Title: Dynamic updating of hippocampal object representations reflects new conceptual knowledge

Authors: Michael L. Mack, Bradley C. Love, and Alison R. Preston

Supplemental Information

Methods

Participants

Twenty-three volunteers (11 females, mean age 22.3 years old, ranging from 18 to 31 years) participated in the experiment. All subjects were right handed, had normal or corrected-to-normal vision, and were compensated $75 for participating.

Stimuli

Eight color images of insects were used in the experiment (Figure 1b). The insect images consisted of one body with different combinations of three features: legs, mouth, and antennae. There were two versions of each feature (thick and thin legs, thick and thin antennae, and shovel or pincer mouth). The eight insect images included all possible combinations of the three features. The stimuli were sized to 300 x 300 pixels.

Task procedures

After an initial screening and consent in accordance with the University of Texas Institutional Review Board, participants were instructed on the classification learning tasks. Participants then performed the tasks in the MRI scanner by viewing visual stimuli back-projected onto a screen through a mirror attached onto the head coil. Foam pads were used to minimize head motion. Stimulus presentation and timing was performed using custom scripts written in Matlab (Mathworks) and Psychtoolbox (www.psychtoolbox.org) on an Apple Mac Pro computer running OS X 10.7.

Participants were instructed to learn how to classify the insects based on the combination of the insects' features. They were instructed to learn by using the feedback displayed on each trial. As part of the initial instructions, participants were made aware of the three features and the two different values of each feature. Before beginning each classification problem, additional instructions that described the cover story for the current task and which buttons to press for the two insect classes were presented to the participants. One example of this instruction text is as follows: "Each insect prefers either Warm or Cold temperatures. The temperature that each insect prefers depends on one or more of its features. On each trial, you will be shown an insect and you will make a response as to that insect's preferred temperature. Press the 1 button under your index finger for Warm temperatures or the 2 button under your middle finger for Cold temperatures." The other two cover stories involved classifying insects into those that live in the Eastern vs. Western hemisphere and those that live in an Urban vs. Rural environment. The cover stories were randomly paired with the

familiarization task and the two learning tasks for each participant. After the instruction screen, the four fMRI scanning runs (described below) for that task commenced, with no further task instructions. After all four scanning runs for a task finished, the next task began with the corresponding cover story description. Importantly, the rules that defined the classification problems were not included in any of the instructions; rather, participants had to learn these rules through trial and error.

Participants first performed a familiarization task, in which they were presented with and learned class association responses to each of the insect stimuli. This task had the same format as the classification learning tasks, but was structured such that all insect features had to be attended in order to respond correctly. The familiarization task was included to familiarize participants with the insect stimuli and task procedures to eliminate any neural activation due to stimulus and task novelty during the learning tasks. Data from the familiarization task was not considered for analysis. In contrast to the familiarization task, the type 1 and type 2 learning tasks were structured such that perfect performance required attending only to a subset of feature dimensions. For the type 1 task, class associations were defined by a rule depending on the value of one dimension. For the type 2 task, class associations were defined by an XOR logical rule that depended on the value of the two dimensions that were not relevant in the type 1 task (Fig. 1b). As such, different dimensions were relevant to the two tasks and successfully learning the classification tasks required a shift in attention to attend to dimensions most relevant for the current task. The binary values of the eight insect stimuli along with the class association for the type 1 and type 2 tasks are depicted in Table S1. The stimulus features were randomly mapped onto the dimensions for each participant. These feature-to-dimension mappings were fixed across the different classification learning tasks within a participant. After the familiarization task, participants learned the type 1 and 2 tasks in sequential order. The learning order of the type 1 and 2 tasks was counterbalanced across participants.

	feature dimension			Class	
stimulus	1	2	3	type 1	type 2
1	0	0	0	A	C
2	0	0	1	A	D
3	0	1	0	A	D
4	0	1	1	A	C
5	1	0	0	B	C
6	1	0	1	B	D
7	1	1	0	B	D
8	1	1	1	B	C

Table S1: Stimulus features and class associations. Each of the eight stimuli are represented by the binary values of the three feature dimensions and their class associations for the type 1 and type 2 classification tasks.

The classification tasks consisted of learning trials (Fig. 1a) during which an insect image was presented for 3.5s. During stimulus presentation, participants were instructed to respond to the insect's class by pressing one of two buttons on an fMRI-compatible button box. Insect images subtended 7.3° × 7.3° of visual space. The stimulus presentation period was followed by a 0.5-4.5s fixation. A feedback screen consisting of the insect image, text of whether the response was correct or incorrect, and the correct class was shown for 2s followed by a 4-8s fixation. The timing of the stimulus and feedback phases of the learning trials was jittered to optimize general linear modeling estimation of the fMRI data. Within one functional run, each of the eight insect images was presented in four learning trials. The order of the learning trials was pseudo randomized in blocks of sixteen trials such that the eight stimuli were each presented twice. One functional run was 194s in duration. Each of the learning problems included four functional runs for a total of sixteen repetitions for each insect stimulus. The entire experiment lasted approximately 65 minutes.

Behavioral analysis

Learning performance during the classification tasks was analyzed using a bounded logistic regression with random effects of repetition, order, and task (Fig. 1c). This analysis was performed using lme4 (version 1.1-12) and psyphy (version 0.1-9) packages in R (version 3.2.5). Participant-specific learning curves were also extracted for each task by calculating the average accuracy across blocks of sixteen learning trials. These learning curves were used for the computational learning model analysis.

Computational learning modeling

Participant behavior was modeled with an established mathematical learning model, SUSTAIN (1). SUSTAIN is a network-based learning model (Fig. 2a) that classifies incoming stimuli by comparing to memory-based knowledge representations of previously experienced stimuli. Sensory stimuli are encoded by SUSTAIN into perceptual representations based on the value of the stimulus features. The values of these features are biased according to attention weights operationalized as receptive fields on each feature dimension. During the course of learning, these attention weight receptive fields are tuned to give more weight to diagnostic features. SUSTAIN represents knowledge as clusters of stimulus features and class associations that are built and tuned over the course of learning. New clusters are recruited and existing clusters updated according to the current learning goals. The formulization of SUSTAIN is provided in the supplemental information.

To characterize the latent attention-biased representations participants formed during learning, we fit SUSTAIN to each participant's learning performance. First, SUSTAIN was initialized with no clusters and equivalent attention weights across the stimulus dimensions. Then, stimuli were presented to SUSTAIN in the same order as what the participants experienced and model parameters were optimized to predict each participant's learning performance in the familiarization task and two learning tasks

through a maximum likelihood genetic algorithm optimization method (2). In the fitting procedure, the model state from the end of the familiarization task (in which attention to features was equivalent) was used as the initial state for the first learning task, and the model state at the end of the first learning task was used as the initial state for the second learning task. In doing so, parameters were optimized to account for learning in the familiarization task and both learning tasks with the assumption that attention weights and knowledge clusters learned from the familiarization task carried over to influence learning in the first task; and similarly, model state from the first task carried over and influenced early learning in the second task. The optimized parameters were then used to extract measures of dimensional attention weights and latent representations of the stimuli during the second half of learning in the two tasks. Specifically, for each participant, the model parameters were fixed to the optimized values and the model was presented with the trial order experienced by the participant. After the model was presented with the first half of trials, the value of the dimensional attention weights, λ_i, were extracted for each participant (Fig. 2b). Latent model representations were also extracted for each stimulus. We did this by presenting the model with each stimulus and saving out vectors of cluster activations, $H^{act}{}_i$ (see below for model formalism). The pairwise similarities of these cluster activation vectors were then calculated with Pearson correlation. The resulting similarity matrices served as the model-based prediction of attention-biased representations (Fig. 2c) used in the multivariate fMRI pattern analysis (Fig. 3).

Computational modeling methods

The following sections describe SUSTAIN's formalism, how the model learns, and how the model was fit to each participant's learning behavior.

Perceptual encoding. An input stimulus is presented to SUSTAIN as a pattern of activation on input units that code for the different stimulus features and possible values that these features can take. For each stimulus feature, *i* (e.g., a beetle's legs), with *k* possible values (two in the present experiment; e.g., thick or thin legs), there are *k* input units. Input units are set to one if the unit represents the feature value or zero otherwise. The entire stimulus is represented by $I^{pos_{ik}}$, with *i* indicating the stimulus feature and *k* indicating the value for feature *i*. "pos" indicates that the stimulus is represented as a point in a multidimensional space. The distance μ_{ij} between the *i*th stimulus feature and cluster *j*'s position along the *i*th feature is

$$\mu_{ij} = 1/2 \sum_{k=1}^{v_i} | I^{pos_{ik}} - H_j^{pos_{ik}} | \qquad (1)$$

wherein v_i is the number of possible values that the *i*th stimulus feature can take and $H_j^{pos_{ik}}$ is cluster *j*'s position on the *i*th feature for value *k*. Distance μ_{ij} is always between 0 and 1, inclusive.

Response selection. After perceptual encoding, each cluster is activated based on the similarity of the cluster to the input stimulus. Cluster activation is given by:

$$H_j^{act} = \frac{\sum_{i=1}^{n_a} (\lambda_i)^{\gamma} e^{-\lambda_i \mu_{ij}}}{\sum_{i=1}^{n_a} (\lambda_i)^{\gamma}} \quad (2)$$

wherein H_j^{act} is cluster *j*'s activation, n_a is the number of stimulus features, λ_i is the attention weight receptive field tuning for feature *i*, and γ is the attentional parameter (constrained to be non-negative). Clusters compete to respond to an input stimulus through mutual inhibition. The final output of each cluster *j* is given by:

$$H_j^{out} = \frac{(H_j^{act})^{\beta}}{\sum_{i=1}^{n_c} (H_j^{act})^{\beta}} H_j^{act} \quad (3)$$

wherein n_c is the current number of clusters and β is a lateral inhibition parameter (constrained to be non-negative) that controls the level of cluster competition. The cluster that wins the competition, H_m, passes its output to the *k* output units of the unknown feature dimension *z*:

$$C_{zk}^{out} = w_{m,zk} H_m^{out} \quad (4)$$

wherein C_{zk}^{out} is the output of the unit representing the *k*th feature value of the *z*th feature, and $w_{m,zk}$ is the weight from the winning cluster, H_m, to the output unit C_{zk}. In the current simulations, the class label is the only unknown feature dimension. Thus, equation 4 is calculated for each of the two values of the class label. Finally, the probability of making a response *k* for a queried dimension, *z*, on a given trial is:

$$P(k) = \frac{e^{(dC_{zk}^{out})}}{\sum_{j=1}^{v_z} e^{(dC_{zk}^{out})}} \quad (5)$$

Cluster recruitment. In the current study, SUSTAIN was initialized with zero clusters. During learning, clusters are recruited in response to a combination of the order of the stimuli presented in the participant-specific trial orders and the error feedback received on each trial. In the current study, SUSTAIN was presented with trial orders from the familiarization task followed by the two learning tasks. We included a cluster recruitment parameter, τ (constrained to be between 0 and 1), that probabilistically determines whether an error will lead to new cluster recruitment. If SUSTAIN makes a prediction

error, and τ exceeds *q*, wherein *q* is a randomly generated value between 0 and 1, a new cluster is recruited. Otherwise, the winning cluster from the cluster competition is updated to reflect current stimulus features and class label according to the learning rules explained next.

Learning. SUSTAIN's learning rules determine how clusters are updated during learning. Only the winning clusters are updated. If a new cluster is recruited on a trial, it is considered the winning cluster. Otherwise, the cluster that is most similar to the current stimulus will be the winner. The winning cluster H_m, is adjusted by:

$$\Delta H_m^{pos_{ik}} = \eta(I^{pos_{ik}} - H_m^{pos_{ik}}) \quad (6)$$

wherein η is the learning rate parameter. The result of the updating is that the winning cluster moves toward the current stimulus. Over the course of learning, each cluster will tend toward the center of its members. Attention weight receptive field tunings for the different feature dimensions are updated according to:

$$\Delta\lambda i = \eta e^{-\lambda_i \mu_{im}}(1 - \lambda_i \mu_{im}) \quad (7)$$

wherein *m* indexes the winning cluster.

The weights from the winning cluster to the output units are adjusted by a one layer delta learning rule.

$$\Delta w_{m,zk} = \eta(t_{zk} - C_{zk}^{out})H_m^{out} \quad (8)$$

Simulations. For the current study, stimuli were presented to SUSTAIN using the same trial order as the participants. To reflect the carryover of the previous learning task on the current learning task, the attention weight receptive field tunings and clusters were not reinitialized between tasks. Rather, model fits were such that a single set of parameters were optimized to describe behavior on both learning tasks. This methodology takes into account each participant's learning experience and allows us to quantify how the first task influenced learning on the second task. Thus, task order effects are considered a natural consequence of our model fitting approach. The free parameters, γ, β, η, *d,* and τ_h, were fit to each participant's learning curve using a maximum likelihood genetic algorithm optimization technique(2). Obtained mean parameter values and 95% confidence intervals were: γ = 3.286 ± 2.064, β = 4.626 ± 0.220, η = 0.308 ± 0.145, *d* = 20.293 ± 5.724, τ_h = 0.112 ± 0.039.

MRI data acquisition

Whole-brain imaging data were acquired on a 3.0T Siemens Skyra system at the University of Texas at Austin Imaging Research Center. A high-resolution T1-weighted MPRAGE structural volume (TR = 1.9s, TE = 2.43ms, flip angle = 9°, FOV = 256mm, matrix = 256x256, voxel dimensions = 1mm isotropic) was acquired for coregistration

and parcellation. Two oblique coronal T2-weighted structural images were acquired perpendicular to the main axis of the hippocampus (TR = 13,150ms, TE = 82ms, matrix = 384x384, 0.4x0.4mm in-plane resolution, 1.5mm thru-plane resolution, 60 slices, no gap). These images were coregistered and averaged to generate a mean coronal image for each participant that was used to localize peak voxels from the model-based RSA results to hippocampal subfields. High-resolution functional images were acquired using a T2*-weighted multiband accelerated EPI pulse sequence (TR = 2s, TE = 31ms, flip angle = 73°, FOV = 220mm, matrix = 128x128, slice thickness = 1.7mm, number of slices = 72, multiband factor = 3) allowing for whole brain coverage with 1.7mm isotropic voxels.

MRI data preprocessing and statistical analysis

MRI data were preprocessed and analyzed using FSL 6.0 (3) and custom Python routines. Functional images were realigned to the first volume of the seventh functional run to correct for motion, spatially smoothed using a 3mm full-width-half-maximum Gaussian kernel, high-pass filtered (128s), and detrended to remove linear trends within each run. Functional images were registered to the MPRAGE structural volume using Advanced Normalization Tools, version 1.9 (4). All analyses were performed in the native space of each participant.

Hippocampus region of interest

The hippocampus was delineated by hand on the 1mm Montreal Neurological Institute (MNI) template brain and reverse-normalized to each participant's functional space using ANTS. Specifically, a nonlinear transformation was calculated from the MNI template brain to each participant's T1-weighted MPRAGE volume. This warp was then concatenated with the MPRAGE to functional space transformation calculated using ANTS. Finally, the concatenated transformation was applied to the anatomical hippocampus ROI to move the ROI into each participant's functional space.

Model-based representational similarity analysis

The goal of the similarity analysis was to assess the extent that attention processes bias neural representations of individual stimuli during the different learning tasks. In contrast to classification techniques that are used to decode activation patterns associated with relatively small number of stimulus classes or conditions, pattern similarity methods allow one to evaluate activation patterns at the level of single events or stimuli (5, 6). In the current study, we used pattern similarity methods to evaluate the similarity between neural patterns for each of the insect stimuli under the different learning contexts.

Pattern similarity analyses were implemented using PyMVPA (7) and custom Python routines and were conducted on preprocessed and spatially smoothed functional data. The decision to perform multivariate analyses on spatially smoothed data is consistent with recent studies employing MVPA (8, 9) and demonstrations that smoothing does not result in information loss (10, 11). First, whole brain activation patterns for each stimulus within each run were estimated using an event-specific univariate general linear model

(GLM) approach (12, 13). In contrast to the classification approach that leverages the variance in neural patterns to learn voxel weights that best discriminate conditions, pattern similarity analyses require stable estimates of neural representations for the conditions of interest. In the current study, the condition of interest was at the level of specific stimuli. Thus, we took a GLM approach to model stable estimates of neural patterns for each of the eight insect stimuli. For each classification task run, a GLM with separate regressors for stimulus presentation of the eight insect stimuli, modeled as 3.5s boxcar convolved with a canonical hemodynamic response function (HRF), was conducted to extract voxelwise parameter estimates to each of the stimuli. Additionally, stimulus-specific regressors for the feedback period of the learning trials (2s boxcar) and responses (impulse function at the time of response), as well as six motion parameters were included in the GLM. Since the majority of participants had reached asymptotic performance by the end of the second run, we focused on learned representations present in the latter half of learning. Thus, a second level GLM analysis was conducted to average the stimulus-specific parameter estimates from the third and fourth runs of the two classification tasks. This procedure resulted in, for each participant, whole brain activation patterns during the later stages of learning for each of the eight stimuli in both classification tasks.

We compared neural measures of stimulus representation during learning to model predictions with a searchlight method (14). Using a searchlight sphere with a radius of 3 voxels, we extracted a vector of activation values across all voxels within a searchlight sphere for each of the eight stimuli. The pairwise similarities between these activation vectors were calculated with Pearson correlation. The resulting similarity matrices captured the similarity structure among the neural representations of the stimuli during learning. We then tested whether or not the neural representations were consistent with model-based predictions of stimulus representations by calculating the Spearman correlation between the values in the upper triangles of the neural and model similarity matrices. A reshuffling randomization test was performed on the resulting correlation coefficient. For each iteration of the randomization test, the rows of the model similarity matrix were randomly shuffled and the Spearman correlation between the shuffled model and neural similarity matrices was calculated. This procedure was repeated 1000 times to create a null distribution. Finally, a test statistic defined as the probability that the correlation coefficient between the actual model and neural similarity matrices was larger than the null distribution was calculated. This entire procedure was performed for each searchlight sphere location resulting in statistical maps that characterized the consistency between attention-biased model predictions (i.e., attention weighting hypothesis) and neural measures of learned stimulus representations for each participant in both tasks. A second analysis using the same methods was also performed that compared the neural measures of stimulus representations to similarity predictions based only on class associations (i.e., associative mapping hypothesis). Specifically, matrices representing whether or not pairs of stimuli were in the same class were constructed and evaluated for consistency with neural similarity matrices in the same manner as the model similarity matrices (Fig. 3). In separate analyses, the

searchlight method was applied to activation patterns present only in the hippocampus ROI.

Group-level analyses were performed on the statistical maps calculated with the pattern similarity searchlight procedure. Each participant's *p*-maps were transformed to z-scores and normalized to MNI space using ANTS. We then performed a one-sample randomization test on the correspondence between attention weighting and neural similarity with voxelwise nonparametric permutation testing (5000 permutations) performed using FSL Randomise (15). To evaluate our hypothesis that the hippocampus builds representations consistent with attentional strategies, we performed a small volume cluster correction analysis restricted only to the hippocampus. Specifically, the resulting statistical maps from the hippocampal ROI (Fig. 4a) were voxelwise thresholded at $p = 0.005$ and cluster corrected at $p = 0.05$ which corresponded to a cluster extent threshold of greater than 149 voxels as determined by AFNI 3dClustSim using the *acf* option, second-nearest neighbor clustering, and 2-sided thresholding. The version of 3dClustSim used was compiled on 1/21/2106 and included fixes for the recently discovered errors of failing to account for edge effects in simulations involving small regions and improperly accounting for spatial autocorrelation in smoothness estimates.

A control analysis was conducted to interrogate the response magnitude across the learning tasks in the left anterior hippocampus region identified in the model-based RSA results (Fig. 4). Specifically, the average signal from the trial-by-trial betaseries within a region defined by the hippocampus cluster was extracted from the stimulus presentation phase of each trial for each participant. Response magnitude differences between the two tasks were evaluated with Wilcoxon signed rank tests and revealed no significant differences between task across the full experiment ($Z = 0.091$, $p = 0.927$), nor the early and late phases (early: $Z = 0.183$, $p = 0.855$; late: $Z = 0.365$, $p = 0.715$). There were also no significant differences in response amplitude across the early and late phases within the tasks (type 1: $Z = 0.395$, $p = 0.693$; type 2: $Z = 0.760$, $p = 0.447$). These null findings suggest the task-related differences in neural activity were not due to differences in overall engagement of the hippocampus, but at the level of neural representations.

As an additional control analysis, we contrasted the model-based RSA results with a separate analysis employing a standard RSA approach (6, e.g., 9) wherein neural similarity is simply predicted to follow class association. This standard RSA approach was operationalized as a similarity matrix where pairs of stimuli in the same class had maximum similarity and pairs in different classes had minimum similarity. No HPC regions were consistent with simple class association, and the left anterior HPC cluster revealed in the model-based RSA remained significant when the model-based and standard RSA results were directly contrasted. These findings suggest that HPC dynamically codes for attention-weighted conceptual representations that are optimized for current learning goals.

Neurally-derived attention weights

To visualize attentional tuning in the hippocampus region identified in the model-based RSA, we estimated attention weights from stimulus-specific neural representations. It is important to note that this analysis is not independent of the RSA findings. To be clear, we are not presenting it as additional evidence, but as a method for visually representing the conceptual coding in the hippocampal activation patterns identified by the RSA. Neurally-derived attention weights (λ^n) were estimated by first extracting the stimulus-specific neural representations from the left anterior hippocampal region from the late phase of learning in both tasks for each participant. These neural representations were extracted from the trial-by-trial betaseries used for the model-based RSA. For each of the three stimulus feature dimensions, the average pairwise similarity between stimuli that shared the same value on the feature (e.g., both had thick legs or both had thin legs) was divided by the average similarity between stimuli that did not share the same value (e.g., one had thick legs, the other thin legs). This ratio served as a neural estimate of the attention weight for that feature. Pairwise similarity was calculated as the exponential of the negative Euclidean distance between stimulus representations. For each participant, neurally-derived attention weights were estimated for each feature dimension in the two learning tasks separately. These attention weights were normalized for each task to sum to 1 (λ^n mean and 95% confidence intervals for type 1: [0.409 ± 0.062, 0.289 ± 0.040, 0.302 ± 0.25]; and type 2: [0.277 ± 0.035, 0.322 ± 0.043, 0.402 ± 0.059]). Finally, the attention weights for the two tasks were averaged across participants and projected into stimulus feature space (as defined in Table 3) to demonstrate how attentional tuning changed across tasks (Fig. 4b).

Functional connectivity analysis

The goal of the functional connectivity analysis was to evaluate the functional coupling between the hippocampal region showing attention-biased representations (Fig. 4) and the rest of the brain. In particular, we were interested in investigating how connectivity with the hippocampus is mediated by early versus late learning. We investigated connectivity with a psychophysiological interaction (PPI) analysis (16). Seed time courses from the left anterior hippocampal region identified in the pattern similarity analysis were extracted for each participant by averaging mean BOLD signal across the region separately for each time point. These seed time courses were then entered into a voxelwise GLM analysis of the functional data across the whole brain. A second level GLM analysis was conducted to contrast voxel time course connectivity with the hippocampal seed region time course in early versus late learning. Specifically, separately for the two tasks, first level parameter estimates from the first two functional runs were labeled as early learning and contrasted with parameter estimates from the last two functional runs. The resulting contrast images were normalized to MNI space using ANTS and submitted to a group analysis using FSL Randomise nonparametric randomization tests (5000 repetitions). The resulting statistic maps (Fig. 4c) were voxelwise thresholded at $p < 0.005$ and cluster corrected at $p < 0.05$ with a cluster

extent threshold of 791 voxels as determined by 3dClustStim using the *acf* option, second-nearest neighbor clustering, and 2-sided thresholding (Table S2).

anatomical region	peak z-value	extent (voxels)	peak location
bilateral medial prefrontal cortex	4.34	2836	10, 43, -6
right inferior lateral occipital cortex	4.44	2386	35, -82, -11
right frontopolar cortex	4.12	1806	36, 57, -10
right dorsolateral prefrontal cortex	6.00	1155	58, -13, 48

Table S2: Results of functional connectivity analysis. Clusters that survived statistical thresholding are described according to their corresponding anatomical region, peak z-value in the group-level statistical maps, cluster extent in voxels, and the location of the peak z-value in MNI coordinates.

Supplemental References

1. Love BC, Medin D, Gureckis TM (2004) SUSTAIN: A network model of category learning. *Psychol Rev* 111(2):309–332.
2. Storn R, Price K (1997) Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *J Glob Optim* 11:341–359.
3. Smith SM, et al. (2004) Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* 23 Suppl 1:S208–19.
4. Avants BB, et al. (2011) A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage* 54(3):2033–2044.
5. Kriegeskorte N, Mur M, Bandettini P (2008) Representational similarity analysis - connecting the branches of systems neuroscience. *Front Syst Neurosci* 2(November):4.
6. Schlichting ML, Mumford JA, Preston AR (2015) Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. *Nat Commun* 6:8151.
7. Hanke M, et al. (2009) PyMVPA: A python toolbox for multivariate pattern analysis of fMRI data. *Neuroinformatics* 7(1):37–53.
8. Kuhl BA, Chun MM (2014) Successful remembering elicits event-specific activity patterns in lateral parietal cortex. *J Neurosci* 34(23):8051–60.
9. Schapiro AC, Kustner L V., Turk-Browne NB (2012) Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Curr Biol* 22(17):1622–1627.
10. Kamitani Y, Sawahata Y (2010) Spatial smoothing hurts localization but not information: Pitfalls for brain mappers. *Neuroimage* 49(3):1949–1952.
11. Op de Beeck HP (2010) Against hyperacuity in brain reading: Spatial smoothing does not hurt multivariate fMRI analyses? *Neuroimage* 49(3):1943–1948.
12. Mumford JA, Turner BO, Ashby FG, Poldrack RA (2012) Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. *Neuroimage* 59(3):2636–2643.
13. Xue G, et al. (2010) Greater neural pattern similarity across repetitions is associated with better memory. *Science* 330(October):97–101.
14. Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. *Proc Natl Acad Sci U S A* 103(10):3863–3868.
15. Winkler AM, Ridgway GR, Webster M a., Smith SM, Nichols TE (2014) Permutation inference for the general linear model. *Neuroimage* 92:381–397.
16. Friston KJ, et al. (1997) Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage* 6(3):218–229.