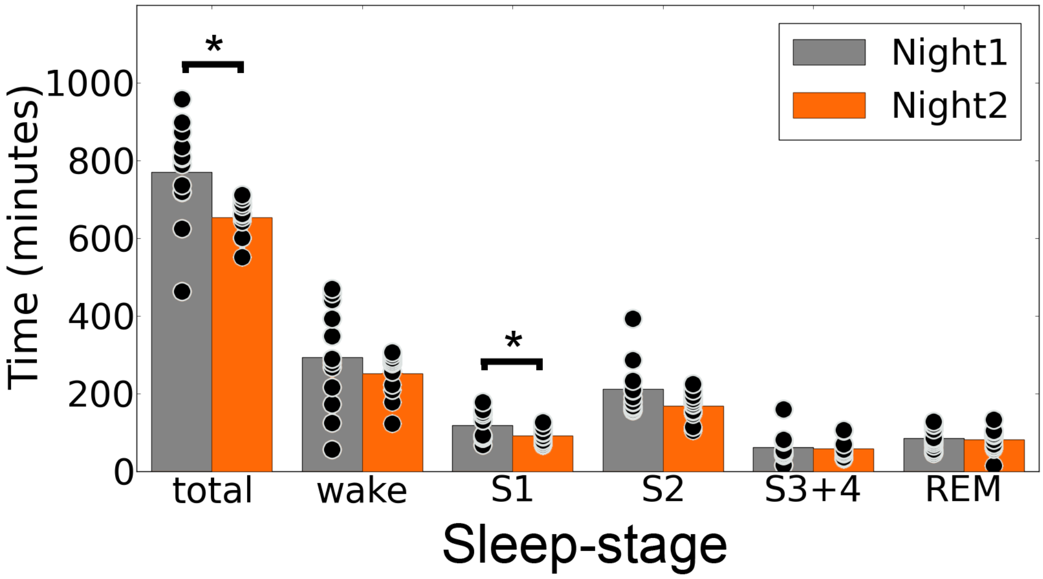

Figure 12.5: Time spent in the different sleep-stages in minutes for Night1 and Night2. All patients reached all sleep-stages in both nights. Stars denote significant differences in a two-sided paired t-test between Night1 and Night2 across patients ($p < 0.05$, uncorrected).

.

in Night2, but this does not survive Bonferroni-correction for multiple comparisons ($t_{11} = 3.105$, $p_{uncorr} = 0.010$).

The difference in duration between the two nights is likely due to the experimental proceedings. In the first night, no paradigm was performed and sleep recording started early in the evening, before the core staff left for the day (often 6pm). In the second night, the learning task was started between 8pm and 9.30pm. Only after the end of the task did recording for the rest of the night start. Also, the memory test was performed early in the morning of the next day, which might have cut recording short further.

Different length of night recordings should not be a problem for the presented analyses, because the analyses were based on ratios (e.g. “In how many of all REM sleep epochs was a high confidence prediction present?”, see below). Therefore, no further steps were taken to assimilate night recording lengths.

12.3.3 Classifier accuracy

Classifier accuracy was low compared to the first two studies in this thesis, especially if the data was filtered into seven different frequency bands (see Figure 12.6). Across

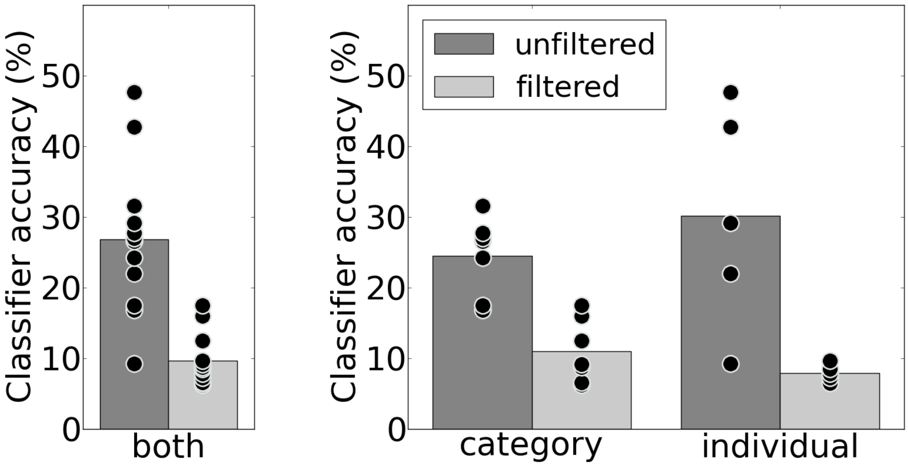

Figure 12.6: Classifier accuracy. Left: For all participants, for classifiers trained on unfiltered (dark gray) or on filtered (light gray) data. Right: Classifier performance, separately for participants who performed the first ("category") or the second ("individual") version of the paradigm, again on unfiltered and filtered data.

.

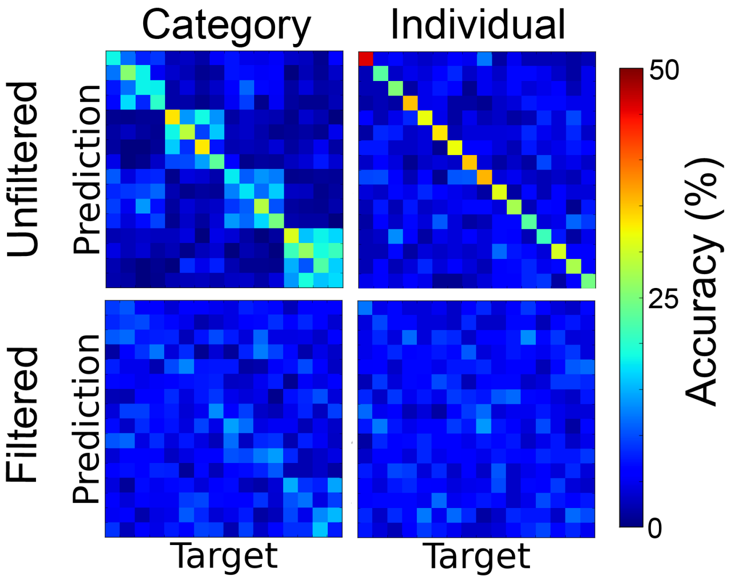

Figure 12.7: Confusion matrices for the crossvalidation runs. Plots the target (correct label) against the prediction. The diagonal thus contains correct predictions. The matrices also inform which stimuli are confused with one another. For participants who performed the "category" version, one can clearly see the category structure. Thus, objects from one category are more likely to be confused with each other.

.

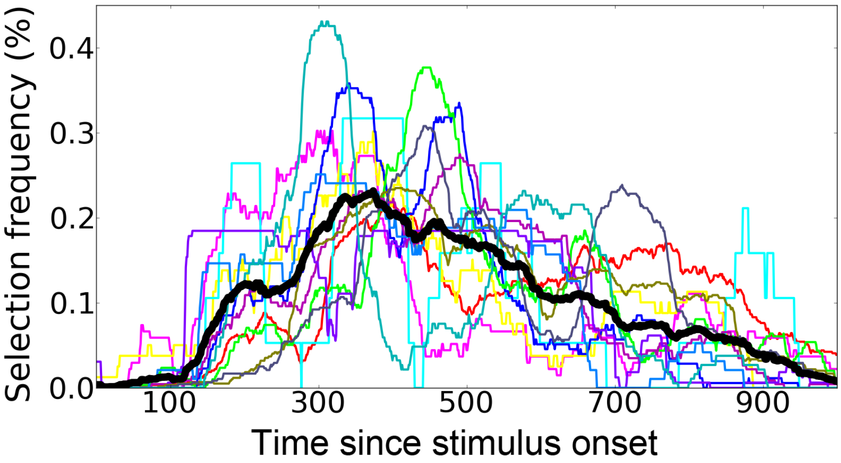

Figure 12.8: Features that were selected during the cross-validation of the unfiltered data approach. For every time point during the 1000ms epoch after stimulus onset it was determined how often the time point was selected as a feature. This amount was divided by the total number of features for each participant and multiplied by 100, yielding a percentage. Thin colored lines represent individual participants. Thick black line is the mean across participants. Note that percentage values are quite small because of the fine bins (1ms bins).

twelve participants, classifier accuracy was 26.9%±10.2% mean±std in the unfiltered data and 9.7%±3.6% (mean±std) in the filtered data, which is significantly different in a two-sided paired t-test ($t_{11} = 5.180$, $p = 0.0003$). Still, classifier performance in both approaches was significantly better than the chance level of $100/16 = 6.25\%$ as ascertained with a two-sided one-sample t-test (unfiltered: $t_{11} = 6.715$, $p_{uncorr} < 0.0001$; filtered: $t_{11} =$, $p_{uncorr} = 0.0003$).

When only those patients were considered who performed the “category” version of the paradigm, classifier accuracy was 24.5% ± 5.1% mean±std in the unfiltered data and 11.0% ± 4.1% mean±std in the filtered data. When only those patients who performed the “individual” version were considered, the classifier accuracy was 30.1% ± 14.0% mean±std in the unfiltered data and 7.9% ± 1.1% mean±std in the filtered data.

Features that were selected during cross-validation are presented in Figure 12.8 for the unfiltered approach. Here, it is evident that the time-points around 400ms are most often selected as features across participants. Features most often selected in the filtered data approach are presented in Figure 12.9. Here, it becomes evident that in different frequency bands, different phases post stimulus presentation serve as features: Earlier time-points are selected in a low frequency band, while later

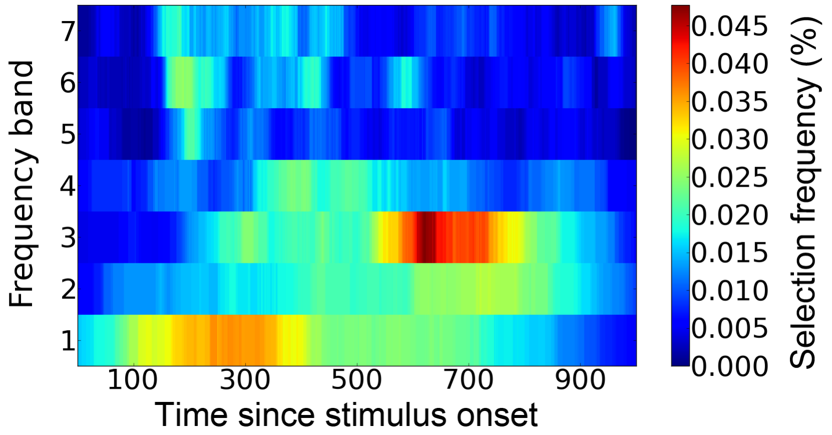

Figure 12.9: Features that were selected during the cross-validation of the filtered data approach. For every time point during the 1000ms epoch after stimulus onset and for each of the seven frequency bands, it was determined how often the data point was selected as a feature. This amount was divided by the total number of features for each participant and multiplied by 100, yielding a percentage. The figure shows the mean across twelve participants. In the lowest frequency band (4 − 8 Hertz), time-points early in processing served as features, while in a higher frequency band (12 − 10 Hertz), later time-points were selected most often.

.

time-points are selected more often in a higher frequency band.

12.3.4 Probability difference

The probability difference between Night2 and Night1 (see Methods section “Classifier predictions on the nights”) was taken in this study as a marker for neuronal replay. If this value is positive, it means that more high confidence classifier predictions were made during Night2 than during Night1. Figure 12.10 shows how this metric presented itself over varying probability thresholds between 0.1 and 0.99. As can be seen, the metric is slightly negative for the unfiltered data and is only slightly positive for the filtered data.

More importantly, compared to the maximum probability difference found in the predictions of surrogate classifiers, which were trained on nonsense data, the probability difference in the real data is smaller at all thresholds, which indicates that the difference is due to noise rather than a real effect.

Statistical testing seems rather pointless in the face of such visually obvious mismatch of data and hypotheses. Still, two probability thresholds were selected

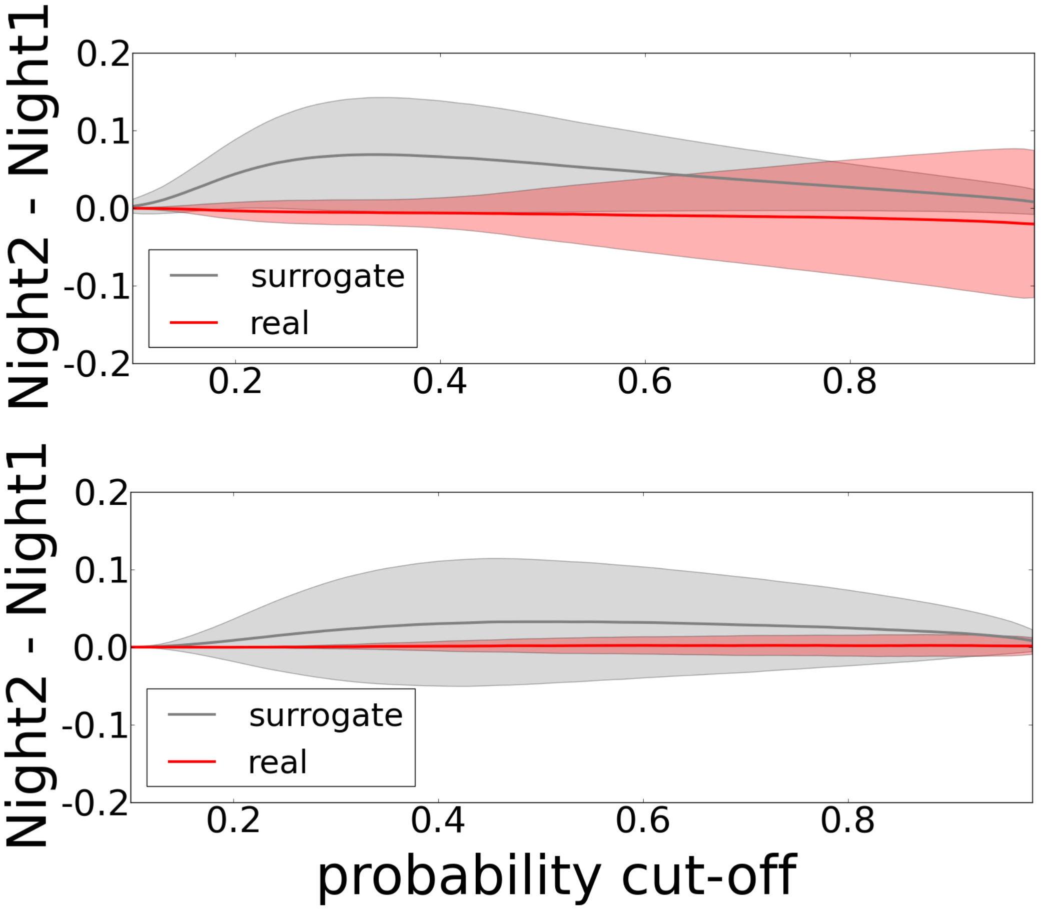

Figure 12.10: Probability difference for unfiltered data (top) and filtered data (bottom). For varying probability cut-offs between 0.1 and 0.99, the difference is shown between the ratio of survivors to all predictions in Night2 and Night1. Bold lines depict means across 12 patients, the shaded areas indicate the standard-deviation at each probability cutoff. Positive values would indicate increased replay in Night2. However, the probability difference in classifier predictions that was based on real data is below that of surrogate classifiers which were trained on shuffled data. In the unfiltered approach, the difference is even numerically negative for high probability cut-offs.

.

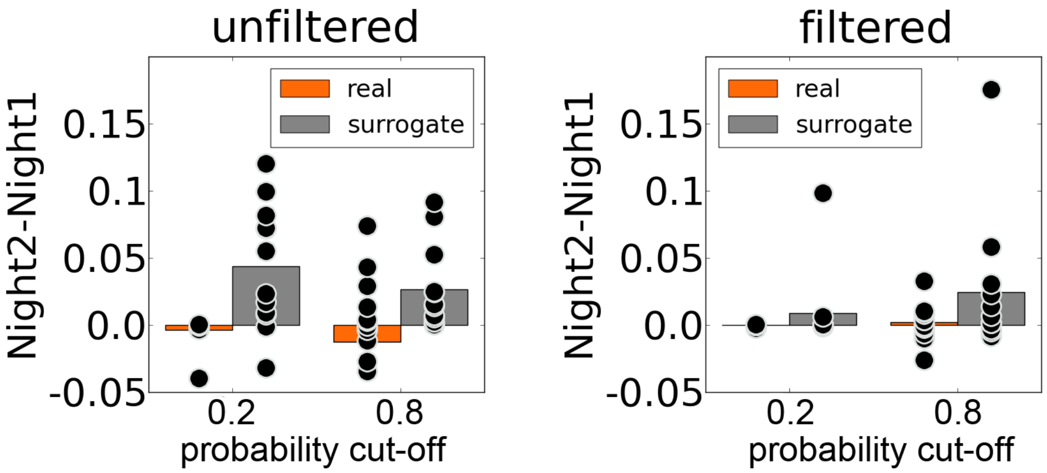

Figure 12.11: Probability difference for filtered and unfiltered data at two representative probability cut-offs. It is obvious that the differences found in the real data do not exceed the maximal differences found in the shuffled data across patients.

.

as representatives, a low threshold of 0.2 and a higher threshold of 0.8. Results for these thresholds are shown in Figure 12.11.

Neither at 0.2 nor at 0.8, neither for the unfiltered data nor for the filtered data were probability differences across 12 patients significantly larger than zero or significantly exceeded the maximal probability difference found in the surrogate classifier predictions of 12 patients (assessed with one-sided t-tests).

The results reported so far were based on classifier predictions on all epochs of Night1 and Night2, irrespective of whether patients were awake or sleeping. Next, the probability difference was investigated in different sleep-stages. The procedure is analogous to the one used above, but for each sleep-stage, only those epochs are included that have been classified as belonging to the sleep-stage.

Figures 12.12 and 12.13 provide an overview of the results for individual sleep-stages. As can be easily seen, the probability difference in real data again did not exceed the probability difference in surrogate classifiers – in any of the sleep-stages, neither for unfiltered nor for filtered data.

Figure 12.14 shows the results of the probability difference at two exemplary thresholds of 0.2 and 0.8 across sleep-stages for the unfiltered data. T-tests again confirmed that in no case did the probability difference found in the real data exceed the difference found in shuffled data.

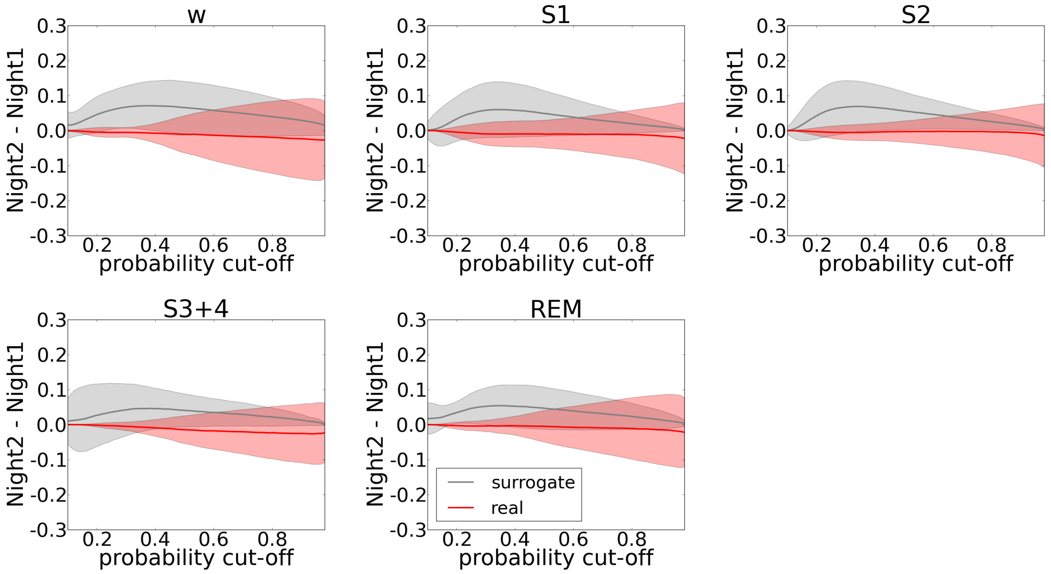

Figure 12.12: Probability difference in the individual sleep-stages for unfiltered data. Bold lines depict means across 12 patients, the shaded areas indicate the standard-deviation at each probability cutoff. Again, differences found in the real data do not exceed the maximal difference found in the shuffled data. Abbreviations: w=wake, S1=sleep-stage 1, S2=sleep-stage 2, S3+4=sleep-stages 3 and 4 (slow-wave sleep), REM=rapid eye movement sleep.

.

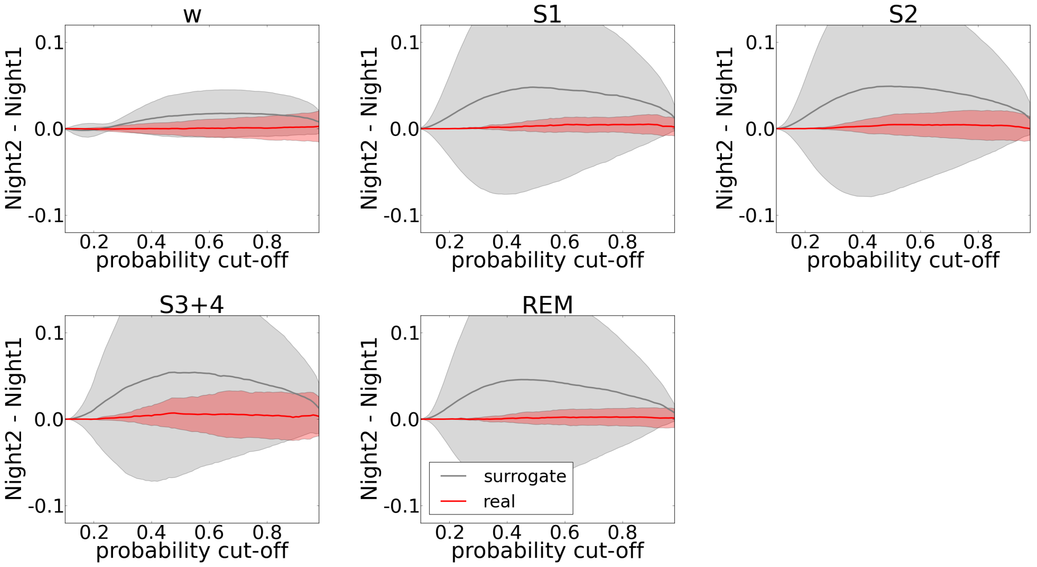

Figure 12.13: Probability difference in the individual sleep-stages for filtered data. Bold lines depict means across 12 patients, the shaded areas indicate the standard-deviation at each probability cutoff. Again, differences found in the real data do not exceed the maximal difference found in the shuffled data. Abbreviations: w=wake, S1=sleep-stage 1, S2=sleep-stage 2, S3+4=sleep-stages 3 and 4 (slow-wave sleep), REM=rapid eye movement sleep.

.

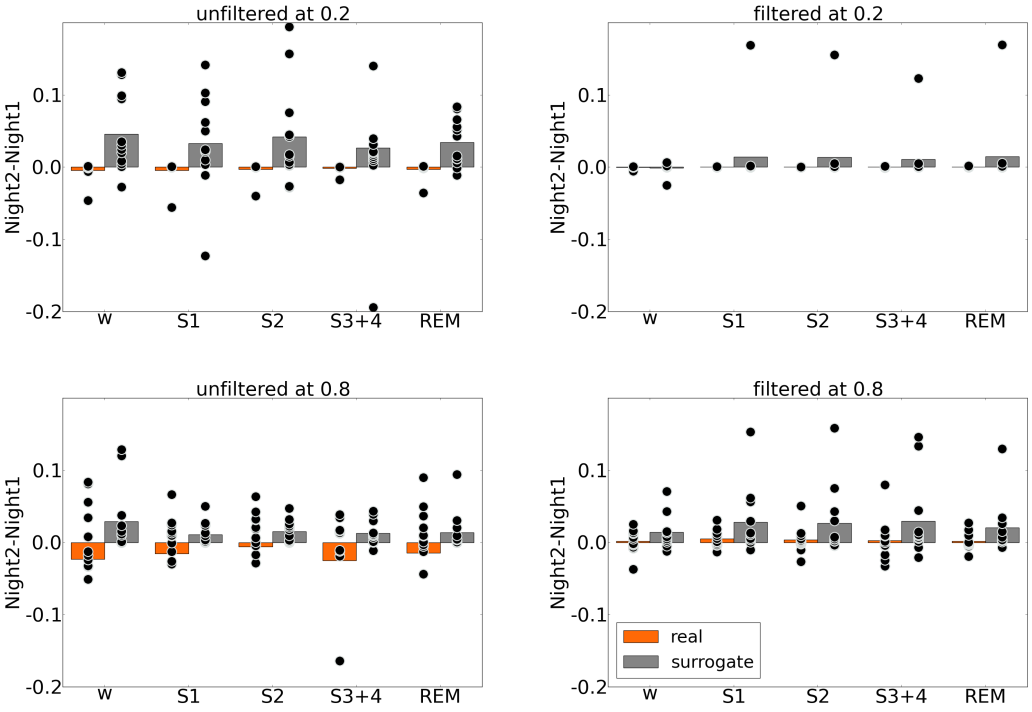

Figure 12.14: Probability difference at two exemplary probability thresholds in individual sleep-stages. Again, differences found in the real data do not exceed the maximal difference found in the shuffled data. Abbreviations: w=wake, S1=sleep-stage 1, S2=sleep-stage 2, S3+4=sleep-stages 3 and 4 (slow-wave sleep), REM=rapid eye movement sleep.

.

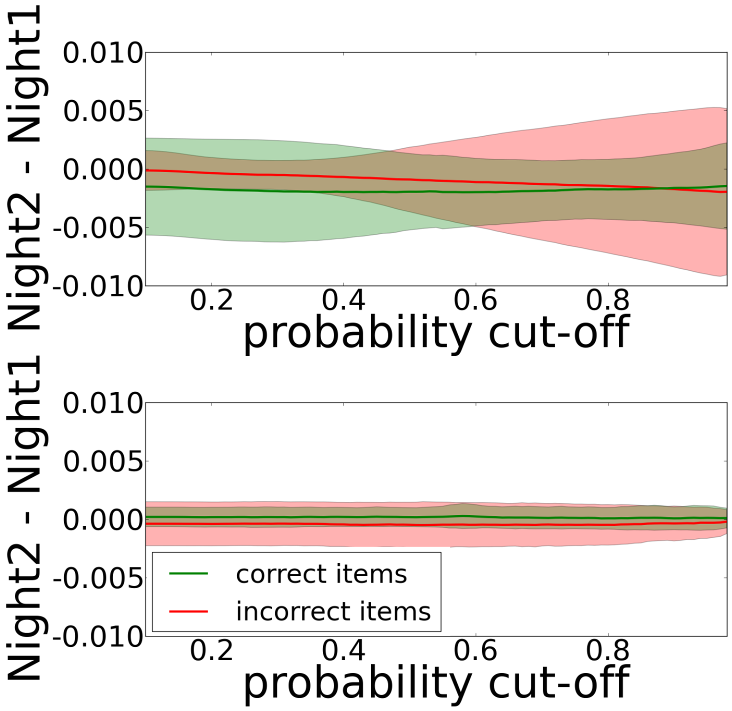

Figure 12.15: Probability difference between Night2 and Night1 plotted separately for those items that were subsequently remembered ("correct") and those that were subsequently forgotten ("incorrect") for the unfiltered and the filtered approach. Bold lines depict means across 10 patients, the shaded areas around the lines indicate the standard-deviation at each probability cutoff. Correct items did not have a higher probability difference at probability thresholds 0.2 and 0.8 for unfiltered or filtered data.

.

Taken together, there is no evidence for more high confidence classifier predictions in Night2 as compared to Night1.

12.3.5 Relationship with behavior

As was done in the previous studies, classifier predictions were investigated with regard to later memory performance. Figure 12.15 shows probability differences that were calculated separately for items that were correctly remembered versus those that were incorrectly remembered at the memory test that took place in the morning after Night2. For this analysis, again only 10 patients were considered because the results of the memory test were missing for 2 patients (see above).

In the unfiltered approach, the probability difference is not higher for remembered ("correct") than for forgotten ("incorrect") items. In the filtered approach, the difference is numerically higher in the remembered than in the forgotten items;

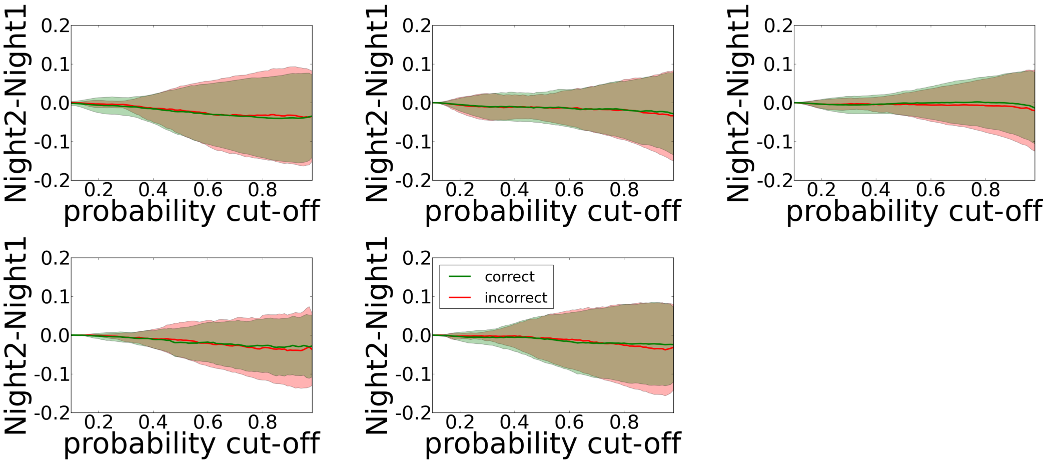

Figure 12.16: Probability difference between Night2 and Night1 for the unfiltered approach, plotted separately for those items that were subsequently remembered ("correct") and those that were subsequently forgotten ("incorrect") across individual sleep-stages. Bold lines depict means across 10 patients, the shaded areas enveloping the lines indicate the standard-deviation at each probability cutoff. Though numerically higher in some parts, correct items did not have a significantly higher probability difference than incorrect items at exemplary probability thresholds 0.2 and 0.8 in any of the sleep-stages.

.

however, the standard deviation clearly overlaps. At a probability level of 0.8, a one-sided paired t-test across 10 patients revealed no significant difference ($t_9 = 1.149$, $p_{uncorr} = 0.140$). Also, t-tests at 0.2 and for the unfiltered approach did not reveal any significant difference.

It is possible that memory-relevant replay happens only in specific sleep-stages. Therefore, the probability difference between Night2 and Night1 for correct and incorrect items was investigated separately for individual sleep-stages. Results are depicted in Figure 12.16 for the unfiltered approach and in Figure 12.17 for the filtered approach. In some sections of the data, the probability difference is numerically larger for correct items, but it is obvious that the standard deviations between the two conditions still overlap. When tested at the two exemplary probability cut-offs of 0.2 and 0.8, the difference was never significant in one-sided paired t-tests.

12.4 Discussion

This study builds upon the previous two studies theoretically and extends them methodologically. While using a near-identical paradigm to that of the first study,

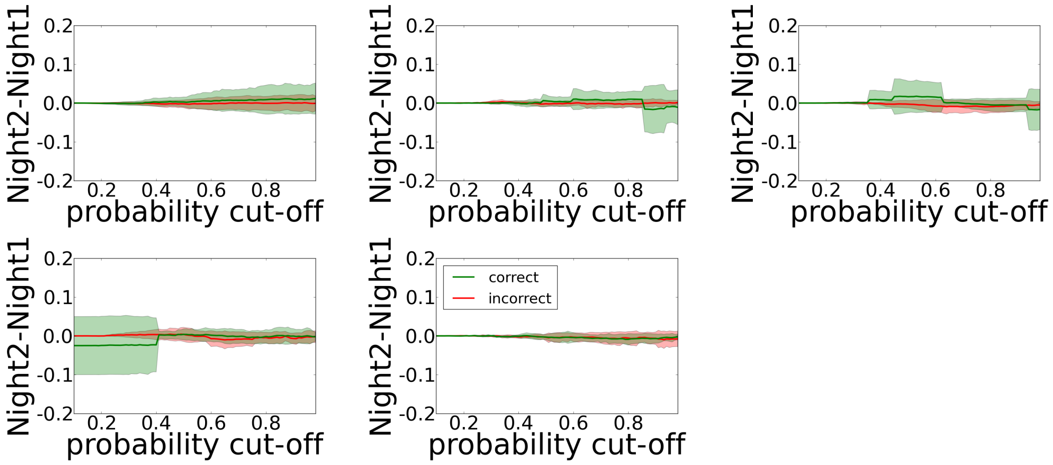

Figure 12.17: Probability difference between Night2 and Night1 for the filtered approach, again plotted separately for items that were subsequently remembered ("correct") or subsequently forgotten ("incorrect") across individual sleep-stages. Bold lines depict means across 10 patients, the shaded areas enveloping the lines indicate the standard-deviation at each probability cutoff. Though numerically higher in some parts, correct items did not have a significantly higher probability difference than incorrect items at exemplary probability thresholds 0.2 and 0.8 in any of the sleep-stages.

.

brain activity was recorded with intracranial EEG instead of fMRI. The use of an electrophysiological method with high temporal resolution puts this study closer in context to place-cell studies in rodents (Skaggs and McNaughton, 1996; Louie and Wilson, 2001; Lee and Wilson, 2002; Foster and Wilson, 2006; Karlsson and Frank, 2009; Carr et al., 2011; Jadhav et al., 2012), which have provided the idea for the project presented in this thesis.

The ability to extract fast-changing stimulus-specific patterns of brain activity and search for them in subsequent sleep adds another dimension to analysis: While the fMRI analysis only took into account the pattern of neuronal activity at one point in time (one fMRI volume), in this study, subsequent points in time could also be used to extract stimulus-specific patterns. In addition, replay could be investigated with higher sampling rate during resting state and sleep: a sliding window approach was used, extracting epochs every $100ms$. In fMRI, only one volume every $2500ms$ could be used. This could pose a problem, e.g. if a replay event starts between two fMRI volumes, it should be harder to detect for a classifier.

In the following, findings from the third and last study of this thesis will be discussed and compared to previous results. In a last section, shortcomings of this study, possible explanations for the lack of findings as well as suggestions for improvements will be presented.

12.4.1 Choice of classification algorithm

In this study, instead of a linear support vector machine, a Sparse Multinomial Logistic Regression (SMLR) classifier was used, which has been described in the Theoretical Part of this thesis. The choice of this particular classifier was mainly made to ensure a seamless transition from prior analysis of this dataset which was started by Thorsten Kranz and in which SMLR was employed. Due to the sheer amount of EEG data (up to 50 gigabyte per patient) and the complex data processing (filtering, frequency decomposition, cluster-based feature selection, cross-validation, prediction on the nights), classification analysis took up to three weeks per participant. Thus, it was important to decide on many parameters, such as choice of classifier, early on, because trying too many different settings would have taken a long time. Comparing different classifiers and settings would certainly be very interesting, but the extent of such analyses was beyond the scope of this thesis.

12.4.2 Feature selection and classifier accuracy

Feature selection was more difficult, or complex, in this study than in the previous two studies, in which either voxels with the largest discriminability were chosen, or voxels within selected regions of interest. A simple, timepoint-by-timepoint F-value-based selection did not yield good cross-validation results for the intracranial EEG. The surrogate cluster approach described above selected only those contiguous parts of the data that were, as a cluster, more significant than clusters found in shuffled data. This approach is resistant to a variety of distortions in the data that could lead to the selection of time-points with high F-values that actually contribute nothing to discriminating between the classes in the test data.

With the surrogate cluster approach, classifier accuracy values were achieved which, compared to the fMRI results, were quite low, but still significantly better than chance across patients.

Three things became evident during feature selection and cross-validation. First, classification accuracy was much better when data were not decomposed into different frequency bands. Second, when the data was decomposed into different frequency bands, features were selected in different frequency bands at different times after stimulus onset – even though, apparently, they did not generalize well to the test dataset. And third, classifier accuracy varied vastly across patients. In some, it was very good, in some it was barely better than chance. It is currently not clear why there are such dramatic differences between patients. It might be related to general cognitive ability level, signal quality or electrode placement. This remains to be further investigated.

One interesting result was that features in the theta-frequency range were selected relatively often across patients. Even though it is not possible, with the current classifier settings, to assess how important features from the theta band are for accurate predictions, the fact that many clusters were selected in these bands warrants the conclusion that information with regard to stimulus identity is present in this band.

This is in accordance with the literature, which proposes an important role for theta oscillations in memory formation (Klimesch, 1999; Osipova et al., 2006). Future analysis of the data could investigate predictions made by classifiers that were trained on features from specific frequency bands only, e.g. theta. If different frequency bands have different, maybe even opposing roles in memory consolidation, including features from all frequency bands (as was done in this study) might obliterate frequency specific effects.

12.4.3 Probability difference

In order to investigate replay of stimulus specific-activity pattern, a sparse multinomial logistic regression was trained on the data from the memory task and made predictions on epochs during a night preceding (Night1) and a night following (Night2) the learning task. The hypothesis was that there would be more evidence for replay during Night2 than during Night1. As the classifier returns one prediction for every sample no matter whether any matching pattern is actually detected, the confidence with which the classifier made the predictions was investigated. For varying probability cut-offs between 0.1 and 0.99, the ratio of “surviving” predictions to all predictions was calculated for Night1 and Night2.

If the difference for these ratios (Night2-Night1) was positive, it was taken as a sign for more confident classifier predictions during Night2, and hence as stronger evidence for replay. Because many factors unrelated to replay could possibly influence this probability difference, surrogate classifiers were trained on shuffled, i.e. randomized, data and also made predictions on epochs during the two nights.

Contrary to the hypothesis, probability differences in the real data were often not even numerically larger than zero, and never in a statistically significant way. Across patients, they were also never significantly larger than the maximum probability difference found in the surrogate data, neither for the unfiltered nor for the filtered approach. This did not change when individual sleep-stages were analysed separately.

The lack of significant findings might be due to several reasons. First, classifier accuracy might have been insufficient to reliably detect reoccurrence of patterns. Even though classifier accuracy was better than chance across patients, in many patients it was not higher than chance level in a relevant way, especially in the filtered approach. Refined filtering, feature selection or better algorithms might help to increase classifier performance.

Second, signal-to-noise ratio might have been too low. Two complete nights of sleep were recorded, and epochs were extracted every 100ms seconds, leading to an average of 427,082 epochs per night. Theoretically, one would expect replay to happen relatively rarely. Thus, searching for rare replay events with an algorithm that is far from perfect, the effect might be missed. Apart from improving classification accuracy, identifying time windows in which replay happens more often could increase signal-to-noise ratio. Such “replay windows” could be connected to sleep spindles (Diekelmann and Born, 2010; Bergmann et al., 2012). An algorithm that detects sleep spindles would be very helpful in this regard. Another option for future experiments would be to experimentally induce replay events, e.g. by exposing patients to odor or subliminal sound cues that have been associated with stimuli during the learning task as it has been done in other studies (Rasch et al., 2007; Rudoy et al., 2009; Diekelmann et al., 2011).

A third explanation for the lack of results might be that replay during sleep happens in a condensed fashion (Skaggs and McNaughton, 1996; Nádasdy et al., 1999) or even in a reversed sequence (Foster and Wilson, 2006; Diba and Buzsáki, 2007), as it has been observed in rats. This problem might be more relevant for electrophysiological data than for fMRI data because of the higher temporal resolution. In fMRI data, the BOLD pattern $5000ms$ after stimulus onset probably looks not much different for a hypothetical neuronal firing sequence “A, B, C” than for “C, B, A” (at least if the sequence happens fast), because the hemodynamic reponse is sluggish and is in any case merely a substitute marker for neuronal activity. In electrophysiological data, the sequence of neuronal activity might play a bigger role – this was one of the reasons for performing the experiment with this different method in the first place.

As has been shown in Figures 12.8 and 12.9, features from different time-points after stimulus onset are part of the pattern; in this sense, neuronal sequences are implicitly used for decoding. If a certain process is identified by a sequence of neuronal events “A, B, C”, which happen $100ms$ apart, a reversed or condensed replay of sequences will not be detected by the algorithm. It will be very exciting to investigate the possibility of replay of such altered or reversed sequences, even though pattern classification is probably not the right method of analysis for this issue.

Lastly, the probability difference found in classifier predictions based on real data was compared to the maximum probability difference found in a set of surrogate classifier predictions, that were based on shuffled data. Taking the maximum probability difference instead of, e.g., the mean probability difference might be an overly conservative test. However, the probability difference in the real data was also not significantly larger than zero, which would be the most basic requirement for supporting the hypothesis that replay activity is more pronounced in Night2 than in Night1.

12.4.4 Relationship with sleep stages

Surprisingly, the probability difference – apart from not displaying the expected effect of being significantly positive – also did not exhibit much variation across different sleep stages or compared to waking state. Different sleep stages have been hypothesized to be differentially involved in replay (Diekelmann and Born, 2010). It might be futile to speculate on reasons for this negative finding, especially since the classification approach might in general not be well suited for detecting replay in this study.

Still, there are many characteristics of sleep which could be investigated for a special relationship with replay: sleep spindles, ripples and slow waves are prominent motifs in sleep. In future analysis of the data of this study, it could be investigated whether replay events occur more frequently in the vicinity of these motifs.

12.4.5 Relationship with behavior

One of the main tenets of two-step models of memory formation is that reactivation improves consolidation which in turn improves memory. Thus, any effect that can be taken as evidence for reactivation should also exhibit a relationship to behavior.

Therefore, the probability difference between Night2 and Night1 were analyzed separately for items that were remembered in a memory test after Night2 and items that were forgotten. There was no significant difference for these two categories, neither over the complete night nor in specific sleep stages.

One factor that might have contributed to this null finding is the great variance in memory performance in this patient sample. In some patients, nearly none of the objects were correctly assigned to their associated location while in others, almost all objects were correctly assigned. Thus, for some patients the task apparently was too easy, for others too difficult. This might be also be influenced by the location of the epileptic focus, the age of patients, severity and duration of epilepsy and general cognitive abilities.

Using a continuous metric of memory performance in form of the error distance as used in the previous two studies might improve this situation. Even in high-performing patients, graded memory for individual items could still be detected. This was done with the second version of the paradigm. As there were only four patients who performed the second version and underwent the memory test, separate statistical analysis did not appear sensible. Accordingly, their memory results were also binarized. As the testing of patients is still being continued at the Clinic for Epileptology, the addition of a few more patients with the continuous memory metric might yield better results.

12.4.6 Critical review of the study and outlook

With the approach used in this study, no evidence for replay of stimulus-specific activity could be detected. Given that a similar mechanism as it was found in rodent studies is also present in humans, an assumption which is supported by the literature and results from the first two studies, the lack of significant findings is probably not due to false hypotheses, but to insufficient methods for investigating the hypotheses.

One of the most important problems is the sheer amount of data which results from recording two complete nights of sleep. This might lead to a low signal-to-noise ratio, especially in combination with relatively low classifier accuracy. One of the first steps in future analyses should be to improve classifier performance by finding different algorithms or data preprocessing.

Another important step would be to identify time-windows of interest in which replay events are thought to occur more frequently. Comparing the time-windows of interest to the epochs outside these windows would also constitute a different, perhaps more sensible statistical test than comparing probability differences between Night2 and Night1.

Also, a change in the experimental setup could help in the future: In the first study of this thesis, two stimulus sets were used. The two sets could also be presented to the patients: one prior to Night1 and the second prior to Night2. Increased replay of Set1 in Night1 and Set2 in Night2 could be taken as a sign for stimulus-specific replay. However, there would be issues regarding temporal asymmetry: Set1 stimuli would be inherently more similar to electrophysiological recordings that happen close in time, and the same is true for Set2 stimuli.

A similar approach as was taken in the first fMRI study would be most convincing: Presenting Set1 stimuli on the evening of a night and Set2 stimuli on the morning after. A memory test could then happen immediately after the second memory task, or in the afternoon/evening of the same day. Finding more classifier predictions for Set1 as compared to Set2 stimuli in the night between the two tasks would then be excellent evidence for replay. This design, however, could be hard to implement in the every-day routine of the ward and task difficulty would have to be adapted to the cognitive abilities of the patients.

Taken together, the complex data presented in this study has so far not provided any evidence for stimulus-specific reactivation of neuronal acitivty, but merits further analysis. In the future, an adapted version of the paradigm in additional patient recordings could also provide more insight into the mechanisms underlying memory consolidation in humans.

13 General discussion

This thesis investigated replay of stimulus-specific memory reactivation during resting state and sleep using a variety of imaging and electrophysiological methods. The main idea was to identify neuronal signatures of specific stimuli during a learning task with multi-variate pattern analysis and track spontaneously reoccuring instances of these neuronal signatures in resting state or sleep after the learning task. The hypotheses in all three studies were that, first, occurrence of learning-related neuronal signatures should increase in frequency after the learning task and, second, that the frequency of replay should be associated with subsequent performance in a memory task. The frequency or confidence level of the detected re-occurrence was always compared to a control condition.

The specific results of the three studies, as well as their benefits and shortcomings, have been discussed at length above. In this general discussion, the studies will be compared to one another with regards to methods and results. Finally, the general merits or disadvantages of applying pattern classification approaches to resting state in order to detect replay events shall be briefly discussed.

13.1 Comparison of the three empirical studies

13.1.1 Classifier accuracy

In the first two studies, fMRI data was recorded and used for pattern classification. In both studies, classifier accuracy was excellent and well above chance level. In the third study, pattern classification analyses were based on electrophysiological data and classifier accuracy was worse than in the other studies, especially if the signal was decomposed into different frequency bands.

There are various explanations for this. First, electrophysiological signals might in general have worse signal-to-noise ratio than fMRI data and pattern classification might simply not work as well on this kind of data.

Second, the lack of good classifier accuracy might be based on subject population. It is possible that decoding data which was recorded from epilepsy patients, even though it has been done before (van Gerven et al., 2013), does not yield classifier accuracies which are as high as those found in healthy participants. This could be related to problems in attention during the experiment or to altered brain function.

Lastly, the optimal approach for this kind of data might not have been found in the current analysis. Possibly, better classification results can be achieved if the data are preprocessed or transformed in a different manner, i.e. working with the first derivative or including phase information as features. This might be explored in future analyses of the data.

13.1.2 Association with behavior

In the first two studies, an association of replay frequency with behavioral memory performance was found. In the third study, no association with behavior was found. This could be due to either the relatively low classifier accuracy or to different participant populations (i.e. patients versus healthy volunteers). For the latter it should be noted that epilepsy patients often have memory impairments and thus, for some, the task might have been too difficult. An indication for this is the high variability in the number of remembered object-place associations across patients. If memory performance is very low, item-wise association between replay frequency and memory performance might not be as meaningful as it is for performance in the medium range.

Another factor for the lack of an association might be that only in the third study, an entire night of sleep was recorded after the memory task. During such long recordings, effects might be too diluted to become significant, especially if they only occur in some parts of the data.

13.1.3 Comparison to control condition

In the first study, a set of stimuli that had been presented prior to the resting period was detected more often by the classifier than a second stimulus set which had been presented after. Such a set-wise reactivation is a good proof of concept and indicates that stimulus-related, not only task-related activity is reactivated: The task was the same for both stimulus sets. Thus, if only task-related activity had been reactivated, there should have been no advantage for Set1 stimuli.

In the second study, such a set-wise difference was not found for emotionally negative as compared to emotionally neutral stimuli. This, however, was also accompanied by a lack of behavioral difference for these two sets, which had been the hypothetical reason for assuming a differential effects for these two sets. If negative items are not remembered better, there really is no reason to expect that they should be detected by the classifier more often in a resting period in which reactivation might occur. Thus, the lack of a set-wise effect in the second study is likely due to a deficit in setup of the paradigm.

In the third study, there were no two stimulus-sets that could be compared against one another. Instead, two nights were compared against each other with regard to classifier confidence and no effect was found. In addition to the various problems with the third study that were already mentioned (classifier accuracy, data size, patient population), the lack of a second, control dataset further limits conclusions that can be drawn from the data.

13.1.4 The role of sleep

In the first and third study, periods of sleep were recorded in addition to periods of wakefulness. Sleep should be the primary window of opportunity for replay to happen because the brain is insulated during sleep and consolidation will not be disturbed (Diekelmann and Born, 2010). However, in the first study the strongest evidence for replay was found during periods of quiet waking state. In the third study, apart from general lack of significant results, looking at sleep-stages separately also did not yield any findings.

From this, one can either draw the conclusion that the current pattern classification approach does not detect replay in periods of sleep or that a different experimental setup is needed to investigate sleep. In the first study, sleep was recorded during an afternoon nap. Deep stages of sleep were only recorded in half of the participants and even then the duration of slow-wave sleep was short. It is quite possible that recording longer periods of night sleep allows more reliable investigation of replay associated with deeper stages of sleep. Methodological improvements of the pattern classification approach might also help in detecting replay events.

13.2 Discussion of the MVPA approach

Using pattern classification to track neuronal activity in paradigm-free periods is, to my best knowledge, a novel approach. One study tracked neuronal activity during free recall (Polyn et al., 2005) and was able to decode the category of the stimulus that was covertly being retrieved by participants. Another study decoded the hidden intention of whether participants were going to add or subtract two numbers (Haynes et al., 2007) in the following trial. Both of these studies decoded brain activity that was, in a way, internally generated. It is encouraging to know that pattern classification can decode activity that is not induced by an external stimulus. However, the two studies above still involved a paradigm in which the episodes of internally generated activity were embedded.

Some theoretical assumptions are made in the approach used in this thesis which need to be treated with caution. For one, the pattern of neuronal activation as a whole is hypothesized to be similar during initial learning and subsequent reactivation across the brain. In rodent studies, there is support for this assumption because the same sequence of place cell firing spontaneously reoccurs during sleep. However, this does not say much about what happens in the rest of the rodent brain at the same time. Possibly, activity in certain brain areas that were not recorded from in these studies could be up- or down-regulated during replay events as compared to initial learning.

A second assumption is that if classifiers can reliably decode stimulus-specific activity during a learning task, they can also reliably detect stimulus-specific reactivation during sleep. Cross-validation was used in all three studies to assess classifier performance. However, the accuracy with which a trained classifier can detect replay events in the paradigm free periods cannot be assessed in our studies because there is no information when such a replay event might occur in truth.

Both of the issues are, in part, addressed by the design of the first study with one stimulus set preceding and one following the task. If the classifier was unable to detect any “familiar” activity, classifier votes during the main resting period, which was recorded at a symmetrical temporal and spatial distance to the two memory task recordings, should be random, i.e. distributed equally across the two stimulus sets. This, however was not the case. In addition, the fact that classifier prediction frequencies during the resting periods after the task were associated with memory performance for the individual stimuli further supports the idea that the classifier is able to detect some reoccuring neuronal activity that is related to the initial encoding of the individual stimuli.

The degree to which a classifier detects real replay events in terms of sensitivity and specificity can, as stated above, not be answered based on the three studies of this thesis because we have no information on what happens during the resting period. However, it seems plausible that good classifier accuracy in the memory task dataset is a necessary prerequisite for good detection of replay events. In the first study, we excluded participants in which classifier performance was not excellent and this decision is supported by the third study, in which classifier accuracy was not very good and there was no evidence for replay in this dataset. Thus, good classifier accuracy is important in this new approach and it is reasonable to assume that in taking measure to increase classifier accuracy, replay detection will also be improved.

13.3 Future directions

Future studies should try to address the problem that the actual occurrence of replay events is unknown in resting periods. One approach would be to identify time windows during which replay events are more likely to occur, for example during spindle or ripple events or in epochs of high hippocampal activity. Another way would be to introduce an experimental manipulation which triggers reactivation as it has been done in previous studies with odor or sound cues (Rasch et al., 2007; Rudoy et al., 2009). If the classifier detected more replay events during these time windows of interest than during others, it would lend strong support for classifier based tracking of memory reactivation.

Lastly, one important aspect has not been addressed at all by the studies in this thesis which should definitely be investigated in the future. In the rodent studies reporting evidence for replay (Skaggs and McNaughton, 1996; Louie and Wilson, 2001; Lee and Wilson, 2002; Foster and Wilson, 2006; Karlsson and Frank, 2009; Carr et al., 2011; Jadhav et al., 2012), *sequences* of neuronal firings were usually investigated. These sequences corresponded to real episodic sequences (i.e. running along one location after the other). This temporal aspect has not been addressed at all in this thesis, but it could be a key aspect in further elucidating reactivation in humans and in establishing the validity of the current methods. For example, some well-classifiable stimuli could be presented in a specific order during a learning task preceding a resting state and in a different order in a learning task following the resting period. If classifier predictions displayed the pre-rest order more than the post-rest order, this would also be strong evidence that the classifier detects reactivation events. In addition, sequences of events are what, in everyday life, usually comprise an episode instead of single events. Thus, investigating sequences of events would have more validity.

From a more general point of view, any study of memory or, by extension, memory consolidation, should strive to achieve high ecological validity, at least in the long run. On the one hand, psychologists try to design well controlled experiments which can be easily replicated. On the other hand, learning object-place associations might not be the most salient or life-like form of memory. By including more complex stimuli such as videos or 3D environments, one might capture more relevant aspects of every-day memory while still retaining experimental control. More salient learning content would probably lead to better memory consolidation, which might increase reactivation frequencies and, in turn, make it easier for the classifier to detect the events due to better signal to noise ratio.

13.4 Summary

In this thesis, three studies were presented that used a novel method for tracking memory reactivation in paradigm-free resting periods and sleep. Results from two of the three studies support the notion that the method is able to detect stimulus-specific replay, even though the reliability of the detection can not be assessed with the current experimental design.

It was found that the frequency of replay events was associated with later stimulus-wise memory performance in two of the three studies. This supports hypotheses derived from two-step models of memory formation which propose that after initial encoding, memory traces become stable by a reactivation of the associated neuronal activity.

14 Abstract

Memory consolidation is a theoretical process by which initially labile memory traces become more stable. The neuronal mechanism supporting this stabilization is thought to include a spontaneous reactivation of the same neuronal activity that was present during learning in task-subsequent periods of resting state and sleep. This reactivation is hypothesized to support information transfer from the hippocampus, which is conceptualized as a “fast learner” and temporary memory storage, to the neocortex, in which slow reconfiguration of neuronal connections leads to an integration of new memories into an existing network of life-time experience and knowledge. In rodents, evidence for reactivation has been found in the form of the coordinated replay of experience-related place-cell firing sequences during sleep and quiet resting state. In humans, increased activation of memory related brain structures has been observed after exposing participants to odor cues which were associated with the task. Subliminal sound cues during sleep have been shown to selectively enhance memory for those stimuli they were associated with during a learning task. In this thesis, a new method of identifying stimulus-specific neuronal activity patterns, multi-variate pattern analysis (MVPA), is employed to search for these neuronal patterns in resting state and sleep after a memory task. Three studies are presented which use functional magnetic resonance imaging (fMRI), alone or combined with simultaneous electroencephalography (EEG), and intracranial EEG recordings in epileptic patients. The impact of reactivation-related neuronal activity on memory performance for normal and emotionally negative stimuli is investigated. In two of the three studies, a relationship between the frequency of stimulus-specific reactivation and later memory performance were found, even though adequate control conditions have to be discussed. The conclusions of this thesis are that MVPA is well suited to decode the neuronal signatures of individual stimuli and can be useful for tracking these neuronal signatures across periods of resting state and sleep.

15 German Summary (Deutsche Zusammenfassung)

Deutscher Titel: Neuronale Korrelate von Gedächtniskonsolidierung während des Wachzustandes und Schlafes

15.1 Einleitung

Gedächtniskonsolidierung ist ein theoretischer Prozess, bei dem zunächst instabile Gedächtnisspuren in einen stabileren Zustand überführt werden. Schon zu Beginn des vorigen Jahrhunderts wurde beobachtet, dass eine zuvor gelernte Liste weniger anfällig für Störung durch eine zweite Liste war, wenn seit dem Lernen der ersten Liste etwas Zeit vergangen war (Müller und Pilzecker, 1900). Auch bei neuropsychologischen Patienten konnte beobachtet werden, dass jüngere Erinnerungen anfälliger für einen Gedächtnisverlust infolge einer Gehirnschädigung waren als ältere Erinnerungen (Ribot, 1882).

Es wird angenommen, dass einer der zugrunde liegenden neuronalen Mechanismen für diese Stabilisierung eine spontane Reaktivierung derjenigen neuronalen Aktivität ist, die während des Lernens auftrat (Frankland und Bontempi, 2005). Diese Reaktivierung tritt, so die Theorie, in Ruhephasen und Schlaf nach einem Lernprozess auf und dient einem Informationstransfer zwischen Hippocampus und Neokortex. Der Hippocampus, der oft als ein "schneller Lerner" konzeptualisiert wird, dient in diesen Modellen als ein vorrübergehender Speicher, in dem neuronale Aktivität aus verschiedenen kortikalen Modulen zu einer Episode gebunden wird.

Studien an Patienten bestätigen, dass eine Schädigung oder Resektion des Hippocampus zu einer anterograden Amnesie führt, also zu einer Unfähigkeit, neue Gedächtnisinhalte zu bilden (Corkin, 2002). Gleichzeitig kommt es oft zu einem Gedächtnisverlust, der einen zeitlichen Gradienten aufweist, so dass ältere Erinnerungen bestehen bleiben, während jüngere Erinnerungen verloren gehen (Squire et al., 2001). Diese Befunde sprechen dafür, dass einerseits der Hippocampus eine zentrale Rolle dabei spielt, neue Gedächtnisinhalte zu bilden, und dass andererseits Gedächtnisinhalte irgendwann unabhängig werden vom Hippocampus und nach einer Schädigung entsprechend erhalten bleiben.

Der Theorie sognannter Zwei-Stufen-Modellen nach können die anfänglich im Hippocampus gespeicherten Gedächtnisspuren erst durch wiederholte Reaktivierung in den langsamer lernenden Neokortex übertragen werden (Marr, 1970; Frankland und Bontempi, 2005), wo die Reaktivierung nach und nach zu einer langsamen Rekonfiguration neuronaler Gewichte und Verbindungen führt und wo neue Gedächtnisinhalte in ein bestehendes Netzwerk lebenslanger Erfahrungen und Wissen integriert werden (Squire und Alvarez, 1995; McClelland et al., 1995; Squire et al., 2004, Hasselmo, 2005). Dieser Integrationsprozess erstreckt sich vermutlich über einen Zeitraum von Monaten und Jahren, wenn man den zeitlichen Gradienten bei Patienten mit Hippocampus-Schädigung betrachtet.

Eine implizite Annahme des Reaktivierungsmodells ist, dass Gedächtnisspuren beim Reaktivieren vorrübergehend wieder instabil werden und der Prozess somit zu Zeiten stattfinden muss, in denen keine Störung durch von außen einströmende Information besteht. Ein wichtiges Zeitfenster für diesen Prozess könnte Schlaf sein, bei dem das Gehirn weitestgehend isoliert ist von äußeren Einflüssen. Obwohl Befunde, dass Schlaf zu einer Verbesserung von Gedächtnis führt, schon lange bestehen (Jenkins und Dallenbach, 1924), hat sich erst in den letzten Jahren die Erkenntnis durchgesetzt, dass Schlaf auf nahezu alle Gedächtnisformen einen positiven Einfluss hat (siehe Review von Diekelmann und Born, 2010) und dass dies damit zusammenhängen könnte, dass Schlaf eine optimale Umgebung für Reaktivierung und somit Konsoliderung bietet.

In Studien an Nagetieren konnten in den letzten zwei Jahrzehnten experimentelle Belege gesammelt werden, die sowohl Zwei-Stufen-Modelle der Gedächtnisbildung stützen als auch die Rolle von Schlaf bei der Gedächtniskonsolidierung: Im Hippocampus von Nagetieren finden sich sogenannte “Ortszellen”, die ihre Feuerrate zuverlässig dann erhöhen, wenn sich das Tier an einem bestimmten Ort in einer Umgebung aufhält (OKeefe und Dostrovsky, 1971). Lernt das Tier eine räumliche Aufgabe, etwa das Navigieren durch ein Labyrinth, so lässt sich in verschiedenen Durchgängen immer wieder die gleiche Sequenz an feuernden Ortszellen beobachten. Im Ruhezustand und Schlaf nach einer solchen Lernaufgabe wurde dann beobachtet, dass es zu einem spontanen Wiederauftreten eben dieser Sequenz in den Ortszellen kam – mehr als man unter Annahme zufälliger Sequenzen erwarten würde. Dies ist mittlerweile sowohl im Schlaf (Louie und Wilson, 2001; Lee und Wilson, 2002) als auch im ruhigen Wachzustand dokumentiert (Foster und Wilson, 2006; Carr et al., 2011; Jadhav et al., 2012).

Bei Menschen kann ein derart eindeutiger neuronaler Kode für Verhalten leider bislang nicht gemessen werden. Jedoch wurden auch hier Belege für die Theorie der Reaktivierung gefunden: In einer Studie wurde ein Rosenduft während einer Lernaufgabe präsentiert (Rasch et al., 2007). Anschließend sollten die Probanden im MRT-Scanner schlafen. Eine erneute Präsentation des Rosenduftes während der Tiefschlafphase führte zu einer Aktivierung des Hippocampus. In einer anderen Studie lernten die Probanden Assoziationen zwischen Objekten und Orten auf einem Schachbrett. Jedes Objekt-Ort-Paar war während des Lernens zusätzlich mit einem semantisch passenden Geräusch assoziiert. In einer anschließenden Schlafphase wurden den Probanden für die Hälfte der gelernten Paare die assoziierten Geräusche subliminal präsentiert. Für diese im Schlaf "getriggerten" Assoziationen zeigte sich in einem anschließenden Gedächtnistest eine bessere Gedächtnisleistung im Vergleich zu den nicht getriggerten Assoziationen. Diese beiden Studien geben Hinweise darauf, dass der Mechanismus der Reaktivierung sehr wahrscheinlich auch beim Menschen eine wichtige Rolle bei der Gedächtniskonsolidierung spielt. Jedoch wurde Reaktivierung in diesen Studien extern induziert (durch Gerüche oder Geräusche) und die dadurch hervor gerufenen Aktivität des Hippocampus war nicht spezifisch für einen bestimmten Lerninhalt, wie es etwa eine Sequenz von Ortszellen ist.

Kann man auch beim Menschen spontan auftretende stimulus-spezifische Reaktivierung beobachten? Zunächst muss man dazu stimulus-spezifische neuronale Aktivität identifizieren. Dies ist in den letzten Jahren zunehmend möglich geworden durch den Einsatz von multi-variaten Mustererkennungsalgorithmen in bildgebenden und elektrophysiologischen Verfahren. Bei diesem auch als "Mindreading" bekannt gewordenen Ansatz werden komplexe Muster neuronaler Aktivität von einem computergestützten Algorithmus einer von mehreren Klassen zugeordnet. Wenn nun ein solcher Algorithmus stimulus-spezifische Muster unterscheiden kann, so kann man mithilfe dessen möglicherweise auch ein Wiederauftreten dieses Musters im Ruhezustand oder Schlaf entdecken.

In der vorliegenden Dissertation wurde genau dies versucht: In drei Studien wurden Probanden mit einer assoziativen Gedächtnisaufgabe konfrontiert. Die neuronale Aktivität für einzelne Stimuli aus dieser Lernaufgabe wurde mit Musterklassifikationsverfahren extrahiert und ein mögliches Wiederauftreten dieser "neuronalen Signatur" wurde in sich an die Lernaufgabe anschließenden Ruhe- und Schlafphasen untersucht. Für eine erfolgreiche Durchführung dieses Vorhabens sind zwei Voraussetzungen von besonderer Bedeutung: Der Musterklassifikationsalgorithmus muss die einzelnen Stimuli mit einer hohen Genauigkeit voneinander trennen können, um überhaupt eine Chance zu haben, ihr Wiederauftreten in einer Ruhe- oder Schlafphase entdecken zu können. Außerdem ist davon auszugehen, dass eine Reaktivierung nicht die ganze Zeit, sondern eher sporadisch während einer Ruhe- oder Schlafphase auftritt. Da der Zeitpunkt eines tatsächlichen Wiederauftretens stimulus-spezifischer Aktivität nicht bekannt ist, sollte ein Mustererkennungsalgorithmus kontinuierlich Vorhersagen treffen, um die tatsächlichen Wiederauftretensereignisse nicht zu verpassen. Dies führt natürlich dazu, dass sehr wahrscheinlich über eine Menge Zeitpunkte Vorhersagen getroffen werden, in denen keine Reaktivierung stattfindet. Dies führt zu einem niedrigen Signal-zu-Rauschen Verhältnis. Es ist daher unbedingt sinnvoll, die Vorhersagen eines solchen Algorithmus nicht bloß als solche zu betrachten, sondern sie mit passenden Kontroll-Bedingungen zu vergleichen.

Die drei in dieser Dissertation vorgestellten Studien verwendeten ähnliche Paradigmen, wurden jedoch mit unterschiedlichen Methoden und mit einem Augenmerk auf unterschiedliche Aspekte durchgeführt. Im Folgenden werden sie kurz dargestellt.

15.2 Zusammenfassung von Studie 1

15.2.1 Hintergrund

In dieser ersten Teilstudie sollte die generelle Machbarkeit des oben skizzierten Ansatzes überprüft werden, also mit einem Musterklassifikationsalgorithmus das Wiederauftreten von stimulus-spezifischer neuronaler Aktivität zu entdecken.

15.2.2 Methoden

17 gesunde Probanden wurden gleichzeitig mit funktioneller Magnetresonanztomographie (fMRT) und Elektroenzephalographie (EEG) untersucht. Sie absolvierten zweimal eine identische Lernaufgabe, die sich nur in den zu lernenden Stimuli unterschied. Hierbei wurden je 16 Objekt-Ort-Assoziationen 30 mal präsentiert und die Probanden sollten sich merken, welches Objekt mit welchem Ort assoziiert war. Eine Lernaufgabe mit 16 Stimuli wurde vor einer Nickerchen-Phase durchgeführt, während der die Probanden versuchen sollten im MRT-Scanner einzuschlafen. Die

andere Lernaufgabe mit 16 neuen Stimuli wurde nach der Nickerchen-Phase durchgeführt. Im Anschluss an die zweite Lernaufgabe gab es einen Gedächtnistest, bei dem die 32 Objekte aus den beiden Lernaufgaben präsentiert wurden und die Probanden jeweils angeben mussten, an welcher Position das Objekt zuvor gezeigt worden war. Anhand der fMRT-Daten der beiden Lernaufgaben wurde ein Musterklassifikationsalgorithmus so trainiert, dass er die 32 verschiedenen Stimuli unterscheiden konnte. Die EEG-Daten wurden lediglich dazu verwendet, festzustellen ob die Probanden wach waren oder schliefen. Der Musterklassifikationsalgorithmus gab dann Vorhersagen auf die fMRT Messung während der Nickerchen-Phase ab. Es wurde erwartet, dass der Musterklassifikationsalgorithmus während der Nickerchen-Phase häufiger Objekte aus der ersten Lernaufgabe detektieren würde (da eine Reaktivierung der Objekte aus der nachfolgend absolvierten zweiten Lernaufgabe in dieser Phase nicht möglich war). Zudem erwarteten wir, dass die Häufigkeit, mit der einzelne Stimuli in der Nickerchen-Phase durch den Musterklassifikationsalgorithmus detektiert wurden, mit der Gedächtnisleistung in einem abschließenden Gedächtnistest zusammen hängen würden.

15.2.3 Ergebnisse

Der Musterklassifikationsalgorithmus konnte die 32 verschiedenen Stimuli mit einer Genauigkeit trennen, die deutlich über dem Zufallsniveau lag ($p < 0.0001$). Dennoch wurden 6 Probanden ausgeschlossen, bei denen der Musterklassifikationsalgorithmus nicht mit einer sehr hohen Genauigkeit die Stimuli trennen konnte. Ein weiterer Proband wurde ausgeschlossen, weil er das Experiment wegen einer Toilettenpause unterbrochen hatte. Somit konnten 10 Probanden für die weitere Analyse betrachtet werden. In Übereinstimmung mit den Vorhersagen zeigte sich in der Nickerchen-Phase, dass die Vorhersagen des Musterklassifikationsalgorithmus deutlich häufiger Objekte aus der ersten Lernaufgabe nannten als dies der Fall war, wenn Surrogat-Musterklassifikationsalgorithmen, die auf durcheinander gewürfelten Daten trainiert waren, Vorhersagen abgaben ($p = 0.006$). Dies kann als erster Beleg gewertet werden, dass der Musterklassifikationsalgorithmus nicht nur zufällige Vorhersagen macht, sondern einen Trend in der neuronalen Aktivität während der Nickerchen-Phase entdeckt, der für eine Reaktivierung spricht. Zudem wurde für Objekte aus der ersten Lernaufgabe die jeweilige Häufigkeit ermittelt, mit der sie vom Musterklassifikationsalgorithmus während der Nickerchen-Phase klassifiziert worden waren. Diese Häufigkeit wurde mit der Genauigkeit in Verbindung gesetzt, mit der ein Proband das Objekt seinem assoziierten Ort zuweisen konnte. Es zeigte sich, dass über die Probanden hinweg höhere Klassifikationshäufigkeit während der Nickerchen-Phase assoziiert war mit genauerem Gedächtnisabruf für die einzelnen Objekte ($p = 0.027$). Dies ist ein weiterer Beleg dafür, dass der Musterklassifikationsalgorithmus Reaktivierung entdeckt und dass die Häufigkeit dieser Reaktivierung Relevanz für die Gedächtnisleistung besitzt. Interessanterweise wurden die beiden hier beschriebenen Effekte hauptsächlich in denjenigen Abschnitten der Nickerchen-Phase gefunden, in denen die Probanden laut Schlafphasen-Einteilung wach waren.

15.2.4 Fazit

Diese erste Teilstudie liefert gute Belege für eine generelle Verwendbarkeit von Musterklassifikationsalgorithmen zur Detektion von stimulus-spezifischer Reaktivierung.

15.3 Zusammenfassung von Studie 2

15.3.1 Hintergrund

Nachdem in der ersten Teilstudie gezeigt werden konnte, dass Musterklassifikationsalgorithmen Reaktivierung in Ruhephasen entdecken können, wurde ein ähnliches Paradigma verwendet, um eine möglicherweise verstärkte Reaktivierung von emotional negativen Bildern zu untersuchen. Die Gedächtnisleistung für emotional negative Bilder ist in der Regel besser als die für emotional neutrale Bilder (Hamann, 2001; Kensinger und Corkin, 2003; Kensinger, 2004). Dies legt nahe, dass sie auch besser konsolidiert werden. Somit sollten sie auch häufiger reaktiviert werden. Daher wurde in dieser Studie erwartet, dass ein Musterklassifikationsalgorithmus, der emotional negative und emotional neutrale Bilder zu unterscheiden gelernt hatte, in einer Ruhephase nach der Lernaufgabe häufiger Vorhersagen für emotional negative Stimuli machen würde und dass die individuelle Häufigkeit der Vorhersagen erneut mit Gedächtnisleistung assoziiert sein würden. Da in der Vorgängerstudie die Effekte hauptsächlich im ruhigen Wachzustand gefunden worden waren, wurde für diese Studie nur eine Wachphase gemessen und folglich konnte auf ein simultan erhobenes EEG verzichtet werden.

15.3.2 Methoden

21 gesunde Probanden wurden mit fMRT gemessen, während sie eine Lernaufgabe durchliefen, die beinahe identisch zu der Lernaufgabe in der ersten Studie war. 12 emotional negative und 12 emotional neutrale Bilder wurden 24 Mal an einer bestimmten Position auf dem Bildschirm gezeigt und die Probanden sollten sich für jedes Bild die entsprechende Position einprägen. Vor der Lernaufgabe und danach wurden jeweils 30 Minuten Ruhemessung durchgeführt. Direkt nach der Lernaufgabe und dann erneut nach der zweiten Ruhemessung wurde ein Gedächtnistest durchgeführt, bei dem jedes Bild in der Mitte des Bildschirmes gezeigt wurde und bei dem die Probanden die Position angeben mussten, an der das Bild während der Lernaufgabe gezeigt worden war. Es wurde erwartet, dass der Musterklassifikationsalgorithmus in der zweiten Ruhephase (nach dem Lernen) aber nicht in der ersten Ruhephase (vor dem Lernen) häufiger Vorhersagen für negative Bilder treffen würde. Zudem wurde erwartet, dass die Vorhersagehäufigkeit für einzelne Bilder während der zweiten Ruhephase zusammenhängen würde mit der Gedächtnisleistung für diese Bilder während des zweiten Gedächtnistests.

15.3.3 Ergebnisse

Der Musterklassifikationsalgorithmus konnte die 24 verschiedenen Bilder mit einer Genauigkeit voneinander trennen, die deutlich über dem Zufallsniveau lag ($p < 0.0001$). Kein Proband musste aufgrund schlechter Musterklassifikationsgenauigkeit ausgeschlossen werden. Ein Proband musste wegen fehlerhafter Darbietung des Experimentes ausgeschlossen werden, so dass 20 Probanden für die Analyse übrig blieben. Überraschend war der fehlende Verhaltenseffekt: Die Position von emotional negativen Bilder wurden nicht besser im Gedächtnis behalten als die von emotional neutralen Bildern. Dies stellte die Erwartung in Frage, dass emotional negative Bilder auch häufiger reaktiviert werden sollten als emotional neutrale Bilder, da diese Annahme sich auf die erwartete bessere Gedächtnisleistung stützte. In der Tat gab es keinen Anhalt dafür, dass emotional negative Bilder vom Musterklassifikationsalgorithmus in der ersten oder der zweiten Ruhephase häufiger erkannt wurden als emotional neutrale Bilder. Das Gegenteil war der Fall: Emotional negative Bilder wurden sowohl in der ersten als auch in der zweiten Ruhephase seltener als in 50% der Fälle vom Musterklassifikationsalgorithmus vorhergesagt ($p = 0.0006$ und

$p = 0.0004$). Es gab auch keinen signifikanten Anstieg der Vorhersagehäufigkeit für emotional negative Bilder von der ersten in die zweite Ruhephase ($p = 0.1$). Es wurde jedoch ein Zusammenhang gefunden zwischen der Häufigkeit, mit der einzelne der gelernten Bilder vom Musterklassifikationsalgorithmus in der Ruhephase nach der Lernaufgabe detektiert wurden und der Gedächtnisleistung im zweiten Gedächtnistest ($p = 0.008$). Allerdings zeigte sich dieser Zusammenhang auch für die Häufigkeit der Vorhersagen in der ersten Ruhephase, in der noch keine Bilder gelernt worden waren ($p = 0.014$). Dies lässt sich nicht unmittelbar erklären.

15.3.4 Fazit

In dieser Studie wurde gezeigt, dass ein Musterklassifikationsalgorithmus 24 unterschiedliche Bilder mit hoher Genauigkeit voneinander trennen kann. Es zeigte sich keine höhere Häufigkeit der Vorhersage für emotional negative Bilder. Allerdings wurde auch kein Gedächtnisvorteil für emotional negative Bilder gefunden. Erneut zeigte sich ein Zusammenhang zwischen der Häufigkeit der Vorhersagen des Musterklassifikationsalgorithmus für einzelne Stimuli in einer Ruhephase nach der Lernaufgabe und der Gedächtnisleistung in einem anschließenden Gedächtnistest. Allerdings konnte ein solcher Zusammenhang auch für die Häufigkeit der Vorhersagen in einer Ruhephase vor der Lernaufgabe beobachtet werden, was die Ergebnisse relativiert.

15.4 Zusammenfassung von Studie 3

15.4.1 Hintergrund

Diese dritte Studie benutzte ein nahezu identisches Paradigma wie die erste Studie, verwendete jedoch statt fMRT als Messmethode intrakranielles EEG (iEEG) in Patienten der Klinik für Epileptologie in Bonn, die zur Abklärung eines chirurgischen Eingriffes die intrakranielle Elektroden implantiert bekommen hatten. Die gleiche Studie mit intrakraniellem EEG durchzuführen war deswegen von hohem Interesse, weil diese Messmethode über eine deutlich bessere zeitliche Auflösung verfügt und somit auch Muster entdeckt werden können, die sich schnell verändern. Zudem kann durch eine Zerlegung des iEEG Signals in verschiedene Frequenzbänder der Einfluss dieser Frequenzbänder auf die Dekodierung und Reaktivierung untersucht werden.

15.4.2 Methoden

12 Patienten mit medikamentös nicht behandelbarer Epilepsie nahmen während ihres Aufenthaltes auf der Station an dieser Studie teil. In einer Lernaufgabe sollten sie 16 Objekt-Ort-Assoziationen lernen. Jedes Paar wurde 30 Mal präsentiert. Vor der Lernaufgabe wurde eine komplette Nacht als Kontrollnacht gemessen. Am Abend darauf erfolgte die Lernaufgabe. Danach wurde eine zweite Nacht gemessen. Am Morgen nach dieser zweiten Nacht erfolgte ein Gedächtnistest, bei dem die Patienten für jedes der 16 Objekte den zugehörigen Ort angeben sollten. Erneut wurde ein Musterklassifikationsalgorithmus auf den Daten der Lernaufgabe so trainiert, dass er die einzelnen Objekte voneinander unterscheiden konnte. Dann machte er Vorhersagen auf die iEEG Daten der beiden Nächte. Es wurde erwartet, dass es zu einem Anstieg von Vorhersagen mit einer besonders hohen Konfidenz des Algorithmus von der ersten auf die zweite Nacht kommen würde. Außerdem wurde erwartet, dass die Vorhersagehäufigkeit für einzelne Objekte während der zweiten, aber nicht während der ersten Nacht, mit der Gedächtnisleistung beim Gedächtnistest assoziiert sein würde.

15.4.3 Ergebnisse

Die Klassifikationsgenauigkeit in dieser Studie war im Vergleich zu der Genauigkeit in den beiden fMRT-Studien eher gering, jedoch trotzdem besser als Zufallsniveau ($p < 0.0001$). Entgegen der Erwartungen fand sich kein Anstieg an Vorhersagen mit besonders hoher Konfidenz von der ersten auf die zweite Nacht. Dies änderte sich auch nicht, wenn einzelne Schlafphasen getrennt betrachtet wurden. Zudem wurden korrekt erinnerte Objekte auch nicht häufiger vom Musterklassifikationsalgorithmus detektiert als nicht korrekt erinnerte Objekte.

15.4.4 Fazit

In dieser Studie war die Genauigkeit des Musterklassifikationsalgorithmus sehr gering, was möglicherweise einen entscheidenden Grund darstellt für den Mangel an Evidenz zugunsten einer Reaktivierung von neuronaler Aktivität, die mit spezifischen Stimuli während einer Lernaufgabe assoziiert war.

15.5 Abschließende Zusammenfassung

Drei Studien wurden in dieser Dissertation durchgeführt, die ein Wiederauftreten von stimulus-spezifischer Aktivität während Schlaf- und Ruhephasen nach einer Lernaufgabe mithilfe von Musterklassifikationsalgorithmen untersuchten. In zwei von drei Studien wurde gefunden, dass ein Musterklassifikationsalgorithmus verschiedene Stimuli, die während einer Lernaufgabe präsentiert wurden, mit sehr guter Genauigkeit voneinander trennen kann. In der ersten Studie gibt es gute Belege dafür, dass Stimuli aus einem zuvor gelernten Set in einer anschließenden Ruhephase häufiger vom Musterklassifikationsalgorithmus detektiert werden als Stimuli aus einem Set, das erst nach der Ruhephase gelernt wurde. Dies ist ein erster Beleg dafür, dass sich die Methode grundsätzlich dazu eignet, Reaktivierung zu entdecken. In zwei von drei Studien konnte außerdem ein Zusammenhang gezeigt werden zwischen der Häufigkeit des vom Musterklassifikationsalgorithmus detektierten Wiederauftretens für einzelne Stimuli und der Gedächtnisleistung für diese Stimuli während eines Gedächtnistests. In der zweiten Studie muss dieser Befund jedoch mit Vorsicht interpretiert werden: Er zeigt sich sowohl für eine Ruhephase nach der Lernaufgabe, als auch für eine Ruhephase vor der Lernaufgabe, in der per Definition kein Wiederauftreten stattfinden kann. Die Ergebnisse der letzten Studie, die als einzige Studie intrakranielles EEG statt fMRT als Messmethode benutzte, weichen stark von den Ergebnissen der ersten beiden Studien ab. Die Genauigkeit des Musterklassifikationsalgorithmus beim Trennen der einzelnen Objekte war nicht zufriedenstellend und es wurde kein Hinweis auf vermehrtes Wiederauftreten gefunden. Dies mag an der anderen Messmethode, der schlechteren Genauigkeit oder eines nicht optimal konstruierten experimentellen Ablaufes liegen.

Zusammenfassend lässt sich jedoch sagen, dass zumindest für fMRT Daten die Untersuchung von stimulus-spezifischer Reaktivierung mittels Musterklassifikationsalgorithmen in Schlaf- und Ruhephasen möglich ist und ein spannendes neues Instrument darstellen könnte, um Konsolidierungsprozesse beim Menschen besser zu verstehen. Die Methode sollte in weiteren Studien genauer validiert und auf ihre Möglichkeiten und Grenzen hin untersucht werden.

Glossary

blood oxygenation level dependent signal

blood with different degrees of oxygenation emits different fMRI signals; changes in the oxygenation of blood are associated with local changes in neuronal activity; the fluctuation of oxygenation over time can be measured with fMRI and related to psychological states.

classifier

any algorithm that is employed to differentiate between two or more distinct classes.

classifier accuracy

a measure of how well a classifier can distinguish classes in a dataset; is yielded by comparing classifier predictions to the actual labels of samples; usually, classifier accuracy is assessed with data that were not included in the training of the classifier.

cross-validation

in MVPA: a method to assess how well a classifier generalizes to new data; usually, a dataset is split into training data and validation data; the classifier is trained on the training data only and makes predictions on the validation data; the degree of overlap between classifier predictions on the validation data and the actual correct labels of the validation data yields a measure of classifier accuracy.

electrocardiography

non-invasive measurement of the electrical activity of the heart.

electroencephalography

scalp measurement of electric potentials which are generated by hundreds of thousands of neurons.

electromyography

non-invasive measurement of muscle activity; in sleep-staging, EMG recording of chin muscle tone can provide information about rapid eye movement sleep.

electrooculography

non-invasive measurement of activity of eye-muscles; can be used to detect eye movements.

event-related potential

method of analysis that is often used for electrophysiological data in which a time series is segmented into epochs that are cut out around specific events of an experiment; all epochs from the same condition are then averaged together, which is thought to eliminate noise and accentuate the real part of the data; ERPs often contain typical components (e.g. peaks and troughs at specific times after stimulus-onset); the averages of different conditions may also be compared to one another.

feature selection

in MVPA: the process by which only those features of a dataset are selected that are thought to be useful for distinguishing the classes; is especially important in datasets with a high number of potential features in comparison to a low number of training samples in order to avoid overfitting; during cross-validation, it should be done on the training dataset only.

features

in MVPA: properties which qualify or quantify aspects of the classes in a classification problem.

functional magnetic resonance imaging

imaging method which uses differences in blood oxygenation as a marker of neuronal activation.

general linear model

statistical analysis which is often used for fMRI data; determines the influence of different factors (regressors) on the activity in a single voxel.

medial temporal lobe

part of the brain which includes several subregions such as the hippocampus, amygdala, parahippocampal cortex, entorhinal cortex and perirhinal cortex;

is acknowledged to play an important part in memory formation, especially episodic memory.

memory consolidation

the progressive postacquisition stabilization of long-term memory (Dudai, 2004).

multi-variate pattern analysis

umbrella term for a number of analysis methods which take into account many properties of a dataset simultaneously; is often used to highlight the difference to a univariate approach; popular methods include pattern classification algorithms (such as linear SVMs) or representational similarity analysis.

rapid eye movement sleep

sleep stage that is characterized by rapid eye movements, flat muscle tone and high-frequency, low-amplitude EEG; occurs more often during the second half of a night.

slow-wave sleep

part of normal sleep, refers to sleep stages 3 and 4 (Rechtschaffen et al., 1968), often called "deep" sleep; is characterized by low-frequency, high-amplitude delta wave-forms; occurs predominantly in the first half of the night.

synaptic consolidation

memory stabilization that happens at a synaptic level; it "is complete within hours after learning, and involves the stabilization of changes in synaptic connectivity in localized circuits" (Frankland and Bontempi, 2005, p. 119).

system consolidation

memory stabilization that happens at a time-scale of weeks and years and "involves gradual reorganization of the brain regions that support memory", which "may involve a time-dependent shift in the circuits that support memory recall (Frankland and Bontempi, 2005, p. 119).

training

in MVPA: the process in which a classification algorithm is confronted with data that consists of already labeled samples; based on this training data, the classifier finds a decision rule for classifying the different classes.

volume

in MRI, refers to a 3D image; in fMRI, one volume is typically recorded every 1-3 seconds.

voxel

in MRI, refers to a unit in a 3D image of the brain – equivalent to a pixel in 2D images.

Acronyms

BOLD	blood oxygenation level dependent.
ECG	electrocardiography.
EEG	electroencephalography.
EMG	electromyography.
EOG	electrooculography.
ERP	event-related potential.
fMRI	functional magnetic resonance imaging.
GLM	general linear model.
MRI	magnetic resonance imaging.
MTL	medial temporal lobe.
MVPA	multi-variate pattern analysis.
REM	rapid eye movement.
SVM	support vector machine.
SWS	slow-wave sleep.

List of Figures

List of Tables

References

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(3):291–300.

Adolphs, R., Tranel, D., Hamann, S., Young, A., Calder, A., Phelps, E., Anderson, A., Lee, G., and Damasio, A. (1999). Recognition of facial emotion in nine individuals with bilateral amygdala damage. *Neuropsychologia*, 37(10):1111–1117.

Aeschbach, D., Cutler, A. J., and Ronda, J. M. (2008). A role for non-rapid-eye-movement sleep homeostasis in perceptual learning. *Journal of Neuroscience*, 28(11):2766–2772.

Anderson, M. C., Bjork, R. A., and Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(5):1063–1087.

Axmacher, N., Haupt, S., Fernandez, G., Elger, C. E., and Fell, J. (2008). The role of sleep in declarative memory consolidation–direct evidence by intracranial eeg. *Cerebral Cortex*, 18(3):500–507.

Baddeley, A. D., Eysenck, M. W., and Anderson, M. C. (2009). *Memory.* Psychology Press, Hove [England], New York.

Barrett, T. R. and Ekstrand, B. R. (1972). Effect of sleep on memory. 3. controlling for time-of-day effects. *Journal of Experimental Psychology*, 96(2):321–327.

Beeman, C. L., Bauer, P. S., Pierson, J. L., and Quinn, J. J. (2013). Hippocampus and medial prefrontal cortex contributions to trace and contextual fear memory expression over time. *Learning & Memory*, 20(6):336–343.

Benca, R. M. (1992). Sleep and psychiatric disorders - a meta-analysis. *Archives of General Psychiatry*, 49(8):651.

Bergmann, T. O., Mölle, M., Diedrichs, J., Born, J., and Siebner, H. R. (2012). Sleep spindle-related reactivation of category-specific cortical regions after learning face-scene associations. *Neuroimage*, 59(3):2733–2742.

Birbaumer, N. and Schmidt, R. F. (2010). *Biologische Psychologie: Mit 44 Tabellen.* Springer-Lehrbuch. Springer-Medizin-Verl, Heidelberg, 7., überarb. und erg edition.

Bishop, C. M. (2009). *Pattern recognition and machine learning.* Information science and statistics. Springer, New York, NY, corr. at 8. printing edition.

Bode, S. and Haynes, J.-D. (2009). Decoding sequential stages of task preparation in the human brain. *Neuroimage*, 45(2):606–613.

Bonnici, H. M., Chadwick, M. J., Kumaran, D., Hassabis, D., Weiskopf, N., and Maguire, E. A. (2012). Multi-voxel pattern analysis in human hippocampal subfields. *Frontiers in Human Neuroscience*, 6:290.

Burnham, W. H. (1903). Retroactive amnesia: illustrative cases and a tentative explanation. *The American Journal of Psychology*, (14):118–132.

Cahill, L., Babinsky, R., Markowitsch, H. J., and McGaugh, J. L. (1995). The amygdala and emotional memory. *Nature*, 377(6547):295–296.

Cahill, L., Haier, R. J., Fallon, J., Alkire, M. T., Tang, C., Keator, D., Wu, J., and McGaugh, J. L. (1996). Amygdala activity at encoding correlated with long-term, free recall of emotional information. *Proceedings of the National Academy of Sciences of the United States of America*, 93(15):8016–8021.

Cahill, L. and McGaugh, J. L. (1998). Mechanisms of emotional arousal and lasting declarative memory. *Trends in Neurosciences*, 21(7):294–299.

Canli, T., Zhao, Z., Brewer, J., Gabrieli, J. D., and Cahill, L. (2000). Event-related activation in the human amygdala associates with later memory for individual emotional experience. *Journal of Neuroscience*, 20(19):RC99.

Canolty, R. T. and Knight, R. T. (2010). The functional role of cross-frequency coupling. *Trends in Cognitive Sciences*, 14(11):506–515.

Carr, M., Jadhav, S., and Frank, L. (2011). Hippocampal replay in the awake state: a potential substrate for memory consolidation and retrieval. *Nature Neuroscience*, 14(2):147–153.

Ciranni, M. A. and Shimamura, A. P. (1999). Retrieval-induced forgetting in episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(6):1403–1414.

Clark, R. E., Broadbent, N. J., Zola, S. M., and Squire, L. R. (2002). Anterograde amnesia and temporally graded retrograde amnesia for a nonspatial memory task after lesions of hippocampus and subiculum. *Journal of Neuroscience*, 22(11):4663–4669.

Corkin, S. (2002). What's new with the amnesic patient h.m.? *Nature Reviews Neuroscience*, 3(2):153–160.

Cox, D. D. and Savoy, R. L. (2003). Functional magnetic resonance imaging (fmri) "brain reading": detecting and classifying distributed patterns of fmri activity in human visual cortex. *Neuroimage*, 19(2 Pt 1):261–270.

Dang-Vu, T. T., Schabus, M., Desseilles, M., Albouy, G., Boly, M., Darsaud, A., Gais, S., Rauchs, G., Sterpenich, V., Vandewalle, G., Carrier, J., Moonen, G., Balteau, E., Degueldre, C., Luxen, A., Phillips, C., and Maquet, P. (2008). Spontaneous neural activity during human slow wave sleep. *Proceedings of the National Academy of Sciences of the United States of America*, 105(39):15160–15165.

Diba, K. and Buzsáki, G. (2007). Forward and reverse hippocampal place-cell sequences during ripples. *Nature Neuroscience*, 10(10):1241–1242.

Diekelmann, S. and Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience*, 11(2):114–126.

Diekelmann, S., Büchel, C., Born, J., and Rasch, B. (2011). Labile or stable: opposing consequences for memory when reactivated during waking and sleep. *Nature Neuroscience*, 14(3):381–386.

Doppelmayr, M., Klimesch, W., Stadler, W., Pöllhuber, D., and Heine, C. (2002). Eeg alpha power and intelligence. *Intelligence*, 30(3):289–302.

Dragoi, G. and Tonegawa, S. (2010). Preplay of future place cell sequences by hippocampal cellular assemblies. *Nature*, 469(7330):397–401.

Drosopoulos, S., Schulze, C., Fischer, S., and Born, J. (2007). Sleep's function in the spontaneous recovery and consolidation of memories. *Journal of Experimental Psychology: General*, 136(2):169–183.

Duda, R. O., Hart, P. E., and Stork, D. G. (2001). *Pattern classification.* Wiley, New York, 2 edition.

Dudai, Y. (2004). The neurobiology of consolidations, or, how stable is the engram? *Annual Review of Psychology*, 55(1):51–86.

Dupret, D., O'Neill, J., Pleydell-Bouverie, B., and Csicsvari, J. (2010). The reorganization and reactivation of hippocampal maps predict spatial memory performance. 13(8):995–1002.

Ebbinghaus, H. (1885). *Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie.*

Ekstrand, B. R. (1967). Effect of sleep on memory. *Journal of Experimental Psychology*, 75(1):64–72.

Ellenbogen, J., Hulbert, J., Stickgold, R., Dinges, D., and Thompson-Schill, S. (2006a). Interfering with theories of sleep and memory: sleep, declarative memory, and associative interference. *Current Biology*, 16(13):1290–1294.

Ellenbogen, J. M., Hu, P. T., Payne, J. D., Titone, D., and Walker, M. P. (2007). From the cover: Human relational memory requires time and sleep. *Proceedings of the National Academy of Sciences of the United States of America*, 104(18):7723–7728.

Ellenbogen, J. M., Payne, J. D., and Stickgold, R. (2006b). The role of sleep in declarative memory consolidation: passive, permissive, active or none? *Current Opinion in Neurobiology*, 16(6):716–722.

Ethofer, T., Van De Ville, D., Scherer, K., and Vuilleumier, P. (2009). Decoding of emotional information in voice-sensitive cortices. *Current Biology*, 19(12):1028–1033.

Ficca, G. and Salzarulo, P. (2004). What in sleep is for memory. *Sleep Medicine*, 5(3):225–230.

Fischer, S., Drosopoulos, S., Tsen, J., and Born, J. (2006). Implicit learning – explicit knowing: a role for sleep in memory system interaction. *Journal of Cognitive Neuroscience*, 18(3):311–319.

Fischer, S., Hallschmid, M., Elsner, A. L., and Born, J. (2002). Sleep forms memory for finger skills. *Proceedings of the National Academy of Sciences of the United States of America*, 99(18):11987–11991.

Fogel, S. M. and Smith, C. T. (2006). Learning-dependent changes in sleep spindles and stage 2 sleep. *Journal of sleep research*, 15(3):250–255.

Fogel, S. M., Smith, C. T., and Cote, K. A. (2007). Dissociable learning-dependent changes in rem and non-rem sleep in declarative and procedural memory systems. *Behavioural Brain Research*, 180(1):48–61.

Foster, D. J. and Wilson, M. A. (2006). Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature*, 440(7084):680–683.

Frankland, P. W. and Bontempi, B. (2005). The organization of recent and remote memories. *Nature Reviews Neuroscience*, 6(2):119–130.

Gais, S. and Born, J. (2004). Low acetylcholine during slow-wave sleep is critical for declarative memory consolidation. *Proceedings of the National Academy of Sciences of the United States of America*, 101(7):2140–2144.

Gais, S., Plihal, W., Wagner, U., and Born, J. (2000). Early sleep triggers memory for early visual discrimination skills. *Nature Neuroscience*, 3(12):1335–1339.

Gennaro, L. d. and Ferrara, M. (2003). Sleep spindles: an overview. *Sleep Medicine Reviews*, 7(5):423–440.

Giuditta, A., Ambrosini, M. V., Montagnese, P., Mandile, P., Cotugno, M., Grassi Zucconi, G., and Vescia, S. (1995). The sequential hypothesis of the function of sleep. *Behavioural Brain Research*, 69(1-2):157–166.

Hamann, S. (2001). Cognitive and neural mechanisms of emotional memory. *Trends in Cognitive Sciences*, 5(9):394–400.

Hanke, M., Halchenko, Y. O., Sederberg, P. B., Hanson, S. J., Haxby, J. V., and Pollmann, S. (2009a). Pymvpa: a python toolbox for multivariate pattern analysis of fmri data. *Neuroinformatics*, 7(1):37–53.

Hanke, M., Halchenko, Y. O., Sederberg, P. B., Olivetti, E., Fruend, I., Rieger, J. W., Herrmann, C. S., Haxby, J. V., Hanson, S. J., and Pollmann, S. (2009b). Pymvpa: a unifying approach to the analysis of neuroscientific data. *Frontiers in Neuroinformatics*, 3(0).

Harrison, S. A. and Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, 458(7238):632–635.

Hasselmo (1999). Neuromodulation: acetylcholine and memory consolidation. *Trends in Cognitive Sciences*, 3(9):351–359.

Hasselmo, M. (2005). What is the function of hippocampal theta rhythm? linking behavioral data to phasic properties of field potential and unit recording data. *Hippocampus*, 15(7):936–949.

Haxby, J., Gobbini, M., Furey, M., Ishai, A., Schouten, J., and Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293(5539):2425–2430.

Haynes, J.-D. and Rees, G. (2005a). Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nature Neuroscience*, 8(5):686–691.

Haynes, J.-D. and Rees, G. (2005b). Predicting the stream of consciousness from activity in human visual cortex. *Current Biology*, 15(14):1301–1307.

Haynes, J.-D. and Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience*, 7(7):523–534.

Haynes, J.-D., Sakai, K., Rees, G., Gilbert, S., Frith, C., and Passingham, R. E. (2007). Reading hidden intentions in the human brain. *Current Biology*, 17(4):323–328.

Hebb, D. O. (1949). The organization of behavior; a neuropsychological theory.

Henson, R. N. A., Shallice, T., and Dolan, R. J. (1999). Right prefrontal cortex and episodic memory retrieval: a functional mri test of the monitoring hypothesis. *Brain*, 122(7):1367–1381.

Hu, P., Stylos-Allan, M., and Walker, M. P. (2006). Sleep facilitates consolidation of emotional declarative memory. *Psychological Science*, 17(10):891–898.

Huber, R., Ghilardi, M. F., Massimini, M., and Tononi, G. (2004). Local sleep and learning. *Nature*, 430(6995):78–81.

Huettel, S. A., Song, A. W., and McCarthy, G. (2008). *Functional magnetic resonance imaging.* Sinauer Associates, Sunderland, Mass, 2nd edition.

Hupbach, A., Gomez, R., Hardt, O., and Nadel, L. (2007). Reconsolidation of episodic memories: A subtle reminder triggers integration of new information. *Learning & Memory*, 14(1-2):47–53.

Iber, C., Ancoli-Israel, S., Chesson, A., and Quan, S. (2007). *The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specification.* American Academy of Sleep Medicine, Westchester, IL.

Jadhav, S. P., Kemere, C., German, P. W., and Frank, L. M. (2012). Awake hippocampal sharp-wave ripples support spatial memory. *Science.*

Jenkins, J. G. and Dallenbach, K. M. (1924). Obliviscence during sleep and waking. *The American Journal of Psychology*, 35(4):605–612.

Jensen, O., Bonnefond, M., and VanRullen, R. (2012). An oscillatory mechanism for prioritizing salient unattended stimuli. *Trends in Cognitive Sciences*, 16(4):200–206.

Jensen, O. and Colgin, L. L. (2007). Cross-frequency coupling between neuronal oscillations. *Trends in Cognitive Sciences*, 11(7):267–269.

Ji, D. and Wilson, M. A. (2007). Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nature Neuroscience*, 10(1):100–107.

Johnson, J. D., McDuff, S. G. R., Rugg, M. D., and Norman, K. A. (2009). Recollection, familiarity, and cortical reinstatement: A multivoxel pattern analysis. *Neuron*, 63(5):697–708.

Káli, S. and Dayan, P. (2004). Off-line replay maintains declarative memories in a model of hippocampal-neocortical interactions. *Nature Neuroscience*, 7(3):286–294.

Karlsson, M. P. and Frank, L. M. (2009). Awake replay of remote experiences in the hippocampus. *Nature Neuroscience*, 12(7):913–918.

Kay, K. N., Naselaris, T., Prenger, R. J., and Gallant, J. L. (2008). Identifying natural images from human brain activity. 452(7185):352–355.

Kensinger, E. A. (2004). Remembering emotional experiences: the contribution of valence and arousal. *Reviews in the Neurosciences*, 15(4):241–251.

Kensinger, E. A. and Corkin, S. (2003). Memory enhancement for emotional words: are emotional words more vividly remembered than neutral words? *Memory & Cognition*, 31(8):1169–1180.

Kim, J. J. and Fanselow, M. S. (1992). Modality-specific retrograde amnesia of fear. *Science*, 256(5057):675–677.

Kleinsmith, L. J. and Kaplan, S. (1963). Paired-associate learning as a function of arousal and interpolated interval. *Journal of Experimental Psychology*, 65(2):190–193.

Klimesch, W. (1999). Eeg alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Research Reviews*, 29(2-3):169–195.

Kloet, E. R. d., Vreugdenhil, E., Oitzl, M. S., and Joëls, M. (1998). Brain corticosteroid receptor balance in health and disease. *Endocrine Reviews*, 19(3):269–301.

Knowlton, B. J., Squire, L. R., and Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learning & memory (Cold Spring Harbor, N.Y.)*, 1(2):106–120.

Kohrman, M. H. (2007). What is epilepsy? clinical perspectives in the diagnosis and treatment. *Journal of Clinical Neurophysiology*, 24(2):87–95.

Kriegeskorte, N., Mur, M., and Bandettini, P. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, 2:4.

Krishnapuram, B., Carin, L., Figueiredo, M. A. T., and Hartemink, A. J. (2005). Sparse multinomial logistic regression: fast algorithms and generalization bounds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(6):957–968.

Kyung Lee, E. and Douglass, A. B. (2010). Sleep in psychiatric disorders: where are we now? *Canadian Journal of Psychiatry*, 55(7):403–412.

LaBar, K. S. and Cabeza, R. (2006). Cognitive neuroscience of emotional memory. *Nature Reviews Neuroscience*, 7(1):54–64.

Lahl, O., Wispel, C., Willigens, B., and Pietrowsky, R. (2008). An ultra short episode of sleep is sufficient to promote declarative memory performance. *Journal of sleep research*, 17(1):3–10.

Lang, P. J., Bradley, M. M., and Cuthbert, B. N. (1999). International affective picture system (iaps): Technical manual and affective ratings.

Lantz, G., Grave Peralta, R. d., Spinelli, L., Seeck, M., and Michel, C. M. (2003). Epileptic source localization with high density eeg: how many electrodes are needed? *Clinical Neurophysiology*, 114(1):63–69.

Lau, H., Tucker, M. A., and Fishbein, W. (2010). Daytime napping: Effects on human direct associative and relational memory. *Neurobiology of Learning and Memory*, 93(4):554–560.

Lee, A. K. and Wilson, M. A. (2002). Memory of sequential experience in the hippocampus during slow wave sleep. *Neuron*, 36(6):1183–1194.

Lisman, J. and Morris, R. G. (2001). Memory. why is the cortex a slow learner? *Nature*, 411(6835):248–249.

Louie, K. and Wilson, M. A. (2001). Temporally structured replay of awake hippocampal ensemble activity during rapid eye movement sleep. *Neuron*, 29(1):145–156.

Luck, Woodman, and Vogel (2000). Event-related potential studies of attention. *Trends in Cognitive Sciences*, 4(11):432–440.

Maguire, E. A., Henson, R. N. A., Mummery, C. J., and Frith, C. D. (2001). Activity in prefrontal cortex, not hippocampus, varies parametrically with the increasing remoteness of memories. *Neuroreport*, 12(3):441–444.

Makeig, S., Debener, S., Onton, J., and Delorme, A. (2004). Mining event-related brain dynamics. *Trends in Cognitive Sciences*, 8(5):204–210.

Maris, E. and Oostenveld, R. (2007). Nonparametric statistical testing of eeg- and meg-data. *Journal of Neuroscience Methods*, 164(1):177–190.

Marr, D. (1970). A theory for cerebral neocortex. *Proceedings of the National Academy of Sciences of the United States of America*, 176(1043):p 161–234.

Marr, D. (1971). Simple memory: a theory for archicortex. *Philosophical Transactions of the Royal Society of London*, 262(841):23–81.

Mather, M. (2007). Emotional arousal and memory binding: An object-based framework. *Perspectives on Psychological Science*, 2(1):33–52.

Mather, M., Mitchell, K. J., Raye, C. L., Novak, D. L., Greene, E. J., and Johnson, M. K. (2006). Emotional arousal can impair feature binding in working memory. *Journal of Cognitive Neuroscience*, 18(4):614–625.

Mather, M. and Nesmith, K. (2008). Arousal-enhanced location memory for pictures. *Journal of Memory and Language*, 58(2):449–464.

McClelland, J. L., McNaughton, B. L., and O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102(3):419–457.

McGaugh, J. L. (2000). Memory–a century of consolidation. *Science*, 287(5451):248–251.

McGaugh, J. L. (2004). The amygdala modulates the consolidation of memories of emotionally arousing experiences. *Annual Review of Neuroscience*, 27:1–28.

Michel, C. M. and Murray, M. M. (2012). Towards the utilization of eeg as a brain imaging tool. *Neuroimage*, 61(2):371–385.

Michel, C. M., Murray, M. M., Lantz, G., Gonzalez, S., Spinelli, L., and Grave Peralta, R. d. (2004). Eeg source imaging. *Clinical Neurophysiology*, 115(10):2195–2222.

Morris, C. D., Bransford, J. D., and Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, 16(5):519–533.

Mueller, G. E. and Pilzecker, A. (1900). Experimentelle beitraege zur lehre vom gedaechtnis. *Zeitschrift fur Psychologie*, 1:1–300.

Nádasdy, Z., Hirase, H., Czurkó, A., Csicsvari, J., and Buzsáki, G. (1999). Replay and time compression of recurring spike sequences in the hippocampus. *Journal of Neuroscience*, 19(21):9497–9507.

Newman, S. M., Paletz, E. M., Rattenborg, N. C., Obermeyer, W. H., and Benca, R. M. (2008). Sleep deprivation in the pigeon using the disk-over-water method. *Physiology & Behavior*, 93(1-2):50–58.

Nishida, M., Pearsall, J., Buckner, R. L., and Walker, M. P. (2009). Rem sleep, prefrontal theta, and the consolidation of human emotional memory. *Cerebral Cortex*, 19(5):1158–1166.

Norman, K. A., Polyn, S. M., Detre, G. J., and Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fmri data. *Trends in Cognitive Sciences*, 10(9):424–430.

Northoff, G. (2011). Self and brain: what is self-related processing? *Trends in Cognitive Sciences*, 15(5):186–187.

Okada, G., Okamoto, Y., Kunisato, Y., Aoyama, S., Nishiyama, Y., Yoshimura, S., Onoda, K., Toki, S., Yamashita, H., Yamawaki, S., and Harrison, B. J. (2011). The effect of negative and positive emotionality on associative memory: An fmri study. *PLoS ONE*, 6(9):e24862.

O'Keefe, J. and Dostrovsky, J. (1971). The hippocampus as a spatial map. preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, 34(1):171–175.

Onoda, K., Okamoto, Y., and Yamawaki, S. (2009). Neural correlates of associative memory: the effects of negative emotion. *Neuroscience Research*, 64(1):50–55.

Osipova, D., Takashima, A., Oostenveld, R., Fernández, G., Maris, E., and Jensen, O. (2006). Theta and gamma oscillations predict encoding and retrieval of declarative memory. *Journal of Neuroscience*, 26(28):7523–7531.

Pascual-Marqui, R. D. (1999). Review of methods for solving the eeg inverse problem. *International Journal of Bioelectromagnetism*, 1(1):75–86.

Payne, J. D., Stickgold, R., Swanberg, K., and Kensinger, E. A. (2008). Sleep preferentially enhances memory for emotional components of scenes. *Psychological Science*, 19(8):781–788.

Peigneux, P., Orban, P., Balteau, E., Degueldre, C., Luxen, A., Laureys, S., and Maquet, P. (2006). Offline persistence of memory-related cerebral activity during active wakefulness. *PLoS Biology*, 4(4):e100.

Penfield, W. and Milner, B. (1958). Memory deficit produced by bilateral lesions in the hippocampal zone. *A.M.A. Archives of Neurology and Psychiatry*, 79(5):475–497.

Phan, K. L., Wager, T., Taylor, S. F., and Liberzon, I. (2002). Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in pet and fmri. *Neuroimage*, 16(2):331–348.

Phan, K. L., Wager, T. D., Taylor, S. F., and Liberzon, I. (2004). Functional neuroimaging studies of human emotions. *CNS Spectrums*, 9(4):258–266.

Plihal, W. and Born, J. (1997). Effects of early and late nocturnal sleep on declarative and procedural memory. *Journal of Cognitive Neuroscience*, 9(4):534–547.

Plihal, W. and Born, J. (1999). Effects of early and late nocturnal sleep on priming and spatial memory. *Psychophysiology*, 36(5):571–582.

Polyn, S. M., Natu, V. S., Cohen, J. D., and Norman, K. A. (2005). Category-specific cortical activity precedes retrieval during memory search. *Science*, 310(5756):1963–1966.

Qin, P. and Northoff, G. (2011). How is our self related to midline regions and the default-mode network? *Neuroimage*, 57(3):1221–1233.

Rasch, B., Buchel, C., Gais, S., and Born, J. (2007). Odor cues during slow-wave sleep prompt declarative memory consolidation. *Science*, 315(5817):1426–1429.

Rasch, B. H., Born, J., and Gais, S. (2006). Combined blockade of cholinergic receptors shifts the brain from stimulus encoding to memory consolidation. *Journal of Cognitive Neuroscience*, 18(5):793–802.

Rauchs, G., Bertran, F., Guillery-Girard, B., Desgranges, B., Kerrouche, N., Denise, P., Foret, J., and Eustache, F. (2004). Consolidation of strictly episodic memories mainly requires rapid eye movement sleep. *Sleep*, 27(3):395–401.

Rechtschaffen, A. and Bergmann, B. M. (2002). Sleep deprivation in the rat: an update of the 1989 paper. *Sleep*, 25(1):18–24.

Rechtschaffen, A., Kales, A., University of California, L. A. B. I. S., and NINDB Neurological Information Network (1968). *A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects.* US Dept. of Health, Education, and Welfare Bethesda, Md.

Ribot, T. (1882). *Diseases of Memory.* Univ. Publ. Am., reprint (1972), Washington, DC.

Rissman, J. and Wagner, A. D. (2012). Distributed representations in memory: Insights from functional brain imaging. *Annual Review of Psychology*, 63(1):101–128.

Ritter, P. and Villringer, A. (2006). Simultaneous eeg–fmri. *Neuroscience & Biobehavioral Reviews*, 30(6):823–838.

Rolls, A., Colas, D., Adamantidis, A., Carter, M., Lanre-Amos, T., Heller, H. C., and Lecea, L. d. (2011). Optogenetic disruption of sleep continuity impairs memory consolidation. *Proceedings of the National Academy of Sciences of the United States of America*, 108(32):13305–13310.

Ross, J. J. (1965). Neurological findings after prolonged sleep deprivation. *Archives of Neurology*, 12(4):399–403.

Roth, T., Benca, R. M., and Erman, M. (2010). An introduction to the clinical correlates of disrupted slow-wave sleep. *The Journal of Clinical Psychiatry*, 71(4):e09.

Rudoy, J. D., Voss, J. L., Westerberg, C. E., and Paller, K. A. (2009). Strengthening individual memories by reactivating them during sleep. *Science*, 326(5956):1079.

Rugg, M. D., Fletcher, P. C., Frith, C. D., Frackowiak, R. S. J., and Dolan, R. J. (1996). Differential activation of the prefrontal cortex in successful and unsuccessful memory retrieval. *Brain*, 119(6):2073–2083.

Sanders, H. I. and Warrington, E. K. (1971). Memory for remote events in amnesic patients. *Brain*, 94(4):661–668.

Scoville, W. B. and Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery, and Psychiatry*, 20(1):11–21.

Sharot, T. and Phelps, E. A. (2004). How arousal modulates memory: disentangling the effects of attention and retention. *Cognitive, Affective & Behavioral Neuroscience*, 4(3):294–306.

Siegel, J. M. (2001). The rem sleep-memory consolidation hypothesis. *Science*, 294(5544):1058–1063.

Siegel, J. M. (2009). Sleep viewed as a state of adaptive inactivity. *Nature Reviews Neuroscience*, 10(10):747–753.

Skaggs, W. E. and McNaughton, B. L. (1996). Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience. *Science*, 271(5257):1870–1873.

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E., Johansen-Berg, H., Bannister, P. R., De Luca, M., Drobnjak, I., Flitney, D. E., Niazy, R. K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J. M., and Matthews, P. M. (2004). Advances in functional and structural mr image analysis and implementation as fsl. *Neuroimage*, 23(Supplement 1):S208 – S219.

Squire, L. R. and Alvarez, P. (1995). Retrograde amnesia and memory consolidation: a neurobiological perspective. *Current Opinion in Neurobiology*, 5(2):169–177.

Squire, L. R., Clark, R. E., and Knowlton, B. J. (2001). Retrograde amnesia. *Hippocampus*, 11(1):50–55.

Squire, L. R., Stark, C. E. L., and Clark, R. E. (2004). The medial temporal lobe. *Annual Review of Neuroscience*, 27:279–306.

Squire, L. R. and Zola, S. M. (1996). Structure and function of declarative and non-declarative memory systems. *Proceedings of the National Academy of Sciences of the United States of America*, 93(24):13515–13522.

Staresina, B. P. and Davachi, L. (2009). Mind the gap: Binding experiences across space and time in the human hippocampus. *Neuron*, 63(2):267–276.

Stein, B. S. (1978). Depth of processing reexamined: The effects of the precision of encoding and test appropriateness. *Journal of Verbal Learning and Verbal Behavior*, 17(2):165–174.

Stickgold, R. and Walker, M. P. (2005). Memory consolidation and reconsolidation: what is the role of sleep? *Trends in Neurosciences*, 28(8):408–415.

Stickgold, R., Whidbee, D., Schirmer, B., Patel, V., and Hobson, J. A. (2000). Visual discrimination task improvement: A multi-step process occurring during sleep. *Journal of Cognitive Neuroscience*, 12(2):246–254.

Takashima, A., Petersson, K. M., Rutters, F., Tendolkar, I., Jensen, O., Zwarts, M. J., McNaughton, B. L., and Fernandez, G. (2006). Declarative memory consolidation in humans: a prospective functional magnetic resonance imaging study. *Proceedings of the National Academy of Sciences of the United States of America*, 103(3):756–761.

Takehara, K., Kawahara, S., and Kirino, Y. (2003). Time-dependent reorganization of the brain components underlying memory retention in trace eyeblink conditioning. *Journal of Neuroscience*, 23(30):9897–9905.

Takeuchi, D., Hirabayashi, T., Tamura, K., and Miyashita, Y. (2011). Reversal of interlaminar signal between sensory and memory processing in monkey temporal cortex. *Science*, 331(6023):1443–1447.

Tambini, A., Ketz, N., and Davachi, L. (2010). Enhanced brain correlations during rest are related to memory for recent experiences. *Neuron*, 65(2):280–290.

Tamminen, J., Payne, J. D., Stickgold, R., Wamsley, E. J., and Gaskell, M. G. (2010). Sleep spindle activity is associated with the integration of new memories and existing knowledge. *Journal of Neuroscience*, 30(43):14356–14360.

Tucker, M., Hirota, Y., Wamsley, E., Lau, H., Chaklader, A., and Fishbein, W. (2006). A daytime nap containing solely non-rem sleep enhances declarative but not procedural memory. *Neurobiology of Learning and Memory*, 86(2):241–247.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124–1131.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., and Joliot, M. (2002). Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage*, 15(1):273–289.

Underwood, B. J. (1957). Interference and forgetting. *Psychological Review*, 64(1):49–60.

van Gerven, M. A., Maris, E., Sperling, M., Sharan, A., Litt, B., Anderson, C., Baltuch, G., and Jacobs, J. (2013). Decoding the memorization of individual stimuli with direct human brain recordings. *Neuroimage*, 70:223–232.

Vertes, R. P. (2004). Memory consolidation in sleep. *Neuron*, 44(1):135–148.

Wacker, M. and Witte, H. (2013). Time-frequency techniques in biomedical signal analysis. a tutorial review of similarities and differences. *Methods of Information in Medicine*, 52(4).

Wagner, U. (2001). Emotional memory formation is enhanced across sleep intervals with high amounts of rapid eye movement sleep. *Learning & Memory*, 8(2):112–119.

Wagner, U. and Born, J. (2008). Memory consolidation during sleep: interactive effects of sleep stages and hpa regulation. *Stress*, 11(1):28–41.

Wagner, U., Gais, S., Haider, H., Verleger, R., and Born, J. (2004). Sleep inspires insight. *Nature*, 427(6972):352–355.

Walker, M. P., Brakefield, T., Seidman, J., Morgan, A., Hobson, J. A., and Stickgold, R. (2003). Sleep and the time course of motor skill learning. *Learning & Memory*, 10(4):275–284.

Walker, M. P. and Stickgold, R. (2004). Sleep-dependent learning and memory consolidation. *Neuron*, 44(1):121–133.

Wilhelm, I., Wagner, U., and Born, J. (2011). Opposite effects of cortisol on consolidation of temporal sequence memory during waking and sleep. *Journal of Cognitive Neuroscience*, 23(12):3703–3712.

Wixted, J. T. (2004). The psychology and neuroscience of forgetting. *Annual Review of Psychology*, 55(1):235–269.

Woolrich, M. W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., Beckmann, C., Jenkinson, M., and Smith, S. M. (2009). Bayesian analysis of neuroimaging data in fsl. *Neuroimage*, 45(1, Supplement 1):S173 – S186.