

CMA

## Mapping working memory retrieval in space and in time: A combined electroencephalography and electrocorticography approach



Qiong Zhang<sup>a,\*</sup>, Marieke van Vugt<sup>b</sup>, Jelmer P. Borst<sup>b</sup>, John R. Anderson<sup>c</sup>

<sup>a</sup> Machine Learning Department, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA, 15213, United States

<sup>b</sup> Department of Artificial Intelligence, University of Groningen, 9747 AG, Groningen, The Netherlands

<sup>c</sup> Psychology Department, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA, 15213, United States

### ARTICLE INFO

#### Keywords:

Spatial-temporal analysis  
Sternberg task  
Hidden semi-Markov model  
Working memory  
Electroencephalography  
Electrocorticography

### ABSTRACT

In this study, we investigated the time course and neural correlates of the retrieval process underlying visual working memory. We made use of a rare dataset in which the same task was recorded using both scalp electroencephalography (EEG) and electrocorticography (ECoG), respectively. This allowed us to examine with great spatial and temporal detail how the retrieval process works, and in particular how the medial temporal lobe (MTL) is involved. In each trial, participants judged whether a probe face had been among a set of recently studied faces. With a method that combines hidden semi-Markov models and multivariate pattern analysis, the neural signal was decomposed into a sequence of latent cognitive stages with information about their durations on a trial-by-trial basis. Analyzed separately, EEG and ECoG data yielded converging results on discovered stages and their interpretation, which reflected 1) a brief pre-attention stage, 2) encoding the stimulus, 3) retrieving the studied set, and 4) making a decision. Combining these stages with the high spatial resolution of ECoG suggested that activity in the temporal cortex reflected item familiarity in the retrieval stage; and that once retrieval is complete, there is active maintenance of the studied face set in the decision stage in the MTL. During this same period, the frontal cortex guides the decision by means of theta coupling with the MTL. These observations generalize previous findings on the role of MTL theta from long-term memory tasks to short-term memory tasks.

### Introduction

Information can be preserved in working memory across a short delay without active maintenance (Ericsson and Kintsch, 1995; Lewis-Peacock et al., 2012; Oberauer, 2002). Our interest is in the process by which this information is later retrieved. This can take place either through an attention-based process that refocuses and refreshes the memory traces (Lewis-Peacock et al., 2012; Souza et al., 2014), or a cue-driven process that is very similar to cue-based retrieval from long-term memory (Nairne, 2002). The objective of this study to map the time course and neural correlates of this retrieval process. We focus on a visual working memory task from a published dataset (Van Vugt et al., 2013). Two experiments using the same task were carried out using scalp electroencephalography (EEG) and electrocorticography (ECoG), which allows us to harness the complementary strengths of these two recording methods. In each trial, participants first studied a list of faces, then, after a short delay, they were cued with a probe face

and asked to judge if it had been among the just-studied faces (i.e., a Sternberg task). To create a detailed mapping of the time course of the retrieval process, we modeled the trial-to-trial variability of this process with a novel method that combines hidden semi-Markov models with multivariate pattern analysis (HSMM-MVPA; e.g., Anderson et al., 2016), which we applied to both EEG and ECoG datasets. To create a detailed mapping of the neural correlates we relied on the ECoG dataset, which has superior spatial resolution. In the rest of this introduction, we will describe three main challenges to isolate the working memory retrieval process in space and in time, and our approaches to tackle these challenges. In the method section, we will provide the technical details of these approaches.

#### *The first challenge: trial-to-trial variability in the timing of cognitive processes*

We focus on the period of the working memory task where

\* Corresponding author.

E-mail addresses: [qiongz@andrew.cmu.edu](mailto:qiongz@andrew.cmu.edu) (Q. Zhang), [m.k.van.vugt@rug.nl](mailto:m.k.van.vugt@rug.nl) (M. van Vugt), [j.p.borst@rug.nl](mailto:j.p.borst@rug.nl) (J.P. Borst), [ja@cmu.edu](mailto:ja@cmu.edu) (J.R. Anderson).

<https://doi.org/10.1016/j.neuroimage.2018.03.039>

Received 23 December 2017; Received in revised form 27 February 2018; Accepted 17 March 2018

Available online 20 March 2018

1053-8119/© 2018 Elsevier Inc. All rights reserved.

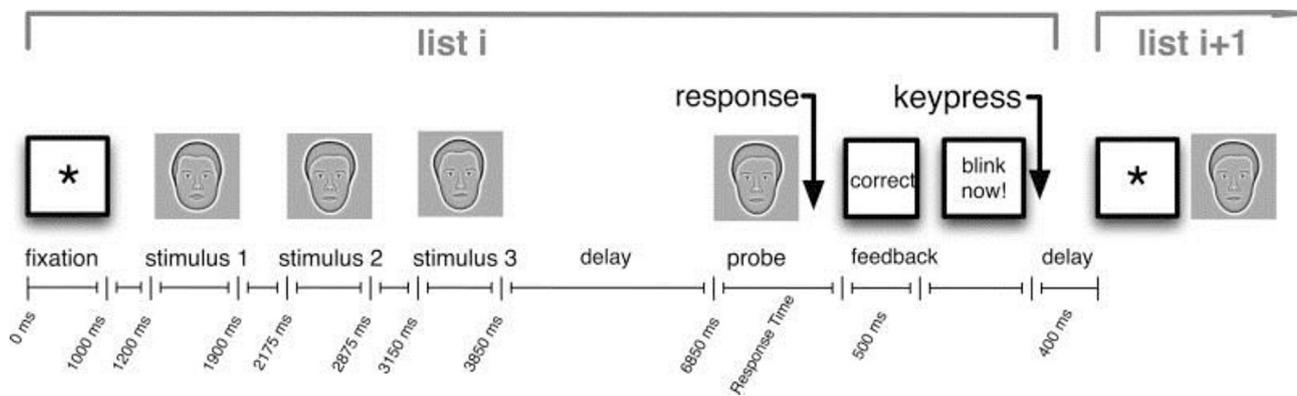


Fig. 1. Trial structure of the Sternberg task. This figure illustrates the sequence and timing of events in a trial with a set size of 3. Adopted from (Van Vugt et al., 2013) with permission.

participants are asked to judge whether the presented probe is one of the recently studied faces. Each trial analyzed starts with encoding a probe face and ends with a motor response (the period marked as ‘probe’ in Fig. 1), but the intermediate cognitive processes (including the retrieval process) can vary greatly from trial to trial given the self-paced nature of the task. Event-related potentials (ERP) are commonly used in EEG literature to identify the occurrence of important cognitive events by averaging stimulus-locked or response-locked signal across trials. However, cognitive events that are further away from stimulus or response will have higher trial-to-trial temporal variability and are often distorted or lost in the averaged waveforms (Luck, 2014). Therefore, in order to isolate the full range of cognitive processes in the visual working memory task, we need to apply a method that can capture the trial-to-trial variability in when they occur.

In the past, we have combined hidden semi-Markov models and multivariate pattern analysis (i.e. HSMM-MVPA) to decompose the neural signal (i.e. EEG, MEG) into a sequence of latent stages (Anderson et al., 2016; Borst and Anderson, 2015; Zhang et al., 2017b). This approach assumes that transitions from one processing stage to the next are accompanied by a significant change in the information processing and therefore in the underlying neural signal. HSMM-MVPA shares the same notion as the microstate analysis developed by Lehmann and colleagues, which is to extract a sequence of non-overlapping mental states from neural data during a specific task (Lehmann, 1987; Pascual-Marqui et al., 1995). In addition, the HSMM-MVPA method also captures the trial-to-trial variability of the durations in each state. This is in contrast to the microstate analysis that was applied to average ERP data (Pascual-Marqui et al., 1995).

The HSMM-MVPA model identifies brief, distinctive profiles of scalp activity (i.e., bumps) with variable latencies in single trial EEG data. A bump is modeled as a half-sine multidimensional peak across the scalp that signifies a significant change in the information processing. This assumption is inspired by two theories of ERP generation (Yeung et al., 2004; Makeig, 2002; Shah et al., 2004; Basar, 1980). According to the classical theory, significant cognitive events generate bursts of activity in discrete brain regions (Shah et al., 2004). Therefore, the EEG signal can be described as a sum of sinusoidal peaks and ongoing neural signal of uncorrelated sinusoidal variation. According to the second theory of ERP generation, the synchronized oscillation theory, significant cognitive events reset the phase of the oscillation at a certain frequency (Basar, 1980). Under both theories, averaging neural signal across trials will reveal the peaks as averaged ERP waveforms that we see, and the ERP waveforms are indistinguishable between the two theories under simulated datasets (Yeung et al., 2007).

In the current work, we will apply this method to both the EEG and ECoG datasets to identify and isolate the cognitive processes in our visual working memory task. Details on parameter estimation and model fitting can be found in the Method section.

### The second challenge: isolate the retrieval process from the discovered stages

The HSMM-MVPA method gives a sequence of latent stages that correspond to different cognitive processes. We are particularly interested in when the retrieval process takes place. In order to isolate this retrieval process among the set of stages uncovered by HSMM-MVPA, we need to find the mapping between the obtained model stages and the series of cognitive processes thought to be involved in our task. Extending an existing stage model based on an HSMM-MVPA analysis of a Sternberg task involving digits (Anderson et al., 2016), we expect to find the following five stages of processing:

- 1) Pre-attention: the time for the visual signal to reach the brain and to be attended to;
- 2) Encoding: encode the probe presented;
- 3) Retrieval: reactivate face(s) from the studied memory set;
- 4) Decision: compare and decide if the probe face is part of the reactivated set;
- 5) Motor response: press a key that reflects the decision.

To confirm this mapping for the current task, we will also examine how stage durations vary given the manipulation of different experimental factors. If a particular stage corresponds to a given cognitive process, the way that different experimental factors alter the duration of this stage should be consistent with our knowledge of the corresponding cognitive process. We use regular scalp EEG with normal populations to identify such condition effects, because there is more power associated with larger number of subjects.<sup>1</sup>

### The third challenge: identify the neural correlates associated with the retrieval process

Processing stages obtained in the HSMM-MVPA method provide us with the fine temporal resolution of when the retrieval process takes place on single trials. We can further examine which brain regions are activated during these discovered periods in a combined spatio-temporal analysis. With EEG data, we can identify the mapping from HSMM-MVPA stages to individual cognitive processes with high temporal precision, but poor spatial resolution limits our ability to determine how different brain regions are engaged during these periods. To achieve a better spatial resolution, we look into the ECoG activity during the periods of interest. ECoG recordings from epileptic patients not only give finer spatial resolution in the cortical regions, but also make it possible to examine subcortical activity such as that of MTL, which plays an important role in

<sup>1</sup> In contrast, there are limited opportunities to collect many subjects of ECoG data.

visual (working) memory tasks (Van Vugt et al., 2010; Ranganath, 2006). Because recording sites in ECoG data vary across subjects in both numbers and locations, we need a method of subject alignment before pooling data from different subjects. Multi-set canonical correlation analysis (M-CCA) is used for this purpose to transform electrode activity in each subject to a common neural representational space, where the inter-subject correlations of the transformed data are maximized across subjects. We have previously demonstrated the reliability of M-CCA in aligning subjects in MEG data (Zhang et al., 2017a). In the current work, we will apply M-CCA to align ECoG data from individual subjects; we will also apply M-CCA to the EEG data, which allows us to compare the obtained common dimensions and the HSMM-MVPA results across two experiments in two different measurement modalities.

## Methods

### Experimental paradigm

Participants completed a Sternberg task: Following the appearance of a fixation stimulus, participants viewed a short series of faces in sequence (Fig. 1). After a retention interval, a probe item appeared and participants indicated with a key press whether the probe was a member of the just presented set (target) or not (foil). After each trial, participants were given accuracy feedback. Fig. 1 illustrates the sequence and timing of trial events in the original report of the data used in this study (Van Vugt et al., 2013). 16 synthetic faces, which varied along four perceptual dimensions, were used as stimuli (Wilson et al., 2002). A multidimensional scaling (MDS) study was carried out on 23 participants to characterize the psychological perceived similarities among these synthetic faces (see Van Vugt et al., 2013; Kahana and Bennett, 1994, for details). Therefore, we can characterize the overall similarity between a probe and the just studied set as the average similarity between the probe and each face on the set (i.e., set similarity). We define two levels of set similarity among the foils (i.e., low-sim and high-sim), dividing foils into two groups with equal numbers of trials.

In the datasets we will use, there are two versions of the visual working memory task, adapted to the cognitive capacities of the respective subject populations. One was administered to healthy undergraduates while their EEG was recorded (Experiment 1), while the other was adapted to epileptic patients from who ECoG was recorded (Experiment 2). The two experiments differed in the number of faces participants viewed in the studied set before a probe in each series (set size). In Experiment 1, set size could be two, three or four; in Experiment 2, set size could be one, two or three (adapted to the capacity of the patients). Sets were constructed so that items could not be repeated on successive sets, and targets were equally likely to match a study item from each serial position. Incorrect trials and trials with RTs shorter than 400 ms or longer than 4000 ms were removed from the analysis.

### Participants

In Experiment 1, 29 adults (ages 20–32) were recruited from the University of Pennsylvania student community. Informed consent was obtained from all participants. Each participant completed two sessions, with each session involving ten blocks of 30 trials. Participants with mean accuracy lower than 60% or with mean RT longer than 4s were excluded. All 29 participants were retained for further analysis.

Participants in Experiment 2 were 16 neurosurgical patients being treated for medically refractory epilepsy and were monitored with arrays of subdural and/or depth electrodes. Patients were recruited from Brigham and Women's Hospital in Boston, the Hospital of the University of Pennsylvania in Philadelphia, and Universitäts Klinikum Freiburg in Germany. Informed consent was obtained from all participants. Under the same accuracy and RT criteria as Experiment 1, 12 out of 16 patients were retained for further analysis. Each patient completed different number of trials for the experiment ( $\mu = 202$ ,  $SEM = 41$ ), depending on

**Table 1**

Number of patients and electrodes in each of the five ROIs.

Numbers	Electrodes	Patients
Temporal Cortex	269	12
Medial Temporal Lobe	69	10
Frontal Cortex	161	9
Parietal Cortex	93	9
Occipital Cortex	22	7

their availability and willingness to participate in the experiments.

### Scalp EEG recordings in experiment 1

Scalp EEG signals were recorded using a 129-channel EGI Inc. system, with an AC-coupled, high-input-impedance amplifier (200 M $\Omega$ , Net Amps, Electrical Geodesics, Inc., Eugene, OR). The sampling rate was 500 Hz, and data were recorded with a 0.1–250 Hz bandpass filter. Individual channels were adjusted until impedances were below 50 k $\Omega$ . EEG signals were filtered with a bandpass of 0.1–70.0 Hz, and then decomposed into independent components using the EEGLAB FastICA algorithm (Delorme and Makeig, 2004). Components associated with eye blinks were automatically identified and projected out of the EEG recording. Epochs (from probe presentation to motor response in each trial) were then extracted from the continuous recording. Epochs containing voltages above +100  $\mu$ V or below –100  $\mu$ V were excluded. Data were down-sampled to 100Hz before further analysis. More details of the experiment can be found in the original report of the EEG data (Van Vugt et al., 2013).

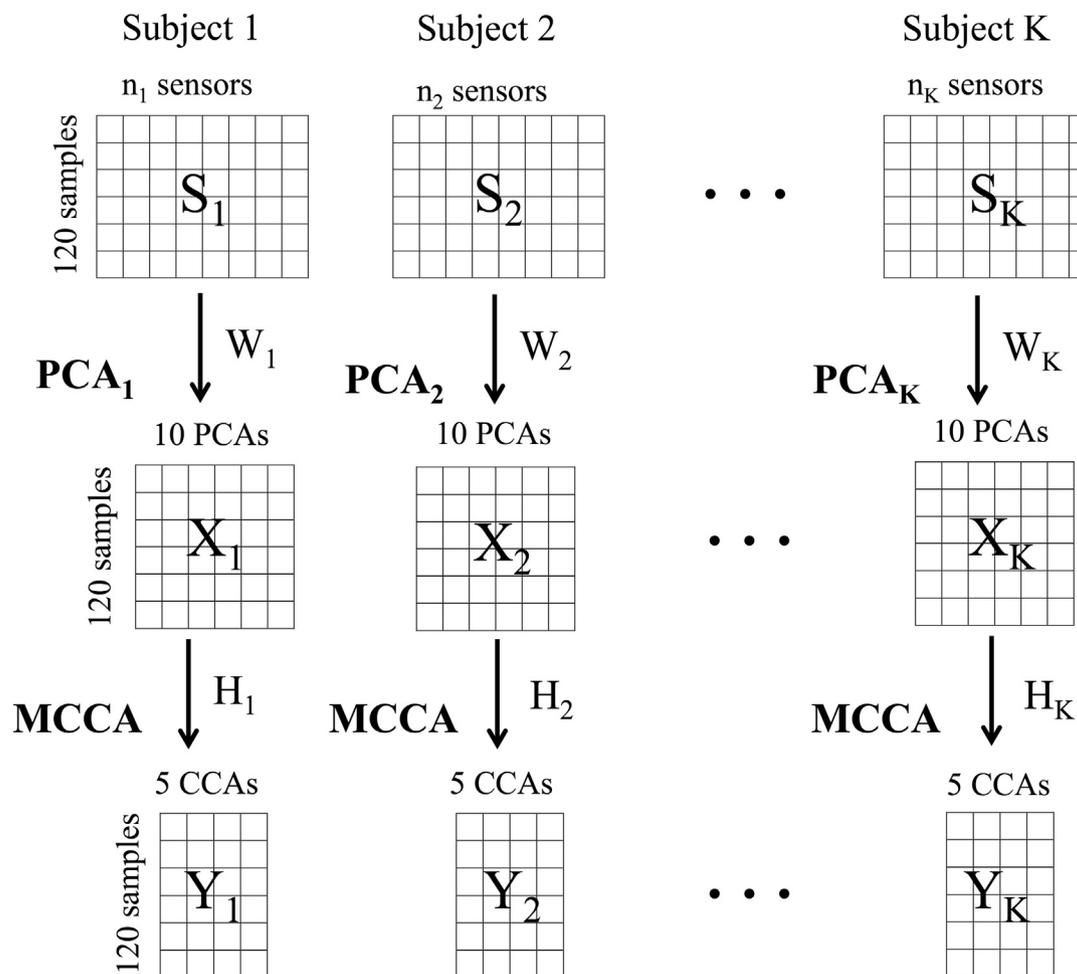
### ECoG and depth electrode recordings in experiment 2

The local field potential was amplified and digitally recorded at sampling rates between 250 and 1024 Hz,<sup>2</sup> and bandpass-filtered between 0.1 and 100 Hz. Data were subsequently notch-filtered with a Butterworth filter with zero phase distortion to eliminate line noise. For all participants, the locations of implanted electrodes were determined by means of co-registered postoperative computed tomographies and preoperative magnetic resonance imaging (MRI) or from postoperative MRIs by an indirect stereotactic technique and converted into MNI (Montreal Neurological Institute) coordinates. Localization of depth electrode contacts in the medial temporal lobe (MTL) was done manually through clinician's inspection of the postoperative MRIs, which includes areas in hippocampus and parahippocampal gyrus. Analysis was done in five pre-defined regions of interest (ROIs; see Table 1), and data were down-sampled to 100Hz. More details of the experiment can be found in the original reports of the ECoG data (Van Vugt et al., 2013; Van Vugt et al., 2010).

### Multi-set canonical correlation analysis (M-CCA)

While it is common practice to assume that the sensors for different subjects in EEG recordings correspond, recording sites in ECoG data vary from subject to subject in both number and locations since localization is driven solely by clinical considerations. Therefore, in Experiment 2, a method of subject alignment is required before pooling data from all subjects together. Previously, we have found success in aligning subjects with multi-set canonical correlations analysis (M-CCA) using only functional information of the neural data in MEG datasets (Zhang et al., 2017a). This method is used to find the optimal transformation for each subject from electrode activity to a common neural representational space, where the inter-subject correlations of the transformed data are

<sup>2</sup> The sampling rates vary greatly as a result of different protocols used in three different hospitals where ECoG data were collected.



**Fig. 2.** An illustration to demonstrate an application of the M-CCA procedure to EEG data in Experiment 1 ( $K = 29$ ) or ECoG data in Experiment 2 ( $K = 12$ ).  $S_k$  is the averaged data (across all trials for each condition) from all electrodes in subject  $K$  with 120 time samples;  $X_k$  has 10 PCA components for subject  $K$ , each with 120 time samples;  $Y_k$  has 5 CCA components, each with 120 time samples;  $W_1, W_2, \dots, W_{12}$  are PCA weights obtained for each subject independently; and  $H_1, H_2, \dots, H_{12}$  are CCA weights obtained jointly from all subjects by maximizing all of the inter-subject correlations.

maximized across subjects.

Fig. 2 illustrates the steps to apply M-CCA. The procedure starts by averaging multiple trials to obtain a highly reliable representation of the change in sensor activity for each ECoG subject (indicated in the figure as  $S_k$ ). This is similar to obtaining event-related potential waveforms in the EEG literature (Picton et al., 2000). However, trials are quite variable in their duration, and temporal alignment is lost when a time sample is further away from stimulus presentation or response emission. Samples locked to response emission in Experiment 2 also have poor temporal alignment, as there is potential delay in response timing given the condition of neurosurgical patients. Therefore, we only use samples from the first 600 ms (60 samples given the sampling rate of 100Hz) of a trial when applying M-CCA in Experiment 2. This averaging process is repeated for the target and foil conditions separately, as we potentially have different latent components for different conditions after averaging. This gives rise to 120 samples (60 samples  $\times$  2 conditions) per subject as the input  $S$  of M-CCA shown in Fig. 2. To reduce dimensionality and remove subject-specific noise, we perform a spatial PCA to obtain the top 10 components on these matrices for each subject first,<sup>3</sup> instead of applying M-CCA directly to the sensor data  $S$ . This results in 12 matrices  $X$  of dimension  $120 \times 10$ . The procedure also yields PCA weights,  $W_s$ ,

obtained independently for each subject. Next, the M-CCA estimates a set of weights,  $H_s$ , that will map each subject's PCAs onto a common set of dimensions. Once  $W_s$  and  $H_s$  are obtained, we can go back to full-time-course data of individual trials and transform them from representation in electrodes to representation in CCA dimensions. The top 5 CCA dimensions are retained for each subject.<sup>4</sup> M-CCA was also applied to the EEG dataset in Experiment 1 using the exact same procedure, allowing us to better compare the two experiments.

#### HSMM-MVPA

The HSMM-MVPA method explicitly models the variability of endogenous ERP components that would otherwise be distorted or lost in the average waveforms. Previous applications of the HSMM-MVPA method to EEG data were effective in recovering the durations of the underlying processing stages (e.g., recollection, decision) and showed predictable changes with experimental factors in an associative recognition task for word pairs and a Sternberg task (Anderson et al., 2016; Zhang et al., 2017b). The HSMM-MVPA model identifies brief, distinctive profiles of scalp activity (i.e., bumps) with variable latencies in the single

<sup>3</sup> 10 PCA retained for each subject explained 85% of the variance for EEG subjects, and 64% for ECoG subjects.

<sup>4</sup> Averaged inter-subject correlations over the left-out half of the data for the first five CCA components is 0.55 for EEG subjects and 0.48 for ECoG subjects. Starting from the sixth CCA component for both EEG and ECoG subjects, inter-subject correlations do not generalize well on the testing data ( $<0.3$ ).

trial EEG data. A bump is modeled as a half-sine multidimensional peak across the scalp that signifies a significant change in the information processing, followed by a flat period where the signal appears as ongoing sinusoidal noise around a mean of 0. HSMM-MVPA models the durations of the flats as gamma distributions. The HSMM-MVPA method was applied to the first 5 CCA components, separately over two experiments. The CCA components were z-scored for each trial. As a result, the data for the analysis consisted of 5 orthogonal CCA components in each experiment, sampled every 10 ms and with constant mean and variability across trials. Only correct trials are considered in the analyses presented below.

As described in more detail in our previous application of the HSMM-MVPA method (Anderson et al., 2016), a  $n$ -bump HSMM requires estimating  $n + 1$  stage distributions to describe the durations of the flats plus the  $n$  5-sample bumps for each CCA component. A different magnitude is estimated for each of the  $n$  bumps along each CCA dimension. A bump extends temporally across 5 samples (50 ms) and is multiplied by weights of 0.309, 0.809, 1.000, 0.809, and 0.309 (i.e., a 10-Hz half sine wave). The best model fit of such HSMMs is given by maximizing the summed log likelihood of the bumps and flats across all trials. For each trial, this log likelihood reflects the combination of two factors: the likelihood of the EEG data given that the bumps are centered at each time point, and the likelihoods that the bumps are centered at those time points given the gamma distributions that constrain their locations. In other words, the HSMM must select bump locations within a trial to maximize the correspondence between the observed and the estimated EEG/ECoG signal, while selecting relatively consistent flat durations across trials to maximize their fit to the gamma distributions.

#### Brain synchrony analysis

Quantification of phase synchrony between two neural signals was done by means of a phase locking value (PLV). PLV measures the consistency in the phase difference at a frequency of interest between two recording sites across multiple trials at the corresponding time points (Lachaux et al., 1999). Phases for single-trial neural signal at different brain regions were measured using Hilbert transform (Tass et al., 1998; Lachaux et al., 1999). It is equivalent to an alternative method by using convolution with a complex wavelet, as demonstrated in a direct comparison study (Le Van Quyen et al., 2001). If the phase difference between signals from two recording sites is very similar from trial to trial at the corresponding time point, then it is considered that at that particular time point, the two brain sites are well synchronized in phase with PLV close to 1; if the phase difference is very variable across trials, PLV is close to 0.

Calculation of phase locking values requires establishing corresponding time points across multiple trials. Typically one assumes that samples correspond when they are at the same delay from the stimulus. This assumption is only approximately correct given the trial-to-trial variability of where the same cognitive event occurs. Alternatively, one can assume that samples correspond when they are at the same offset from the same bump in the HSMM-MVPA analysis (Portoles et al., 2018). In this way, phase locking value calculates how synchronized two brain sites are around the cognitive event signified by a particular bump. In this experiment, we are interested in the synchronization between the MTL and cortical regions around when retrieval completes. Following the procedure outlined by Lachaux et al. (1999), we bandpass-filtered ECoG signal with finite impulse response (i.e., theta: 4–9 Hz, alpha: 9–15 Hz, and beta: 15–30 Hz) and extracted instantaneous phases using Hilbert transform. Then we focused on the 100-ms time period right before and the 100-ms time period right after the bump that signifies completion of retrieval, and compared the phase locking values averaged within each period to examine changes upon transitioning of cognitive processing stages. This procedure was done for each subject separately, and repeated for every combination of electrode pairs, with one electrode from the MTL and one from the cortical region.

## Results

### Behavioral analyses (Fig. 3)

**Probe type.** Participants are faster and more accurate when the presented probe is a target than a foil. The effect is significant in Experiment 1, as revealed by a repeated measures ANOVA (accuracy:  $F(1,28) = 17.5$ ,  $p < .001$ ; RT:  $F(1,28) = 28.6$ ,  $p < .001$ ) but not in Experiment 2 (accuracy:  $F(1,11) = 0.04$ ,  $p = .84$ ; RT:  $F(1,11) = .08$ ,  $p = .78$ ).

**Set similarity.** There is an effect of set similarity on both RT and accuracy in Experiment 1, as revealed by a repeated measures ANOVA (accuracy:  $F(1,28) = 34.8$ ,  $p < .001$ ; RT:  $F(1,28) = 4.83$ ,  $p < .05$ ). In Experiment 2, set size has a significant effect on accuracy but not on RT (accuracy:  $F(1,11) = 43.7$ ,  $p < .001$ ; RT:  $F(1,11) = .078$ ,  $p = .77$ ).

**Set size.** There is a significant effect of set size on both RT and accuracy in Experiment 1, as revealed by a repeated measures ANOVA (accuracy:  $F(2,56) = 28.0$ ,  $p < .001$ ; RT:  $F(2,56) = 27.8$ ,  $p < .001$ ). In Experiment 2, set size had a significant effect on accuracy but not on RT (accuracy:  $F(2,22) = 25.4$ ,  $p < .001$ ; RT:  $F(2,22) = .08$ ,  $p = .92$ ).

### Consistency in CCA dimensions across two experiments

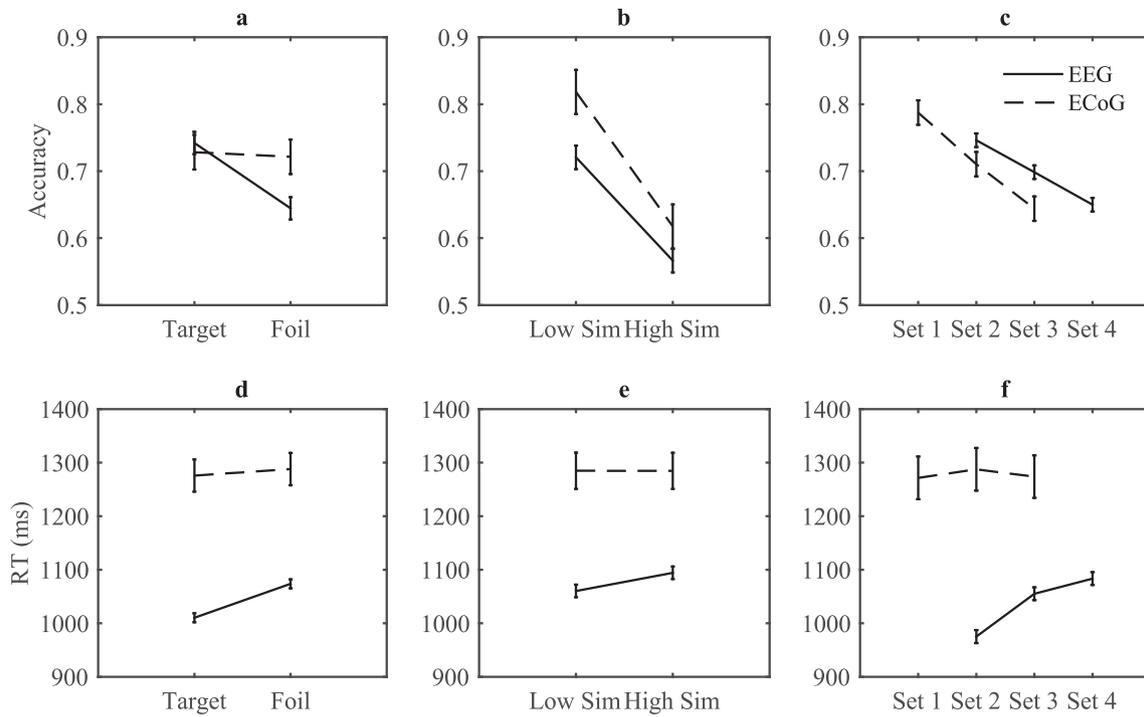
Before pooling data across different ECoG subjects, we first identified dimensions that correspond across subjects by applying the M-CCA method. M-CCA serves to transform the data for each subject from the number of recording sites into a reduced number (5 in our case) of common dimensions (i.e., CCA components) shared across subjects. Two M-CCAs were applied to the two experiments independently. We then compare the temporal dynamics of the top 5 CCAs from the two experiments. Being recorded from the same task, the two experiments are expected to share similar CCAs.

Although the recording sites from individual subjects in ECoG data were quite varied, the M-CCA obtained common dimensions of variation. Moreover, Fig. 4 shows that the top 5 CCA components are comparable to the top 5 CCAs obtained from the EEG dataset – only the order is different. We matched each CCA component in ECoG with the one in EEG that had the highest absolute correlation (value noted in the legend; matching indicated with colors). There appears to be a one-to-one mapping from the 5 CCAs in ECoG to the 5 CCAs in EEG, with their corresponding correlations of time courses ranging from 0.4 to 0.87.

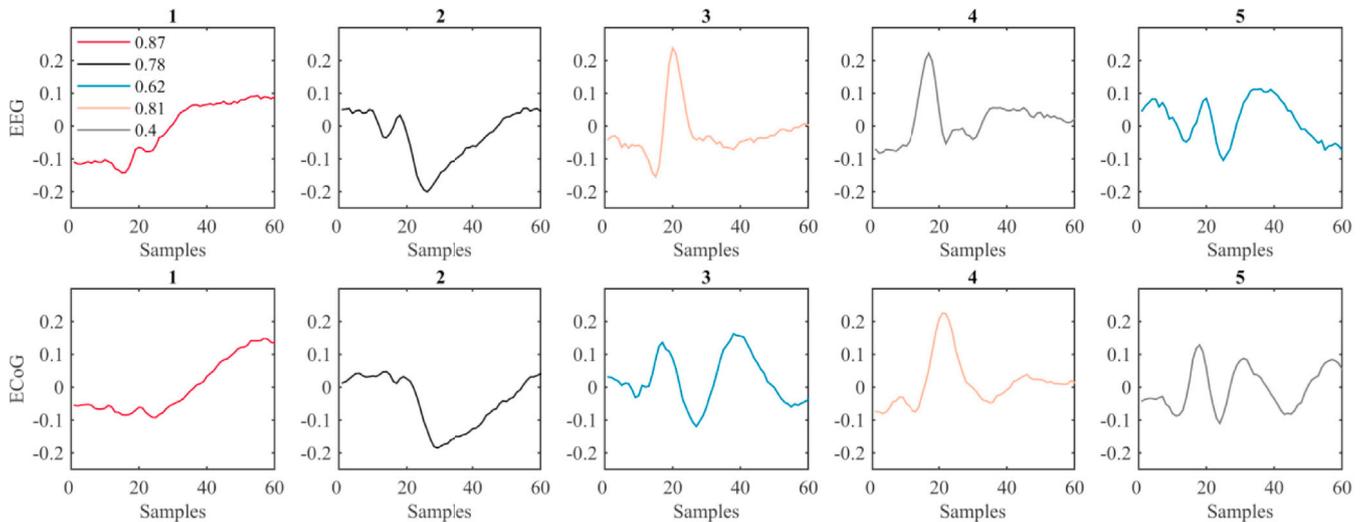
### Identification of the stage durations and the bump profiles in HSMM-MVPA

First, the number of stages in the HSMM-MVPA was determined. Two HSMM-MVPAs were applied to the top 5 CCAs of the two experiments independently. HSMM-MVPA identifies bumps in the ongoing EEG signal related to significant changes in information processing. In this study, the number of stages in HSMM was decided on the basis of between-experiment predictions. Model parameters of the stage distributions were obtained from the 29 subjects in Experiment 1, and used to calculate the likelihood for each of the 12 subjects in Experiment 2 while re-estimating the bump magnitudes; and vice versa. We prefer a more parsimonious model with fewer bumps: we only select a model with  $m+1$  bumps over that with  $m$  bumps if there is improvement in model likelihood over a significant proportion of the total number of subjects. A 3-bump model is significantly ( $p < .0001$ ; two-tailed sign test) better than a 2-bump model in 34 out of 41 subjects (specifically, 24 of 29 in Exp 1 and 10 of 12 in Exp 2). A 4-bump model is significantly ( $p = .01$ ; two-tailed sign test) better than a 3-bump model in 29 out of 41 subjects (specifically, 19 of 29 in Exp 1 and 10 of 12 in Exp 2), but a 5-bump model is only better ( $p = 1.0$ ; two-tailed sign test) than a 4-bump model in 21 (17 in Exp 1 and 4 in Exp 2) out of 41 subjects. Therefore, the 4-bump model is the preferred solution.

Second, we compared consistency in HSMM-MVPA results across two experiments. Two 4-bump HSMMs were applied to the top 5 CCAs of the



**Fig. 3.** Accuracy (a–c) and RT (d–f) of EEG and ECoG data as a function of probe type, set similarity and set size. SEMs are shown in the error bars with between-subjects variance removed (Loftus and Masson, 1994).



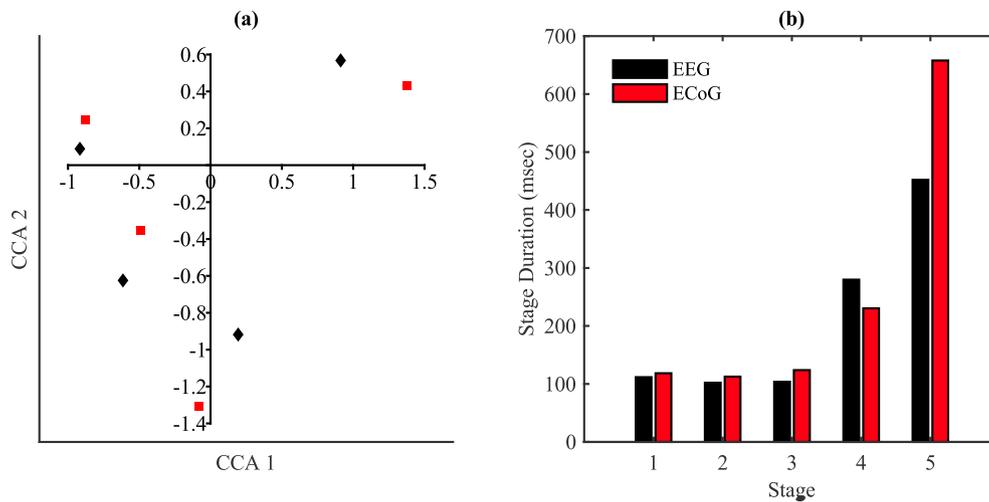
**Fig. 4.** First 5 CCA components obtained over EEG data in Experiment 1 (top row) and over ECoG data in Experiment 2 (bottom row). 60 samples correspond to 600 ms of stimulus-locked data in target condition. Corresponding CCAs are highlighted in the same color, with their absolute correlations noted in the legend. CCAs in EEG with large negative correlations with CCAs in ECoG have been flipped, so that all signs are positive.

two experiments independently. We then compared bump magnitudes and stage durations of the resulting two models. Despite that two HSMMs were applied to EEG and ECoG separately, there is considerable consistency in bump magnitudes with the correlation of the 20 values (5 CCAs x 4 bumps/CCA) being 0.62. Fig. 5a further demonstrates this consistency by comparing representations of the 4 bumps in the top two CCA dimensions across two experiments, which are highly similar.

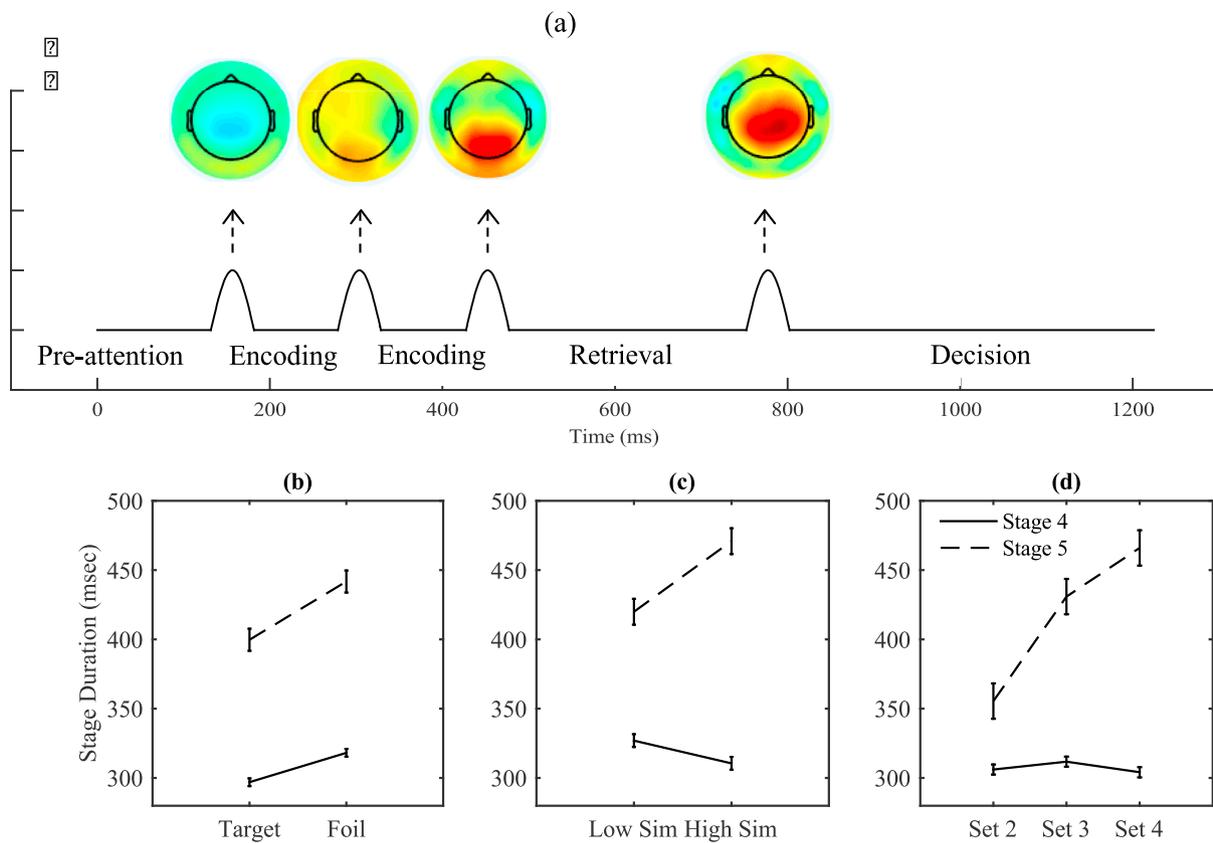
Next, we interpreted the recovered stages and mapped them to corresponding cognitive processes. Fig. 5b shows the stage durations of the two HSMMs. Consistent across the two experiments are three briefer periods at the beginning of the trial followed by two longer periods towards the end. ECoG subjects, who have overall longer RTs, are markedly

slower in the last stage. Fig. 6a shows the reconstructed scalp profiles of the four bumps in Experiment 1 (EEG), which were created by averaging the observed voltages at the time of the maximum-likelihood samples for each bump and during each trial. The scalp profiles are plotted against the five stage durations.

Guided by the process model of Anderson et al. (2016), which decomposed the recognition memory task into a encoding process, a retrieval process, a decision process, and a motor response process, we can interpret the HSMM-MVPA stages as follows: In both latency and topographical distribution, the first bump resembles the N1, which is typically interpreted as an index of visual attention (Luck et al., 2000). Therefore, we interpret the first stage is a pre-attention stage before



**Fig. 5.** Representation of the 4 bumps in the top two CCA dimensions in the obtained HSMs for both EEG and ECoG data (a). Durations of the five processing stages identified using the HSM-MVPA method for both EEG and ECoG data (b).



**Fig. 6.** (a). HSM stages labeled with corresponding cognitive processes, interleaved by four reconstructed bumps for EEG data. Mean EEG electrode activity of the reconstructed bumps were obtained by averaging the observed voltages at the time of the maximum-likelihood samples for each bump and during each trial. Electrode voltages have been normalized for each trial. (b–d). Duration of Stage 4 (Retrieval stage) and Stage 5 (Decision stage) in EEG as a function of probe type, set similarity and set size. SEMs are shown in the error bars with between-subjects variance removed (Loftus and Masson, 1994).

actual encoding takes place. It is then followed by two brief encoding stages associated with the second and the third bumps that show a posterior positivity. We believe these two bumps mark the encoding of the probe face similar to a pair of encoding bumps identified in a word recognition task (Anderson et al., 2016). Given its stage duration (i.e. ~200 ms) and the process model, Stage 4 is identified as the Retrieval stage. This is consistent with the frontal-central distribution of the fourth bump. Of the 5 stages, the last stage is not likely to be a motor

response stage given its long duration of around 600 ms. It is more likely that this stage represents the combination of the decision and motor response stage and that we failed to detect a bump separating the decision and motor stages. The interpretations of the five stages as labeled in Fig. 6a will be further verified in the following sections by examining the effect of different experimental factors (probe type/set similarity/set size) on the durations and brain activity associated with each stage.

### Stage durations by condition with EEG

RTs varied by condition for EEG subjects in Experiment 1. These differences must show up in the durations of some of the stages. To determine which stages were affected by the experimental manipulations and examine if these effects are consistent with our interpretation of the stages, we fit HSMMs with different stage durations to each condition. That is, we estimated parameters for the gamma distributions of each stage separately for the different conditions while constraining the bump magnitudes to be the same. The HSMM methods return the probabilities of each bump occurring at each time point on a trial-by-trial basis. These probabilities can be used to calculate the most likely location of each bump in a trial. Mean stage durations for a particular subject can then be calculated as the average time between bumps across all trials within that subject. Fig. 6b–d shows the resulting mean time durations across all EEG subjects for Stage 4 and Stage 5. We submitted them to a repeated-measures ANOVA for each condition (probe type/set similarity/set size) and for each stage. The stage durations do not differ between conditions for the first three stages. Consistent to the impression conveyed in the figure, there is an effect of probe type in Stage 4 - the Retrieval stage ( $F(1,28) = 29.210$ ;  $p < .001$ ) and in Stage 5 - the Decision stage ( $F(1,28) = 14.013$ ;  $p = 0.001$ ). There is also effect of set similarity in the Retrieval stage ( $F(1,28) = 6.227$ ;  $p = .019$ ) and the Decision stage ( $F(1,28) = 14.943$ ;  $p = 0.001$ ). There is effect of set size in the Decision stage ( $F(2,56) = 29.923$ ;  $p < 0.001$ ) but not in the Retrieval stage ( $F(2,56) = 1.260$ ;  $p = .29$ ). To summarize, the Retrieval stage is shorter for targets than for foils, consistent with the idea that it is easier to retrieve the memory set given a member of that set. In addition, the Retrieval stage is faster when the similarity between foil and memorized items is high. The duration of the Decision stage is affected by all three experimental factors: probe type, set similarity and set size.

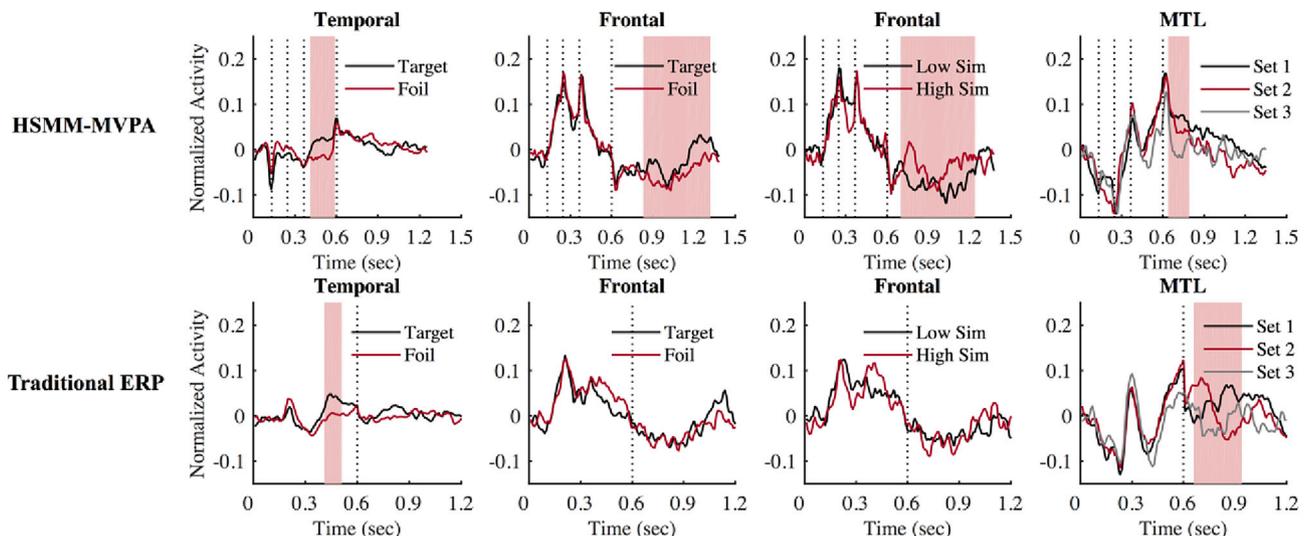
### Stage-locked brain activity by condition with ECoG

We exploit the spatial resolution of the ECoG data to examine where the differences in brain activity across different conditions occur. In event-related potential (ERP) analyses, the neural signal is anchored to observable events such as the presentation of a stimulus. The bumps obtained from the HSMM signal latent points of change in information processing. These events can be used, along with observable events, to align the neural data. We anchored ECoG data from each trial according

to stimulus onset, response, and the maximum likelihood locations of each of the four bumps during that trial. We then expanded or contracted the resulting five intervals in every trial to have durations equal to mean durations specified by HSMM-MVPA. In this way, the stimulus, the locations of the four bumps, and the response are aligned across all trials. Over such stage-locked data, we can examine where and when different conditions lead to different patterns of brain activity. Fig. 7 (top row) highlights time periods and brain regions where there are significant differences of stage-locked brain activity between any two conditions using paired t-tests over a sliding window of 100 ms (dashed line represents average bump positions). Multiple comparisons across regions and time windows were corrected with the Bonferroni correction. The bottom row in Fig. 7 is a comparison using traditional ERP analysis with 600-ms stimulus-locked and 600-ms response-locked data (separated by the dashed line).

In the stage-locked brain activity, we observe more positivity for targets than foils in the temporal cortex before the Retrieval stage completes. The temporal cortex has previously been associated with face familiarity (Gainotti, 2007) and is known to be instrumental in determining stimulus familiarity (e.g., Borst et al., 2016; Diana et al., 2007; Henson et al., 1999; Gonsalves et al., 2005; Rugg and Yonelinas, 2003). During the Decision stage, when the items are in active maintenance, targets are associated with more positive deflections than foils in the frontal cortex, which is known to be involved in post-retrieval monitoring and maintenance (e.g., Achim and Lepage, 2005; Borst et al., 2016; Mitchell et al., 2004; Rugg and Yonelinas, 2003). Similar patterns in the frontal cortex also show up in different levels of set similarity among foils, where a higher set similarity corresponds to more positive ECoG amplitudes. During the same period in the Decision stage, there is an ordered effect of set size in MTL, with set size 1 being the most positive. This observation extends what we know about the MTL in maintaining items in working memory. Increasing working memory load (set size) was associated with elevated negativity of evoked response potentials of hippocampus during the delay period (Axmacher et al., 2007). In our study, this pattern also extends to the period after memory retrieval (Bump 4), when the retrieved face(s) need to be actively maintained during the Decision stage.

Comparison with the traditional ERP analysis demonstrates the power of the HSMM-MVPA method in modeling cognitive events on a trial-by-trial basis. In the ERP analysis, we observe more positivity for targets than foils in the temporal cortex before the Retrieval stage



**Fig. 7.** Time periods (over sliding window of 100 ms) over stage-locked average activity (top row) and stimulus/response-locked average activity (bottom row) where there are significant differences for different probe types, set similarity and set sizes in ECoG data, adjusted for multiple comparisons with Bonferroni correction. The dotted lines mark average positions of four bumps in the stage-locked activity (top row), and separation between stimulus-locked activity and response-locked activity (bottom row).

completes, similar to the stage-locked activity with the HSMM-MVPA method. Such consistency is attributed to low trial-to-trial variability when data are closely time-locked to the stimulus. However, when moving further away from the stimulus, in contrast to the stage-locked brain activity, there are no significant effects across conditions in the frontal cortex. There is a significant effect of set size across conditions in the MTL in the Decision stage, but this difference is not ordered by the set size.

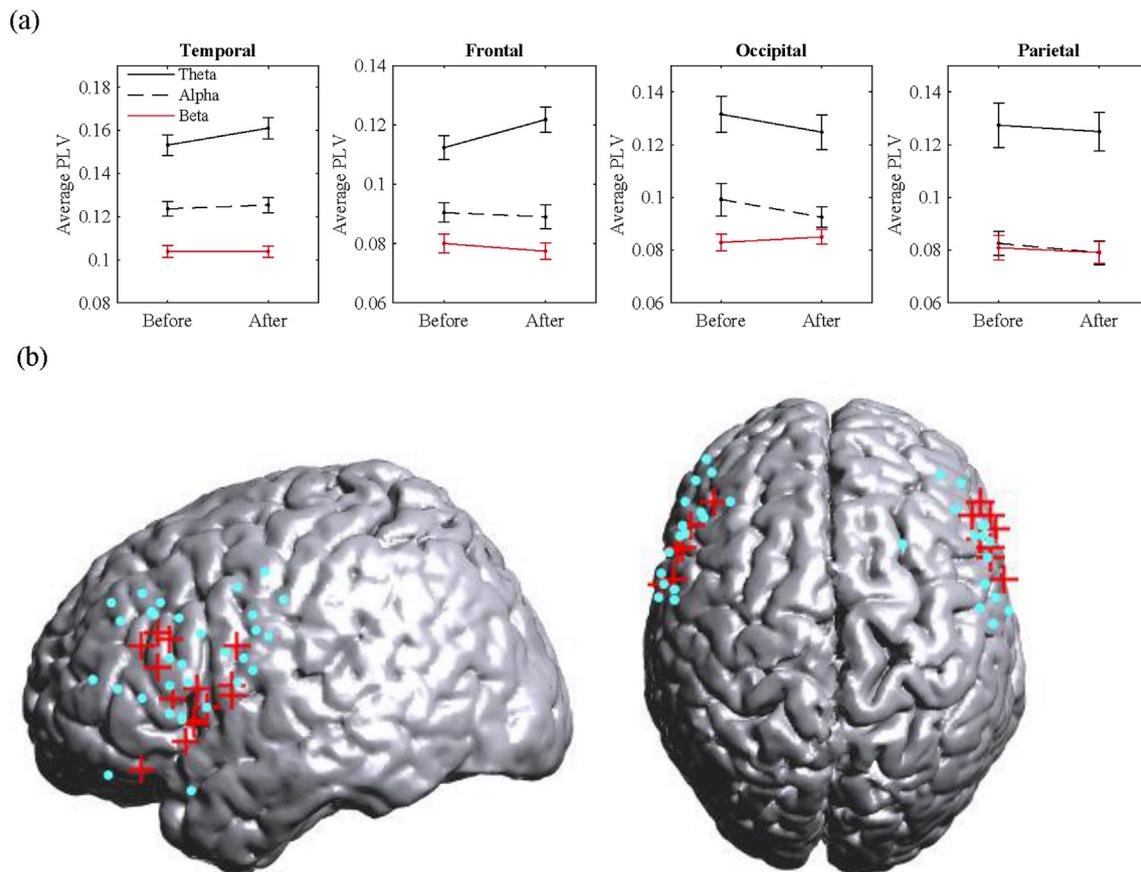
#### Phase synchrony between medial temporal lobe and cortical areas

ECoG data provides the unique opportunity to study the properties of oscillations in sub-cortical regions such as the MTL. In this section, we examine the synchrony between MTL and different cortical regions during the retrieval process.

Brain synchrony measures the relation between the temporal structures of the brain signals, and is considered as an important mechanism for integrating activity from across distributed brain areas into coherent perception and behavior (Varela et al., 2001; Fries, 2009; Fell and Axmacher, 2011). Simultaneous recordings in the hippocampus and prefrontal regions in animal studies have revealed synchronized theta oscillations during working memory tasks (Siapas et al., 2005). To examine whether those also occur in humans around the memory retrieval stage, we focus on phase synchrony between MTL and different cortical areas in the current experiment. Specifically, we examine if there is any transient change in phase synchrony upon completion of memory retrieval (i.e. when transitioning from the Retrieval stage to the Decision stage; cf. Portoles et al., 2018).

We measured synchrony with phase locking values (Lachaux et al., 1999). Fig. 8a plots the phase locking values between MTL and 4 cortical regions in different frequency bands. We compared synchronization during two distinct periods (each of a 100 ms duration) across all trials: 1) samples right before the completion of the Retrieval stage; 2) samples right after completion of the Retrieval stage (i.e. the onset of the Decision stage). Synchrony for a particular cortical electrode was averaged across phase locking values calculated with each of the MTL electrodes of the same subject. Electrodes in a cortical region from different subjects were then pooled together to obtain the standard error of the means shown in the figure. We were able to include 7 out of 12 subjects with more than 2 MTL electrodes in this analysis. Comparing the period right before and right after when the retrieval completes, there is increased theta phase synchrony between MTL and frontal recording sites (34 out of 47 electrodes;  $p = .003$ , binomial test). This increase is specific to theta band, and is not significant to alpha band (19 out of 47 electrodes;  $p = .24 > .05$ , binomial test) or beta band (22 out of 47 electrodes;  $p = .77 > .05$ , binomial test). We also observed increased phase synchrony between MTL and recording sites in the temporal cortex in the theta band (125 out of 190 electrodes;  $p < .0001$ , binomial test), but not in alpha band (105 out of 190 electrodes;  $p = .17 > .05$ , binomial test) or beta band (101 out of 190 electrodes;  $p = .42 > .05$ , binomial test). There is no significant theta coupling between MTL with either the occipital cortex (22 out of 55 electrodes;  $p = .18 > .05$ , binomial test) or the parietal cortex (10 out of 18 electrodes;  $p = .81 > .05$ , binomial test).

To further test if the theta coupling with MTL is significant on the level of individual electrodes, we built surrogate data by randomly shuffling the two time periods examined (i.e. before and after retrieval



**Fig. 8.** (a). Phase synchrony in theta, alpha and beta bands between MTL and cortical regions in 100-ms periods before and after retrieval completes. Synchrony for a particular cortical electrode is averaged across phase locking values calculated with each of the MTL electrodes of the same subject. Cortical electrodes across 7 subjects are then pooled together to obtain the standard error of the means. (b). Frontal electrodes with significant increase in phase-locking values with MTL in the theta band during the 100-ms after retrieval compared with before (red marker). The remaining non-significant frontal electrodes are in blue.

completes), and calculated the resulting phase locking values. This procedure is repeated 1000 iterations for between each pair of electrodes examined, and its corresponding increase in theta coupling with MTL is only considered significant if the amount of increase is larger than 95% of the times in the surrogate data. In total, there are 40 out of 190 electrodes in the temporal cortex that demonstrate significant increase in theta coupling with MTL, and 16 out of 47 electrodes in the frontal cortex. Fig 8b shows that, despite that the 16 frontal electrodes are from 5 out of the 7 subjects examined, there is considerable across-subject consistency in their locations (after being mapped to a common brain).

## General discussion

In this study, we provided a detailed mapping of the time course and neural correlates of the retrieval process underlying visual working memory. Detailed mapping of the time course was achieved by capturing the trial-to-trial variability of different cognitive processes using the HSMM-MVPA method, instead of examining averaged neural activity across trials locked to observable events in traditional ERP analysis. HSMM-MVPA decomposed each trial into a sequence of latent stages. By examining how the duration of the HSMM-MVPA stages differed between task conditions, combined with our knowledge of a process model, we found evidence for the existence of the following cognitive stages in the visual working memory task: After a brief period for the visual signal to reach the brain (i.e. pre-attention), participants first encode the probe face (i.e. encoding), then retrieve faces from the memorized set (i.e. retrieval), and lastly, compare the retrieved set of faces with the probe to make a decision (i.e. decision). Once a detailed temporal mapping of the task was achieved with HSMM-MVPA, fine spatial resolution in ECoG was used to examine the neural correlates associated with each identified cognitive stage. Main effects centered around the transition from the Retrieval stage to the Decision stage (after completion of retrieval). These effects are summarized in Table 2 and will be discussed in the following sections.

### Isolation of a retrieval process prior to the decision-making

It has been widely assumed that retention of information in the working memory relies on maintenance of an active memory trace (Fuster and Alexander, 1971). However, the maintained information can be fragile if attention is temporally directed away or if the information is not amenable to rehearsal (Jeneson and Squire, 2011). In that case, memory can be preserved across a short delay without active maintenance (Lewis-Peacock et al., 2012), and retrieved later.

In this study, we were able to identify such a retrieval process, with EEG and ECoG data yielding converging results on a Retrieval stage isolated from the Decision stage. In the Retrieval stage, we observed more positivity for targets than foils in the temporal cortex, reflecting a fast and automatic process that does not require item details (Clark and Gronlund, 1996; Raaijmakers and Shiffrin, 1992). Once retrieval is complete, in the decision stage, we observed evidence of active maintenance of the just-retrieved items, with an effect of set size in MTL similar to that in Axmacher et al. (2007), and an effect of probe type in the frontal cortex in supporting post-retrieval monitoring (e.g., Achim and Lepage, 2005; Borst et al., 2016; Mitchell et al., 2004; Rugg and Yonelinas, 2003). This provides support for cue-based retrieval theories of working memory.

### Duration of the retrieval stage and the decision stage

How the duration of a particular cognitive stage changes across different experimental conditions provides important information on the nature of the underlying process. Typically, the effect of an experiment factor on a particular stage is reflected in the overall reaction times (RTs). However, given that RT only provides a cumulative measure of all the cognitive stages involved in a particular trial, there is not enough

**Table 2**

A summary of identified effects on the duration and the brain activity across different processing stages.

	Retrieval Stage	Decision Stage
Duration	i Probe type ii Set similarity ↓	i Probe type ii Set size ↑ iii Set similarity ↑
Temporal Cortex	i Probe type	i Synchrony with MTL i Set size ↓ ii Synchrony with frontal cortex iii Synchrony with temporal cortex
MTL		
Frontal Cortex		i Probe type ii Set similarity ↑ iii Synchrony with MTL

↑ indicates more positivity or longer duration; ↓ indicates the opposite.

information in RT alone to isolate the effect of a particular experimental factor when there is more than one cognitive stage affected. In this study, we applied the HSMM-MVPA method to obtain durations of individual cognitive stages, and identified effects in stage durations when both the Retrieval stage and the Decision stage vary across conditions.

In particular, we observed from the EEG data that the duration of the Retrieval stage depends on both probe type and set similarity. When the probe is one of the previously memorized faces (i.e., targets), retrieval is faster than when the probe is a foil because it is easier to reactivate the memory set given a member of that set. In addition, retrieval speed also depends on the set similarity. According to the ACT-R theory, greater similarity between foil and memorized item will result in greater activation for the retrieved item, and consequently shorter retrieval time (Anderson et al., 1998). The Retrieval stage does not depend on set size, which is comparable to the Sternberg task analyzed in Anderson et al. (2016).

The duration of the Decision stage is affected by all three factors: probe type, set similarity and set size. There are two possible mechanisms underlying the Decision stage: One possibility is that there is a serial comparison procedure where the probe is compared with each of the items in the retrieved set to find a match. The other possibility is that there is an evidence-accumulation procedure where a decision is driven by the strength of the overall similarity between the probe and the retrieved set of items—basically the summed similarity discussed in Van Vugt et al. (2013), van Vugt et al. (2009), Kahana and Sekuler (2002) and Exemplar-Based Random Walk models in Nosofsky et al. (2011).

A serial self-terminating decision process would be consistent with the data: Targets will yield a faster Decision stage due to earlier termination once a match is found (Fig. 6b). In foils, higher set similarity slows down each comparison thus giving rise to a longer total Decision time (Fig. 6c). Larger set size corresponds to a larger number of comparisons, and therefore a longer Decision stage (Fig. 6d). A 2 (probe type) x 3 (set size) repeated-measures ANOVA shows significant effect of interaction between probe type and set size ( $F(2,56) = 4.68, p = 0.01$ ). This is also consistent with the assumption that the serial comparison is self-terminating because the difference between targets and foils increases with set size.

If the Decision stage involves evidence accumulation, that would also be consistent with the data: Low similarity will lead to faster responses in correct foils (Nosofsky et al., 2011; Ratcliff, 1978; Bogacz et al., 2006). The effect of set size on the duration of the Decision stage can also be explained by set similarity, as a smaller set size corresponds to higher set similarity which leads to faster response in correct targets. In addition, smaller set size also increases average memory strength with briefer time lags, which gives rise to faster decision time under the evidence accumulation account (Nosofsky et al., 2011). Under either mechanism, the observation that increased set similarity speeds up the Retrieval stage while slowing down the Decision stage is consistent with an earlier study

with an associative recognition task (Zhang et al., 2017b).

### Role of frontal cortex and MTL during the decision stage

fMRI studies have identified that the MTL and prefrontal cortex support the general retrieval process of working memory tasks (Bledowski et al., 2006; Öztekin et al., 2009). However, given the low temporal resolution of fMRI, isolation of detailed processing stages such as the retrieval stage and the decision stage is difficult (but see Borst and Anderson, 2013, 2017, for an application of model-based fMRI analysis that attempts to tease apart working memory updating and retrieval). In contrast, the high temporal resolution of applying the HSMM-MVPA method to EEG and ECoG data enables us to divide the working memory task into a sequence of stages. In particular, retrieval of the studied face set marks the transition to a period of active maintenance of items, where information in the retrieved set is compared with the probe before a decision is made (Fig. 6a). The active maintenance is supported by MTL, where we observed more negativity when the set size is large. Negativity in hippocampal activity has been associated with increasing workload (i.e., set size) in visual working memory task during the delay period, when there is active maintenance of items right after they are encoded (Axmacher et al., 2007). With the finer temporal resolution of HSMM-MVPA, the current experiment extended this result to the decision period of a working memory task after the items are just retrieved.

During the Decision stage, we also observed more positive activity in frontal cortex for targets than foils (comparable to the process described in Borst et al., 2016). In working memory tasks, there is debate concerning whether the involvement of prefrontal cortex contributes to the maintenance of items or the selection of an item from memory to guide a response (Curtis & D'Esposito, 2003; Rowe, 2000). Our experiment supports the latter, with an effect for target/foil but not for different set sizes in frontal cortex. This interpretation is consistent with the process during the Decision stage, where an item needs to be selected to match with the probe. When there is a match (i.e., targets), a more positive response is triggered in the frontal cortex compared with that of non-matches or foils. An alternative mechanism for the Decision stage could be that instead of selecting and comparing each item in the retrieved face set against the probe, the frontal cortex guides the decision in an evidence-accumulation procedure driven by the similarity between the probe and the entire retrieved face set (i.e., set similarity; Nosofsky et al., 2011). This latter account is supported by more positivity in frontal cortex for high-similarity trials than for low-similarity trials, in addition to more positivity for targets than for foils.

We also found evidence that the frontal cortex guides the decision by means of theta coupling with the MTL once retrieval has been completed. Previously, it has been suggested that 4–9 Hz theta power in the hippocampus is associated with encoding and retrieval of episodic memories (Lega et al., 2011). Functional coupling between prefrontal cortex and medial temporal cortex is considered one of the key connections in the neural circuitry underlying working memory tasks, with theta oscillations proposed to mediate this interaction (Mitchell et al., 2008; Anderson et al., 2009). This is supported by multiple animal studies that demonstrate phase-locking in the theta band between prefrontal cortex and hippocampus (Siapas et al., 2005; Hyman, 2010; for a review see Colgin, 2011). In human studies, PFC-hippocampal coupling, both structural (Colgin, 2011) and functional (Campo et al., 2011), has been shown to correlate with individual differences in task performance. However, PFC-hippocampal communication through theta oscillations in human working memory tasks has not been directly observed, though it is shown to be important in free recall and associative recognition memory tasks (Anderson et al., 2009). The observation of frontal-MTL theta coupling generalizes the role of MTL theta from retrieval in episodic memory to retrieval of previous items in working memory. This is consistent with recent studies that MTL is not uniquely involved in long-term memory, but also critical to short-term memory even when the retention period is as short as 2–10s (Holdstock et al., 1995; Owen et al.,

1995; Holdstock et al., 2000; Aggleton et al., 1992; Hannula et al., 2006, Van Vugt et al., 2010).

There are several limitations to our research. First, the analysis of cortical-subcortical interaction was limited by the number and locations of recording sites in each ECoG subject. Therefore, only pairs of electrodes were examined at one time and pooled across all subjects in the end. Second, the HSMM-MVPA method successfully isolated different cognitive processing stages, but it did not provide enough evidence to distinguish whether the decision stage undergoes a serial comparison process or an evidence accumulation process. It would take more experimental studies with targeted manipulations to make this distinction in the future.

### Conclusion

In this study, we extended a previous account of the visual working memory task using a summed-similarity model (Van Vugt et al., 2013) to one that comprises multiple sequential cognitive stages. Combining the temporal resolution of the EEG data, the spatial resolution of the ECoG data, and the application of the HSMM-MVPA method, we were able to identify the time duration and brain activity associated with these stages. In contrast to traditional ERP analyses which only models the effects that are closely locked to the beginning and the end of a trial, the HSMM-MVPA method applied to both experiments isolated a Retrieval stage where memorized items were re-activated, followed by a Decision stage. In contrast to examining only the overall RT, the HSMM-MVPA method applied to EEG data revealed how durations of the Retrieval stage and the Decision stage vary across different experimental conditions. Combined with fine spatial resolution of ECoG data, it was identified that frontal cortex and MTL play a key role in response selection and item maintenance respectively. The effect of set size observed in MTL generalizes its role in actively maintaining items from the delay period in working memory tasks to the decision period once items are re-activated. The theta coupling between frontal cortex and the MTL generalizes previous findings that were considered unique to long-term memory tasks to working memory tasks. In addition, they provide support for a cue-based retrieval account of visual working memory.

### Acknowledgements

This work was supported by the National Science Foundation (grant number: 1420009) to J.R.A.; the James S. McDonnell Foundation (grant number: 220020162) to J.R.A.; the Office of Naval Research (grant number: N00014-15-1-2151) to J.R.A.; and the Netherlands Organisation for Scientific Research Veni (grant number: 451-15-040) to J.P.B.. The authors declare that there is no conflict of interest.

### References

- Achim, A.M., Lepage, M., 2005. Dorsolateral prefrontal cortex involvement in memory post-retrieval monitoring revealed in both item and associative recognition tests. *NeuroImage* 24 (4), 1113–1121. <https://doi.org/10.1016/j.neuroimage.2004.10.036>.
- Aggleton, J., Shaw, C., Gaffan, E., 1992. The performance of postencephalitic amnesic subjects on two behavioural tests of memory: concurrent discrimination learning and delayed matching-to-sample. *Cortex* 28 (3), 359–372. [https://doi.org/10.1016/S0010-9452\(13\)80146-3](https://doi.org/10.1016/S0010-9452(13)80146-3).
- Anderson, J.R., Bothell, D., Lebiere, C., Matessa, M., 1998. An integrated theory of list memory. *J. Mem. Lang.* 38 (4), 341–380. <https://doi.org/10.1006/jmla.1997.2553>.
- Anderson, J.R., Zhang, Q., Borst, J.P., Walsh, M.M., 2016. The discovery of processing stages: extension of sternberg's method. *Psychol. Rev.* 123 (5), 481–509. <https://doi.org/10.1037/rev0000030>.
- Anderson, K.L., Rajagovindan, R., Ghacibeh, G.A., Meador, K.J., Ding, M., 2009. Theta oscillations mediate interaction between prefrontal cortex and medial temporal lobe in human memory. *Cereb. Cortex* 20 (7), 1604–1612. <https://doi.org/10.1093/cercor/bhp223>.
- Axmacher, N., Mormann, F., Fernandez, G., Cohen, M.X., Elger, C.E., Fell, J., 2007. Sustained neural activity patterns during working memory in the human medial temporal lobe. *J. Neurosci.* 27 (29), 7807–7816. <https://doi.org/10.1523/jneurosci.0962-07.2007>.

- Basar, E., 1980. EEG Brain Dynamics: Relation between EEG and Brain Evoked Potentials. Elsevier, Amsterdam, The Netherlands.
- Bledowski, C., Kadosh, K.C., Wibral, M., Rahm, B., Bittner, R.A., Hoehstetter, K., Linden, D.E., 2006. Mental chronometry of working memory retrieval: a combined functional magnetic resonance imaging and event-related potentials approach. *J. Neurosci.* 26 (3), 821–829.
- Bogacz, R., Brown, E., Moehlis, J., Holmes, P., Cohen, J.D., 2006. The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychol. Rev.* 113 (4), 700–765. <https://doi.org/10.1037/0033-295x.113.4.700>.
- Borst, J.P., Anderson, J.R., 2013. Using model-based functional MRI to locate working memory updates and declarative memory retrievals in the fronto-parietal network. *Proc. Natl. Acad. Sci. U. S. A.* 110 (5), 1628–1633. <https://doi.org/10.1073/pnas.1221572110>.
- Borst, J.P., Anderson, J.R., 2015. The discovery of processing stages: analyzing EEG data with hidden semi-Markov models. *NeuroImage* 108, 60–73. <https://doi.org/10.1016/j.neuroimage.2014.12.029>.
- Borst, J.P., Anderson, J.R., 2017. A step-by-step tutorial on using the cognitive architecture ACT-R in combination with fMRI data. *J. Math. Psychol.* 76, 94–103.
- Borst, J.P., Ghuman, A.S., Anderson, J.R., 2016. Tracking cognitive processing stages with MEG: a spatio-temporal model of associative recognition in the brain. *NeuroImage* 141, 416–430. <https://doi.org/10.1016/j.neuroimage.2016.08.002>.
- Campo, P., Garrido, M.L., Moran, R.J., Maestu, F., Garcia-Morales, I., Gil-Nagel, A., Friston, K.J., 2011. Remote effects of hippocampal sclerosis on effective connectivity during working memory encoding: a case of connective diachisis? *Cereb. Cortex* 22 (6), 1225–1236. <https://doi.org/10.1093/cercor/bhr201>.
- Clark, S.E., Gronlund, S.D., 1996. Global matching models of recognition memory: how the models match the data. *Psychon. Bull. Rev.* 3 (1), 37–60. <https://doi.org/10.3758/bf03210740>.
- Colgin, L.L., 2011. Oscillations and hippocampal–prefrontal synchrony. *Curr. Opin. Neurobiol.* 21 (3), 467–474. <https://doi.org/10.1016/j.conb.2011.04.006>.
- Curtis, C.E., D'Esposito, M., 2003. Persistent activity in the prefrontal cortex during working memory. *Trends Cogn. Sci.* 7 (9), 415–423. [https://doi.org/10.1016/s1364-6613\(03\)00197-9](https://doi.org/10.1016/s1364-6613(03)00197-9).
- Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134 (1), 9–21. <https://doi.org/10.1016/j.jneumeth.2003.10.009>.
- Diana, R.A., Yonelinas, A.P., Ranganath, C., 2007. Imaging recollection and familiarity in the medial temporal lobe: a three-component model. *Trends Cogn. Sci.* 11 (9), 379–386. <https://doi.org/10.1016/j.tics.2007.08.001>.
- Ericsson, K.A., Kintsch, W., 1995. Long-term working memory. *Psychol. Rev.* 102 (2), 211.
- Fell, J., Axmacher, N., 2011. The role of phase synchronization in memory processes. *Nat. Rev. Neurosci.* 12 (2), 105–118. <https://doi.org/10.1038/nrn2979>.
- Fries, P., 2009. Neuronal gamma-band synchronization as a fundamental process in cortical computation. *Annu. Rev. Neurosci.* 32 (1), 209–224. <https://doi.org/10.1146/annurev.neuro.051508.135603>.
- Fuster, J.M., Alexander, G.E., 1971. Neuron activity related to short-term memory. *Science* 173 (3997), 652–654.
- Gainotti, G., 2007. Face familiarity feelings, the right temporal lobe and the possible underlying neural mechanisms. *Brain Res. Rev.* 56 (1), 214–235. <https://doi.org/10.1016/j.brainresrev.2007.07.009>.
- Gonsalves, B.D., Kahn, I., Curran, T., Norman, K.A., Wagner, A.D., 2005. Memory strength and repetition suppression: multimodal imaging of medial temporal cortical contributions to recognition. *Neuron* 47 (5), 751–761. <https://doi.org/10.1016/j.neuron.2005.07.013>.
- Hannula, D.E., Tranel, D., Cohen, N.J., 2006. The long and the short of it: Relational memory impairments in amnesia, even at short lags. *J. Neurosci.* 26 (32), 8352–8359. <https://doi.org/10.1523/jneurosci.5222-05.2006>.
- Henson, R.N.A., Shallice, T., Dolan, R.J., 1999. Right prefrontal cortex and episodic memory retrieval: a functional MRI test of the monitoring hypothesis. *Brain* 122 (7), 1367–1381. <https://doi.org/10.1093/brain/122.7.1367>.
- Holdstock, J., Gutnikov, S., Gaffan, D., Mayes, A., 2000. Perceptual and mnemonic matching-to-sample in humans: contributions of the hippocampus, perirhinal and other medial temporal lobe cortices. *Cortex* 36 (3), 301–322. [https://doi.org/10.1016/s0010-9452\(08\)70843-8](https://doi.org/10.1016/s0010-9452(08)70843-8).
- Holdstock, J., Shaw, C., Aggleton, J., 1995. The performance of amnesic subjects on tests of delayed matching-to-sample and delayed matching-to-position. *Neuropsychologia* 33 (12), 1583–1596. [https://doi.org/10.1016/0028-3932\(95\)00145-x](https://doi.org/10.1016/0028-3932(95)00145-x).
- Hyman, 2010. Working memory performance correlates with prefrontal-hippocampal theta interactions but not with prefrontal neuron firing rates. *Front. Integr. Neurosci.* <https://doi.org/10.3389/fneuro.07.002.2010>.
- Jenison, A., Squire, L.R., 2011. Working memory, long-term memory, and medial temporal lobe function. *Learn. Mem.* 19 (1), 15–25. <https://doi.org/10.1101/lm.024018.111>.
- Kahana, M.J., Bennett, P.J., 1994. Classification and perceived similarity of compound gratings that differ in relative spatial phase. *Atten. Percept. Psychophys.* 55 (6), 642–656.
- Kahana, M.J., Sekuler, R., 2002. Recognizing spatial patterns: a noisy exemplar approach. *Vis. Res.* 42 (18), 2177–2192. [https://doi.org/10.1016/s0042-6989\(02\)00118-9](https://doi.org/10.1016/s0042-6989(02)00118-9).
- Lachaux, J.-P., Rodriguez, E., Martinerie, J., Varela, F.J., 1999. Measuring phase synchrony in brain signals. *Hum. Brain Mapp.* 8 (4), 194–208. [https://doi.org/10.1002/\(sici\)1097-0193\(1999\)8:4<194::aid-hbm4>3.0.co;2-c](https://doi.org/10.1002/(sici)1097-0193(1999)8:4<194::aid-hbm4>3.0.co;2-c).
- Le Van Quyen, M., Foucher, J., Lachaux, J.-P., Rodriguez, E., Lutz, A., Martinerie, J., Varela, F.J., 2001. Comparison of hilbert transform and wavelet methods for the analysis of neuronal synchrony. *J. Neurosci. Methods* 111 (2), 83–98. [https://doi.org/10.1016/s0165-0270\(01\)00372-7](https://doi.org/10.1016/s0165-0270(01)00372-7).
- Lega, B.C., Jacobs, J., Kahana, M., 2011. Human hippocampal theta oscillations and the formation of episodic memories. *Hippocampus* 22 (4), 748–761. <https://doi.org/10.1002/hipo.20937>.
- Lehmann, D., 1987. Principles of spatial analysis. Methods of analysis of brain electrical and magnetic signals. In: EEG Handbook, pp. 309–354.
- Lewis-Peacock, J.A., Drysdale, A.T., Oberauer, K., Postle, B.R., 2012. Neural evidence for a distinction between short-term memory and the focus of attention. *J. Cogn. Neurosci.* 24 (1), 61–79. [https://doi.org/10.1162/jocn\\_a.00140](https://doi.org/10.1162/jocn_a.00140).
- Loftus, G.R., Masson, M.E., 1994. Using confidence intervals in within-subject designs. *Psychon. Bull. Rev.* 1 (4), 476–490.
- Luck, S.J., 2014. An Introduction to the Event-related Potential Technique. MIT press, 2014.
- Luck, S.J., Woodman, G.F., Vogel, E.K., 2000. Event-related potential studies of attention. *Trends Cogn. Sci.* 4 (11), 432–440. [https://doi.org/10.1016/s1364-6613\(00\)01545-x](https://doi.org/10.1016/s1364-6613(00)01545-x).
- Makeig, S., 2002. Dynamic brain sources of visual evoked responses. *Science* 295 (5555), 690–694. <https://doi.org/10.1126/science.1066168>.
- Mitchell, D.J., McNaughton, N., Flanagan, D., Kirk, I.J., 2008. Frontal-midline theta from the perspective of hippocampal “theta”. *Prog. Neurobiol.* 86 (3), 156–185. <https://doi.org/10.1016/j.pneurobio.2008.09.005>.
- Mitchell, K.J., Johnson, M.K., Raye, C.L., Greene, E.J., 2004. Prefrontal cortex activity associated with source monitoring in a working memory task. *J. Cogn. Neurosci.* 16 (6), 921–934. <https://doi.org/10.1162/0899929041502724>.
- Nairne, J.S., 2002. Remembering over the short-term: the case against the standard model. *Annu. Rev. Psychol.* 53 (1), 53–81. <https://doi.org/10.1146/annurev.psych.53.100901.135131>.
- Nosofsky, R.M., Little, D.R., Donkin, C., Fific, M., 2011. Short-term memory scanning viewed as exemplar-based categorization. *Psychol. Rev.* 118 (2), 280–315. <https://doi.org/10.1037/a0022494>.
- Oberauer, K., 2002. Access to information in working memory: exploring the focus of attention. *J. Exp. Psychol. Learn. Mem. Cogn.* 28 (3), 411.
- Owen, A.M., Sahakian, B.J., Semple, J., Polkey, C.E., Robbins, T.W., 1995. Visuo-spatial short-term recognition memory and learning after temporal lobe excisions, frontal lobe excisions or amygdalo-hippocampectomy in man. *Neuropsychologia* 33 (1), 1–24. [https://doi.org/10.1016/0028-3932\(94\)00098-a](https://doi.org/10.1016/0028-3932(94)00098-a).
- Öztekin, I., McElree, B., Staresina, B.P., Davachi, L., 2009. Working memory retrieval: contributions of the left prefrontal cortex, the left posterior parietal cortex, and the hippocampus. *J. Cogn. Neurosci.* 21 (3), 581–593.
- Pascual-Marqui, R.D., Michel, C.M., Lehmann, D., 1995. Segmentation of brain electrical activity into microstates: model estimation and validation. *IEEE Trans. Biomed. Eng.* 42 (7), 658–665.
- Picton, T., Bentin, S., Berg, P., Donchin, E., Hillyard, S., Johnson, R., et al., 2000. Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. *Psychophysiology* 37 (2), 127–152. <https://doi.org/10.1111/1469-8986.3720127>.
- Portoles, O., Borst, J.P., van Vugt, M.K., 2018. Characterizing synchrony patterns across cognitive task stages of associative recognition memory. *Eur. J. Neurosci.* <https://doi.org/10.1111/ejn.13817>.
- Raaijmakers, J.G.W., Shiffrin, R.M., 1992. Models for recall and recognition. *Annu. Rev. Psychol.* 43 (1), 205–234. <https://doi.org/10.1146/annurev.ps.43.020192.001225>.
- Ranganath, C., 2006. Working memory for visual objects: complementary roles of inferior temporal, medial temporal, and prefrontal cortex. *Neuroscience* 139 (1), 277–289. <https://doi.org/10.1016/j.neuroscience.2005.06.092>.
- Ratcliff, R., 1978. A theory of memory retrieval. *Psychol. Rev.* 85 (2), 59–108. <https://doi.org/10.1037/0033-295x.85.2.59>.
- Rowe, J.B., 2000. The prefrontal cortex: response selection or maintenance within working memory? *Science* 288 (5471), 1656–1660. <https://doi.org/10.1126/science.288.5471.1656>.
- Rugg, M.D., Yonelinas, A.P., 2003. Human recognition memory: a cognitive neuroscience perspective. *Trends Cogn. Sci.* 7 (7), 313–319. [https://doi.org/10.1016/s1364-6613\(03\)00131-1](https://doi.org/10.1016/s1364-6613(03)00131-1).
- Shah, A.S., Bressler, S.L., Knuth, K.H., Ding, M., Mehta, A.D., Ulbert, I., Schroeder, C.E., 2004. Neural dynamics and the fundamental mechanisms of event-related brain potentials. *Cereb. Cortex* 14, 476–483. <https://doi.org/10.1093/cercor/bhh009>.
- Siapas, A.G., Lubenov, E.V., Wilson, M.A., 2005. Prefrontal phase locking to hippocampal theta oscillations. *Neuron* 46 (1), 141–151. <https://doi.org/10.1016/j.neuron.2005.02.028>.
- Souza, A.S., Renko, L., Oberauer, K., 2014. Refreshing memory traces: thinking of an item improves retrieval from visual working memory. *Ann. N. Y. Acad. Sci.* 1339 (1), 20–31. <https://doi.org/10.1111/nyas.12603>.
- Tass, P., Rosenblum, M.G., Weule, J., Kurths, J., Pikovsky, A., Volkman, J., Freund, H.-J., 1998. Detection of n:m Phase locking from noisy data: application to magnetoencephalography. *Phys. Rev. Lett.* 81 (15), 3291–3294. <https://doi.org/10.1103/physrevlett.81.3291>.
- Varela, F., Lachaux, J.-P., Rodriguez, E., Martinerie, J., 2001. The brainweb: phase synchronization and large-scale integration. *Nat. Rev. Neurosci.* 2 (4), 229–239. <https://doi.org/10.1038/35067550>.
- Van Vugt, M.K., Schulze-Bonhage, A., Litt, B., Brandt, A., Kahana, M.J., 2010. Hippocampal gamma oscillations increase with memory load. *J. Neurosci.* 30 (7), 2694–2699. <https://doi.org/10.1523/jneurosci.0567-09.2010>.
- Van Vugt, M.K., Sekuler, R., Wilson, H.R., Kahana, M.J., 2013. An electrophysiological signature of summed similarity in visual working memory. *J. Exp. Psychol. General* 142 (2), 412–425. <https://doi.org/10.1037/a0029759>.

- Wilson, H.R., Loffler, G., Wilkinson, F., 2002. Synthetic faces, face cubes, and the geometry of face space. *Vis. Res.* 42 (27), 2909–2923. [https://doi.org/10.1016/S0042-6989\(02\)00362-0](https://doi.org/10.1016/S0042-6989(02)00362-0).
- Yeung, N., Bogacz, R., Holroyd, C.B., Cohen, J.D., 2004. Detection of synchronized oscillations in the electroencephalogram: an evaluation of methods. *Psychophysiology* 41 (6), 822–832. <https://doi.org/10.1111/j.1469-8986.2004.00239.x>.
- Yeung, N., Bogacz, R., Holroyd, C.B., Nieuwenhuis, S., Cohen, J.D., 2007. Theta phase resetting and the error-related negativity. *Psychophysiology* 44, 39–49.
- Zhang, Q., Borst, J.P., Kass, R.E., Anderson, J.R., 2017a. Inter-subject alignment of MEG datasets in a common representational space. *Hum. Brain Mapp.* 38 (9), 4287–4301. <https://doi.org/10.1002/hbm.23689>.
- Zhang, Q., Walsh, M.M., Anderson, J.R., 2017b. The effects of probe similarity on retrieval and comparison processes in associative recognition. *J. Cogn. Neurosci.* 29 (2), 352–367. [https://doi.org/10.1162/jocn\\_a\\_01059](https://doi.org/10.1162/jocn_a_01059).