Visual imagery of familiar people and places in category selective cortex

Catriona L. Scrivener*, Jessica A. Teed, & Edward H. Silson

Department of Psychology, School of Philosophy, Psychology and Language Sciences, University of Edinburgh, Edinburgh, UK.

*corresponding author: catriona.scrivener@ed.ac.uk, 7 George Square, Edinburgh, EH8 9JZ

Abstract

Visual mental imagery is a dynamic process that involves a network of multiple brain regions. We used an electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) fusion approach to ask how the neural dynamics of category selective mental imagery in EEG related to activity within medial parietal, ventral temporal, and primary visual cortex (V1). Subjects attended separate EEG and fMRI sessions where they were asked to recall personally familiar people and place stimuli. The fMRI contrast of places versus people successfully replicated previous findings of category-selectivity in the medial parietal cortex during visual recall (Silson et al., 2019), as well as other regions including the ventral and lateral place memory areas, the fusiform face area, and frontal eye fields. Using multivariate decoding analysis in pre-defined fMRI regions of interest (ROIs), we tested the hypothesis that we would be able to decode individual stimuli within the preferred category for each region. This was largely the case in the ventral temporal ROIs, but a more complex pattern emerged in the medial parietal cortex; these regions represented information during imagery that was not restricted to their preferred category. EEG-fMRI fusion indicated that the timing of both medial parietal and ventral temporal involvement peaked early on during recall but did not clearly differ from each other. However, place-selective regions generally peaked earlier than people-selective regions, suggesting that representations of place stimuli evolved more quickly for these subjects. This mirrors the results from EEG stimuli decoding, where individual places were decodable earlier in time than people. In contrast, fusion correlations in V1 occurred later during the recall period, possibly reflecting the top-down progression of mental imagery from category-selective regions to primary visual cortex.

Keywords: mental imagery, recall, EEG-MRI fusion, category selectivity

1. Introduction

The ability to picture something in your 'minds-eye' in the absence of an external stimulus is referred to as mental imagery; an everyday task with surprising variation across individuals (Zeman, 2024). One theory of visual mental imagery describes the phenomenon as perception in reverse (Pearson et al., 2019). Perception of visual stimuli activates a hierarchy of brain regions, extending from early visual cortex into dorsal and ventral streams, further recruiting parietal and frontal regions in a task dependent manner. Visual mental imagery might then be characterised by activity in similar regions but in the reverse order, culminating in a reinstatement of 'perception-like' activity in primary visual cortex (V1) (Dijkstra et al., 2017). This hypothesis has led to a focus of exploration in V1, with groups searching for similar functional magnetic resonance imaging (fMRI) blood oxygenation level dependent (BOLD) activity or representational patterns across perception and recall of the same stimuli. There has been some success in this approach for visual imagery of simple stimuli such as gratings, letters, and objects (Albers et al., 2013; Ragni et al., 2020 Cortex; Lee et al., 2012). However, successful comparison of perception and imagery in V1 may be influenced by the stimuli used and the task given to participants (Dijkstra et al., 2019; Kosslyn et al., 2003).

An alternative theory is that mental imagery recruits a flexible network of category-selective, memory, and semantic processing regions that vary with the specific imagery content and task performed (Spagna et al., 2024). It is plausible that regions of the brain outside of early visual cortex are also involved in generating mental images, especially for images that require other cognitive processes such as engaging memory representations or mental rotation (Dijkstra et al., 2019; Zeman, 2024). One example of this is the visual imagery of specific people and places, which has been found to activate several regions within medial parietal cortex (MPC) (Silson et al., 2019). These regions display category-selectivity, with increased BOLD activation during recall of stimuli from their preferred category (people or places) that is elevated when the stimuli are personally familiar. Several regions in the ventral temporal cortex (VTC) are also active during recall, including the parahippocampal place area (PPA, Epstein & Kanwisher, 1998) and fusiform face area (FFA, Kanwisher et al., 1997), despite mostly being studied in the context of visual perception. The link between these category-selective regions in medial parietal and ventral temporal cortex is indicated by their increased resting-state connectivity. For example, the anterior parahippocampal place area (aPPA, Baldassano et al., 2016) is most functionally connected with the largest place selective region in medial parietal cortex (Places 1; also referred to as ventral medial parietal cortex (MPCv) in Silson et al., 2019), whereas the anterior fusiform face area (FFA2) is most functionally connected with the analogous face selective region (People 1; also referred to as dorsal medial parietal cortex (MPCd) in Silson et al., 2019).

Although these category-selective regions in medial parietal and ventral temporal cortex are all engaged during recall of people and places, it is currently unclear if they contribute similarly to the generation of visual imagery or if they play distinct roles. During perception, medial parietal regions typically demonstrate a decrease in activation, which is related to their proximity to the default mode network (Raichle et al., 2001; Silson et al., 2019) and differentiates them from their ventral temporal counterparts. It has also been suggested that category-selective regions within medial parietal cortex form part of a semantic network, based on their overlap with regions representing the semantic content of visual images (Koslov et al., 2024). It is therefore plausible that the activity of medial parietal and ventral temporal regions during imagery will differ. Given that these ROIs are distributed along a ventral-dorsal axis within cortex, it is possible that their activity may emerge at different timepoints during mental imagery, and that this may differ from V1.

Prior work using electroencephalography (EEG) and magnetoencephalography (MEG) provides evidence for the time course of visual mental imagery across the brain as a whole. Significant decoding of imagery stimuli has been found to emerge between 540ms (Dijkstra et al., 2017) and 1000ms (Corriveau et al., 2022) post recall onset, with sustained representations throughout the recall period. Further, cross-decoding using perceptual responses to the same stimuli suggests that perception data at 150ms, 200ms, and 300ms is most predictive of imagery representations (Dijkstra et al., 2017; Xie et al., 2020), and that 'perception-like' imagery emerges from 400ms (Dijkstra et al., 2017) or 700ms (Xie et al., 2020) onwards (the differences in the exact timings across these studies are likely caused by variation in design and analysis choices). While these results provide valuable information regarding the emergence of visual imagery through time, it is difficult to assess the

source of this information in the brain, or how different regions may contribute to these imagery representations. To address this, we used an EEG-fMRI fusion approach (Cichy & Oliva, 2020) to examine how the time course of visual mental imagery recorded in EEG relates to activity across several key regions; medial parietal cortex (People and Place regions), ventral temporal cortex (PPA and FFA), and V1. We hypothesised that we may see different timings across medial parietal and ventral temporal category-selective ROIs, reflecting their different roles in the generation of mental images. We also hypothesised that we would find evidence of a top-down progression of mental imagery, with representations of people and places emerging earlier in higher-level visual cortex than in V1.

2. Methods

2.1 Experimental procedure

A total of 12 subjects participated in two experimental sessions, one with EEG recording and the other with fMRI (mean age = 27, 3 males). Ethical approval was provided by the University of Edinburgh (420-2122/5 and 420-2122/6). All participants gave informed consent and were given monetary compensation for their time. Separate EEG and MRI recordings were appropriate for this experiment as we did not intend to look at trial-by-trial fluctuations in a time-dependent manner (Scrivener, 2021). Instead, we planned to capture the shared information between conditions using a representational similarity analysis (RSA) approach (Nili et al., 2014). Before attending their first session, subjects were asked to provide the researchers with images of six personally familiar people and six personally familiar places that they were confident they could recall easily. We asked subjects to save these images with a name that would serve as a cue for them to recall the stimuli, such as ‘Sophie’, or ‘Office’. For the exact details of our request for stimuli, refer to the supplementary material (S1).

During the EEG session, subjects completed two runs of a perception task lasting 4.5 minutes each, where each of the 12 stimuli were presented 25 times per run in a randomised order (50 total repetitions of each stimuli, stimuli presented for 500ms, with a jittered inter-trial interval between 300 and 500ms). Stimuli were presented at 12 by 12 degrees of visual angle. Subjects were asked to look at a central fixation cross and press ‘space’ when it turned from green to red during a random subset of trials. They then completed four runs of the recall task lasting 6.8 minutes each **(Figure 1)**, where each recall cue was presented 10 times per run in a randomised order (40 total repetitions). One subject only completed 3 out of the 4 recall runs due to a technical error during the last run. Subjects were specifically instructed to create a mental image of the picture they had seen during the perception runs. This was an attempt to standardise their mental imagery across repetitions of the same stimuli, and to make it clear that they should be thinking of the images rather than about general aspects of the people and places (as much as possible). Subjects were also instructed to keep their eyes open and remain fixated on the cross. In this manuscript we focus on analysis of the recall data only, leaving consideration of the perception data for future work.

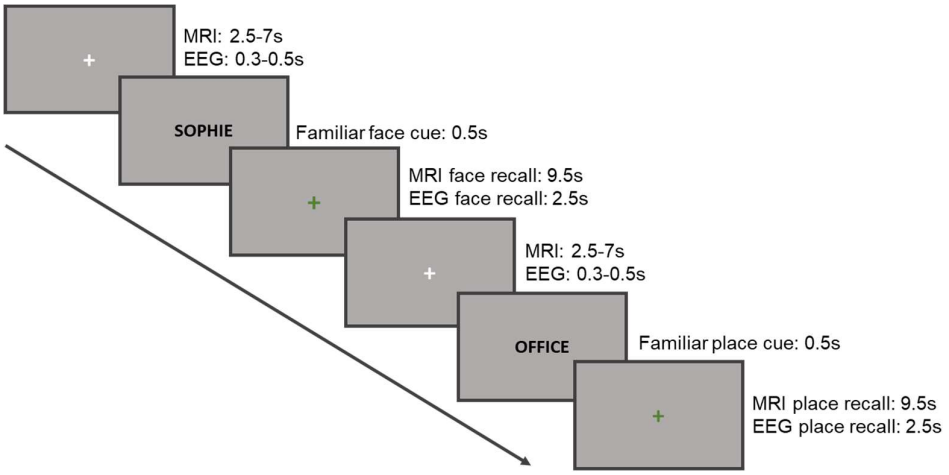

Figure 1: Recall paradigm during EEG and fMRI sessions. Each trial began with a white fixation cross with a jittered duration. Subjects were then presented with a visual cue for 0.5 seconds telling them which of their images they should recall. This was followed by a green fixation cross which indicated that they should visualise the cued image.

During the MRI session, subjects completed two hours of scanning with a break in between (ranging from 15-60 minutes, depending on scanner timetabling constraints). In session 1, we acquired T1 weighted and T2 weighted structural scans. These were followed by three runs of the perception task, lasting 4.4 minutes each, where each of the 12 stimuli were presented four times per run (12 total repetitions, stimuli presented for 500ms, with a jittered inter-trial interval between 2500 to 7000ms). This was followed by four functional runs of population receptive field (pRF) mapping (Dumoulin & Wandell, 2008), lasting six minutes each. For pRF mapping we used a bar aperture that made eight sweeps through the visual field (18 positions, 36s per sweep). Scene fragments from five images were presented inside the bar aperture during each position (400ms per scene), sampled randomly from a list of 90 possible scenes (Silson et al., 2015). During both pRF and perception tasks, subjects were asked to fixate a central cross and respond with a right-hand button press when it turned from green to red during a random subset of trials. In session 2, we acquired a second T1w structural scan to facilitate across session alignment and segmentation if needed. This was followed by two runs of a scene/face localiser task, lasting 5 minutes each. Subjects viewed 18 blocks of 20 scene or face stimuli that lasted 16s each (300ms on, 500ms off), and were asked to complete a one-back task by making a right-hand button press if the same image was repeated sequentially (twice per block). Finally, they completed six runs of the recall task, lasting six minutes each. Each recall cue was presented twice per run (12 total repetitions). One participant only completed four of the six recall runs. See **Figure 1** for the stimulus timing information during recall.

At the end of both EEG and MRI sessions, subjects were asked to complete the same short questionnaire regarding each stimulus presented during the session. This included an estimate of the overall vividness of their imagery per stimulus (1 = no imagery, 5 = moderately strong imagery, 10 = very clear imagery). All subjects also completed a paper version of the Vividness of Visual Imagery Questionnaire (VVIQ) during their first experimental session.

2.2 EEG acquisition and pre-processing

EEG data was recorded in ActiView using a BioSemi ActiveTwo AD-box amplifier and 64 pin-type active electrodes. Two additional flat-type free electrodes were secured to the left and right mastoids underneath the EEG cap. Data was acquired at a sampling rate of 512Hz. Subjects were seated approximately 57cm from a computer monitor that displayed the task (size 1920 x 1080). EEG data was pre-processed using EEGLab (2022.0, Delorme & Makeig, 2004) in MATLAB (2023b). The data was re-referenced to the left and right mastoid electrodes (pop_biosig) and then filtered using a Hamming window bandpass filter between 0.1 and 40Hz (pop_eegfiltnew). We then downsampled

the data to 256Hz to reduce computation time (pop_resample). Noisy channels were interpolated using a spherical interpolation (pop_interp) for two subjects only (005, 3 electrodes; 006, 7 electrodes). Noisy trials were identified by eye before removal, using a conservative criterion of any time period with large movement or muscle artifacts that affected more than 50% of the channels. On average, 12 trials were removed per subject across the whole EEG session (M = 12, SD = 12). Break periods were also removed as they often contained large artifacts. ICA was then performed to identify blink and lateralized eye movements (pop_runica, runica with extended informax). One blink component was removed for all subjects, and one lateralised eye movement was removed for 11 of the 12 subjects (pop_subcomp) after manual inspection of the component topologies and timeseries. Epochs were created from -200ms:3000ms in the recall runs in relation to the onset of the cue (pop_epoch). Baseline correction was performed using the timepoints -200:0ms (pop_rmbase).

2.3 MRI acquisition

MRI scans were acquired using a Siemens 3T Prisma scanner and a 32-channel head coil at the University of Edinburgh Imaging Facility RIE, Edinburgh Royal Infirmary. We acquired two types of structural image; T1 weighted (TR = 2.5s, TE = 4.37ms, flip angle = 7 deg, FOV = 256mm x 256mm x 192mm, resolution = 1mm isotropic, acceleration factor = 3), and T2 weighted (TR = 3.2s, TE = 408ms, flip angle = 9 deg, FOV = 256mm x 240mm x 192mm, resolution = 0.9mm isotropic, acceleration factor = 2). Functional scans were acquired using a multiecho multiband echo planar imaging sequence (TR = 2, TEs = 14.6ms, 35.79ms, 56.98ms, MB factor = 2, acceleration factor = 2, 48 interleaved slices, phase encoding anterior to posterior, transversal orientation, slice thickness = 2.7mm, voxel size = 2.7mm x 2.7mm, distance factor = 0%, flip angle = 70 degrees).

2.4 MRI pre-processing

MRI scans were processed using AFNI, SPM, Freesurfer, and SUMA (Cox, 1996; Saad & Reynolds, 2012). Before pre-processing, the origin of all MRI data was manually reset to the anterior commissure (AC) using SPM display (SPM12). For the structural images, the AC origin was defined in the T1w image and the transformation applied to the T2w image, to ensure that they were aligned to each other. For the EPIs in each run, the origin was defined in the first TR of the second echo of the first run, and applied to all other echoes and scans in the same session. This procedure was to facilitate co-registration of EPIs and structural scans across the two sessions, given that the origins of the acquired data could be far apart in space.

During pre-processing, dummy scans were first removed from the start of each run (AFNI 3dTcat). Large deviations in signal were removed (3dDespike) followed by slice time correction (3dTshift), aligning each slice with a time offset of zero. The skull was removed from the first echo 1 scan (TE = 14.6ms) and used to create a brain mask (3dSkullStrip and 3dAutomask), as this echo contained the most signal. The first echo 2 scan (TE = 35.79ms) was used as a base for motion correction and registration with the T1 structural scan (3dbucket), as this echo was the most similar to standard EPI acquisition. For session 1 we used the pRF scans as a base and for session 2 we used the scene localiser scans. Motion parameters were estimated for the echo 2 scans (3dVolreg) and applied to the other echos (3dAllineate). After completing the standard pre-processing, the data were processed using Tedana (DuPre et al., 2021; Kundu et al., 2012, 2013) to denoise the multi-echo scans (version 0.0.12, using default options). The Tedana optimally combined and denoised output was then scaled so that each voxel had a mean value of 100 (3dTstat and 3dcalc, max value of 200). This means that the fMRI values can be interpreted as a percentage of the mean signal, and effect estimates can be viewed as percentage change from baseline (Chen et al., 2017). For the pRF scans, an average was then calculated across runs to leave a single time series for further analysis.

The session 1 structural scans were aligned to the functional data collected in session 1 (align_epi_anat) with a multi-cost function (including LPC, MNI, and LPA) and manually checked for accuracy. For all subjects we used the default LPC method output (Localised Pearson Correlation), as this worked sufficiently well. Functional data collected in session 2 was aligned to the session 1 structural to ensure that all functional data had the same alignment. Freesurfer reconstructions were

estimated using both the T1 and T2 scans (recon-all) from session 1, and the output used to create surfaces readable in SUMA (SUMA_Make_Spec_FS). The SUMA structural was then aligned to the session 1 experimental structural to ensure alignment with the functional images (SUMA_AlignToExperiment). Surface based analyses were conducted using the SUMA standard cortical surface (std.141).

2.5 EEG analysis

2.5.1 Multivariate timeseries decoding

Decoding analyses were performed using CoSMoMVPA (Oosterhof et al., 2016) and custom MATLAB scripts. A linear support vector machine (libSVM) classifier was trained to distinguish between all individual stimuli (1-12) across all channels and timepoints. For category-level decoding, we took an average of all across-category decoding timeseries (e.g. person1 versus place1 etc.). For stimulus-level decoding, we took an average of all within-category decoding timeseries (e.g. person1 versus person2 etc.). We used a leave-one-pseudotrial-out cross-validation approach, where pseudotrials were created using chunks of trials that were averaged together before classification. We chose to use 10 trials per pseudotrial for category level decoding, and 5 trials for stimuli level decoding. This was based on the total number of trials available for each condition, and on previous research evaluating the trade-off between reduced signal-to-noise provided by averaging with the subsequent increases in between-subject variation (Grootswagers et al., 2016; Scrivener et al., 2023). To ensure that the results were not dependent on a specific division of trials, we repeated the random allocation of single trials to pseudotrial chunks 50 times per subject and took the average across all iterations. The choice of 50 iterations was to balance sensitivity with computation time. For each subject, we used the average across these 50 iterations. Classification was run using a moving time-window of 3 samples to increase the number of datapoints available to the classifier. To assess the stability of category representation across time (person or place), we also ran a time-generalised analysis (King & Dehaene, 2014). Here the classifier was trained and tested across all pairs of time points. Given the increased computation required for time-generalised analysis, we only ran 10 iterations of the chunking procedure. Time points with decoding significantly above chance were assessed using Bayesian t-test calculated using the 'bayesFactor Toolbox' in MATLAB (https://github.com/klabhub/bayesFactor). This was implemented using a half-Cauchy prior with a default width of 0.707 and a range of 0.5 to increase the detection of small effects under the null hypothesis (Kass & Raftery, 1995; Morey & Wagenmakers, 2014; Rouder et al., 2012; Teichmann et al., 2022).

2.6 MRI analysis

2.6.1 Defining regions of interest

Population receptive fields were estimated using AFNI's non-linear fitting algorithm (3dNLfim) and the GAM basis function (Silson et al., 2015). The outputs were used to delineate subject-specific retinotopic maps which were drawn manually on the SUMA standard (std.141) surface using the polar angle and eccentricity estimates (in this work we use V1 only). These ROIs were then converted into 1D files (ROI2dataset) to facilitate future node selection.

To localise the PPA and FFA, a general linear model was estimated for the scene/face localiser scans using a block design with a 16 second GAM basis function (GAM: 8.6, .547, 16, 3dDeconvolve and 3dREMLfit). The output of the model was then projected onto the SUMA cortical surface (3dVol2Surf), and smoothed with a FWHM of 2mm (SurfSmooth). The PPA and FFA ROIs were drawn manually for each subject on the surface (SUMA draw ROI) after thresholding the contrast of scenes versus faces at $t > 3.5$ ($p<0.0001$, uncorrected). PPA and FFA ROIs were divided into anterior and posterior portions, either by identifying two separate peaks in the t-maps, or by dividing a single large cluster equally along the posterior-anterior axis (Silson et al., 2019; Baldassano et al., 2016). Not all ROIs were identifiable across all subjects (left hemisphere: FFA1 = 10, FFA2 = 12, pPPA = 9, aPPA = 12; right hemisphere: FFA1 = 10, FFA2 = 11, pPPA = 12, aPPA = 12). Our surface-based ROIs were also converted to volume masks by projecting them from the surface to

the volume, using the same parameters that were used to project the localiser data to the surface (3dSurf2Vol).

Medial parietal ROIs, also defined on the surface, were taken from previous work to avoid issues of circularity in our analysis (Kriegeskorte et al., 2009; Silson et al., 2019). Here, two ROIs on the medial parietal surface with increased BOLD activation during recall of place stimuli were named as Places1 and Places2. A further two ROIs with greater activation for people stimuli were named People1 and People2 (Silson et al., 2019).

2.6.2 Recall univariate activation

The activity associated with each stimulus in the recall scans was deconvolved using a BLOCK basis function (BLOCK(9.5,1), 3dDeconvolve and 3dREMLfit) aligned to the onset of the recall period. Estimates of subject motion were included as regressors of no interest. We ran separate stimuli and category level deconvolutions, producing either one beta value per category (people or place), or one per stimulus (stimuli 1 to 12). In version one of the stimulus level deconvolutions, the six runs were concatenated and each stimulus regressor included 12 onsets (two from each run). The output of the model was projected onto the SUMA standard cortical surface (3dVol2Surf), and the data within each subject specific ROI was extracted for further analysis (pPPA, aPPA, FFA1, FFA2, Places1, People1, Places2, People2). In a second version of the deconvolutions, the runs were modelled separately so that we could check for an effect of run, and to facilitate a leave-one-run-out cross-validation approach for decoding. Group level analysis was conducted on the surface data using a linear mixed model (3dLME). We included subject as a random effect and modelled the main effect of category (people vs. places), the main effect of run (runs 1 to 6), and an interaction between category and run effects.

2.6.3 ROI univariate analysis

We entered the averaged beta values per ROI from the category deconvolution into a linear mixed model with factors Region and Category (people and places), keeping the medial parietal and ventral temporal regions separate (R studio package v1.3, lme4 v27.1, lmerTest v3.1). As there were no effects of hemisphere in any of our analyses, we averaged across the left and right hemispheres. Subject was entered as a random factor. As the timeseries were scaled, the beta values can be interpreted as BOLD percent signal change relative to the voxel baseline (Chen et al., 2017). We also extracted the average contrast values for each ROI (places versus people). We used both Frequentist and Bayeisan t-tests to identify category selectivity across our ROIs; i.e. contrast values significantly different from zero.

2.6.4 ROI decoding analysis

ROI decoding using the beta values was implemented in CoSMoMVPA (Oosterhof et al., 2016) using a leave-one-run-out cross-validation approach and a support vector machine classifier (libSVM). To mimic the ROI univariate analysis, we first decoded the stimulus category across our ROIs (person or place). Next, we decoded the stimulus identity separately within each category (person 1 or person 2 etc.). For this we took the average decoding across all pairs of stimuli, meaning that chance level decoding was 50%. Decoding analyses were run in the volume and in each hemisphere separately. We entered the average decoding value across hemispheres for each ROI and each subject into a linear mixed model with factors Region and Category, keeping the medial parietal and ventral temporal regions separate (R studio package v1.3, lme4 v27.1, lmerTest v3.1). Subject was entered as a random factor. Post-hoc tests for decoding above chance within each ROI were assessed using both Frequentist and Bayeisan one-way t-tests.

2.6.5 EEG-fMRI fusion

To facilitate EEG-fMRI fusion, RDMs were also constructed in the EEG recall data. First, an average was calculated across all trials for each of the 12 stimuli. At each time point, RDMs were then constructed using the activity across all sensors for each stimulus (1 – Pearson correlation r values).

These 12x12 RDMs had the same order and structure as the corresponding fMRI RDMs for each subject (people then places). For EEG-fMRI fusion, we then correlated (Spearman's rho) the EEG RDMs at each time point with the corresponding fMRI RDM for each ROI in each hemisphere (pPPA, aPPA, FFA1, FFA2, Places1, Places2, People1, People2, & V1). We then averaged the fusion correlations across hemispheres for each subject. Time points with fusion correlations above zero were assessed using Bayesian t-tests as in the EEG decoding analyses.

3. Results

3.1 Vividness scores

We asked subjects to rate the average vividness of their mental imagery for each stimulus during each session (EEG, fMRI). This was to assess the range in vividness scores across subjects, and to identify any differences across stimulus categories or sessions. For each subject, we computed the average overall vividness score for the recall of each stimulus during each session. During the EEG session, the mean vividness for people was 7.08 out of a maximum of 10 (SD = 1.11) and 6.74 for places (SD = 1.52). There was no significant difference between categories, $t(11) = 0.881$, $p = 0.40$. During the MRI session, the mean vividness for people was 6.88 out of 10 (SD = 1.13) and 6.67 for places (SD = 1.36). Again, there was no significant difference between categories, $t(11) = 0.453$, $p = 0.659$. The overall average vividness score was 6.91 for the EEG session and 6.77 for the MRI session. There was no significant difference between sessions, $t(11) = 0.828$, $p = 0.425$. The VVIQ scores ranged from 38 to 74, with an average score of 58.58.

3.2 MRI: whole-brain univariate contrast

First, we ran a whole-brain linear mixed model to quantify the univariate contrast of people versus places during mental imagery. At the group level, we identified significant clusters of activation in the medial parietal cortex, corresponding to people and place areas (Silson et al., 2019), as well as the lateral and ventral place memory areas (Steel et al., 2021, 2024), PPA, FFA, and the frontal eye fields **(Figure 2)**. Replicating previous findings, we found anterior shifts in the location of the lateral and ventral place memory areas in relation to their perceptual counterparts; the occipital place area (OPA) and PPA, respectively (Steel et al., 2021; Bainbridge at al., 2021). We found a smaller cluster of activation in the fusiform face area for recall compared to the perceptual localiser scans, but unlike for places, perception and recall of people activated very similar cortical locations. This mirrors recent findings reporting a lack of consistent shift in location for recall of people (Chen et al., 2024). We also found a main effect of run in the peripheral visual cortex, where activity decreased across time, but no voxels where the interaction between run and category was significant.

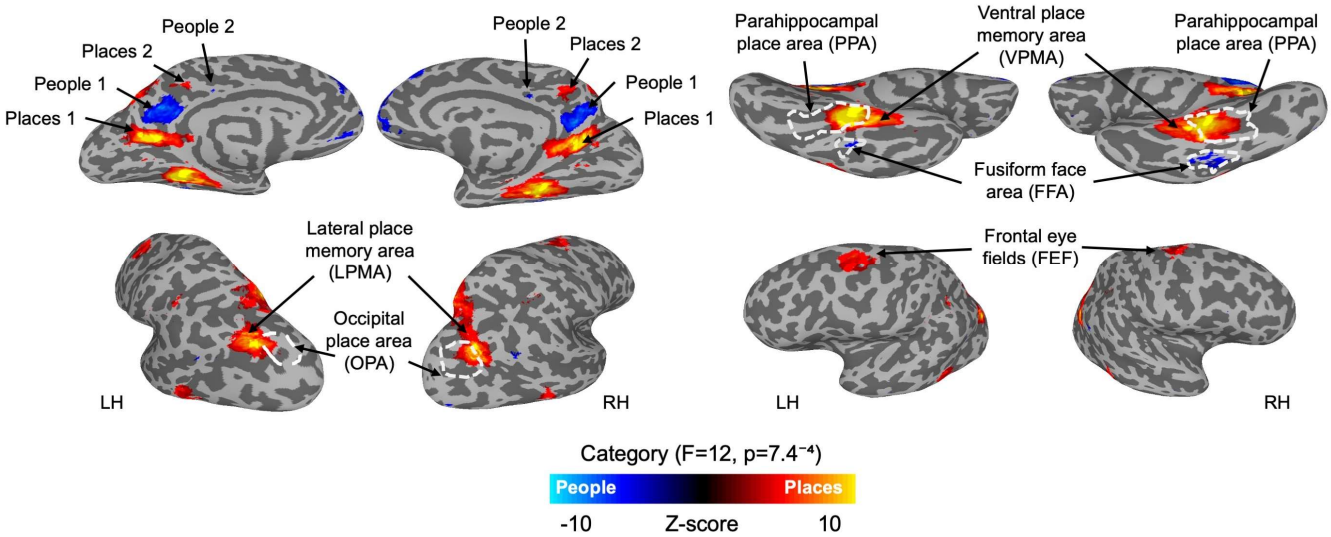

Figure 2: Group level results from a whole-brain linear mixed model contrast for the recall of place versus people stimuli. The results are displayed on partially-inflated cortical surfaces of a sample subject, separately for the left and right hemispheres. Activation was thresholded with a minimum F value of 12 for the main effect of category ($p < 7.4^{-4}$). Regions with greater activation for places are displayed in hot colours, and those with greater activation for people displayed in cold colours. Additional regions outlined in white dashed lines correspond to the group level activation from the perceptual scene/face localiser scans, contrasting perception of places versus people. All regions were identifiable in both hemispheres.

3.3 MRI: region of interest univariate analysis

Next, we extracted the univariate GLM beta values for people and place stimuli across medial parietal and ventral temporal ROIs. The medial parietal ROIs were taken from a prior dataset (Silson et al., 2019), and the ventral temporal regions were identified using independent scene-face localiser scans acquired on the same day as the recall **(see Methods)**. We entered the average beta values (across hemispheres) within each ROI for each subject into a linear mixed model with factors Region (People1, People2, Places1, Places2), and Category (people and places), keeping the medial parietal and ventral temporal regions separate. Across both sets of ROIs, activation was largest for the preferred category. In the medial parietal regions, there was a significant main effect of ROI ($F(3) = 10.23$, $p = 9.50\text{-}10$), and an interaction between Category and ROI ($F(1, 3) = 26.89$, $p = 5.38\text{-}12$). In the ventral temporal regions (FFA1, FFA2, pPPA, aPPA), there was a significant main effect of ROI ($F(3) = 19.98$, $p = 1.26\text{-}9$), a main effect of Category ($F(1) = 8.63$, $p = 0.004$), and an interaction between Category and ROI ($F(1, 3) = 24.26$, $p = 4.41\text{-}11$). See supplementary material (S4) for results from the hippocampus and amygdala.

As a post-hoc analysis, we also tested whether the contrast values for places versus people stimuli during recall were significantly different from zero **(Figure 3)**. The contrast was significantly greater than zero in Places 1, Places 2, aPPA, and pPPA ROIs, suggesting greater activation for places. The opposite was found in People 1, People 2, FFA1, and FFA2 ROIs, suggesting greater activation during the recall of people (results from frequentist and Bayesian t-tests can be found in supplemental material S2). Thus, even regions typically considered as visual (FFA and PPA) demonstrated category selectivity during recall of familiar stimuli.

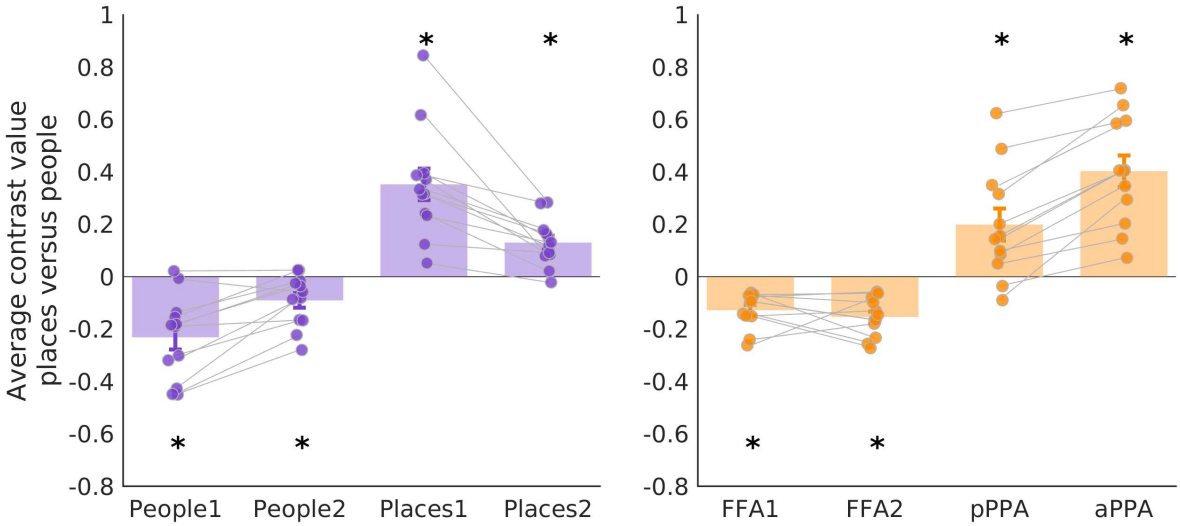

Figure 3: Average fMRI contrast values for places versus people recall in the ventral temporal and medial parietal regions of interest, averaged across hemispheres. All contrast values were significantly different from zero, with values above zero in place selective regions and below zero in face selective regions (* = $p < .05$). Each data point represents a single subject.

3.4 MRI: region of interest multivariate decoding

Our univariate ROI analysis confirmed that category selective regions in the medial parietal and ventral temporal cortex had greater response amplitudes during the recall of their preferred category. However, differences in amplitude are not necessarily informative about the information content of these regions. First, we used multivariate pattern analysis to test whether we could also successfully decode the category of imagined stimuli across our ROIs (person or place). As expected, we were able to decode the category significantly above chance (50%) in all of our ROIs (see supplemental material S5 for Frequentist and Bayesian t-test results).

Next, we asked whether we could successfully decode the individual stimuli from each other within each category. We entered the average decoding (across hemispheres) within each ROI into two linear mixed models with factors Region and Category, keeping the medial parietal and ventral temporal regions separate for consistency with the statistical approach above. Our prediction was that we should be able to decode people stimuli and place stimuli from people- and place-selective regions, respectively.

In line with this prediction, in the ventral temporal regions there was only a significant interaction between Category and ROI ($F(1, 3) = 6.37$, $p = 0.0006$), as decoding was found for each region's preferred category (see **Figure 4**). In the medial parietal regions, the interaction between Category and ROI was not significant ($F(1, 3) = 2.03$, $p = 0.11$). Instead, we found only significant main effects of ROI ($F(3) = 3.15$, $p = 0.03$), and Category ($F(1) = 8.13$, $p = 0.005$). In People1, we were able to decode stimuli belonging to both stimulus categories, and in People2 and Places2 we could decode stimuli in the non-preferred category. Therefore, despite the clear category preferences in the medial parietal ROIs (indexed by differences in BOLD amplitude), the representation of different stimuli within these categories may be more complex (results from one-way t-tests against chance level decoding can be found in supplemental material S5). The ability to decode stimuli across different categories suggests that there may be information pertaining to both types of stimuli within these ROIs during recall, which requires further investigation (see Discussion).

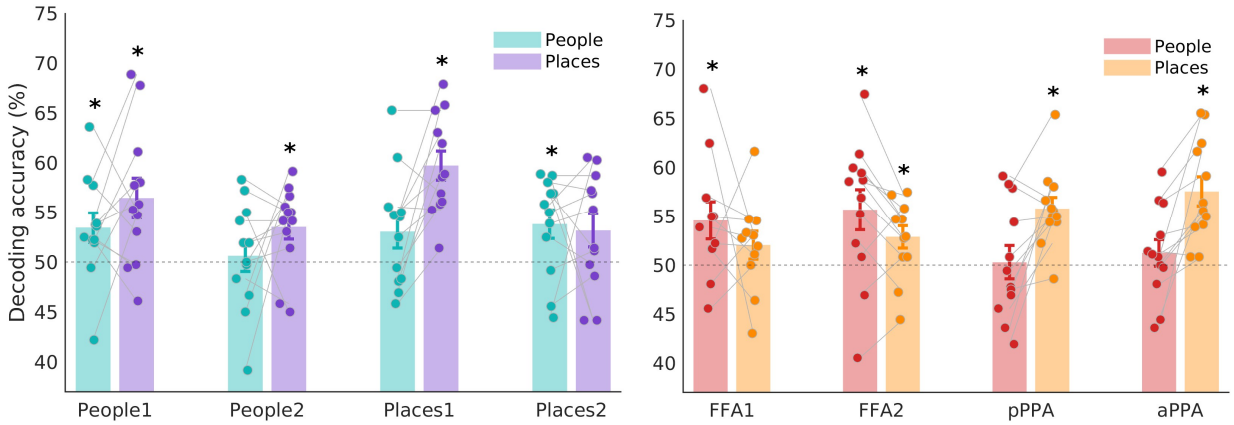

Figure 4: Average fMRI multivariate decoding of individual people and place stimuli in the medial parietal and visual temporal regions of interest, averaged over left and right hemispheres. Chance level decoding was 50% as we took the average decoding across all pairs of stimuli (indicated by the dashed line). Asterisks indicate stimulus decoding significantly above chance accuracy ($p < .05$). Each data point represents a single subject.

To summarise the fMRI results, we successfully identified people and place memory regions within medial parietal cortex in a whole brain univariate analysis (Silson et al., 2019), and confirmed that activation within these regions was greatest during imagery of their preferred category. We also demonstrated that regions within ventral temporal cortex that are typically considered visual (FFA1/2 and p/aPPA) were also active during recall, and had the expected category selectivity. Using multivariate decoding analysis, we tested the hypothesis that we would be able to successfully decode individual stimuli within the preferred category for each ROI. This was largely the case in the ventral temporal ROIs (with the exception of FFA2 where both categories could be decoded), but a

more complex pattern emerged in the medial parietal cortex; these regions represented information during imagery that was not restricted to their preferred category. In comparison, we were only able to decode place stimuli in foveal and peripheral V1, which we hypothesise may be due to the increased visual variability in the to-be-recalled place stimuli compared to the people (see supplemental material S6). Overall, our MRI results confirmed the roles of our key ROIs in visual mental imagery, and hint that there might be shifts in the information content across medial parietal, ventral temporal, and primary visual ROIs. Next, we used the EEG data to investigate the timecourse of this recall information, and identify timepoints where we could successfully decode both the category and identity of mental images across the recall period.

3.5 EEG: multivariate decoding through time

Multivariate pattern analysis (MVPA) was first used to identify timepoints where the category of personally familiar stimuli (people or places) could be decoded above chance accuracy across all EEG electrodes. Recall category was highly decodable for the first 700ms of the recall period, suggesting an early difference in the activity associated with mental imagery for these categories **(Figure 5, left panel)**. This early peak in decoding was followed by sustained but weaker category information with a lower accuracy and variable Bayesian evidence. Next, we took the average decoding within each category (6 stimuli in each category) **(Figure 5, middle and right panels)**. We present the average across all comparisons so that all plots have chance level decoding of 50%. Here, individually recalled stimuli were decodable from each other for a shorter time period following recall onset, lasting roughly 100ms. However, there was a second peak in decoding for both categories of stimuli at ~700ms into the recall period for places and ~1100ms for people.

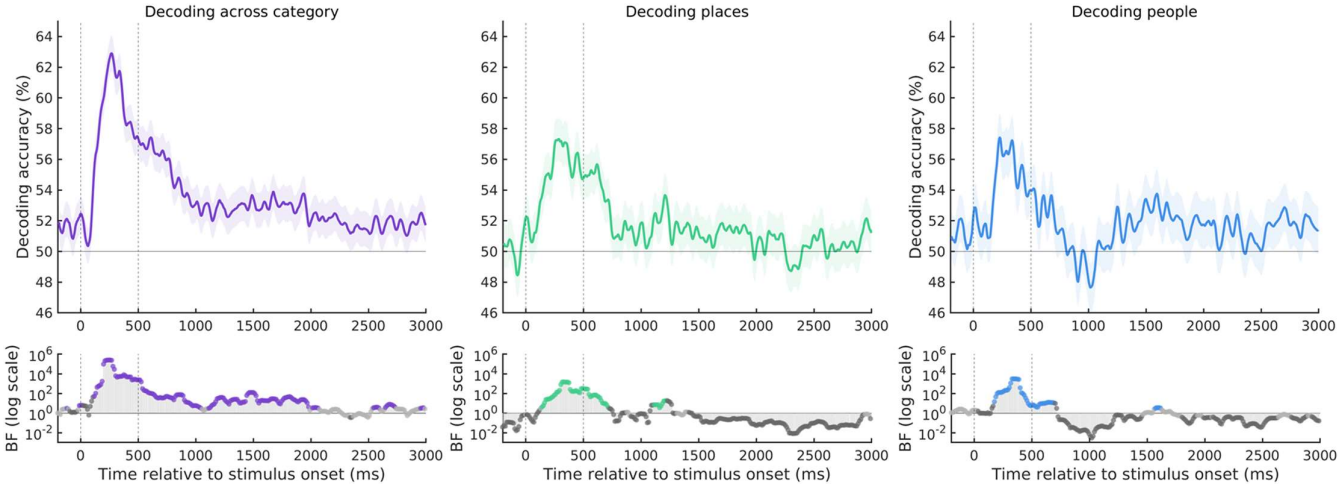

Figure 5: Top: average multivariate EEG timeseries decoding during recall. Left: across category decoding of people versus place stimuli. Middle: within category decoding of places. Right: within category decoding of people. The cue was presented during the first 500ms, followed by 2500ms of recall. Here we plot both average decoding across subjects and the standard error around the mean. Chance level decoding was 50% for all analyses as we took the average decoding across all pairs of stimuli. Bottom: Bayes Factors are plotted on a log scale. Timepoints with Bayes Factors that were greater than 3 are highlighted by the coloured points, suggesting that the alternative hypothesis of decoding above chance level was at least three times more likely than the null hypothesis. Dark grey points indicate Bayes Evidence in favour of the null hypothesis that was greater than 3.

We then used time-generalised decoding to examine the stability of recall representations across time. Increased decoding along the diagonal indicates that the classifier can only generalise to its neighbouring timepoints and that activity is changing over time. In comparison, off-diagonal decoding suggests sustained representations that are more stable through time (King et al., 2014). Here category decoding (people or places) throughout the trial successfully generalised to both an early (~600-1000ms) and late test window (~1600-2100ms), indicating two peaks of sustained recall representation throughout the trial **(Figure 6)**. Importantly, there was little evidence for the cue period

time-generalising to the recall period, reassuring us that category decoding was not driven solely by the presence of the cue.

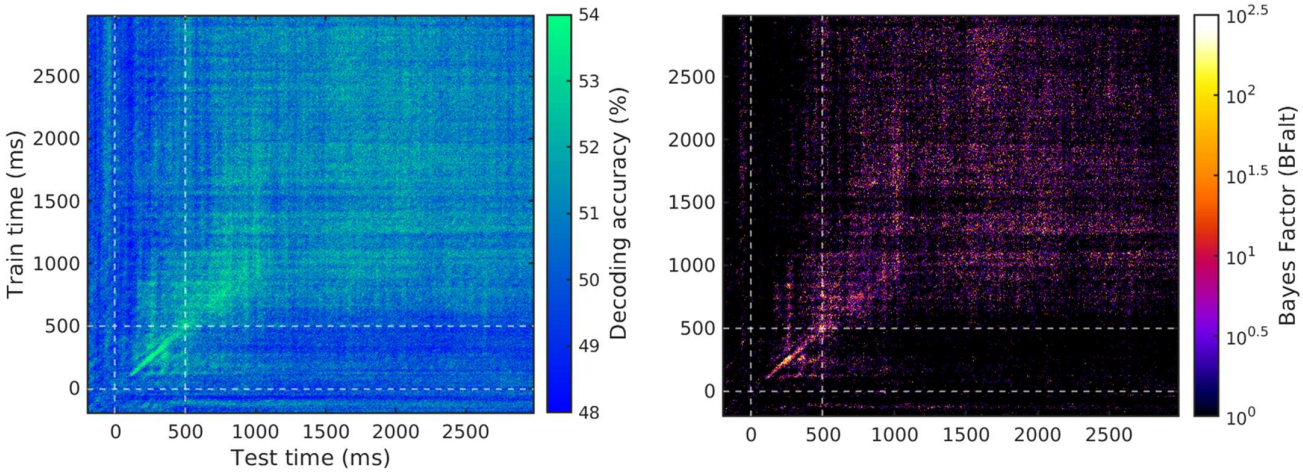

Figure 6: Average time-generalised EEG decoding during recall. Left: Category level time-generalised decoding of people versus place stimuli across all timepoints within the recall period. The cue was presented during the first 500ms, followed by 2500ms of recall. Right: Bayes Factors indicating the evidence for the alternative hypothesis, plotted on a log10 scale. Brighter colours indicate decoding above chance level (50%).

To summarise the EEG analysis, we were able to successfully decode category (places versus people) throughout the recall period. Decoding of within-category stimuli was weaker, possibly due to the reduced number of trials available for classification and the subtler differences between stimuli. However, we were still able to decode both people and place stimuli during two distinct time windows. Stimuli in both categories were decodable during the first ~100ms of the recall period, followed by a re-emergence of decodable activation at ~700ms into the recall period for places and ~1100ms for people. Time-generalised decoding indicated that the imagery was sustained throughout the recall period, with relatively stable representations that appeared to form early and late clusters. These results confirmed that information pertaining to both the category and stimulus identity was captured in the EEG data, but provide no spatial information. Next, we combined the EEG and fMRI data to identify when in time the representational structures present in the fMRI ROIs matched those captured in the EEG.

3.6 EEG and fMRI fusion

Lastly, we calculated the EEG-fMRI fusion time series by correlating the fMRI RDMs in each ROI with the EEG RDMs at each time point. As shown in **Figure 7**, all regions had at least some timepoints with strong Bayesian evidence for a correlation between the EEG and fMRI RDMs (although the evidence for fusion in People2 and FFA2 was weaker). In the medial parietal ROIs, evidence for clusters of significant fusion (BF > 3) occurred first in Places2 (500ms) and Places1 (539ms), followed by People1 (691ms). In People2, there was only a small window with significant fusion (at 852ms). In V1, sustained fusion started later at 1371ms (with a small window at 984ms lasting only 12ms). This fusion pattern may indicate a feedback sweep from place selective regions within the medial parietal cortex to V1.

In the ventral temporal ROIs, a similar pattern emerged. Fusion arose first in aPPA (500ms), followed by pPPA (586ms), FFA1 (883ms). FFA2 had a few small windows with BF > 3 (676ms, 1383ms, and 2207ms), but little evidence for any sustained fusion over time. In general, this finding mirrors the result above with fusion peaking first in place selective regions, followed by face selective regions, and then early visual cortex.

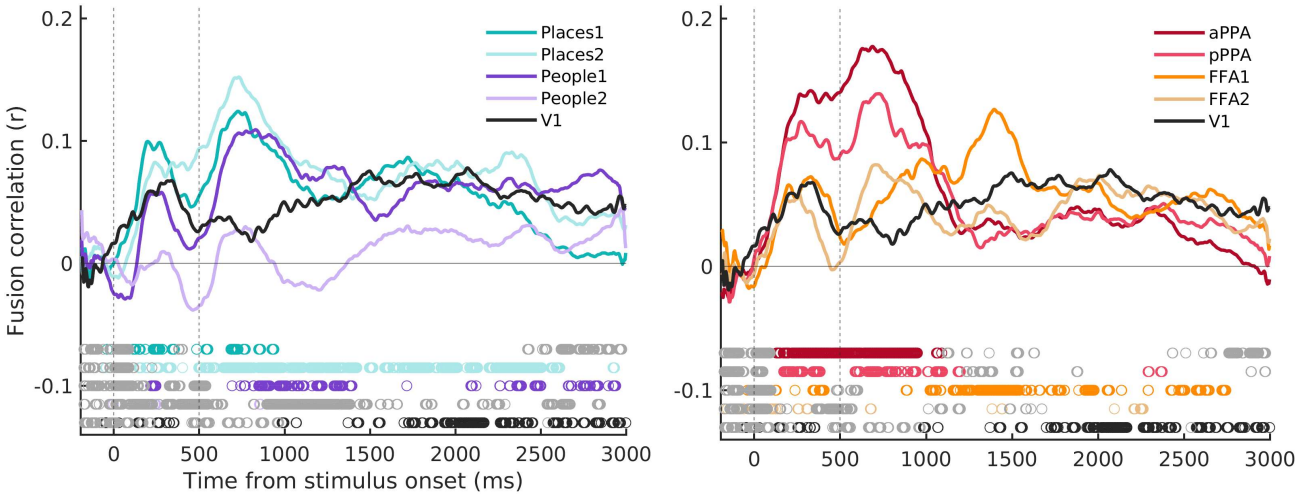

Figure 7: Average EEG-fMRI fusion timeseries for the ventral temporal and medial parietal regions of interest, averaged over hemispheres. The V1 timeseries (plotted in black) is replicated across plots to aid comparison of the significant time points across ROIs. Timepoints with Bayes Factors that were greater than 3 are indicated by the coloured circles along the bottom bar, suggesting that the alternative hypothesis of a fusion correlation is at least three times more likely than the null hypothesis. Grey circles indicate Bayes Evidence in favour of the null hypothesis that was greater than 3.

4. Discussion

Here we used an EEG-fMRI fusion approach to investigate the involvement of medial parietal and ventral temporal category-selective regions during mental imagery. We also compared the activity within these higher-level regions to that in V1. We hypothesised that we may find differences in the fusion time courses for medial parietal and ventral temporal ROIs during mental imagery, assuming that they may play distinct roles. If mental imagery is supported by a mechanism of perception in reverse, then we may expect to see activity peak earlier in medial parietal cortex than ventral temporal given both their anatomical locations on the cortex (MPC regions are more anterior) and the prominent role that VTC regions play in visual perception. However, we did not find any clear differences in the fusion timecourse across these ROIs. This temporal similarity between two spatially distinct cortical regions mirrors recent electrocorticography (ECoG) data during recognition of famous people and places, where medial parietal and medial temporal regions had similar timecourses of activation (Woolnough et al., 2020). One possible explanation for this is that both sets of regions received input from elsewhere, driving all regions to represent visual imagery simultaneously. Candidate regions include those identified in our whole-brain univariate analysis, for example the lateral place memory area. Further, the recruitment of subcortical brain structures such as the hippocampus and amygdala are known to support memory recall (Silson et al, 2019), and could also be driving the activity across our main ROIs. Prior work has suggested that the hippocampus (in particular the anterior portion, anterior PPA), and portions of the medial parietal cortex form a distinct scene memory network that links visual processing with long-term memory and additional scene context (Baldessano et al., 2016). Here we replicated findings of category-selectivity in the hippocampus and amygdala (reported in supplementary material S3), which had greater activity during the recall of place and people stimuli respectively.

Instead of a difference in fusion correlation driven by ROI location, we found a difference driven by the category; in both cases, regions selective for places (Places1/2, p/aPPA) generally peaked earlier than those selective for people (People1/2, FFA1/2). This echoed the results found in the EEG timeseries analysis, where decoding of individual place stimuli peaked earlier in time than for people. It is possible that, in this sample of subjects, mental imagery of place stimuli was quicker to emerge than people. Although we did not formally assess this, some subjects did report that place

stimuli took less time to visualise in their end of session questionnaires, despite there being no overall difference in reported subjective vividness. Future behavioural investigation is required to establish if this is a consistent finding and the extent of any between-subject variation.

Using multivariate pattern analysis, we were able to significantly decode not just category-level information, but also the individual within-category stimuli being visualised during the recall period. We predicted that we would be able to decode imagery of individual people in regions that are 'selective' for face stimuli (i.e. FFA1/2 and People1/2) in our fMRI data. Similarly, we expected to decode individual place stimuli in regions selective for place stimuli (i.e. p/aPPA and Places1/2). This prediction was largely confirmed in the ventral temporal ROIs, where we found a significant interaction between ROI and category. Post-hoc tests for significant decoding within each ROI identified decoding that was mostly confined to the preferred stimulus category. However, there were exceptions to this, with decoding of both categories within FFA2. The contradiction to our hypothesis was more pronounced in the medial parietal regions, where we only found a main effect of category with no significant ROI interaction. While there was clear category-selectivity in the mean univariate activation for people and place selective ROIs, the information content within medial parietal ROIs was not solely related to their preferred category during mental imagery.

Our fMRI ROI stimulus decoding, together with the lack of clear differences in the fusion onset timings across medial parietal and ventral temporal cortex, leaves an open question about what information is represented across high-level medial parietal ROIs during mental imagery. Rather than representing similar information at different points in time, MPC regions may instead contain different information about the stimuli being visualised and have a different representational structure. Stimulus category was not the only difference between the stimuli provided by our subjects; other dimensions such as familiarity, emotional relevance, semantic associations etc. may also structure the representations during mental imagery within medial parietal cortex specifically. Indeed, similar regions in medial parietal cortex have been associated with the semantic structure of people and place images (Koslov et al., 2024) and varying spatial scales (Peer et al., 2019) to name only a few (for a review see Foster et al., 2023). Future work could focus on sampling a larger range of stimuli across other possible dimensions and evaluate how this links to stimulus decoding in medial parietal cortex.

Using EEG-fMRI fusion, we found that the representations of people and place imagery in V1 correlated with EEG signals at a later timepoint than either high-level category-selective regions in medial parietal or ventral temporal cortex. This suggests that the activity in V1 may be related to feedback or recurrent processing that emerged later in time than medial parietal and ventral temporal ROIs. Does this finding mean that recall was visual in nature? Using multivariate pattern analysis in the fMRI data, we were able to decode the individual place stimuli during imagery in both central and peripheral V1. This indicates that there were quantifiably distinct patterns of activation associated with imagery of each place in V1, which could suggest 'perception-like' imagery and the mechanism of perception in reverse. However, the classifier could use any number of features to distinguish between stimuli, and decodable signals within V1 could reflect any number of processes that are not necessarily identical to visual perception (Nesalaris et al., 2015; Favilla at al., 2022; Pace et al., 2023).

Unlike the place stimuli, we were not able to significantly decode people stimuli, either in central or peripheral V1. One possible explanation for this is that the patterns in V1 during imagery of places were more varied in structure than for people, resulting in better decoding for these stimuli. Indeed, the visual distinctiveness of the place stimuli was greater than that of the people when we ran the subjects' images through an LGN visual model, which may explain the differences in V1 decoding across categories. However, this explanation makes the explicit assumption that recall activity in V1 is 'perception-like', and that subjects were faithfully recreating the images that they provided during recall (which may not have always been the case).

Another question is whether V1 activity is *necessary* for mental imagery (Spagna et al., 2024). Patient findings indicate that mental imagery is possible despite visual cortex damage, strongly suggesting that V1 is not a critical region in supporting imagery (Della Salla & Spinnler, 1988; Hajhajate at al., 2022). Given the nature of our analysis and the timing of the activity in V1, we do not make any claims about causality; our results simply highlight differences in the timing of stimuli

representations in medial parietal, ventral temporal, and primary visual cortex during mental imagery in healthy individuals.

In summary, we used EEG-fMRI fusion to investigate the time course of mental imagery representations within category-selective regions in medial parietal, ventral temporal cortex, and V1. The fusion correlations peaked early on during recall and did not clearly differ across medial parietal and ventral temporal ROIs. However, place-selective regions generally peaked earlier than people-selective regions, mirroring EEG stimulus decoding results where individual places were decodable earlier in time than people. In V1, fusion correlations occurred later during the recall period, possibly reflecting the top-down progression of mental imagery from category selective regions to primary visual cortex. Within the ventral temporal ROIs, fMRI decoding was highest for the preferred category. In contrast, a more complex pattern of decoding was identified in the medial parietal cortex. These regions represented information during imagery that was not restricted to their preferred category, prompting future investigation to establish how these regions support the generation of mental images. Overall, our results confirm the roles of category-selective cortex during visual mental imagery, and evidence shifts in information content across medial parietal, ventral temporal, and primary visual cortex during this complex and dynamic process. Understanding the extent to which medial parietal regions represent category information according to different dimensions (visual, semantic, emotional) remains an avenue for future research.

Acknowledgements

This work was funded by the Biotechnology and Biological Sciences Research Council (BBSRC) BB/V003917/1, and the School of Philosophy, Psychology and Language Sciences at the University of Edinburgh (UoE). The MRI data was collected at the Edinburgh Imaging Facility. For the purpose of open access, the author has applied a creative commons attribution (CC BY) licence to any author accepted manuscript version arising.

Data availability

Data will be made available on the Open Science Framework at publication: https://osf.io/ypk6c/

CRediT contributor roles

CLS: Conceptualisation, funding acquisition (UoE), investigation, data curation, formal analysis, visualisation, writing - original draft, writing - reviewing and editing. **JAT:** writing - review and editing. **EHS:** Conceptualisation, funding acquisition (BBSRC and UoE), formal analysis, writing - review and editing, supervision.

References

Albers, A. M., Kok, P., Toni, I., Dijkerman, H. C., & de Lange, F. P. (2013). Shared representations for working memory and mental imagery in early visual cortex. *Current Biology: CB*, *23*(15), 1427–1431. https://doi.org/10.1016/j.cub.2013.05.065

Bainbridge, W. A., Hall, E. H., & Baker, C. I. (2021). Distinct Representational Structure and Localization for Visual Encoding and Recall during Visual Imagery. *Cerebral Cortex*, *31*(4), 1898–1913. https://doi.org/10.1093/cercor/bhaa329

Baldassano, C., Esteva, A., Fei-Fei, L., & Beck, D. M. (2016). Two Distinct Scene-Processing Networks Connecting Vision and Memory. *eNeuro*, *3*(5). https://doi.org/10.1523/ENEURO.0178-16.2016

Chen, G., Taylor, P. A., & Cox, R. W. (2017). Is the statistic value all we should care about in neuroimaging? *NeuroImage*, *147*, 952–959. https://doi.org/10.1016/j.neuroimage.2016.09.066

Chen, Y. Y., Areti, A., Yoshor, D., & Foster, B. L. (2024). Perception and memory reinstatement engage overlapping face selective regions within human ventral temporal cortex. *The Journal of Neuroscience*, e2180232024. https://doi.org/10.1523/JNEUROSCI.2180-23.2024

Cichy, R. M., & Oliva, A. (2020). A M/EEG-fMRI Fusion Primer: Resolving Human Brain Responses in Space and Time. *Neuron*. https://doi.org/10.1016/j.neuron.2020.07.001

Corriveau, A., Kidder, A., Teichmann, L., Wardle, S. G., & Baker, C. I. (2022). Sustained neural representations of personally familiar people and places during cued recall. *Cortex*. https://doi.org/10.1016/j.cortex.2022.08.014

Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research, an International Journal*, *29*(3), 162–173. https://doi.org/10.1006/cbmr.1996.0014

Della Sala, S., & Spinnler, H. (1988). Anton's (-Redlich-Babinski's) syndrome associated with Dide-Botcazo's syndrome: A case report of denial of cortical blindness and amnesia. *Schweizer Archiv Fur Neurologie Und Psychiatrie (Zurich, Switzerland*, *139*(1), 5–15.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Dijkstra, N., Bosch, S. E., & Gerven, M. A. J. van. (2017). Vividness of Visual Imagery Depends on the Neural Overlap with Perception in Visual Areas. *Journal of Neuroscience*, *37*(5), 1367–1373. https://doi.org/10.1523/JNEUROSCI.3022-16.2016

Dijkstra, N., Bosch, S. E., & van Gerven, M. A. J. (2019). Shared Neural Mechanisms of Visual Perception and Imagery. *Trends in Cognitive Sciences*, *23*(5), 423–434. https://doi.org/10.1016/j.tics.2019.02.004

Dumoulin, S. O., & Wandell, B. A. (2008). Population receptive field estimates in human visual cortex. *NeuroImage*, *39*(2), 647–660. https://doi.org/10.1016/j.neuroimage.2007.09.034

DuPre, E., Salo, T., Ahmed, Z., Bandettini, P. A., Bottenhorn, K. L., Caballero-Gaudes, C., Dowdle, L. T., Gonzalez-Castillo, J., Heunis, S., Kundu, P., Laird, A. R., Markello, R., Markiewicz, C. J., Moia, S., Staden, I., Teves, J. B., Uruñuela, E., Vaziri-Pashkam, M., Whitaker, K., & Handwerker, D. A. (2021). TE-dependent analysis of multi-echo fMRI with *tedana*. *Journal of Open Source Software*, *6*(66), 3669. https://doi.org/10.21105/joss.03669

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*(6676), 598–601. https://doi.org/10.1038/33402

Favila, S. E., Kuhl, B. A., & Winawer, J. (2022). Perception and memory have distinct spatial tuning properties in human visual cortex. *Nature Communications*, *13*(1), 5864. https://doi.org/10.1038/s41467-022-33161-8

Foster, B. L., Koslov, S. R., Aponik-Gremillion, L., Monko, M. E., Hayden, B. Y., & Heilbronner, S. R. (2023). A tripartite view of the posterior cingulate cortex. *Nature Reviews Neuroscience*, *24*(3), 173–189. https://doi.org/10.1038/s41583-022-00661-x

Groen, I. I. A., Ghebreab, S., Prins, H., Lamme, V. A. F., & Scholte, H. S. (2013). From Image Statistics to Scene Gist: Evoked Neural Activity Reveals Transition from Low-Level Natural Image Structure to Scene Category. *The Journal of Neuroscience*, *33*(48), 18814–18824. https://doi.org/10.1523/JNEUROSCI.3128-13.2013

Grootswagers, T., Wardle, S. G., & Carlson, T. A. (2016). Decoding Dynamic Brain Patterns from Evoked Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series Neuroimaging Data. *Journal of Cognitive Neuroscience*, *29*(4), 677–697. https://doi.org/10.1162/jocn_a_01068

Hajhajate, D., Kaufmann, B. C., Liu, J., Siuda-Krzywicka, K., & Bartolomeo, P. (2022). The connectional anatomy of visual mental imagery: Evidence from a patient with left occipito-temporal damage. *Brain Structure and Function*, *227*(9), 3075–3083. https://doi.org/10.1007/s00429-022-02505-x

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience*, *17*(11), 4302–4311.

Kass, R. E., & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, *90*(430), 773–795. https://doi.org/10.1080/01621459.1995.10476572

King, J.-R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: The temporal generalization method. *Trends in Cognitive Sciences*, *18*(4), 203–210. https://doi.org/10.1016/j.tics.2014.01.002

Koslov, S. R., Kable, J. W., & Foster, B. L. (2023). *Dissociable contributions of the medial parietal cortex to recognition memory* [Preprint]. Neuroscience. https://doi.org/10.1101/2023.09.12.557048

Kosslyn, S., & Thompson, W. (2003). When Is Early Visual Cortex Activated During Visual Mental Imagery? *Psychological Bulletin*, *129*, 723–746. https://doi.org/10.1037/0033-2909.129.5.723

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. F., & Baker, C. I. (2009). Circular analysis in systems neuroscience: The dangers of double dipping. *Nature Neuroscience*, *12*(5), 535–540. https://doi.org/10.1038/nn.2303

Kundu, P., Brenowitz, N. D., Voon, V., Worbe, Y., Vértes, P. E., Inati, S. J., Saad, Z. S., Bandettini, P. A., & Bullmore, E. T. (2013). Integrated strategy for improving functional connectivity mapping using multiecho fMRI. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(40), 16187–16192. https://doi.org/10.1073/pnas.1301725110

Kundu, P., Inati, S. J., Evans, J. W., Luh, W.-M., & Bandettini, P. A. (2012). Differentiating BOLD and non-BOLD signals in fMRI time series using multi-echo EPI. *NeuroImage*, *60*(3), 1759–1770. https://doi.org/10.1016/j.neuroimage.2011.12.028

Lee, S.-H., Kravitz, D. J., & Baker, C. I. (2012). Disentangling visual imagery and perception of real-world objects. *NeuroImage*, *59*(4), 4064–4073. https://doi.org/10.1016/j.neuroimage.2011.10.055

Levy, I., Hasson, U., Avidan, G., Hendler, T., & Malach, R. (2001). Center–periphery organization of human object areas. *Nature Neuroscience*, *4*(5), 533–539. https://doi.org/10.1038/87490

Morey, R. D., & Wagenmakers, E.-J. (2014). Simple relation between Bayesian order-restricted and point-null hypothesis tests. *Statistics & Probability Letters*, *92*, 121–124. https://doi.org/10.1016/j.spl.2014.05.010

Naselaris, T., Olman, C. A., Stansbury, D. E., Ugurbil, K., & Gallant, J. L. (2015). A voxel-wise encoding model for early visual areas decodes mental images of remembered scenes. *NeuroImage*, *105*, 215–228. https://doi.org/10.1016/j.neuroimage.2014.10.018

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014). A Toolbox for Representational Similarity Analysis. *PLoS Computational Biology*, *10*(4), e1003553. https://doi.org/10.1371/journal.pcbi.1003553

Oliva, A., & Torralba, A. (2006). Chapter 2 Building the gist of a scene: The role of global image features in recognition. In S. Martinez-Conde, S. L. Macknik, L. M. Martinez, J.-M. Alonso, & P. U. Tse (Eds.), *Progress in Brain Research* (Vol. 155, pp. 23–36). Elsevier. https://doi.org/10.1016/S0079-6123(06)55002-2

Oosterhof, N. N., Connolly, A. C., & Haxby, J. V. (2016). CoSMoMVPA: Multi-Modal Multivariate Pattern Analysis of Neuroimaging Data in Matlab/GNU Octave. *Frontiers in Neuroinformatics*, *10*, 27. https://doi.org/10.3389/fninf.2016.00027

Pace, T., Koenig-Robert, R., & Pearson, J. (2023). Different Mechanisms for Supporting Mental Imagery and Perceptual Representations: Modulation Versus Excitation. *Psychological Science*, *34*(11), 1229–1243. https://doi.org/10.1177/09567976231198435

Pearson, J. (2019). The human imagination: The cognitive neuroscience of visual mental imagery. *Nature Reviews Neuroscience*, *20*(10), Article 10. https://doi.org/10.1038/s41583-019-0202-9

Peer, M., Ron, Y., Monsa, R., & Arzy, S. (2019). Processing of different spatial scales in the human brain. *eLife*, *8*, e47492. https://doi.org/10.7554/eLife.47492

Ragni, F., Tucciarelli, R., Andersson, P., & Lingnau, A. (2020). Decoding stimulus identity in occipital, parietal and inferotemporal cortices during visual mental imagery. *Cortex*, *127*, 371–387. https://doi.org/10.1016/j.cortex.2020.02.020

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences*, *98*(2), 676–682. https://doi.org/10.1073/pnas.98.2.676

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*(5), 356–374. https://doi.org/10.1016/j.jmp.2012.08.001

Ryan, J. D., & Shen, K. (2020). The eyes are a window into memory. *Current Opinion in Behavioral Sciences*, *32*, 1–6. https://doi.org/10.1016/j.cobeha.2019.12.014

Saad, Z. S., & Reynolds, R. C. (2012). SUMA. *NeuroImage*, *62*(2), 768–773. https://doi.org/10.1016/j.neuroimage.2011.09.016

Scrivener, C. L. (2021). When Is Simultaneous Recording Necessary? A Guide for Researchers Considering Combined EEG-fMRI. *Frontiers in Neuroscience*, *15*. https://doi.org/10.3389/fnins.2021.636424

Scrivener, C. L., Grootswagers, T., & Woolgar, A. (2023). *Optimising analysis choices for multivariate decoding: Creating pseudotrials using trial averaging and resampling* (p. 2023.10.04.560678). bioRxiv. https://doi.org/10.1101/2023.10.04.560678

Silson, E. H., Chan, A. W.-Y., Reynolds, R. C., Kravitz, D. J., & Baker, C. I. (2015). A Retinotopic Basis for the Division of High-Level Scene Processing between Lateral and Ventral Human Occipitotemporal Cortex. *Journal of Neuroscience*, *35*(34), 11921–11935. https://doi.org/10.1523/JNEUROSCI.0137-15.2015

Silson, E. H., Steel, A., Kidder, A., Gilmore, A. W., & Baker, C. I. (2019). Distinct subdivisions of human medial parietal cortex support recollection of people and places. *eLife*, *8*, e47391. https://doi.org/10.7554/eLife.47391

Spagna, A., Heidenry, Z., Miselevich, M., Lambert, C., Eisenstadt, B. E., Tremblay, L., Liu, Z., Liu, J., & Bartolomeo, P. (2024). Visual mental imagery: Evidence for a heterarchical neural architecture. *Physics of Life Reviews*, *48*, 113–131. https://doi.org/10.1016/j.plrev.2023.12.012

Steel, A., Billings, M. M., Silson, E. H., & Robertson, C. E. (2021). A network linking scene perception and spatial memory systems in posterior cerebral cortex. *Nature Communications*, *12*(1), Article 1. https://doi.org/10.1038/s41467-021-22848-z

Steel, A., Silson, E. H., Garcia, B. D., & Robertson, C. E. (2024). A retinotopic code structures the interaction between perception and memory systems. *Nature Neuroscience*, *27*(2), 339–347. https://doi.org/10.1038/s41593-023-01512-3

Teichmann, L., Moerel, D., Baker, C., & Grootswagers, T. (2022). An Empirically Driven Guide on Using Bayes Factors for M/EEG Decoding. *Aperture Neuro*, *2*, 1–10. https://doi.org/10.52294/ApertureNeuro.2022.2.MAOC6465

Wandell, B. A., Dumoulin, S. O., & Brewer, A. A. (2007). Visual Field Maps in Human Cortex. *Neuron*, *56*(2), 366–383. https://doi.org/10.1016/j.neuron.2007.10.012

Woolnough, O., Rollo, P. S., Forseth, K. J., Kadipasaoglu, C. M., Ekstrom, A. D., & Tandon, N. (2020). Category Selectivity for Face and Scene Recognition in Human Medial Parietal Cortex. *Current Biology*, *30*(14), 2707-2715.e3. https://doi.org/10.1016/j.cub.2020.05.018

Xie, S., Kaiser, D., & Cichy, R. M. (2020). Visual Imagery and Perception Share Neural Representations in the Alpha Frequency Band. *Current Biology*, *30*(13), 2621-2627.e5. https://doi.org/10.1016/j.cub.2020.04.074

Zeman, A. (2024). Aphantasia and hyperphantasia: Exploring imagery vividness extremes. *Trends in Cognitive Sciences*, *0*(0). https://doi.org/10.1016/j.tics.2024.02.007