Chapter 5. Predictors of successful memory encoding

5.1 Introduction

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

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

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

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

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

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

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

5.2 Methods

5.2.1 Task

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

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

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

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

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

5.2.2 Data analysis — LFP

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

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

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

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

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

5.2.3 Data analysis — LFP decoding

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

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

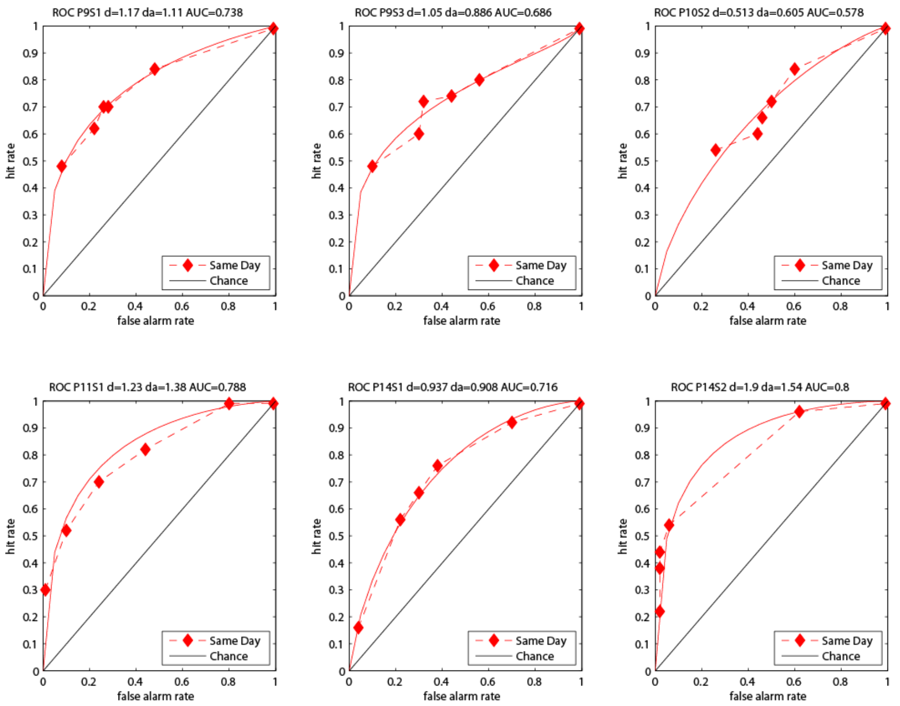

Figure 5-1. Retrieval performance (behavior) shown as a receiver operator characteristic (ROC) curve.

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

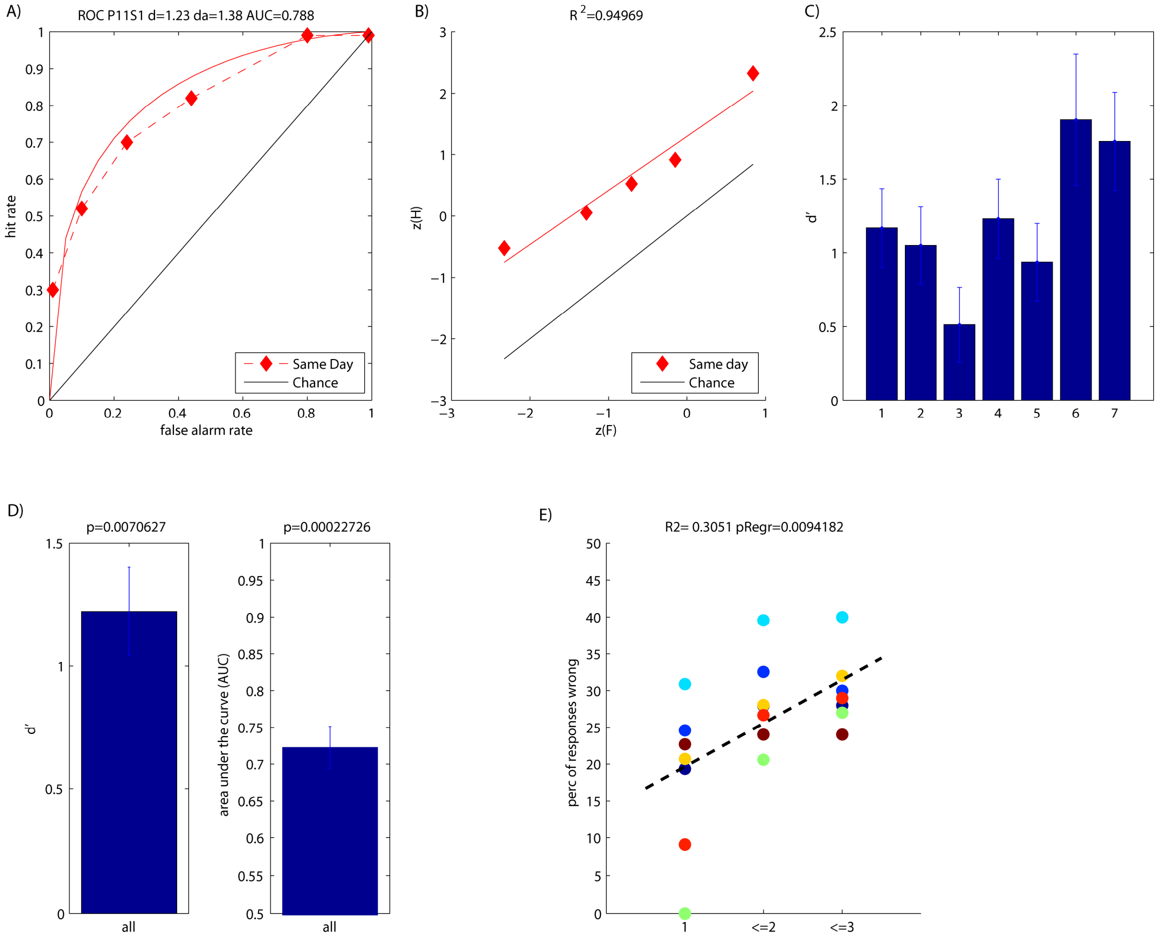

Figure 5-2. Retrieval performance for all subjects.

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

5.3 Results

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

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

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

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

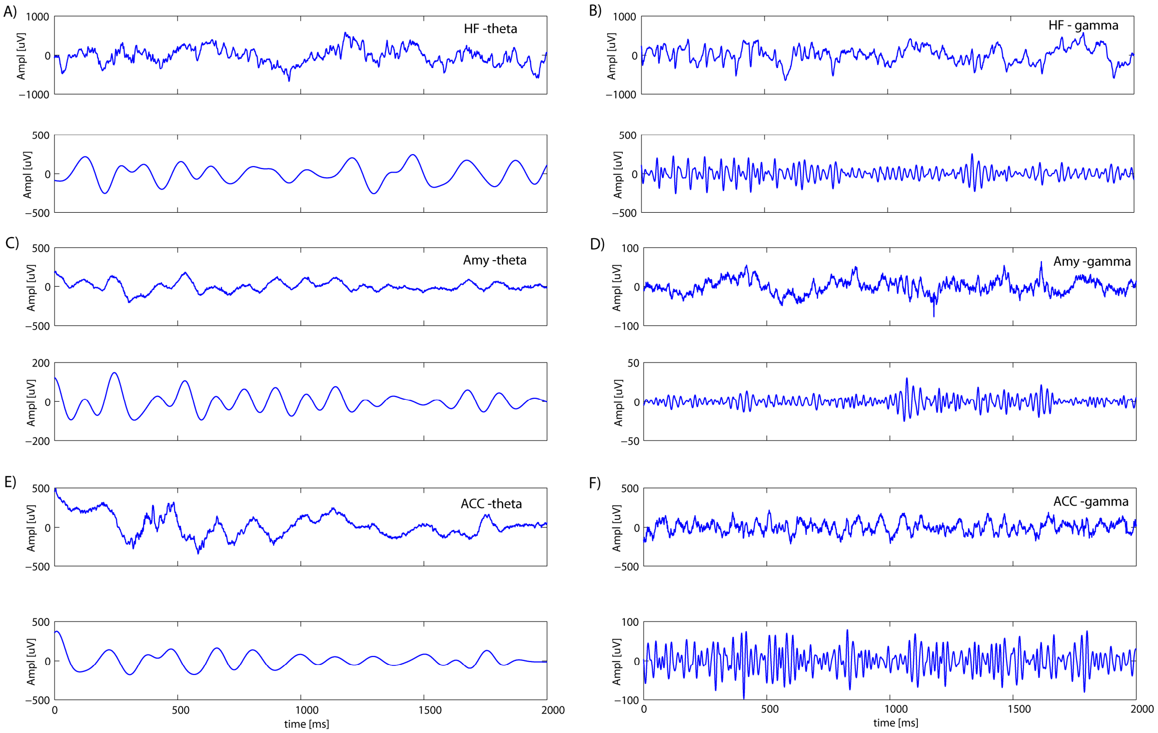

Figure 5-3. Example LFP traces (raw, theta, gamma).
Shown are 2 s of data from the hippocampus (HF; A+B), amygdala (Amy; C+D) and anterior cingulate (ACC; E+F). Traces were from data recorded during the learning part of the task (stimulus onset at 500 ms). Each panel shows the raw fullband trace (high-pass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

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

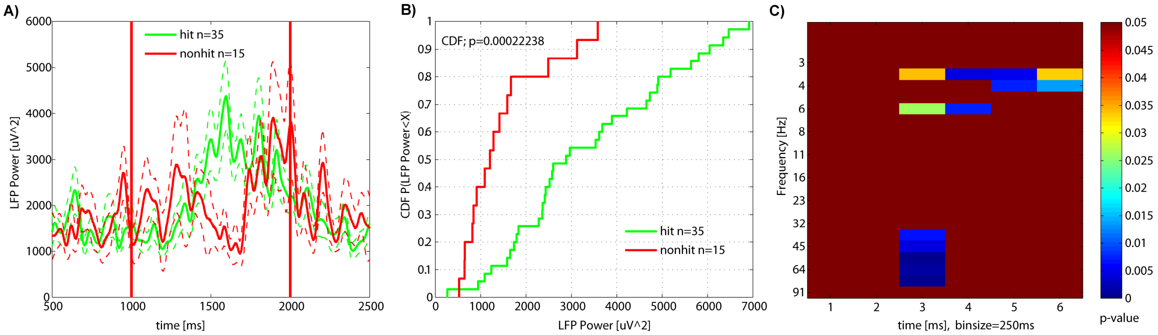

Figure 5-4. Example of LFP power difference due to memory.
All data in this figure is from a microwire in the left hippocampus. The frequency band illustrated is 53 Hz (gamma). (A) shows the LFP power (at 53 Hz) as a function of time for all learning trials. Trials for stimuli which were later remembered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

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

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

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

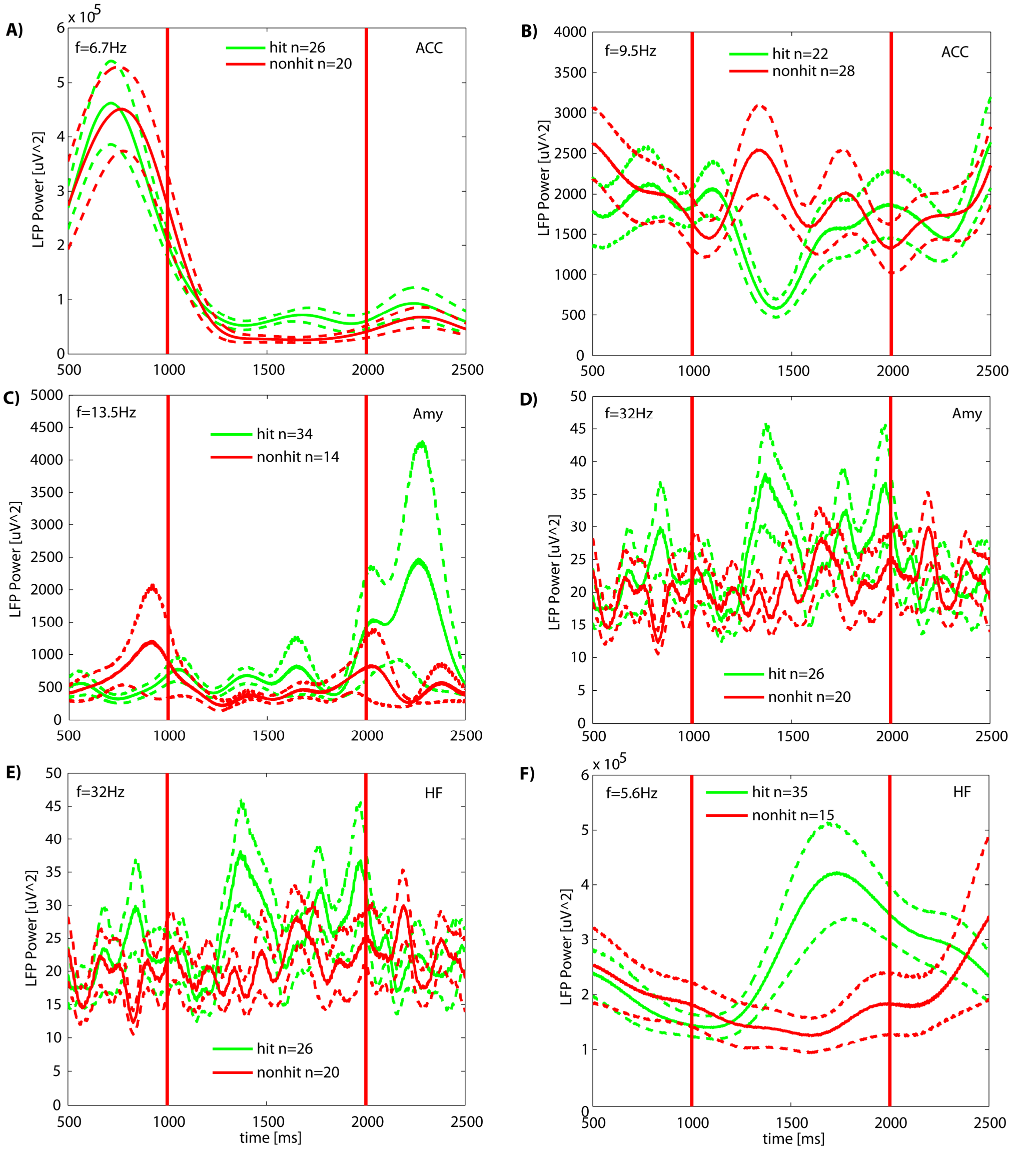

Figure 5-5. Examples of DM effects from all three brain areas as well as different frequency ranges.

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

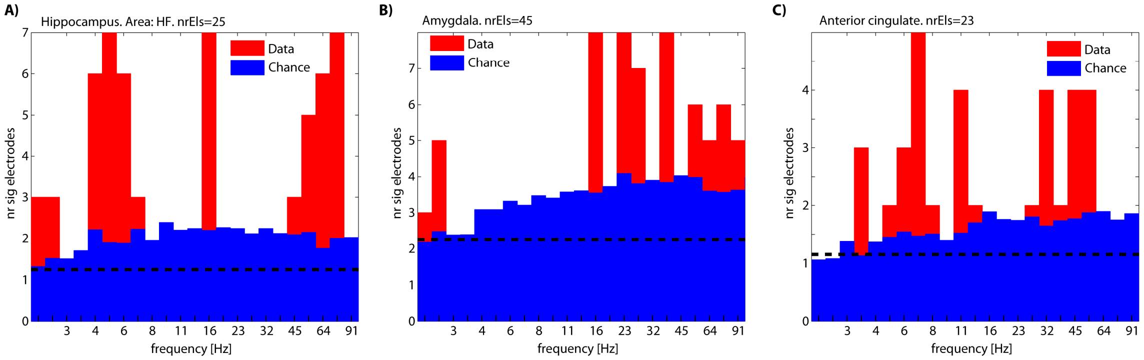

Figure 5-6. Summary of DM effects for all brain areas and frequencies. Shown are the number of electrodes as a function of frequency that have a DM effect for at least one timebin. Red bars show the real data, blue bars the bootstrapped chance level and the black line the theoretical chance level. All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

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

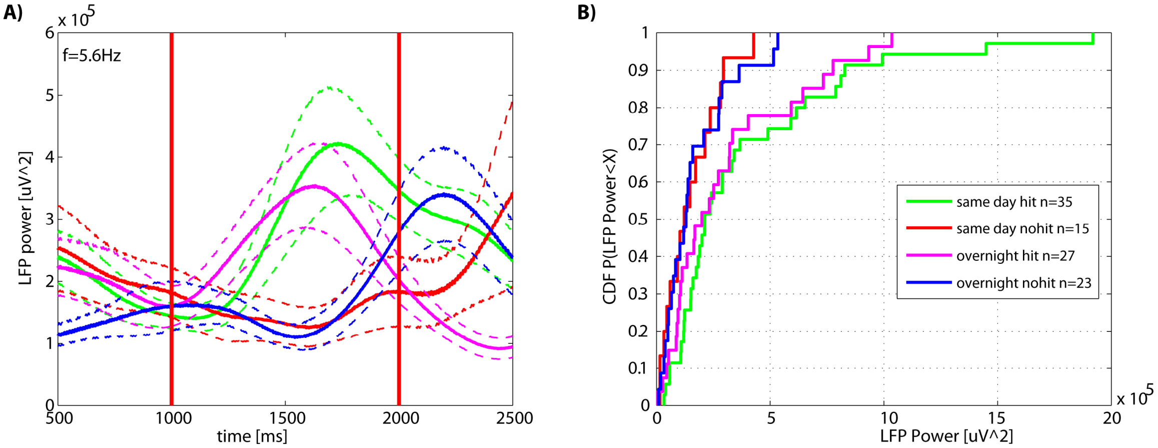

Figure 5-7 Example of LFP power difference, shown for learning trials that were retrieved on the same day (green and red) and overnight (magenta and blue). This channel was selected entirely based on the statistics for the same-day trials. (A) raw trace of LFP power in the 5.6 Hz band. Note that the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

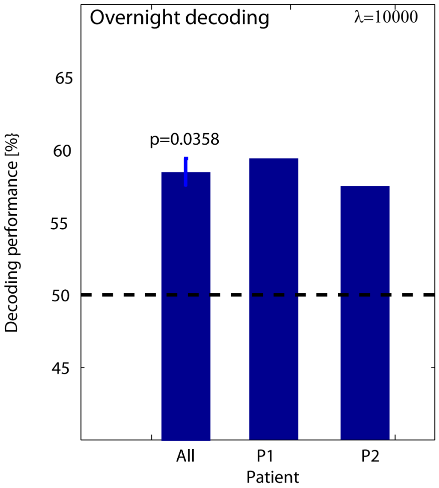

Figure 5-8. LFP power can be used to predict overnight memory.
Channels were identified that correlate with success of retrieval after the short delay (same day) and were then used to train a classifier. This classifier is able to predict overnight memory successfully if used on the learning trials for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

5.4 Discussion

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

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

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

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

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

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