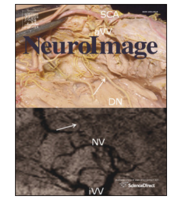
NeuroImage xxx (2013) xxx–xxx

Contents lists available at SciVerse ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg



Review

Classification aided analysis of oscillatory signatures in controlled retrieval

Nicholas Ketz *, Randal C. O'Reilly, Tim Curran

Department of Psychology and Neuroscience, University of Colorado Boulder, 345 UCB, Boulder, CO, USA

A R T I C L E I N F O

Article history:
Accepted 28 June 2013
Available online xxxx

Keywords:
Memory
EEG
Oscillations
Beta
Control
Classification

A B S T R A C T

Control processes are critical for both facilitating and suppressing memory retrieval, but these processes are not well understood. The current work, inspired by a similar fMRI design (Detre et al., in press), used a modified Think/No-Think(TNT) paradigm to investigate the neural signatures of volition over enhancing and suppressing memory retrieval. Previous studies have shown memory enhancement when well-learned stimulus pairs are restudied in cued recall ("Recall or think of studied pair item"), and degradation when restudied with cued suppression ("Avoid thinking of studied pair item"). We used category-based (faces vs. scenes) multivariate classification of electroencephalography signals to determine if individual target items were successfully retrieved or suppressed. A logistic regression based on classifier output determined that retrieval activation during the cued recall/suppression period was a predictor for subsequent memory. Labeling trials with this internal measure, as opposed to their nominal Think vs. No-Think condition, revealed the classic TNT pattern of enhanced memory for successful cued-retrieval and degraded memory for cued-suppression. This classification process enabled a more selective investigation into the time-frequency signatures of control over retrieval. Comparing controlled retrieval vs. controlled suppression, results showed more prominent Theta oscillations (3 to 8 Hz) in controlled retrieval. Beta oscillations (12 to 30 Hz) were involved in high levels of both controlled retrieval and suppression, suggesting it may have a more general control-related role. These results suggest unique roles for these frequency bands in retrieval processes.

Contents

* Corresponding author.
E-mail addresses: nicholas.ketz@colorado.edu (N. Ketz), randy.oreilly@colorado.edu (R.C. O'Reilly), tim.curran@colorado.edu (T. Curran).

1053-8119/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neuroimage.2013.06.077

Introduction

Control over memory retrieval is a powerful tool for effective learning, however the mechanisms supporting this process are not well understood. Successful retrieval of previously-studied material is known to improve long-term retention, a finding referred to as the testing effect (Karpicke and Roediger, 2008). Conversely, successful suppression of previously-learned material has shown the reverse relationship, such that successful suppression of retrieval leads to diminished long-term retention (Anderson and Green, 2001; Depue et al., 2007). The objective for the current work is to focus on the underlying neural mechanisms supporting these control processes, using oscillatory dynamics within Electroencephalography (EEG) as a window into the coordinated processing dynamics across different neural networks.

Neural oscillations in intracranial recordings measure fluctuations in the local field potential, reflecting the excitatory and inhibitory input into different neuronal assemblies. These oscillations can thus provide a measure of the postsynaptic potentials responsible for shaping cell assemblies involved in the storage and retrieval of long-term memories (Buzsaki and Draguhn, 2004; Fell and Axmacher, 2011). If these oscillations are distinct enough, they can be picked up by scalp based EEG/MEG sensors, and several studies have shown that scalp-recorded neural oscillations play an important role in long-term memory (Hanslmayr et al., 2012; Nyhus and Curran, 2010). These studies suggest that fluctuations in synchronized activity in the Theta (approximately 3–8 Hz), Alpha (approximately 8–12 Hz), Beta (approximately 12–30 Hz), and Gamma (approximately 30–50 Hz) frequency ranges may play differential roles in memory formation and retrieval processes.

In particular, Theta and Gamma power are the most prominent oscillatory markers of successful retrieval (Burgess and Gruzelier, 1997; Düzel et al., 2003; Nyhus and Curran, 2010; Osipova et al., 2006). Several of the studies supporting this claim use recognition memory tests to show that Theta power is increased for Hits (i.e. correctly remembered items) compared to correct rejections (new items presented at test that are correctly identified as new) from EEG data recorded during test (Burgess and Gruzelier, 1997; Düzel et al., 2003; Osipova et al., 2006). Similarly, several studies have used a source judgment task (i.e., testing individuals on the particular context in which an item was studied) within the subsequent memory paradigm and found greater Theta power, in EEG data recorded during study, for correctly identified sources compared to incorrect sources (Gruber et al., 2008; Osipova et al., 2006; Sederberg et al., 2003). Gamma results have been more varied than Theta, but they are often found to covary with each other and are perhaps coordinated with more general retrieval processes (Hanslmayr et al., 2009; Osipova et al., 2006). Although these results can provide insight in the oscillatory signatures surrounding successful memory retrieval, these studies were unable to contrast the amount of retrieval control required to perform these tasks, and therefore are unable to selectively reveal the general control processes involved in memory retrieval.

A collection of recent work focused on the suppression of task-irrelevant information to gain more direct insight into memory control processes, and found that that Alpha and Beta power are prominent under these conditions (Hanslmayr et al., 2012; Klimesch, 2012). One such study used a recognition memory task and found a greater decrease in Beta power for Hits compared to Misses from EEG data recorded at test (Düzel et al., 2003).

A pair of more recent studies manipulated the number of items required to be retrieved for successful execution of a given trial (Hanslmayr et al., 2010; Khader and Rösler, 2011). Results from this study, using EEG data recorded during testing, show that Theta power was positively correlated with the number of successfully retrieved items, while Alpha and Beta power were negatively correlated with the number of successfully retrieved items. Although the authors framed their manipulation in reference to the number of items retrieved, one could also interpret this as manipulating the amount of control required for successful retrieval. A similar study also showed that Alpha/Beta power decreased when trying to retrieve a target memory, however the design was also able to show that Alpha/Beta power increased for items competing with the target memory (Waldhauser et al., 2012). A recent review interpreted these and other results to suggest that Alpha/Beta decreases provide a marker for the amount of information retrieved, but also correlates with active suppression in the retrieval of competing/unwanted items (Hanslmayr et al., 2012).

There are at least two open questions that emerge from these results. First, are oscillations within the Alpha and Beta bands always negatively coupled to successful retrieval as suggested by Waldhauser et al. (2012) or can they play separable roles? Results highlighted above suggest Alpha and Beta to be more or less inseparable within the domain of controlled retrieval. These studies, however, have not directly targeted difficulty of retrieval, or required control over retrieval as an independent variable. Second, what role does Theta play in the control over memory? Historically the role of Theta in memory (either encoding or retrieval) has been confounded with the successful execution of those memory processes. It is unclear whether Theta is correlated with the successful execution of encoding and retrieval, or to a more general control processes regulating memory.

These questions were targeted in the current work through the use of a modified Think/No-Think (TNT) paradigm (Anderson and Green, 2001), while participants underwent scalp recorded EEG. The standard TNT paradigm trains participants on a series of paired associates, and then manipulates these associates in a cue recall task where one item within the originally-studied pair is presented and participants are asked to either actively recall (Think) or actively suppress recall (No-Think) of the corresponding pair item. The modified paradigm adopted in the current work uses paired associates with multiple levels of interference to further manipulate the level of control required to retrieve a given pair. Here, we increased the number of interfering traces for a given paired associate to likewise increase the demands on retrieval processes required to successfully identify the originally studied pair. Our manipulation of interfering traces, described below, is derived from the presentation of distracting pair images during the Controlled Retrieval portion of the TNT task. This paradigm was used so that contrasts of low vs. high control over memory suppression and retrieval can be made. This augmented TNT paradigm then allows for the more subtle investigation into the oscillatory dynamics during controlled retrieval processing.

In addition to the manipulation of multiple levels of required control, the current work also takes advantage of multivariate pattern classification methods to sub-select trials in which participants were successfully able to elicit control over retrieval processes. One major concern with using the Think/No-Think paradigm is that previous studies have found varying degrees of success in eliciting the desired effect (Bulevich et al., 2006). Similarly, there has been contention as

Please cite this article as: Ketz, N., et al., Classification aided analysis of oscillatory signatures in controlled retrieval, NeuroImage (2013), http://dx.doi.org/10.1016/j.neuroimage.2013.06.077

to whether the memory systems involved in retrieval are suppressed during No-Think trials, or whether a distracting memory is retrieved in place of the learned associate (Tomlinson et al., 2009). These issues raise concern as to the trial-by-trail efficacy of participants in performing the task. A recent fMRI study by Detre et al. (in press) has addressed these issues through the inclusion of multivariate pattern classification into the study design. This allows for the trial-by-trial detection of how successful participants were at retrieving or suppressing the image category of the target associate. Detre et al. (in press) found that the TNT effects were much stronger when sorted based on this internal measure of memory retrieval. In the present study, we adapted this analysis method to the EEG context to provide a continuous measure of how successful or unsuccessful control over retrieval was on individual trials. This allows us to dissociate subsequent memory from successful control over retrieval, which are ubiquitously confounded in previous studies. With this dissociation we are able to investigate the neural signatures related to retrieval control without the assumptions built into subsequent memory and Think/No-Think studies, i.e., that greater retrieval success leads to better subsequent memory.

We carried out the study in three major phases as shown in Fig. 1, all within one experimental session. First, in the Paired Associate task, we trained individuals on word–image pairs to be later remembered. The second phase was done while scalp EEG was recorded and consisted of two tasks separated by a short break. During this phase participants first performed a 1-Back task using the same types of images used with the word–image pairs studied in the first phase. This data was later used to train a classifier to identify when participants were holding a particular image category in mind (faces or scenes). Following the 1-Back task participants carried out a modified Think/No-Think task, called the *Controlled Retrieval* task. During this task participants performed either cued recall (Think), or cued suppression of recall (No-Think) using pairs learned in the Paired Associate task. Interference within pairs was manipulated after the presentation of the cuedword by either presenting the originally studied image (low interference *Green* condition), or by presenting a distracting image (high interference *Red* and *Blue* conditions). It is expected that these two types of trials either reinforce the original word–image pair (i.e., the Green condition) or interfere with that original association (i.e., Blue and Red conditions). Here, Red trials are similar to the standard No-Think condition, and Blue/Green trials are similar to the Think condition. Finally, in the third phase of the experiment, participants performed a standard cued recall task to assess their memory for the original word–image pairs after the Controlled Retrieval manipulation.

The specific goal of the current work is to target varying levels of control over both enhancement and suppression of memory retrieval with the intention of targeting Theta, Alpha, Beta, and Gamma oscillations as dependent measures. Based on previous literature cited above we expected to find Theta band power to be positively correlated with successful retrieval, and Alpha/Beta band power to be negatively correlated with successful suppression. Further, we targeted varying levels of

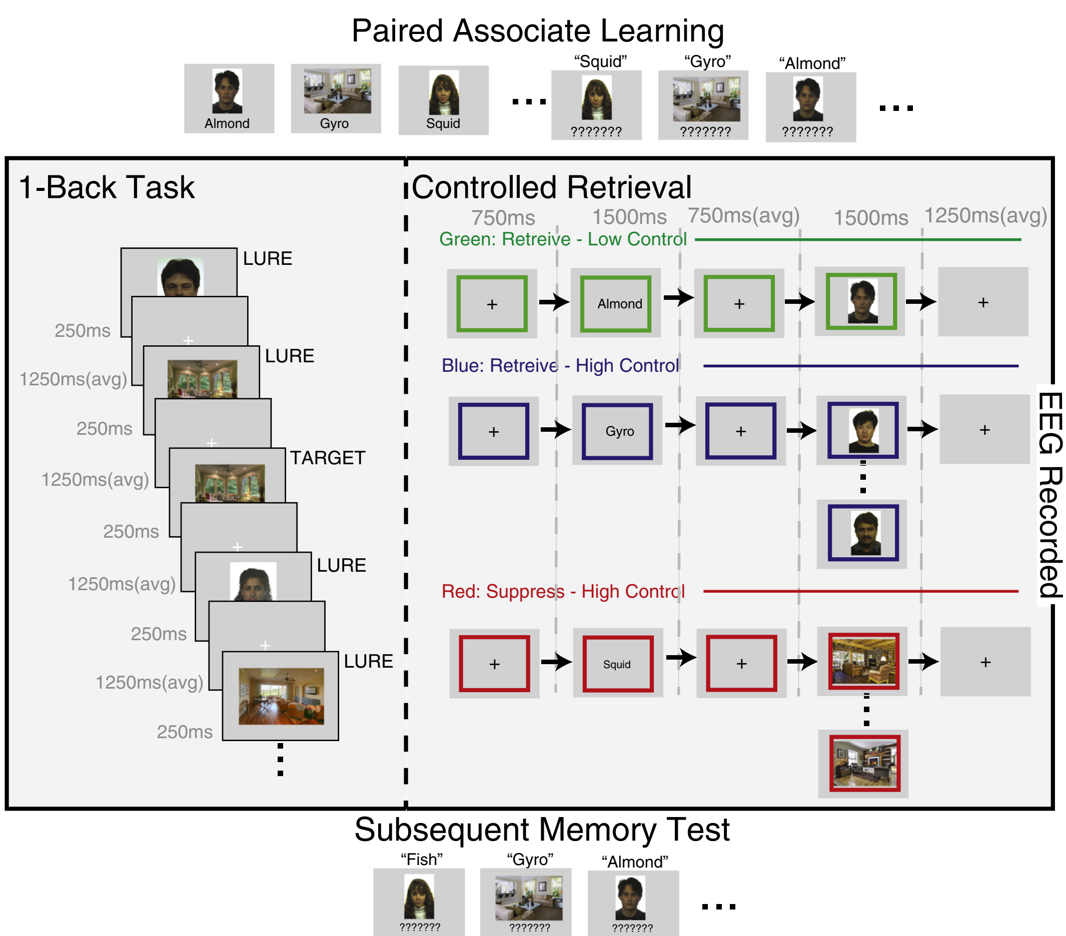


Fig. 1. Schematic of experiment tasks. Black rectangle shows tasks in which EEG data was recorded. Temporal order of tasks graphically proceeds down and left to right. *First*, in the Paired Associate task, we trained individuals on a series of word–image pairs to be later remembered. *Second*, while EEG was recorded, participants performed a 1-Back task using images of faces and scenes. *Third* participants carried out the Controlled Retrieval task. Here the temporal sequence of the three trial types is shown. The Blue and Red conditions also show the presentation of multiple interfering images following the cue word. *Finally*, after removing the EEG recording equipment, a cued recall test of participants' memory for the word–image pairs learned in the Paired Associate Learning task completed the experimental session.

required control by manipulating interference to further define the specific functional correlates within these frequency bands.

Materials and methods

As shown in Fig. 1, the overall structure of the experiment consisted of 4 main tasks: initial Paired Associate learning, a perceptual 1-Back task using faces and scenes, the modified Think/No-Think task referred to as the Controlled Retrieval task, and finally the Subsequent Memory Test for the originally learned paired associates. The initial Paired Associate task was self paced by participants and took approximately 2 rounds of study/test and a total of 20 min to complete.The 1-Back task had constrained stimulus timing but self paced blink breaks and took approximately 20 min to complete. The Controlled Retrieval task also had constrained stimulus timing but self paced blink breaks and electrode impedance adjustments, in total taking approximately 90 min to complete. The final Subsequent Memory Test was self paced and took approximately 10 min to complete. In total, the full course of the experiment took approximately 3.5 h including EEG setup.

Participants

Thirty University of Colorado undergraduates participated in the experiment and received payment of \$15 per hour (ages 18–25, M = 20; 14 males, 16 females). All participants were right handed, had normal or corrected-to-normal vision, and all but three were native English speakers. Informed consent was obtained from each participant, and the study conformed to the Institutional Review Board (IRB) guidelines.

Materials

A list of 96 common food nouns was generated for the study constrained by the food item being only one word long. Words were presented in 52 point Geneva font just below the paired image during the Paired Associate Learning and the Subsequent Memory portions of the study, and in 30 point Geneva font in the center of the screen during the Controlled Retrieval task. The image pool consisted of 1032 color images equally split into faces and scenes. Face images were photographs focused from the shoulder up with the center of the face generally in the center of the image, and taken in front of an off-white background (Phillips et al., 2000). Scene images were photographs taken from the SUN image database within the ‘Living Room’ category Xiao et al. (2010). Each image was scaled to match its original proportions with a maximum size of 350 pixels on either the length or the width corresponding with the largest dimension of the image before resizing. On average across images faces were presented at a size of 325 × 217 pixels, and scenes were presented at a size of 226 × 322 pixels. As is discussed in the EEG classification section this size difference between image categories is likely to be uninformative to the classification process due to the feature-selection procedure shown in Fig. 2, and the optimal time points ultimately used in classifier training shown in Fig. 3A. The experiment was presented on a 17-in flat-panel display with a resolution of 1024 × 768 (60 Hz frame rate) placed 1 m

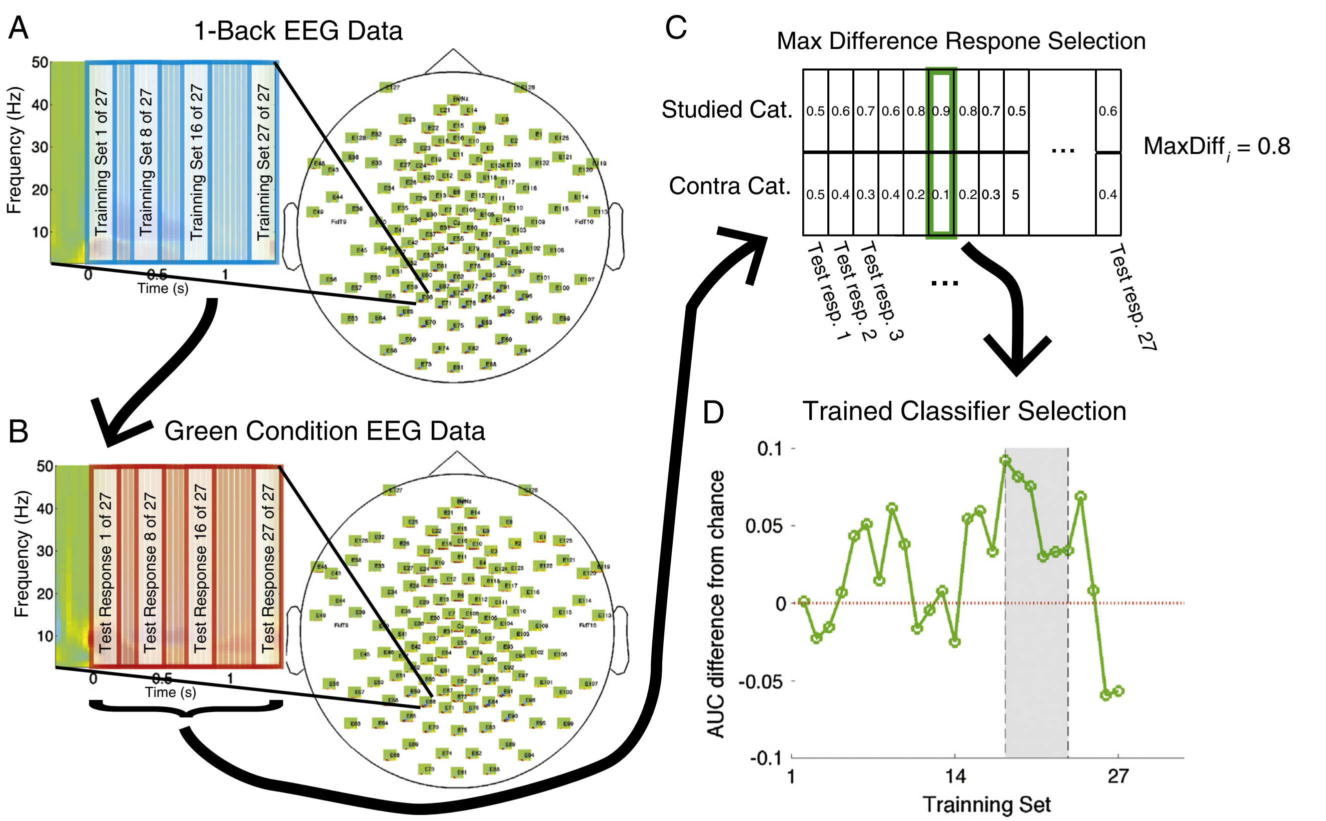


Fig. 2. Schematic of classifier selection procedure. Procedure used to select the best performing classifier. First: 27 individual classifiers are trained on 200 ms time windows of spectral EEG data from the 1-Back task. Second: Each of those classifiers are tested using the spectral EEG data from the Green condition trials across 27 individual time windows yielding a set of 27 test responses for each of the 27 trained classifiers. Third: From these 27 individual test responses for each trial, the single largest *Max Difference* score is selected as the classifier output. This Max Difference score is defined as the absolute value in the difference between classifier output for the image category originally studied vs. the contra-category. Fourth: The classifier output used in the selected Max Difference score for each trial are used in a Area Under the ROC-Curve(AUC) to determine which original training set yields the highest AUC score. Gray shaded area highlights the 200 ms time window of training data that yielded the best AUC score; this is the classifier which is then used in further analyses to determine trial selection.

Please cite this article as: Ketz, N., et al., Classification aided analysis of oscillatory signatures in controlled retrieval, NeuroImage (2013), http://dx.doi.org/10.1016/j.neuroimage.2013.06.077

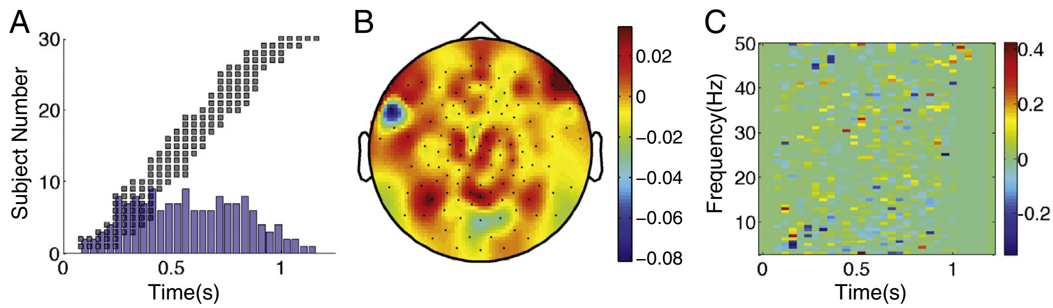


Fig. 3. Selected classifier time points and average weights across subjects. *A)* Subject specific 200 ms time windows (composed of 5 consecutive 40 ms time bins) used to train classifier weights shown in grey dots. Each row of the y-axis shows the 5 40 ms time points (200 ms total) used for a given subject. A histogram summarizing the time points used to train subject specific classifiers is super imposed in blue where the y-axis maintains the same numeric value but now represents the number of subjects which used a given time bin for image classification. *B)* Topographic map of averaged z-scored weights across time (0.08 to 1.16 s) and frequency (3 to 50 Hz). *C)* Average z-score of weights across subjects and electrodes.

in front of the participants. All portions of the display not occupied by stimuli or text were filled with grey pixels.

Design

Paired associates were generated at the time of the experiment for each participant. Each list of stimuli consisted of 96 word–image pairs where each of the 96 words was paired with a given image, randomly drawn without replacement from the stimulus pool of 96 images equally divided into faces and scenes. These pairs were used for the initial Paired Associate learning.

Stimuli for the 1-Back task were taken from a separate image pool of 80 faces and 80 scenes. Each of these images were repeated 3 times throughout the task, and a total of 12 target trials were used giving a total trial count of 492. Target trials are trials in which the image which was just presented is repeated and participants are instructed to identify these trials with a button press. The images used throughout the task were perceptually similar to those used in the Paired Associate learning task, however no image was used in both tasks. Image presentation was blocked by image type (faces or scenes) into a length of 20 trials. These blocks were presented in alternating order based on image category (i.e., a face block always followed a scene block and vice versa). The image category of the starting block was counterbalanced across participants.

In the Controlled Retrieval task the word–image pairs studied in the initial Paired Associate learning task were randomly divided into 4 conditions: retrieval-low (retrieval with low interference), retrieval-high (retrieval with high interference), suppression-high (retrieval suppression with high interference), and Baseline (no Controlled Retrieval manipulation). These conditions will be referred to in shorthand from Fig. 1 as Green (retrieval-low), Blue (retrieval-high), and Red (suppression-high). Each condition contained equal number of stimuli, and equal number of faces and scenes. The Baseline pairs were omitted from the Controlled Retrieval task to be used as a measure of subsequent memory that had no Controlled Retrieval manipulation. Each stimulus from the other three conditions was repeated 10 times throughout the course of the Controlled Retrieval task yielding a total of 720 trials (72 paired associates across Green, Blue and Red conditions by 10 repetitions each). Words from the Blue and Red conditions were shown with a unique new image for each repetition during the task. These distracting new images were randomly selected within image category such that the image presented in the Controlled Retrieval task was contra-category to the image originally studied in the Paired Associate task. For example, a word that was originally studied with an image of a scene would be re-paired with images of 10 different faces throughout the course of the Controlled Retrieval task, and vice versa for a word that was originally paired with a face image. This distractor pair presentation in the Red and Blue conditions provided the increased interference relative to the Green condition pairs which displayed the original studied image during the controlled retrieval task. The use of contra-category images within the interference images was intentionally done to increase the sensitivity of our classifier to detect when interfering items as opposed to target items are being processed. In total 480 new images were used (48 paired associates by 10 repetitions). The presentation order of stimuli was randomized within each cycle of the full list of originally studied pairs.

The final subsequent memory test used only the images from the original Paired Associate learning. Presentation order of the images was random. No testing of the new images shown in the Controlled Retrieval task for Blue and Red conditions was done.

Procedure

All stimulus presentation was done using the Psychophysics Toolbox within Matlab (Brainard, 1997). During the Paired Associate task participants saw 96 unique word–image pairs; each presented in random order with the image in the center of the screen and the word just below it. During this study portion each word–image pair was presented for 7 s. After all word–image pairs were presented once, participants were given a cued recall memory test for all the studied pairs. The test consisted of randomly ordered studied images presented in the center of the screen and a prompt below it where participants were asked to type the word they studied with the given image. During the cued recall testing, if no response was made in 30 s participants were prompted to make a response by displaying 'Please make a response!' above the cued image. The Inter Stimulus Interval (ISI) between both study and test items was 1 s. If an item was misidentified in the cued recall test participants were shown the correct pairing before moving on to the next test pair. If there was less than a 10% accuracy rate in the test portion, then the pairs which were incorrectly labeled were re-studied in random order in the same fashion as the original study portion. Following this a test of the complete list of pairs was again carried out. This cycle of study and test continued until the accuracy rate on the test portion was above 10%. Average accuracy for studied pairs on the final test round of the Paired Associate task was 30 $\pm$ 10%, and the number repetitions (i.e., cycles of study and test) required to complete the task was an average of 1.1 $\pm$ 0.5 cycles, where 1 cycle implies the task was completed on the first round of study and test. This liberal learning criteria was adopted to ensure that participants understood the associations they were instructed to learn and that they had at least a minimal recall rate going into the Controlled Retrieval task, similar to the scenario where only a single round of study were allowed on the word–image pairs.

Following the initial Paired Associate learning, participants were fit with an electrode net and given general instructions regarding proper handling of the net, and to avoid blinking during the tasks unless during designated blink breaks which were spaced no more than 1 minute apart. The 1-Back task was then carried out in which

Please cite this article as: Ketz, N., et al., Classification aided analysis of oscillatory signatures in controlled retrieval, NeuroImage (2013), http://dx.doi.org/10.1016/j.neuroimage.2013.06.077

participants viewed a series of images presented in the center of the screen and were told to press the 'J' key whenever a given image was repeated immediately in sequential order. Images were presented for 250 ms and a jittered ISI was randomly sampled from 1000 to 1500 ms where a '+' symbol would appear in the center of the screen in 62 point Geneva font. If participants correctly identified a target image the word 'Correct' appeared in green superimposed over the image, and if participants false alarmed to a non-target image a 'X' appeared in red over the image. Feedback text was presented in 32 point Geneva font and remained on the screen until the end of the trial which depended upon how quickly participants responded. In total 492 images were presented, 12 of which were targets.

Participants then began the Controlled Retrieval task, as shown in Fig. 1, which consisted of the sequential presentation of previously studied words followed by either novel or previously studied images. Each trial began with the presentation of a 350 by 350 pixel square with a 20 pixel colored border which remained on the screen until the completion of the trial. The color of the square's boarder indicated the trial's condition type; either Green, Blue or Red. After a 750 ms cue from the colored square, a previously studied word appeared in the center of the square in 30 point Geneva font, which remained on the screen for 1500 ms. This was followed by a jittered ISI between 500 and 1000 ms during which a '+' symbol of 62 point Geneva font appear in the center of the screen. An image was then presented in the center of the colored square for 1500 ms. In the Green condition this image was the same image originally studied with the currently presented word in the Paired Associate learning. In the Blue and Red conditions this image was a novel image and was contra-category to the image that was studied with the currently presented word in the Paired Associate learning. Following the image presentation the color square was removed from the screen, and a 62 point Geneva font '+' symbol appear in the center of the screen for a jittered ISI between 1000 and 1500 ms which completed the extent of a single trial. Each of the conditions consisted of 24 word–image pairs taken from the Paired Associate learning, yielding a list of 72 stimuli. This list was repeated 10 times through the course of the task with Red and Blue stimuli using a unique image on each repetition. The presentation order was randomized within each repetition such that each of the 72 stimuli was presented once before any single stimuli was presented in the next repetition.

In the Green and Blue conditions participants were instructed to: "Retrieve or think about the image previously studied with the presented word". In the Green condition the previously studied image appeared following the word presentation. In the Blue condition, a new image appeared following the word presentation, and participants were instructed to "…simply maintain thinking about the originally studied image while holding your gaze on this new image". In the Red condition participants were instructed to: "Try and not think of the image previously studied with the presented word, and to simply try to not think of anything while the word is on the screen". Following the word presentation participants were instructed to "…actively associate this new image with the word that was just presented". Through this design it was intended that the Green condition required the least amount of control over retrieval as the word–image pairs are presented several times together during the course of the Controlled Retrieval task. The Blue condition is analogous to a classic Think condition with the exception that an increasing number of distractor images are presented with a given cue word during the course of the task. Finally the Red condition is analogous to a No-Think condition with the exception that participants are actively trying to encode new associations with a given cue word during the course of the task. The addition of this active encoding task was included to look at active enhancement vs suppression of encoding, however the current analysis only focuses on the interfering effects of these distractor images and leaves these effects of volitional control over encoding for future work. The Red and Blue conditions, due to their presentation with distractor images, were intended to require more control over retrieval than the Green condition.

After the Controlled Retrieval task the electrode net was removed from the participants head, and the final subsequent memory task was carried out. The testing procedure was identical to the testing in the original Paired Associate learning with the exception that no feedback was given for participants' responses, and only one testing of the items was carried out regardless of the participants' accuracy level.

Electrophysiological recordings and data processing

A 128-channel HydroCel Geodesic Sensor Net TM(GSN 200, v. 2.1) was used to measure the EEG at the scalp using a central vertex reference (Cz) with a sampling rate of 250 Hz. The net was connected to an AC-coupled, high-input impedance amplifier (300 MΩ, Net Amps TM; Electrical Geodesics, Inc., Eugene, OR). The electrodes were adjusted until impedance measurements were less than 40 kΩ, and were checked to meet this criteria at an interval of 15 min at most.

Post recording, Net Station software (Electrical Geodesics, Inc.) was used to digitally high-pass filter at 1 Hz, low-pass filter at 100 Hz, and notch filter at 60 Hz. Data was epoched into 3000 ms segments, 1 s before the onset of each test stimulus and 2 s after. The Net Station artifact detection was used to detect trials that contained sufficient eye-blink artifacts. Subsequently, the Net Station's bad channel interpolation and trial rejection algorithms were used. Bad channels were identified by looking at the full epoch time window and evaluating whether it met any of the following criteria: average amplitude exceeds 100 μV, differential average amplitude exceeds 50 μV, or the channel had zero variance. On average across participants and conditions 40 ± 15% of trials were removed due to blink artifacts. All analyses were based on referencing to the average of all electrodes (Dien, 1998) using Net Station's PARE correction (Junghfer et al., 1999). All subsequent data processing and analyses were done in MATLAB (version R2011b; The MathWorks, Inc., Natick, MA) using the FieldTrip toolbox (Oostenveld et al., 2011) and in-house scripts.

The spectral decomposition was performed using a set of 71 Morlet wavelets that were equally spaced in 0.67 Hz intervals from 3 to 50 Hz. Each wavelet had a width that was 4 times the period of its center frequency. After decomposition, the analyzable window was reduced to a 1680 ms window, starting −340 ms prior to stimulus onset and extending to 1340 ms after stimulus onset; stimulus onset here implies word onset in the Controlled Retrieval task and image onset in the 1-Back task. The power, i.e., the magnitude of each complex coefficient, was then computed for every 40 ms time bin within the 1680 ms analyzable window. In total, the spectral decomposition transformed each of the 42 time bins of a trial into the power values of 71 frequency bands for each of the 128 electrodes, yielding a potential 381,696 analyzable features per trial. For each trial, the average power across the 200 ms period between −300 and −100 ms prior to stimulus onset was used to baseline correct each 1640 ms epoch within each electrode/frequency bin. These spectral features were used as inputs to the EEG pattern classifiers and cluster based analyses.

EEG classification

The overall EEG classification goal was to train a pattern classifier to predict when participants were thinking about the studied image category (e.g., faces) vs. the contra image category (e.g., scenes) on each Controlled Retrieval trial. Classifiers were first trained to discriminate face and scene images during the 1-Back task, and were then used to estimate whether participants were thinking about faces or scenes during Controlled Retrieval task. The EEG analysis was composed of the following steps which are outlined in Fig. 2.

First, shown in panel A of Fig. 2, a set of classifiers was trained to detect the patterns of spectral features associated with processing

Please cite this article as: Ketz, N., et al., Classification aided analysis of oscillatory signatures in controlled retrieval, NeuroImage (2013), http://dx.doi.org/10.1016/j.neuroimage.2013.06.077

each image category in the 1-Back task. Spectrally decomposed EEG data from the 1-Back task was used to train the category specific classifiers. 200 ms time windows, selected in steps of 40 ms between 0 and the 1040 ms, were used to train 27 independent classifiers (one for each 200 ms window) per subject.

Second, shown in panel B of Fig. 2, each of those 27 classifiers provide 27 different test responses to each of the Green condition trials. These responses are based on the same running 200 ms time windows in which the classifier was originally trained and span the full 0 to 1040 ms epoch, however the data being input to the already trained classifier is now coming from the Green condition trials.

Third, the classifier features which perform optimally for the intended task of identifying latent processing of face and scene images were selected using what will be referred to as the *Max Difference* method. This Max Difference method, shownin panel 3 of Fig. 2, measures the difference in classifier response between the image category of the originally studied paired-associate, for a given cue word, and the contrary image category. This difference is calculated across the full 0 to 1040 ms word presentation using the 27 test responses acquired in panel B of Fig. 2, yielding 27 difference scores. The time window that showed the maximum absolute value in this difference of response is then taken as the classifiers' categorization for that trial, and is labeled it's Max Difference score.

Finally, the results of this Max Difference score, across trials, are then passed to an Area Under the Curve (AUC) test. Here an AUC score of .5 indicates chance levels of discrimination and an AUC score of 1.0 indicates perfect separation of the two categories (Fawcett, 2006). Using this method, the classifier that was trained on a specific 200 ms time window within the 1-Back EEG data that showed the largest AUC score to Green condition trials was selected within each subject, and was then used in all subsequent analysis for that particular subject.

These selected classifiers provide an estimate of image category processing in Blue and Red Controlled Retrieval trials using the same Max Difference measure. These Max Difference scores were used in two ways. The first was a predictor within a logistic regression model regressing subsequent memory on a composite *Memory Activation* score, which is an average Max Difference score across stimulus repetitions. The second use for the classifier determined Max Difference scores was a trial selection criteria within a cluster based spectral power analysis. Both of these analyses are described in more detail in the Classifier training section.

This multi-step procedure was used because the timing of perceiving a face vs. scene in the 1-Back test is likely to differ from the timing of thinking about a face vs. scene in the Controlled Retrieval task. This selection procedure allows us to select the point in time during which faces and scenes are best discriminated in a trial-by-trial basis during the Controlled Retrieval task. Because the target category is well remembered in the Green condition, performance on that condition is used to determined the training data most sensitive to the latent processing of faces and scenes before applying the selected classifier to the Blue and Red conditions of interest.

Classifier training

The training of category specific image classification of EEG data was implemented using the Donders Machine Learning toolbox as integrated within the FieldTrip toolbox. An Elastic Net classifier was used because of its automated feature selection which allows for parameterized balance between L1 and L2 regularization Zou and Hastie (2005). Like standard multiple linear regression, the elastic net algorithm adjusts feature weights to minimize the squared error between the predicted label and the correct label. Unlike standard multiple linear regression, elastic net also includes both L1 and L2 regularization terms which bias the classification process to find a solution that minimizes the feature weights. Regularized regression algorithms (such as elastic nets) use a parameter (λ) that determines the impact of the regularization term, which in this case was fit across cross-validated training sets to find the best performance within subject. The balance between L1 and L2 regularization is adjusted using another parameter (α) which was loosely fit across individuals to a value of 0.08, emphasizing L1 regularization over L2. In effect this emphasizes the minimization of the absolute values of the weights (L1 regularization) as opposed to the squared weight values (L2 regularization).

Memory activation and trial selection

Following classifier selection a regression on the classifiers' relationship with subsequent memory was carried out. The measure of image processing used in this regression will be termed *Memory Activation* and was constructed by averaging the largest absolute Max Difference measure (as described in the EEG classification section) across all repetitions of the same stimulus for each subject. Each word–image pair had a potential 10 repetitions throughout the Controlled Retrieval task, however due to blink artifacts not every word–image pair had its full 10 repetitions present, and thus an average across repetitions was used. The Memory Activation response is measured relative to the originally associated image for each word–image pair, i.e., positive values in the Memory Activation scores reflect classifier response on average biased towards the original image association, and negative values reflect classifier response on average biased towards the contra-category of the originally associated image. This Memory Activation score was then used in two analyses. The first coarse analysis adjusted trial condition labels regardless of their original condition (Red or Blue), and solely based on the trials' Memory Activation score;specifically positive Memory Activation scores were moved to a relabeled Blue condition referred to as *Blue-Neural* throughout the text, and negative scores were moved to a relabeled Red condition referred to as *Red-Neural*. It was predicted that these re-assigned trials would show stronger memory enhancement (Blue-Neural) and suppression (Red-Neural) effects, as measured by subsequent memory scores, than the original assignments (Detre et al., in press). A second, more robust, analysis was done using the Memory Activation scores as a continuous predictor in a mixed effects logistic regression, with fixed effects grouped across participants, on the binary values of subsequent memory (memory or no-memory) for each word–image pair. We predicted that higher Memory Activation scores would be associated with successful subsequent memory.

In addition to the regression analysis, classifier output from the Controlled Retrieval task was used to sub-select trials that showed successful control, relative to task instructions, over image processing. For example, trials were selected for later analysis when the task was to think of the associated face image, and the classifier showed high levels of face processing. Similarly, trials were selected for later analysis when the task was to not think of the associated face image, and the classifier showed high levels of scene processing. These conditions are referred to as Blue_{succ} for successfully executed Blue trials and Red_{succ} for successfully executed Red trials. This selection was done on a trial-by-trial basis, such that the Max Difference score for a given trial was used in the selection process, regardless of the aggregate Memory Activation score across repetitions of a given cue word. The goal of this process was to select successfully executed trials in an attempt to remove unwanted noise from the power analysis of controlled retrieval effects.

Spectral power analysis

An analysis of the spectral power differences between the levels of control over retrieval was carried out using the trial selection described in the Classifier training section. Average number of trials, both before and after classifier determined sub-selection, are shown in Table 1. In general there were a minimum of 30 trials per condition, per subject used in the EEG analysis which combined across face and scene trial

Table 1
Trial numbers and subsequent memory accuracies for conditions used in the spectral power analysis. Subscripted 'succ' indicates classifier determined sub-selection of trials based on the successful execution of task. Variation in trial numbers in non sub-selected conditions reflect trial loss due to artifact rejection (120 possible in each condition if no loss). Errors shown are standard error in the mean.

Condition	EEG trial counts		Percent correct	
	Faces	Scenes	Faces	Scenes
Green	74 ± 3	73 ± 3	84 ± 3	78 ± 4
Blue	72 ± 3	72 ± 4	53 ± 4	44 ± 4
Red	73 ± 3	73 ± 3	53 ± 4	42 ± 4
Baseline	–	–	51 ± 5	45 ± 4
$\text{Blue}_{\textbf{succ}}$	38 ± 2	34 ± 2	55 ± 5	52 ± 5
$\text{Red}_{\textbf{succ}}$	34 ± 3	39 ± 3	44 ± 5	43 ± 5

types. No significant differences in trial numbers between conditions existed across subjects before or after sub-selection. The three levels of control over retrieval are those described in the Design section, i.e., Green, Blue, and Red which are short hand for retrieval-low interference, retrieval-high interference, and suppression-high interference respectively. It's assumed, based on behavioral measures of subsequent memory, that the Green condition required the least amount of control over retrieval, while the Blue and Red conditions required more. No a-priori differences were assumed in the amount of control over retrieval required in the Blue and Red conditions. The power differences between these conditions were assessed using a cluster based analysis (Maris and Oostenveld, 2007). Frequency data were averaged into the Theta (3–8 Hz), Alpha (8–12 Hz), Beta (12–30 Hz), and Gamma (30–50 Hz) bands. These averaged frequency bands were analyzed across the time window of 200–1000 ms, and across all electrodes except the 4 surrounding the eyes. Clustering was done by performing a *t*-test for conditions of interest within each time/electrode bin across subjects, followed by grouping together the adjacent bins which yielded a p value of less than 0.01. Cluster significance was calculated using a Monte-Carlo style permutation test of the summed *t*-values within a given cluster. Each observed cluster was subject to 10,000 random permutations of condition labels where its significance was estimated by the proportion of random permutations which yielded clusters that had a summed *t*-value as larger or larger than the observed cluster. It should be noted that due to the random nature of these permutation tests the reported p-values are non-stationary and will have 95% confidence intervals approximately equal to $p \pm 0.004$ (Ernst, 2004; Maris and Oostenveld, 2007). Clusters found to have a p value less than 0.1 were inspected manually, and included in the results. Any cluster which did not reach this threshold went unnoticed within our analysis stream.

Results

Classifier selection and performance

Classifier training performance from selected time windows in the 1-Back task, assessed using a 5-fold cross validation, across subjects was significantly greater than chance ($\mu = 0.63 \pm 0.01$, $t(29) = 9.22$, $p < 0.01$). Classifier performance within the Controlled Retrieval conditions was validated across subjects using the Area Under the ROC Curve (AUC) for each condition type. Performance in these conditions were evaluated with a *t*-test against chance (0.5) showing a significant difference for the Green condition trials ($\mu = 0.573 \pm 0.005$, $t(29) = 13.58$, $p < 0.01$), and no difference from chance for the Blue ($\mu = 0.510 \pm 0.009$, $t(29) = 1.07$, $p = 0.29$) or Red ($\mu = 0.50 \pm 0.01$, $t(29) = 0.10$, $p = 0.92$) conditions. It should be noted that AUC measures, which take into account both sensitivity and specificity, within our two category classification problem provide an equally unbiased estimate of classification performance for both faces and scenes. The tested classifiers were selected based on the best performance using the Green condition trials, and it should be noted that this will inflate performance on these trials relative to the Red and Blue conditions, however the average performance across individuals suggests that the classifiers were at least able to categorize the Green trials above chance. The chance level performance within the Blue and Red conditions is assumed to reflect participants' failure to successfully execute the task on a sizable number of trials.

Descriptive data from selected classifiers are shown in Fig. 3. There was a wide range of training data time windows used in selected classifiers across subjects, which can be seen in Fig. 3A. It's interesting to note that there was no strong bias towards any particular time, which can be seen in the overlaid histogram, as the stimulus was only visually present in the 1-Back task for the first 250 ms of the trial. This suggests that, in general, non-perceptual features are being used in the selected classifier. Similarly, averaged z-scored classifier weights across subjects, shown in Figs. 3B and C, suggest no clear pattern of informative electrodes or frequencies. This complex interaction across time, space, frequency and participant is very difficult parse, and there may be systematic patterns underlying classification weights. However, if we assume that similar bottom-up perceptual processes would produce similar time-locked spectral signatures, these results suggest no clear bias across participants towards bottom-up perceptual features being the most informative features for classification.

Subsequent memory

Subsequent memory was assessed for stimulus type (i.e., Faces or Scenes) and Controlled Retrieval conditions (i.e., Green, Blue, and Red). These accuracy values were corrected by subtracting a within-subject baseline condition accuracy which consisted of word–image pairs studied in the original Paired Associated Learning task but not included in the Controlled Retrieval task. Fig. 4A shows the baseline corrected proportion of stimulus pairs correctly identified in the subsequent memory test. Mean uncorrected accuracy measures for the various conditions are listed in Table 1. A 2(Faces, Scenes) × 3(Green, Blue, Red) Linear Mixed Effects ANOVA showed significant fixed effects across subjects for stimulus type and condition type ($\chi^2(2, N = 30) = 8.16$, $p < 0.05$). Main effects were found in accuracies for Faces greater than Scenes (t = 3.28, Markov Chain Monte Carlo(MCMC) estimated $p < 0.01$), and across condition type (t = −10.04, MCMC $p < 0.01$), with no interaction ($\chi^2(5, N = 30) = 1.04$, $p = 0.95$). Follow-up paired *t*-tests showed the Green condition accuracies to be on average larger than the Blue ($t(29) = 9.20$, $p < 0.01$) and Red ($t(29) = 10.32$, $p < 0.01$) conditions, with no significant difference between the Blue and Red condition accuracies ($t(29) = 0.52$, $p = 0.61$).

Adjusted condition labels for the Blue and Red conditions (denoted as Blue-Neural and Red-Neural in Figs. 4 and 5A) were calculated using classifier output to determine a Memory Activation score, as described in the Classifier training section. The subsequent memory for these adjusted condition labels can be seen in Fig. 4. These adjusted condition labels show a significant difference in the baseline corrected accuracies between the Blue-Neural and Red-Neural conditions ($t(29) = 2.66$, $p < 0.05$). Blue-Neural and Red-Neural conditions, however are not significantly different from baseline accuracy (Blue-Neural: $t(29) = 1.26$, $p = 0.22$, Red-Neural: $t(29) = -1.29$, $p = 0.21$). Differences in accuracy across stimulus type were present in the Blue-Neural condition, where Blue-Neural Face accuracy was greater than Scene accuracy ($t(27) = 2.66$, $p < 0.05$), while no differences by stimulus type in the Red-Neural condition were observed ($t(29) = 1.02$, $p = 0.31$).

Memory activation regression

Using a generalized linear mixed effects logistic regression, an analysis of the impact of Memory Activation scores on subsequent

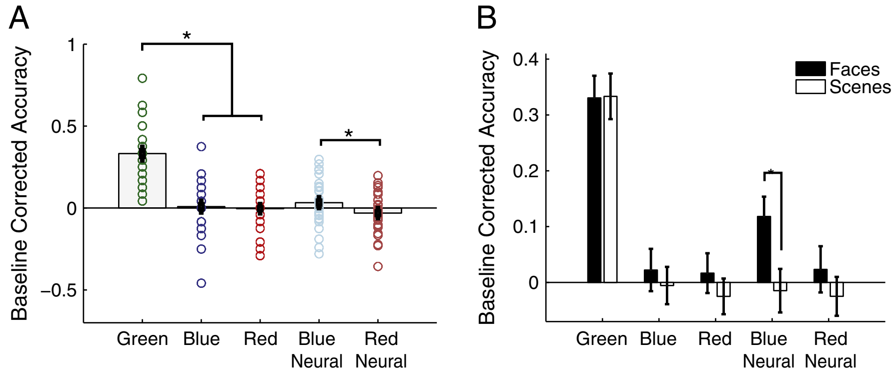


Fig. 4. Subsequent memory related results. *A)* Shows the baseline corrected accuracies across condition type, including classifier adjusted Blue-Neural and Red-Neural conditions. Dots show performance from individual subjects and error bars show Standard Error. *B)* Shows baseline corrected subsequent memory accuracies broken down by stimulus type (Faces in black and Scenes in white).

memory was carried out within the R analysis software using the Linear and Non-Linear Mixed Effects package (Pinheiro et al., 2013; R Core Team, 2012). These results, as seen in Fig. 5B, show the classifier determined Memory Activation scores to have a significant positive relationship with subsequent memory ($\beta = 0.38$, $z = 2.01$, $p = 0.04$). Critically, this estimate was assessed while controlling for any differences in accuracies between stimulus type (i.e., face image pairs or scene image pairs) described in the Subsequent memory section. Similarly, a regression of stimulus type, controlling for subsequent memory, showed no relationship with Memory Activation scores ($z = 0.84$, $p = 0.40$). Therefore the relationship between Memory Activation and subsequent memory cannot be attributed to memory related differences based solely on stimulus type. Finally a fixed effect covariate was also included for each individual's performance on the final cycle of the Paired Associate learning. This covariate allows us to control for any memory differences which might impact individuals' ability to successfully engage in the Controlled Retrieval task.

To provide further support for this relationship we used the Max Difference scores directly in the same mixed effects model used previously and found a significant relationship for the Max Difference score with subsequent memory ($\beta = 0.09$, $z = 2.44$, $p = 0.01$). This shows that Memory Activation scores have a significant positive relationship with subsequent memory even on a trial-by-trial level.

Spectral power analysis

Based on the trial selection described in the Classifier training section, a cluster based analysis of spectral power difference between conditions was carried out. The results from this analysis are broken down into two major contrasts. The first is Suppression vs. Retrieval which is instantiated in $\text{Red}_{succ} > \text{Blue}_{succ}$ contrast. The second targeted contrast is High vs. Low control conditions, which is instantiated in $\text{Blue}_{succ} > \text{Green}$, and $\text{Red}_{succ} > \text{Green}$ contrasts.

Retrieval vs. suppression

The relationship between memory retrieval and suppression of retrieval was probed by contrasting Blue_{succ} and Red_{succ} conditions which are assumed to be equated in control of retrieval demands. This contrast, as seen in Fig. 6, showed a marginally significant cluster within the Theta band averaged over 3 to 8 Hz. The cluster had a temporal extent from 780 to 980 ms and a permutation significance value of $p = 0.058$. The average Green power over time within those

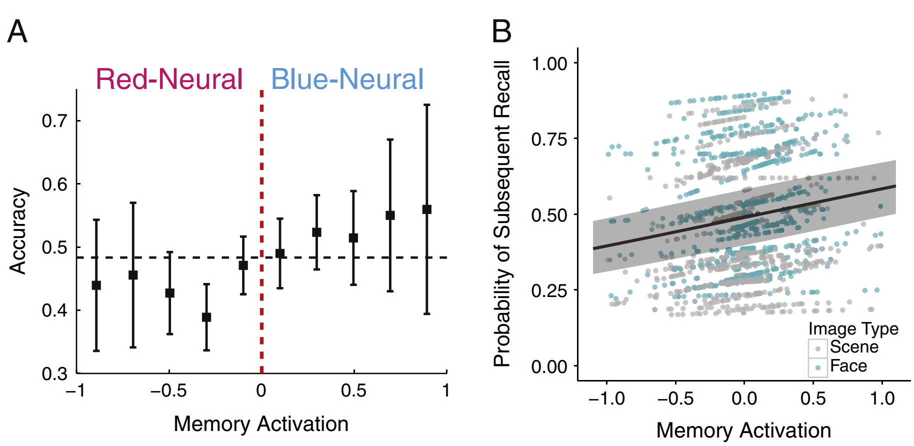


Fig. 5. Memory activation scores as a function of recall. *A)* Subsequent memory accuracies shown as a function of binned Memory Activation scores. Dividing line for adjusted trials labels, Blue-Neural and Red-Neural, shown as the red dotted line. Black dotted line shows mean subsequent memory across Memory Activation scores. Error bars show standard error of the mean. *B)* Results from logistic regression predicting probability of subsequent recall based on classifier determined Memory Activation scores. The color of the dots corresponds to the originally studied image category (faces in cyan and scenes in grey) of the stimuli from which the Memory Activation score was calculated. Fixed effects line shown in black with the transparent bar showing the 95% confidence interval.

Please cite this article as: Ketz, N., et al., Classification aided analysis of oscillatory signatures in controlled retrieval, NeuroImage (2013), http://dx.doi.org/10.1016/j.neuroimage.2013.06.077

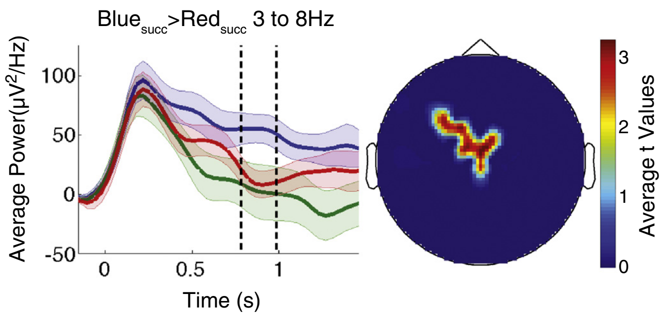


Fig. 6. Significant clusters resulting from the Retrieval vs. Suppression (i.e., $\text{Blue}_{succ} > \text{Red}_{succ}$) conditions. On the left is shown the average power for all three conditions across the frequency band identified, within significant electrodes identified by the cluster analysis. The time window of significance is identified by the dashed line. On the right is the corresponding heat map, for the identified contrast, of the t values within a significant electrode averaged across the significant time window.

electrodes is also shown. However, because of the relatively larger standard error in the Green condition, there is no significant difference between Blue_{succ} and Green trials.

High vs. low control

It is assumed, and supported in subsequent memory accuracies, that Green trials required less control than Blue_{succ} or Red_{succ} trials which both had concurrently presented interfering stimuli during the Controlled Retrieval task. Therefore there are two contrasts that best target the construct of high vs. low retrieval control: $\text{Blue}_{succ} >$ Green and $\text{Red}_{succ} >$ Green.

In the $\text{Blue}_{succ} >$ Green contrast the Beta frequency band, averaged over 12 to 30 Hz, was the only one to show a significant cluster. This cluster, as seen in Fig. 7, had a temporal extent from 460 to 540 ms, and permutation test significance value of $p = 0.032$. The $\text{Red}_{succ} >$ Green contrast showed a significant cluster within the Alpha frequency band averaged over 8 to 12 Hz, and in the Beta band averaged over 12 to 30 Hz, as seen in Fig. 7. The Alpha cluster had a temporal extent from 340 to 660 ms and a permutation significance value of $p = 0.014$. The Beta cluster had a temporal extent from 340 to 500 ms and permutation significance value of $p = 0.020$. To try and differentiate these two clusters the full range of un-averaged frequencies was subject to a cluster analysis and found a single cluster. The significant frequencies of this cluster extended from 7.6 to 37.6 Hz, temporally extended from 220 to 660 ms, and had a permutation significance of $p = 0.008$.

Discussion

The above highlighted results have three major areas of discussion: Memory Activation results, Theta power analysis results, and Alpha vs. Beta power analysis results. Each will be addressed in turn, with three major implications: First, that Think/No-Think results can be at least partially explained by a linear relationship between controlled retrieval success and subsequent memory. Second is a reinforcement of previous results showing Theta power correlated with successful retrieval, and further clarification that suggests Theta effects are unrelated to the controlled suppression of retrieval. Finally the high vs. low retrieval control contrasts help define the functional correlates of Beta oscillations, suggesting they may be related to more general control processes.

Think vs. No-Think

Results show that, on average, classifier-determined Memory Activation scores have a positive linear relationship with subsequent memory. Specifically this implies that the more a given individual recalled a previously-studied associate image, the more likely they are to remember that word–image association later. In the context of Think/No-Think studies this also carries the implication that the

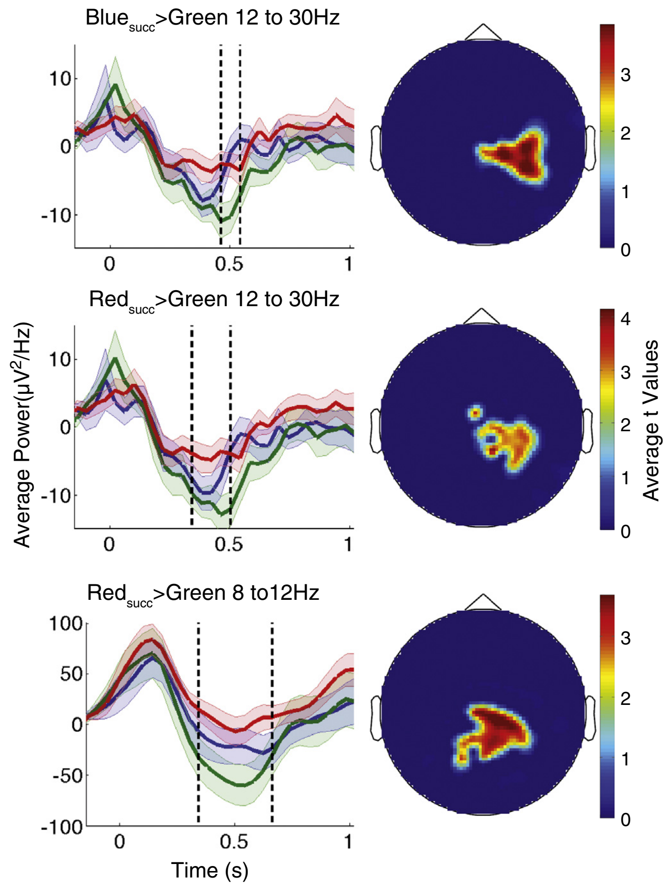


Fig. 7. Three significant clusters resulting from the high vs. low control conditions. On the left is shown the average power for all three conditions across the frequency band identified, within significant electrodes identified by the cluster analysis. The time window of significance is identified by the dashed line. On the right is the corresponding heat map, for the identified contrast, of the t values within a significant electrode averaged across the significant time window.

act of suppressing retrieval reduces an individual's ability to later recall a given item. This relationship, however, could only be found when memory for studied paired associates was assessed through Memory Activation scores, which aggregates the amount of reactivation a given paired-associate experienced across the Controlled Retrieval task. The adjusted trial labels, which were derived from these Memory Activation scores, shown in Fig. 4A do indeed show the classic Think/No-Think pattern, where Think/Blue-Neural trials had a larger baseline corrected accuracy than No-Think/Red-Neural trials. Individually, however, these two conditions are not significantly different from Baseline accuracy.

It should also be noted here that our classifier based regression is only sensitive to image categories, and more specifically to the images of faces and images of scenes used in this experiment. Because our methods do not have the resolution to target individual stimuli (i.e., a particular face or scene) we cannot be certain that the relationship of Memory Activation to subsequent memory is attributable to the reactivation of the originally studied image pair, or just to the general control over retrieval processes. For the purposes of the spectral analysis we need only show that this Memory Activation score is indicative of control over retrieval processes, however further work within this methodology could attempt to use representational similarity to try and gain stimuli specific resolution.

The lack of a significant difference between Baseline subsequent memory and the adjusted condition labels may be attributed to the hypothesis that the Memory Activation scores that determine these labels do not have a simple linear relationship with subsequent

Please cite this article as: Ketz, N., et al., Classification aided analysis of oscillatory signatures in controlled retrieval, NeuroImage (2013), http://dx.doi.org/10.1016/j.neuroimage.2013.06.077

memory (Detre et al., in press; Newman and Norman, 2010; Norman et al., 2007). The proposed hypothesis from Norman et al. (2007) is that feedback inhibition can be used to increase the separation between representations by decreasing the strength of representations that, when cued via some input, only become moderately active. Functionally this serves to reduce the competition for representations that are strongly activated given the same cue. Detre et al. (in press) used a similar classification method implemented in the current study to evaluate the claim that memory re-activation has this non-linear relationship with subsequent memory, such that only moderate levels of re-activation will lead to later deficits in recall. Although our logistic regression of Memory Activation did show a positive linear relationship with subsequent memory, as shown in Fig. 5B, this effect may only be a first order explanation which doesn't fully take into account the moderate levels of activation hypothesis originally proposed in Norman et al. (2007). To capture this effect more sophisticated methods would need to be adopted as our regression approach is insensitive to these more subtle relationships. It is critical to note that our approach does not exclude the possibility that the data is better fit by a non-monotonic model; speculative interpretation of Fig. 5A, indeed, allows room for potential non-monotonic effects within our data.

The results from the logistic regression shown in Fig. 5B can, however, provide a solid grounding for the sub-selection of trials used in the spectral analysis. Further, we have shown that the Max Difference scores also have a positive relationship with subsequent memory. Because of this positive relationship between Memory Activation scores (which are averaged across the trial-by-trial Max Difference scores) and subsequent memory, we can confidently interpret the successfully executed trial labels (e.g. Blue_{succ} and Red_{succ}) as meaningfully related to later memory, and defined by an individual's ability to exert control over retrieval processes.

Retrieval vs. suppression

Results from the spectral power analysis of retrieval vs. suppression (i.e., $\text{Blue}_{succ} > \text{Red}_{succ}$) showed a single significant cluster within the Theta (3 to 8 Hz) frequency range. This is consistent with previous results suggesting a more prominent role for Theta in retrieval. The motivation behind this comparison of active memory retrieval vs. retrieval suppression was in part due to computational models of the hippocampus suggesting that Theta oscillations may be more generally related to control over retrieval processes (both enhancement and suppression) rather than just a marker of successful retrieval or encoding (Hasselmo and Eichenbaum, 2005; Hasselmo et al., 2002). Similarly, several empirical studies have suggested that Theta oscillations are positively correlated with a general control over memory processes (Hanslmayr et al., 2010; Khader and Rösler, 2011). The results of the current study reinforce these previous empirical interpretations on the functional correlates of Theta oscillations in humans, namely that Theta power is correlated with successful retrieval (Nyhus and Curran, 2010). These results, however, also provide clarifying evidence with regards to levels of control over retrieval. In particular, our results suggest that Theta power is correlated with enhancement of retrieval more so than suppression, and is less involved in a general control over retrieval processes.

An open question from these results is the lack of a Theta cluster within the Red_{succ} vs. Green contrast (i.e., high control suppression vs. low control retrieval). If Theta power is positively correlated with retrieval, we might expect to see a cluster related to Green > Red_{succ}. The lack of a significant cluster in the contrast may be explained from previous results which show that higher levels of interference in successfully retrieved memory traces show higher levels of Theta power (Hanslmayr et al., 2010; Khader and Rösler, 2011). The Green condition was subject to the least amount of interference relative to the Blue and Red conditions which both had distractor images presented throughout the course of the experiment. This lack of interference within the Green condition seems to have equated the Green and Red_{succ} conditions in terms of Theta power, which can be seen in Fig. 6. Indeed, a followup paired *t*-test of the average Theta power across the cluster identified significant time/electrode bins shown in Fig. 6, reveals that the Green condition is significantly different from the Blue_{succ} condition ($t(29) = 2.38, p = 0.024$).

A recent study, done within our research group, used a more standard Think/No-Think paradigm to investigate oscillatory signatures of retrieval and suppression of retrieval (Depue et al., in press). Those results show a positive Theta cluster for a No-Think > Think contrast, while the results from the current work show a Theta cluster for Think/Blue_{succ} > No-Think/Red_{succ}. Two informative differences exist between the current work and Depue et al. (in press). The first is the use of interfering/distracting images in the modified Think/No-Think task of the current work, and the second is the sub-selection of successfully executed Think/Blue_{succ} and No-Think/Red_{succ} trials in the current work. It seems the lack of an interference manipulation in Depue et al. (in press) diminished Theta power within the Think condition relative to the current work, and would be similar to our Green condition. Crucially, without trial sub-selection, No-Think trials in Depue et al. (in press) are likely to be more related to retrieval of the cued associate as compared to the current work's Red_{succ} condition, suggesting their No-Think effects are potentially driven by the monitoring of retrieval processes and less driven by a suppression of retrieval processes. These contrasting factors between our study and Depue et al. (in press) provide further constraints on the specificity of cognitive level processes associated with Theta oscillations.

In general the results discussed above suggest that the relatively low levels of interference in the Green condition provide similar changes in Theta power as compared to the controlled suppression of retrieval shown in the Red_{succ} condition. This suggests that suppression of retrieval is similar to retrieval processes with very little interference. We interpret these results to suggest that Theta power does not necessarily positively correlate with all aspects of retrieval processes but rather is more directly related to interference resolution within target representations.

Alpha vs. Beta

Power analysis contrasts of high vs. low control over retrieval showed two main frequency bands of interest, Alpha (8 to 12 Hz), and Beta (12 to 30 Hz). This comparison itself is composed of two specific contrasts, which can be thought of as high controlled retrieval vs. low controlled retrieval (in the Blue_{succ} > Green case), and high control suppression vs. low control retrieval (in the Red_{succ} > Green case). Looking at Fig. 7 shows the Beta cluster was consistent across both contrasts, while the Alpha cluster was witnessed only in the Red_{succ} > Green contrast.

The most straightforward interpretation of these results would suggest that the Green condition follows previously reported patterns of retrieval, i.e., Alpha and Beta bands show a larger decrease in power for well remembered trials (e.g. Green trials) compared to relatively less remembered trials (e.g. Blue and Red) (Depue et al., in press; Hanslmayr et al., 2010; Klimesch, 2012). As can been seen in Fig. 7 our results replicate these previous findings, which are shown in the contrasts highlighting suppression of retrieval compared to retrieval success (i.e., Red_{succ} > Green). The Beta band power changes, however, show a novel relationship to control over retrieval.

As shown in Fig. 7, the Beta frequency band shows an interesting pattern of results within the high vs. low levels of retrieval control. Specifically, a significant cluster was found in the Beta band for both Red_{succ} > Green, and Blue_{succ} > Green. Based on previous findings, the two general expectations for Alpha and Beta power are: (a) that they are negatively correlated with successful retrieval; and (b) that this negative relationship can be facilitated by increasing the amount of interference or control over retrieval processes required (Hanslmayr

Please cite this article as: Ketz, N., et al., Classification aided analysis of oscillatory signatures in controlled retrieval, NeuroImage (2013), http://dx.doi.org/10.1016/j.neuroimage.2013.06.077

et al., 2012; Klimesch, 2012). Our results, in contrast, suggest that there is a diminished decrease in Beta power for high levels of control compared to low. This holds true for both suppression and enhancement of retrieval suggesting Beta to be involved in a more general control related process.

A significant Alpha cluster was found only within the Red_{succ} > Green contrast, suggesting it may be more related to suppression of retrieval as opposed to retrieval control more generally. It must be noted, however, that the potential interpretation of Alpha and Beta clusters as playing differing roles in this instance is dependent on a null effect; that is to say the key $Blue_{succ}$ vs. Green contrast was not associated with a significant cluster within the Alpha band. Follow up tests found tentative evidence to support this interpretation by comparing the difference in power between Red_{succ} and Green with the difference between $Blue_{succ}$ and Green within the electrode/time cluster identified by the Red_{succ} > Green contrast in the Alpha band. This t-test showed a marginally significant difference ($t(29) = 1.82$, $p = 0.078$) suggesting that the Red_{succ} condition shows a larger difference from the Green condition as compared to the difference between $Blue_{succ}$ and Green. However, due to the marginal significance of this interaction, the lack of a Alpha cluster within the high vs. low controlled retrieval contrast should not be over-interpreted in this instance to suggest differing roles for Alpha and Beta. Further testing is necessary to properly establish this relationship.

The $Blue_{succ}$ and Red_{succ} trials used within this design provide a unique condition type that has not previously been investigated. The sub-selection of successfully executed retrieval on a trial-by-trial basis within these conditions allows for an investigation into control over retrieval that is not necessarily bound to subsequent memory. In light of this unique role it is hard to place it precisely in terms of previous results which almost invariably are related to successful retrieval during the subsequent test of memory. This later recall can be disrupted for a various reasons unrelated to prior retrieval success, and does not necessarily reflect the processes occurring during the Think/No-Think manipulations, or in subsequent memory tasks more generally. Results discussed above suggest that previous findings relating the role of Alpha and Beta to successful retrieval may not hold constant when considering successful retrieval independent of subsequent memory, and thatBeta power may play a more general role in control over retrieval processes.

Conclusion

In conclusion, the current work provides three main contributions. The first is that Think/No-Think results can be at least partially explained by a linear relationship between controlled retrieval success and subsequent memory. The second is a reinforcement of previous results showing Theta power correlated with successful retrieval. These results also further clarify this point by suggesting Theta effects are unrelated to the controlled suppression of retrieval, and within successful retrieval Theta is more related to the control over interference. Finally the high vs. low retrieval control contrasts help define Beta functional correlates. These results suggest Beta power shows a diminished level of desynchronization in cases of high levels of required control compared to low, and that Beta power may be more broadly related to control processes. The more general implications of this work suggest that stimuli which later show successful subsequent memory may have differential neural signatures from those that show instantaneous successful retrieval.

The relative novelty of scalp based EEG classification should also be noted. Very few studies have shown success within this domain (however see Morton et al. (in press) for a similar approach) and we hope our methods can contribute to furthering this analysis approach in more studies. In general, however, much more work is required to better understand the neural sources of these oscillations. We use these scalp based oscillatory measures as indicators of differential processes, however we can make no strong claims as to where or how these processes are dissociated outside of their functional correlates. We feel the mechanistic modeling of these neural signals is crucial in advancing the understanding of these cognitive level signatures, and have targeted it for future work.

Acknowledgments

We would like to thank Catherine Baker, Cameron Barton, Christopher Bird, Ryan Guild, Shaina Martis, Samantha Rubeck, Katherine Russman, Leah Tinsley, and Rosalyn Wong for testing participants, as well as Matt Mollison, Yuko Munakata, Brendan Depue, and Ken Norman for general design and analysis support.

Portions of the research in this paper use the Color FERET database of facial images collected under the FERET program (Phillips et al., 2000). This research was funded by NIH Grants MH64812, and MH079485; NSF grant #SMA-1041755 to the Temporal Dynamics of Learning Center, an NSF Science of Learning Center, NSF Research Fellowship grant #DGE-0707432; Office of Naval Research grants N00014-13-1-0067, and N00014-10-1-0177.

Conflict of Interest

The authors have no conflicts of interest.

References

Anderson, M.C., Green, C., 2001. Suppressing unwanted memories by executive control. Nature 410, 366 (03, URL http://www.ncbi.nlm.nih.gov/pubmed/11268212).

Brainard, D., 1997. The psychophysics toolbox. Spat. Vis. 10 (4), 433–436 (URL http://www.ncbi.nlm.nih.gov/pubmed/9176952).

Bulevich, J.B., Roediger, H.L., Balota, D.A., Butler, A.C., 2006. Failures to find suppression of episodic memories in the think/no-think paradigm. Mem. Cognit. 34 (8), 1569–1577 (Dec, URL http://www.ncbi.nlm.nih.gov/pubmed/17489284).

Burgess, A.P., Gruzelier, J.H., 1997. Short duration synchronization of human theta rhythm during recognition memory. Neuroreport 8, 1039 (07, URL http://www.ncbi.nlm.nih.gov/pubmed/9141088).

Buzsaki, G., Draguhn, A., 2004. Neuronal oscillations in cortical networks. Science 304 (5679), 1926–1938 (December).

Depue, B.E., Curran, T., Banich, M.T., 2007. Prefrontal regions orchestrate suppression of emotional memories via a two-phase process. Science 317 (5835), 215–219 (Jul, URL http://www.ncbi.nlm.nih.gov/pubmed/17626877).

Depue, B., Ketz, N., Mollison, M., Nyhus, E., Banich, M., Curran, T., 2013. Event related potentials and neural oscillations during volitional suppression of memory retrieval. J. Cogn. Neurosci. (in press).

Detre, G.J., Natarajan, A., Gershman, S.J., Norman, K.A., 2013. Moderate levels of activation lead to forgetting in the think/no-think paradigm. Neuropsychologia (in press).

Dien, J., 1998. Issues in the application of the average reference: review, critiques, and recommendations. Behav. Res. Methods Instrum. Comput. 30, 34–43 (January).

Düzel, E., Habib, R., Schott, B., Schoenfeld, A., Lobaugh, N., McIntosh, A.R., Scholz, M., Heinze, H.J., 2003. A multivariate, spatiotemporal analysis of electromagnetic time-frequency data of recognition memory. NeuroImage 18, 185–197 (Feb, URL http://www.ncbi.nlm.nih.gov/pubmed/12595175).

Ernst, M.D., 2004. Permutation methods: a basis for exact inference. Stat. Sci. 19 (4), 676–685.

Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recognit. Lett. 27 (8), 861–874.

Fell, J., Axmacher, N., 2011. The role of phase synchronization in memory processes. Nat. Rev. 12 (2), 105–118 (Feb, URL http://www.ncbi.nlm.nih.gov/pubmed/21248789).

Gruber, T., Tsivilis, D., Giabbiconi, C.-M., Müller, M.M., 2008. Induced electroencephalogram oscillations during source memory: familiarity is reflected in the gamma band, recollection in the theta band. J. Cogn. Neurosci. 20, 1043–1053 (Jun, URL http://www.ncbi.nlm.nih.gov/pubmed/18211247).

Hanslmayr, S., Spitzer, B., Buml, K.-H., 2009. Brain oscillations dissociate between semantic and nonsemantic encoding of episodic memories. Cereb. Cortex 19, 1631–1640 (Jul, URL http://www.ncbi.nlm.nih.gov/pubmed/19001457).

Hanslmayr, S., Staudigl, T., Aslan, A., Bäuml, K.-H., 2010. Theta oscillations reflect the dynamics of interference in episodic memory retrieval. J. Neurosci. 30, 11356–11362 (Aug, URL http://www.ncbi.nlm.nih.gov/pubmed/20739556).

Hanslmayr, S., Staudigl, T., Fellner, M.-C., 2012. Oscillatory power decreases and long-term memory: the information via desynchronization hypothesis. Front. Hum. Neurosci. 6 (URL http://www.ncbi.nlm.nih.gov/pubmed/22514527).

Hasselmo, M.E., Eichenbaum, H., 2005. Hippocampal mechanisms for the context-dependent retrieval of episodes. Neural Netw. 18, 1172–1190 (Nov, URL http://www.ncbi.nlm.nih.gov/pubmed/16263240).

Hasselmo, M.E., Bodelon, C., Wyble, B.P., 2002. A proposed function for hippocampal theta rhythm: separate phases of encoding and retrieval enhance reversal of prior learning. Neural Comput. 14 (4), 793–818 (04, URL http://www.ncbi.nlm.nih.gov/pubmed/11936962).

Junghfer, M., Elbert, T., Tucker, D.M., Braun, C., 1999. The polar average reference effect: a bias in estimating the head surface integral in eeg recording. Clin. Neurophysiol. 110, 1149–1155 (Jun, URL http://www.ncbi.nlm.nih.gov/pubmed/10402104).

Karpicke, J.D., Roediger, H.L., 2008. The critical importance of retrieval for learning. Science 319, 966–968 (Feb, URL http://www.ncbi.nlm.nih.gov/pubmed/18276894).

Khader, P.H., Rösler, F., 2011. Eeg power changes reflect distinct mechanisms during long-term memory retrieval. Psychophysiology 48, 362–369 (Mar, URL http://www.ncbi.nlm.nih.gov/pubmed/20624249).

Klimesch, W., 2012. Alpha-band oscillations, attention, and controlled access to stored information. Trends Cogn. Sci. 16, 606–617 (Dec, URL http://www.ncbi.nlm.nih.gov/pubmed/23141428).

Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. J. Neurosci. Methods 164 (1), 177–190 (Aug, URL http://www.ncbi.nlm.nih.gov/pubmed/17517438).

Morton, N.W., Kahana, M.J., Rosenberg, E.A., Baltuch, G.H., Litt, B., Sharan, A.D., Sperling, M.R., Polyn, S.M., 2013. Category-specific neural oscillations predict recall organization during memory search. Cereb. Cortex (Aug in press, URL http://www.ncbi.nlm.nih.gov/pubmed/22875859).

Newman, E.L., Norman, K.A., 2010. Moderate excitation leads to weakening of perceptual representations. Cereb. Cortex (New York, N.Y.) 20 (11), 2760–2770 (Feb, URL http://www.ncbi.nlm.nih.gov/pubmed/20181622).

Norman, K.A., Newman, E.L., Detre, G., 2007. A neural network model of retrieval-induced forgetting. Psychol. Rev. 114 (4), 887–953 (Oct, URL http://www.ncbi.nlm.nih.gov/pubmed/17907868).

Nyhus, E., Curran, T., 2010. Functional role of gamma and theta oscillations in episodic memory. Neurosci. Biobehav. Rev. 34 (7), 1023–1035 (Jun, URL http://www.ncbi.nlm.nih.gov/pubmed/20060015).

Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.-M., 2011. Fieldtrip: Open source software for advanced analysis of meg, eeg, and invasive electrophysiological data. Comput. Intell. Neurosci. 2011 (URL http://www.ncbi.nlm.nih.gov/pubmed/21253357).

Osipova, D., Takashima, A., Oostenveld, R., Fernández, G., Maris, E., Jensen, O., 2006. Theta and gamma oscillations predict encoding and retrieval of declarative memory. J. Neurosci. 26 (Jul, URL http://www.ncbi.nlm.nih.gov/pubmed/16837600).

Phillips, P.J., Moon, H., Rizvi, S.A., Rauss, P.J., 2000. The feret evaluation methodology for face-recognition algorithms. Pattern Anal. Mach. Intell., IEEE Trans. 22 (10), 1090–1104.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., R Core Team, 2013. nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-109.

R Core Team, 2012. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria3-900051-07-0 (URL http://www.R-project.org/).

Sederberg, P.B., Kahana, M.J., Howard, M.W., Donner, E.J., Madsen, J.R., 2003. Theta and gamma oscillations during encoding predict subsequent recall. J. Neurosci. 23, 10809–10814 (11, URL http://www.jneurosci.org/cgi/content/abstract/23/34/10809).

Tomlinson, T.D., Huber, D.E., Rieth, C.A., Davelaar, E.J., 2009. An interference account of cue-independent forgetting in the no-think paradigm. Proc. Natl. Acad. Sci. U. S. A. 106 (37), 15588–15593 (Sep, URL http://www.ncbi.nlm.nih.gov/pubmed/19717438).

Waldhauser, G.T., Johansson, M., Hanslmayr, S., 2012. Alpha/beta oscillations indicate inhibition of interfering visual memories. J. Neurosci. 32 (6), 1953–1961 (Feb, URL http://www.ncbi.nlm.nih.gov/pubmed/22323708).

Xiao, J., Hays, J., Ehinger, K., Oliva, A., Torralba, A., 2010. Sun database: large-scale scene recognition from abbey to zoo. Computer vision and pattern recognition (CVPR), 2010 IEEE conference on. IEEE, pp. 3485–3492.

Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. J. R. Stat. Soc. Ser. B 67 (2), 301–320 (xx, URL http://econpapers.repec.org/article/blajorssb/v3a673ay3a20053ai3a23ap3a301-320:htm).