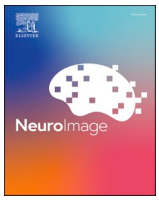
NeuroImage 311 (2025) 121171

Contents lists available at ScienceDirect

NeuroImage

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



Covariance-based decoding reveals a category-specific functional connectivity network for imagined visual objects

Francesco Mantegna [a,b,e,*], Emanuele Olivetti [c,e], Philipp Schwedhelm [d,e], Daniel Baldauf [e]

[a] Department of Psychology, New York University, New York, NY 10003, USA
[b] Department of Engineering Science, Oxford University, Oxford, Oxfordshire, United Kingdom
[c] NeuroInformatics Laboratory (NILab), Bruno Kessler Foundation (FBK), Mattarello, TN 38100, Italy
[d] Functional Imaging Laboratory, German Primate Center - Leibniz Institute for Primate Research, Goettingen, 37077, Germany
[e] CIMeC - Center for Mind and Brain Sciences, Mattarello, TN 38100, Italy

ARTICLE INFO

Keywords:
Imagery
Face
Place
Decoding
Covariance
Magnetoencephalography

ABSTRACT

The coordination of different brain regions is required for the visual imagery of complex objects (e.g., faces and places). Short-range connectivity within sensory areas is necessary to construct the mental image. Long-range connectivity between control and sensory areas is necessary to re-instantiate and maintain the mental image. While dynamic changes in functional connectivity are expected during visual imagery, it is unclear whether a category-specific network exists in which the strength and the spatial destination of the connections vary depending on the imagery target. In this magnetoencephalography study, we used a minimally constrained experimental paradigm wherein imagery categories were prompted using visual word cues only, and we decoded face versus place imagery based on their underlying functional connectivity patterns as estimated from the spatial covariance across brain regions. A subnetwork analysis further disentangled the contribution of different connections. The results show that face and place imagery can be decoded from both short-range and long-range connections. Overall, the results show that imagined object categories can be distinguished based on functional connectivity patterns observed in a category-specific network. Notably, functional connectivity estimates rely on purely endogenous brain signals suggesting that an external reference is not necessary to elicit such category-specific network dynamics.

1. Introduction

Our brain has a remarkable capacity to internally generate vivid mental representations in the complete absence of external sensory stimulation. Imagery is very useful whenever we need to process information that is not accessible in the present. For example, imagery allows us to re-instantiate information encountered in the past or to anticipate information that we will encounter in the future, without constantly requiring an external reference. In particular, visual imagery involves the internal generation of mental images (Pearson, 2019). We can generate rich, vivid, and detailed images in our mind's eye, which can contain precise color and shape information. For example, we can internally visualize a well-known place or a familiar person's face. In both cases, the imagined percept may involve visual details with particular shapes, colors, hues, textures, and shading. The coordination of different brain areas is required for internal image generation. For example, control areas communicate with sensory areas to guarantee the construction of the mental image. While dynamic changes in functional connectivity during visual imagery are expected based on the existing literature (Mechelli et al., 2004; N. Dijkstra et al., 2017), it is unclear whether a functional connectivity network exists in which the strength and the spatial location of the connections vary depending on the imagery target (e.g., face vs. place).

Visual areas play a crucial role for internal image generation. The visual system is retinotopically organized providing a one-to-one mapping which accurately captures the spatial organization of the visual input. Moreover, the visual system is organized into a hierarchy of increasingly complex representations (e.g., orientation, shape, color, luminance, etc.…) which accurately capture the attributes of the visual input. Neuroimaging studies have shown that areas that locally represent specific features during visual perception also represent the same features during visual imagery. For example, the primary visual cortex

* Corresponding author.
E-mail address: fm1672@nyu.edu (F. Mantegna).

https://doi.org/10.1016/j.neuroimage.2025.121171

Available online 24 March 2025

represents spatiotopic information (Bartolomeo et al., 2020; Kosslyn et al., 1995), area V4 represents colors (Bannert and Bartels, 2018; Bartolomeo et al., 2024), area MT represents motion (Goebel et al., 1998), while specialized areas in the inferior temporal cortex represent faces and places, respectively (O'Craven and Kanwisher, 2000). The crosstalk between visual areas may be the very basis of complex mental image formation because it allows for the integration of different parts into a coherent whole.

Visual areas alone are presumably not sufficient to achieve internal image generation. Instead, visual areas may be supported by cognitive control mechanisms, such as memory and attention, exerting top-down influences during imagery, as suggested in the model proposed by Sakai and Miyashita (1994). In particular, their model suggests that different objects or parts of a scene must be retrieved from a memory storage and are visualized using focal attention during imagery. This model is supported also by neuroimaging evidence suggesting that not only visual areas but also frontal, temporal and parietal areas are activated during imagery (Ishai et al., 2000; Ragni et al., 2021). Neuropsychological studies have shown that certain patients may exhibit a selective impairment in visual imagery while maintaining intact visual perception when lesions are confined to the frontal or temporal regions, leaving the visual areas unaffected (Guariglia et al., 1993; Moro et al., 2008). The coordination of frontal, temporal, parietal, and visual areas is necessary to re-instantiate and maintain a mental image in the absence of an external reference.

The visual system is selectively tuned to process information about object categories that define our daily experiences, such as faces and places. A category selective brain network - extending well beyond visual areas (e.g., FFA, PPA) and including temporal, parietal and frontal areas - is involved in perception and action directed towards specific object categories (Chen et al., 2017; Hutchison et al., 2014; Mahon, and Almeida, 2024; Saygin et al., Saxe, 2012; Walbrin, and Almeida, 2021). There is evidence that anatomical and functional connectivity constraints can also be engaged to influence visual perception through category selective attention. Previous studies have shown that connections between inferior temporal regions display category specific patterns when selectively attending to complex objects. (Norman-Haignere et al., 2012). Object categories can be distinguished not only by which region exhibits a stronger response but also by how different regions interact reciprocally. Moreover, Baldauf and Desimone (2014) observed functional connectivity patterns having distinct spatial destinations depending on whether participants were instructed to pay attention to faces or places during visual perception. Collectively, these studies demonstrate that functional connectivity networks vary in strength, spatial destinations, or both, depending on the focus of category selective attention. Crucially, while these studies involved a combination of exogenous and endogenous signals, it is unknown whether the same pattern can be observed for purely endogenous signals, without an external reference, such as during visual imagery.

Neural decoding is an excellent tool to address questions about content-specific representations (Haynes, and Rees, 2006). It uses machine learning algorithms to read out different stimulus categories from brain signals. Content-specific information has been successfully decoded during visual perception, for example, by deciphering single visual features (e.g., orientation, shape, color) but also more complex object information from recorded brain signals (Brandman, and Peelen, 2017; Cichy et al., 2014; Isik et al., 2014). Similar approaches have been used to decode visual imagery, too. For instance, previous studies tried to decode different types of content-specific information (e.g., perceptual, conceptual) during visual imagery from the sparse activation of various brain areas over time (Bainbridge et al., 2021; Dijkstra et al., 2018; Linde-Domingo et al., 2019).

Different coding schemes emerged from previous decoding studies. A *modular* coding scheme in which activation in single brain areas can distinguish between complex object representations (Cichy et al., 2012). A *distributed* coding scheme in which the concurrent co-activation of multiple brain areas can distinguish between complex object representations (Reddy et al., 2010). Here, we explore the possibility that a *network* coding scheme is also suitable (Sporns, 2002). This last coding scheme involves functional connectivity patterns in which the strength and spatial destination of the connections can distinguish between complex object representations (e.g., face vs. place). Crucially, these different coding schemes do not exclude, but rather complement, each other.

In this magnetoencephalography study, we test the hypothesis that different imagery categories are associated with distinct category-specific functional connectivity patterns. Participants were asked to imagine two different types of targets: faces and places. In contrast to previous studies, we instructed the two imagery categories by using word cues only, rather than showing any concrete pictorial aids. The rationale was that - in the absence of any pictorial references - participants have to internally generate mental images purely based on memory and attentional control. Consequently, any differences between imagery categories would be fully attributable to an internally driven effort to (re-)instantiate mental images rather than being confounded with low-level visual information artificially introduced by a pictorial aid. We hypothesized that the imagery of faces and places involves distinct connectivity patterns of different strength, spatial destination, or both. To test this hypothesis, we used neural decoding to read out imagined categories from the connectivity patterns measured across MEG sensors as well as the reconstructed cortical sources. To achieve that goal, we used a connectivity decoding method based on spatial covariance that was originally applied to motor imagery for brain computer interface (BCI) applications (Barachant et al., 2013). This decoding method was designed to capitalize on relative changes in brain activity measured from M/EEG sensor pairs. Here, we tested whether it will allow us to capture connectivity patterns distinguishing face versus place imagery, i.e. a type of internal signal that is notoriously hard to decode with classic time-domain decoders due to the temporal misalignment across trials. Moreover, in order to disentangle the contributions of local vs. distant neural computations within the category-specific functional connectivity network, we test to what extent the decoding is driven by short-range and long-range connections.

2. Methods

2.1. Participants

Eleven healthy participants (mean age = 28.45, range 24–33, 4 female) with no history of psychiatric or neurological disorders took part in this MEG experiment. All of them reported normal or corrected-to-normal vision. Participants signed an informed consent form before the recording session. Ethical approval to conduct the study was provided by the University of Trento ethical committee.

2.2. Vividness of visual imagery questionnaire

The Vividness of Visual Imagery Questionnaire (VVIQ) is a psychometric test that has been designed to measure individual differences in the vividness of visual imagery (Marks, 1973). The VVIQ consists of 16 experimental items organized into four groups. For each group, participants are instructed to imagine a scenario like a familiar person, a familiar shop, or a natural landscape. For each item, participants provide vividness ratings reflecting the visual resolution that they can achieve when they imagine specific details for each scenario (e.g., face contour, characteristic poses, clothes color). Vividness ratings range on a scale from 1 (poor imagination) to 5 (vivid imagination).

Before taking part in the MEG experiment, we asked participants to complete the VVIQ online on an open-source survey platform (LimeSurvey, GmbH, Hamburg, Germany).

2.3. Experimental procedure

We used the Psychophysics Toolbox (PTB-3, Kleiner et al., 2007), MATLAB release R2017b, for stimulus generation and stimulus delivery. The stimuli were projected on a translucent whiteboard using a DLP LED projector (ProPixx, VPixx Technologies Inc., Saint-Bruno, Canada) at a 120 Hz refresh rate. The whiteboard was located at 1 m distance from the participant, and it provided a projection area of 51 × 38 cm (width x height) and 1440 × 1080 pixel resolution.

The experimental paradigm is shown in Fig. 1A. Each trial began with an instruction screen ("Imagine a…"). Then, participants were presented with a visual word cue ("Face" or "Place") instructing a

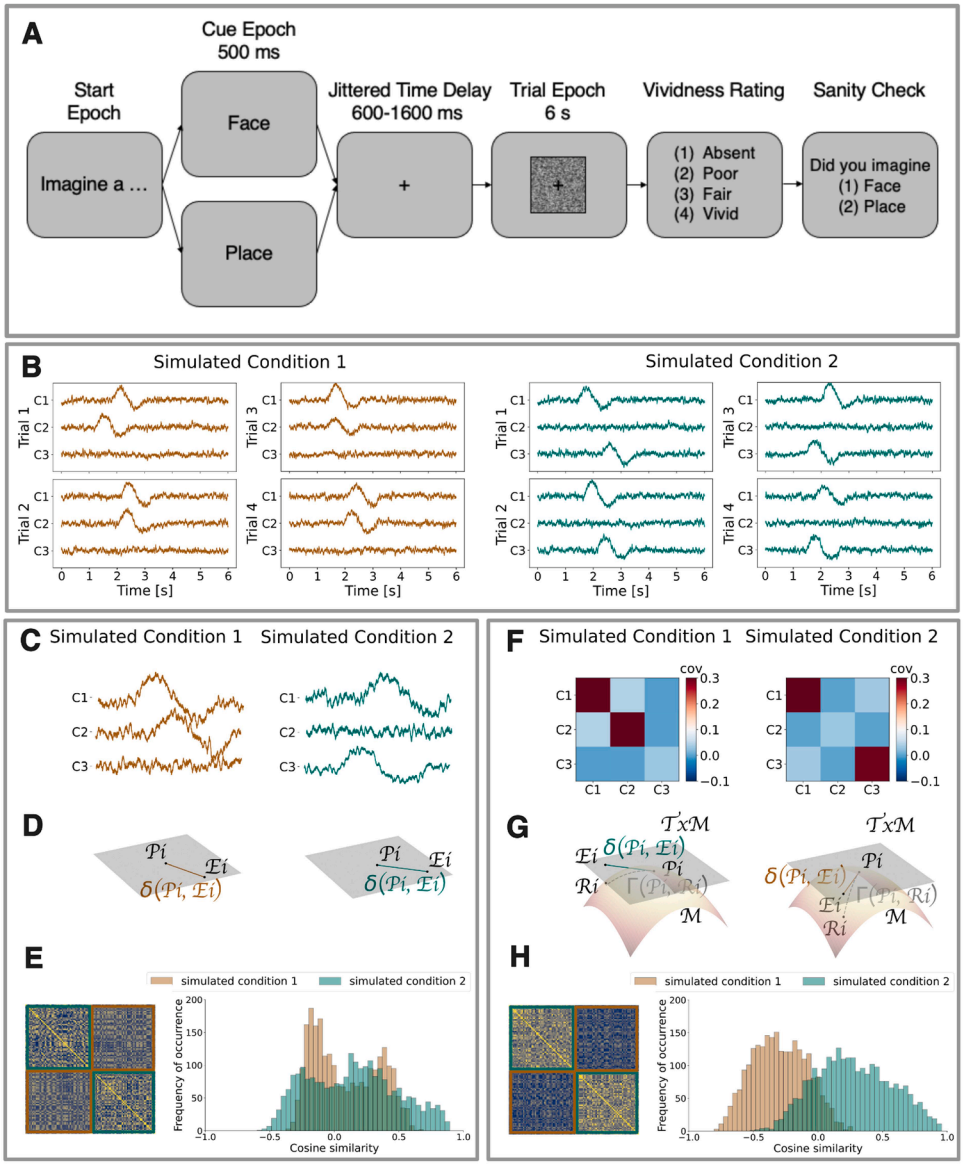


Fig. 1. Experimental procedure, decoding methods and simulated data. The experimental procedure **(A)** was structured as follows: each trial began with a visual word cue instructing one category as the imagery target; then, a jittered time delay ensued after which the subject had to imagine a familiar instance of the cued category while a dynamic phase-scrambled mask was presented on the screen for 6 s. At the end of each trial, participants were asked to rate the vividness of their imagination and confirm the category they had imagined during the trial. In a computer simulation **(B-H)** we tested whether relevant information is captured by covariance-based decoding or classic time-domain decoding. For this purpose, we simulated time series belonging to two different conditions **(B)** (here, we show only 4 representative trials in 3 simulated channels per condition). Each trial contained a signal simultaneously embedded in noise of different channels. However, the onsets of the signal were misaligned across trials and across channels. For illustration purposes, linear separability was assessed by comparing the geometric properties of the vectors used as input data for different decoding methods. To prepare input data for time-domain decoding **(C-E)**, we concatenated the time series of various channels into one single vector for each simulated condition **(C)**. The distance between an exemplar vector (*Ei*; i.e., single trial) and a prototype vector (*Pi*; i.e., average across trials) for each simulated condition can be estimated using an Euclidean metric (δ) **(D)**. Cosine similarity **(E)** across all exemplar vectors - corresponding to all simulated trials - shows that simulated conditions are not linearly separable when using time-domain decoding. To prepare input data for covariance-based decoding **(F-H)** we estimated spatial covariance matrices for each condition **(F)**, measuring the interdependence between channel pairs. The distance between an exemplar matrix (*Ri*) and a prototype matrix (*Pi*) on a Riemannian manifold (M) can be estimated using a Riemannian metric (Γ). Then, the covariance matrix can be projected to an Euclidean tangent space (TxM) obtaining a tangent vector **(G).** After that, the distance between an exemplar tangent vector (*Ei*; single trial) and a prototype tangent vector (*Pi*; average across trials) for each simulated condition can be estimated using an Euclidean metric (δ). Cosine similarity **(H)** across all tangent vectors - corresponding to all simulated trials - shows that simulated conditions are linearly separable when using covariance-based decoding.

category for imagination. After that, a fixation cross was shown in the middle of the screen and there was a 600–1600 ms jittered time delay. At this point, the trial epoch started and lasted for 6 s. A 15 × 25 cm picture frame containing a dynamic phase-scrambled mask centered around the fixation cross was displayed on the screen. The picture frame was meant to constrain participants' imagination to a constant portion of the screen such that the size of the imagined object was consistent across trials and across imagery conditions. Participants were instructed to fill the picture frame with their visual imagination. In particular, they were asked to imagine a familiar face or place of their choosing. Even though participants were allowed to choose the object of their imagination, they were instructed to always imagine the same face and the same place throughout the experiment to reduce within-subject variability. Following the trial epoch, participants were asked to rate the vividness of their imagination on a scale from 1 (poor imagination) to 4 (vivid imagination). Finally, we presented participants with a catch question (i.e., "Did you imagine a face or a place?") to make sure they were following the instructions. We used an MEG-compatible response collection system (ResponsePixx Dual Handheld, VPixx Technologies Inc., Saint-Bruno, Canada) to keep track of participants' responses. Before starting the experiment, participants performed 10 practice trials to familiarize themselves with the task. The experiment consisted of 240 trials evenly distributed over 4 blocks. The presentation order of the instructed categories was randomized.

2.4. Data acquisition

Prior to data acquisition, individual head shapes were digitized with a Polhemus Fastrak digitizer (Polhemus, Vermont, USA), including fiducial landmarks (nasion, right and left pre-auricular points) and about 200 additional points spread out all over the scalp. Five Head Position Indicator (HPI) coils were placed on participant's mastoid bones and forehead to keep track of participant's head position inside the dewar through electromagnetic induction before and after each recording block. Landmarks and HPI coils were digitized twice to ensure that their spatial accuracy was less than 1 mm.

MEG recordings were obtained in a magnetically shielded room (AK3B, Vacuum Schmelze, Hanau, Germany) using a 306-channel (204 first order planar gradiometers, 102 magnetometers) VectorView MEG system (Neuromag, Elekta Inc., Helsinki, Finland). The MEG signal was sampled at 1 kHz, with a low-pass anti-aliasing filter at 330 Hz and a high-pass filter at 0.1 Hz. Before entering the experiment room, we ensured that participants were not wearing or carrying any metallic object and other potential sources of electromagnetic interference. Participants performed the task in a seated position. When positioning participants in the MEG scanner, we ensured tight contact with the dewar. Participants were instructed to avoid head, body and limb movements during the trial epoch.

Moreover, participants were instructed to avoid eye blinks and keep strict eye fixation as much as possible during the trial epoch. Binocular pupil size and eyes' position were continuously monitored by an MEG-compatible eye-tracking device (Eyelink 1000 Plus, SR-Research Ltd. Mississauga, Ontario, Canada). In the beginning of each experimental session, participants performed an eye-tracking calibration task aimed at verifying the correspondence between pupil position in the image recorded from the camera and gaze position on the screen. Calibration was repeated if drift was noticed during the experimental session.

2.5. Eye-Tracking analysis

To rule out potential confounds, we removed all trials in which we measured oculomotor noise. In particular, we identified three types of oculomotor noise associated with potential confounds at the brain level: eye blinks, saccades. Eye blinks consist in the rapid opening and closure of the eyelids. Saccades are fast, voluntary eye movements whose amplitude can be up to 15–20°.

We used co-registered eye-tracking data to exclude trials contaminated by oculomotor noise. The eye-tracker measured left and right pupil size (i.e., pupillometry) as well as left and right, horizontal (x) and vertical (y) gaze coordinates. Thus, the eye-tracking output consisted of 6 channels. The analog output was in voltage (−5 V to +5 V range). The raw eye-tracking signal was sampled at 1 kHz. We segmented eye-tracking data around the trial epoch (0–6 s). Moreover, we down-sampled the raw signal to 250 Hz and applied a notch filter to remove 50 Hz power-line noise. Then, we converted the analog output (in voltage) to digital units (pixels) and we used the physical specificities of the eye tracking device (i.e., data range, voltage range, screen proportion, screen distance) to convert pixels to millimeters. Binocular pupil size was measured in mm^2. Vertical and horizontal (x, y) binocular gaze coordinates were measured in mm.

For eye blink detection, we used an automatic artifact rejection method based on a pupil size threshold. During blinks the eye-tracking device loses track of the pupil, resulting in missing values in the output file. However, eye blinks are preceded and followed by a sharp decrease in pupil size measurements, because the closure and opening of the eyelids is not instantaneous. For each subject, we computed the absolute value and z-normalized (mean subtracted and divided by standard deviation) pupil area measured from left and right eye. We defined 3 standard deviations from the mean as a threshold for eye blink detection. Trials in which pupil size measurements exceeded the threshold were excluded from further analysis.

For saccade detection, we used an automatic artifact rejection method based on a velocity-threshold identification (VT-I) algorithm (Salvucci and Goldberg, 2000). This algorithm separates fixations and saccades based on their point-to-point velocities using binocular x, y gaze coordinates. We computed the tangent of the rotation angle of the eye relative to the head and we used that measure to calculate eye movement velocities (degrees/second). Velocity profiles typically show two distributions: low velocities for fixations (i.e., <100 deg/*sec*), and high velocities for saccades (i.e., >300 deg/*sec*). Trials exceeding the high velocity threshold (i.e., >300 deg/*sec*) were excluded from further analysis.

2.6. MEG pre-processing

MEG pre-processing was performed using MNE-Python (Gramfort et al., 2013), Python release 3.6.7, combined with custom routines. First, we removed external and internal sources of noise from the MEG signal. Then, we performed basic signal processing operations like filtering and epoching.

External noise (e.g., environmental noise, stationary noise) was removed from MEG recordings offline using a MaxFilter software (tsss-filters; Taulu and Simola, 2006). In particular, we used a temporally non-extended spatial Signal Source Separation (SSS) algorithm in order to suppress external sources of magnetic interference. Whenever head movements exceed 1 cm within or between blocks, we used the Max-Move algorithm to spatially co-register MEG recordings across blocks to the median head position. HPI movement correction was applied to MEG data collected from 6 over 11 subjects. Then, continuous data was visually inspected for system related artifacts (e.g., SQUID jumps), and contaminated sensors were interpolated. Up to 10 sensors per experimental block were interpolated.

Internal noise was reduced using independent component analysis (ICA; Makeig et al., 1995) while preserving signals originating from the brain. Among the potential sources of internal noise there are heartbeat, muscular activity and any residual oculomotor activity (e.g., eye blinks, eye movements) that was not removed based on the eye-tracking data. We used a fixed-point algorithm to estimate 15 independent components in the trial epoch time window (0–6 s). Up to 5 components per block were excluded based on visual inspection of spatial topographies and latent sources' time course.

A two-pass zero-phase infinite impulse response (IIR) band-pass filter

was applied to raw data between 1 and 150 Hz. This IIR filter was based on a Butterworth forward-backward filter. Time series were down-sampled to 250 Hz. Then, we segmented trial epochs from picture frame onset to picture frame offset (0–6 s). We further segmented the trial epoch using different time-window segmentation schemas. In particular, we used a short segmentation scheme (100 ms time-windows), an intermediate segmentation scheme (500 ms time-windows) and a long segmentation scheme (1 s time-windows).

Trials were excluded from further analyses according to different criteria. We excluded all trials in which the vividness rating was poor ($\leq$2) or participants provided a wrong answer to the catch question about which category they had just imagined. In both cases, the entire trial epoch was discarded. Moreover, we excluded all trials containing oculomotor noise (eye blinks, saccades, predictive microsaccades). In this case, we discarded only noisy time-windows rather than the entire trial epoch. Importantly, the remaining number of trials for each time window was not systematically different between experimental conditions (face vs. place) after trial exclusion.

2.8. Covariance estimation

We used the pyRiemann toolbox for covariance estimation (Barachant, 2023). We estimated covariance as a measure of joint variability between a pair of time series using Eq. (1):

$$\mathrm{cov}(x^i, x^j) = \frac{1}{N}\sum_{t=1}^{N_t}\left(x_t^i - \bar{x}^i\right)\left(x_t^j - \bar{x}^j\right) \quad (1)$$

Where x^i and x^j are time series recorded from different sensors summed across multiple timepoints t divided by the total number of timepoints N.

Spatial covariance matrices (SCMs) were computed as the set of pairwise covariance estimates between all sensors (i.e., 306 × 306 sensors, including both gradiometers and magnetometers), all reconstructed sources (i.e., 5124 × 5124 sources), and all regions of interest (i.e., 360 × 360 parcels). Covariance estimation can be unstable when the sample size (i.e., trial number) is small and the number of variables (i.e., sensors or sources) is large. Therefore, we used a shrinkage method for covariance estimation (OAS; Chen et al., 2010) that improves numerical stability and ensures that the matrix is symmetric, positive definite, and thus invertible.

2.9. Decoding analysis

We used two different decoding methods: classic *time-domain decoding* and *covariance-based decoding*. From a methodological point of view, these two decoding methods differ in terms of the brain features used for classification. Time-domain decoding features were obtained by concatenating the raw MEG time-series measured from different sensors into a vector. Covariance-based decoding features were obtained by using a kernel transformation to project spatial covariance matrices from a Rimannian manifold to a locally homeomorphic Euclidean tangent space.

We built a decoding pipeline using scikit-learn toolbox (Pedregosa et al., 2011). This decoding pipeline was applied to MEG data collected from individual subjects. Trial epochs were segmented using a sliding time-window. For each time window, we obtained a classification score. We used three different time-window sizes: 100 ms, 500 ms, 1 s. For time-domain decoding, we standardized the MEG signal by estimating the mean and the standard deviation for each trial and each time-window. For covariance-based decoding, we estimated the spatial covariance matrices for each trial and each time-window. After that, we vectorized our input features following two alternative approaches. For time-domain decoding, we concatenated MEG time series from different recording channels into a single vector for each trial. For covariance-based decoding, we approximated geodesic distances in the Riemannian manifold to Euclidean distances in the tangent space (see Fig. 1G) obtaining a tangent vector for each trial. Then, we used a logistic regression model for binary classification of imagined faces and places trials. In this model, the probabilities of the possible outcomes for each trial are modeled using a logistic function. L2 regularization was applied to improve numerical stability. Optimization was performed using a coordinate descent (CD) algorithm that minimizes the cost function by adjusting weights and regularization parameters. Finally, we used the Area Under the Receiver Operating Characteristic Curve (ROC AUC) as a scoring metric. This scoring metric accounts for the tradeoff between true and false positive rates.

2.10. MRI-based source reconstruction

High-resolution T1-weighted anatomical scans were acquired for most of participants (seven over eleven) in a 4T Bruker MedSpec Biospin MR scanner with an 8-channel birdcage head coil (MP-RAGE; 1 × 1 × 1 mm; FOV, 256 × 224; 176 slices; TR = 2700 ms; TE = 4.18 ms; inversion time (TI), 1020 ms; 7-degrees flip angle). When the anatomical scans were not available (four over eleven participants) we used a template brain to perform source reconstruction (Douw et al., 2018). This template brain was the average of the anatomical scans collected from 40 subjects ('fsaverage'). The template brain was deformed to match the head shape of the participants that we measured using the Polhemus Fastrak digitizer (Polhemus, Vermont, USA). For group analysis, we computed a linear interpolation (i.e., morphing) between the individual source model and the template brain for each subject.

The anatomical scans were 3D reconstructed using Freesurfer software (Fischl et al., 1999). A Boundary Element Model (BEM) was estimated using the watershed algorithm. MRI and MEG coordinate systems were co-registered by manually matching digitized anatomical fiducial landmarks on the participant's T1 scan. The resulting whole brain surface reconstruction (5124 vertices; 6.2 mm average source spacing), the BEM model and the aligned coordinate frames were used to compute the 3D forward model for MEG source reconstruction. The inverse operator was estimated using the noise-covariance matrix, the forward solution and the source covariance matrix. We used the Minimum-norm Estimates (MNE, Hämäläinen and Ilmoniemi, 1994) for reconstruction of neuronal sources. We used a loose orientation constraint for source reconstruction. In particular, for each source location we estimated a gain matrix having three columns corresponding to magnetic fields *x*, *y*, and *z* orientations. Then, we computed the norm of these three vectors to obtain one single vector for each source location.

2.11. Statistical analysis

Decoding performance was evaluated using statistical tests to establish whether classification was significant both at the single subject level and at the group level.

At the single subject level, we used cross-validation and permutation tests to assess the decoding performance for each time window. In particular, we used a stratified k-fold cross-validation procedure. Data were divided into five folds and classification scores were obtained for each fold. Then, cross-validated decoding performance was estimated by averaging the scores obtained for each fold. Moreover, we ran a permutation test to evaluate the statistical significance of cross-validated scores. This test consisted in repeating the cross-validated classification procedure 1000 times permuting condition labels. We computed the p-value as the percentage of tests for which the classification score obtained with unpermuted labels was greater than the classification score obtained with permuted labels.

At the group level, we evaluated cross-validated decoding performance across multiple subjects using Bayesian hypothesis testing (Olivetti et al., 2012). To account for the different number of trials per participant resulting from trial exclusion, we used a statistical test that weighs the classification scores for each participant depending on the

number of trials used to train and test the classifier. By doing so, we account for the inherent variability in the dataset and avoid overestimating individual participant results when making inferences about the population. When we performed more than one test for one single decoding analysis (e.g., sub-network analysis) we corrected for multiple comparisons using False Discovery Rate (FDR) correction.

To investigate which nodes provided most information to covariance-based decoding, we ran a cluster-based permutation test (CBPT, Maris, and Oostenveld, 2007) both in sensor space and source space. CBPT consists of two different stages: a cluster formation stage and an inferential stage. In the cluster formation stage, the unit-level statistic is computed for each sensor or source. We used a two-sample covariance matrix unit-level statistic (Cai et al., 2013) that was estimated as follows: first, we estimated the spatial covariance matrices for each trial; then, we computed the element-wise mean covariance matrix and the element-wise variance covariance matrix for each condition; finally, we computed an M (i.e., matrix) standardized statistic that is defined as the squared difference of the mean covariance matrices divided by the sum of the variance covariance matrices. The test statistic is reported in Eq. (2):

$$M_{ij\,(s\times s)} := \frac{\left(\bar{c}_{ij}^{y_1} - \bar{c}_{ij}^{y_2}\right)^2}{\frac{\sigma\left(\bar{c}_{ij}^{y_1}\right)}{N_{y_1}} + \frac{\sigma\left(\bar{c}_{ij}^{y_2}\right)}{N_{y_2}}}, \quad 1 \le i \le j \le s \tag{2}$$

Where M is a sxs (i.e., sensors-by-sensors or sources-by-sources) matrix, c bar is the averaged element-wise covariance estimated between recording channels i and j belonging to either condition y1 or condition y2, and sigma squared is the averaged element-wise variance divided by the number of trials N in each condition. Once we obtained the M matrix, we summed across rows to obtain one single score for each sensor or source measuring the difference in covariance between the two conditions. Given that the distribution of the M standardized statistic is unknown, we run a permutation test under the null hypothesis of exchangeability. We computed the unit-level test statistic 1000 times. For each iteration, assignment to experimental conditions was randomized. Then, the original M values were compared to permuted M values yielding uncorrected p-values. Sensors or sources were selected according to an a priori defined alpha criterion (i.e., $p < 0.05$) and adjacent sensors or sources not exceeding this value were grouped together into clusters. Finally, we summed all the M values within each cluster obtaining one single number. Minimum cluster size was set to 5 sensors or 50 vertices. A spatial adjacency matrix containing information about sensors or sources proximity was taken into account in the cluster formation stage. In the inferential stage, the stored unit-level permutation values summed within clusters were used to compute the cluster-level statistical distribution under the null hypothesis of exchangeability. We calculated the percentage of clusters for which the un-permuted cluster-level statistic was larger than the permuted cluster-level statistic. If the cluster p-value was smaller than 0.05 then we assumed that the data in the two experimental conditions were significantly different.

3. Results

3.1. Decoding performance evaluation on simulated data

First, we ran a simulation to investigate whether relevant information is captured by covariance-based decoding or classic time-domain decoding (see Fig. 1B-H). There are major challenges associated with visual imagery signals - or, more in general, with any internally generated brain signal: On the one hand, relevant information is temporally misaligned because there is a high variability in the onsets of imagination events across trials. On the other hand, some relevant information is presumably encoded in the reciprocal interconnections between channel pairs that give rise to specific spatial configurations. To account for these two aspects, we simulated data as follows. We generated time series for one hundred trials in three different simulated channels. Each time series consisted of a combination of signal and noise. To account for temporal misalignment, we added random delays to signal onsets. To account for specific spatial configurations, we simulated data such that trials belonging to the first simulated condition had higher amplitude modulation in the first and the second channel while trials belonging to the second simulated condition had higher amplitude modulations in the first channel and the third channel (see Fig. 1B). Then, we evaluated different decoding methods by inspection of geometric properties instead of linear classification, for illustration purposes. To do so, we prepared input data by using two different vectorization procedures. To prepare data for classic time-domain decoding, we concatenated time series corresponding to different channels into one single vector (Fig. 1C). To prepare data for covariance-based decoding, we estimated a spatial covariance matrix measuring the interdependence between channel pairs, we estimated the matrix's position in a Riemannian manifold, and we projected the matrix's position on an Euclidean tangent space obtaining a tangent vector (Fig. 1F-G). The rationale of this operation is to capture global covariance properties of the whole matrix (which will be associated with a specific position in the manifold) instead of considering pairwise covariance estimates independently (You and Park, 2021). Finally, we measured cosine similarity across vectors to show that tangent vectors that are used as input for covariance-based decoding are linearly separable, while concatenated vectors that are used as input for classic time-domain decoding are not linearly separable (Fig. 1E and H). Crucially, when the same analysis is run on temporally aligned simulated data (i.e., no delays added to signal onsets) we observed no difference in linear separability between the two decoding methods (Fig. S1). Overall, this simulation revealed that the covariance-based decoding has a specific advantage in decoding temporally misaligned and reciprocally interconnected signals, like the ones driven by cognitive processes such as visual imagery.

3.2. Decoding performance evaluation on MEG data

Then, we used both the covariance-based decoding method and the time-domain decoding method to read out imagined categories from MEG signals. Both decoding methods included a single subject level and a group level statistical test. At the single subject level, we computed cross-validated Area Under the Receiver Operating Characteristic Curve (ROC AUC) scores using a sliding window approach. At the group level, we tested whether classification scores were significant across subjects regardless of different amounts of trials used for classification. This second step is necessary because participants presented a different number of trials after preprocessing (e.g., artifacts, eye movements, low vividness ratings). In line with our expectations based on the simulations, the results obtained on empirical data show a stark difference in performance between the two decoding methods. Classic time-domain decoding did not perform significantly above chance level across subjects, in any time window (Fig. 2A). In contrast, covariance-based decoding achieved correct classifications significantly above chance level ($p < 0.05$, BF>3) across subjects, in three time windows spanning from 2.5 to 5.5 s (Fig. 2B). To test whether these results were determined by the choice of the time window size we also tested shorter time windows. We obtained similar results when using 100 ms (Fig. S2) and 500 ms time windows for classic time-domain decoding and 500 ms time windows for covariance-based decoding (Fig. 2C-D). This suggests that decoding performance is not strictly dependent on the time window size. Even though significant time windows are more sparse when using a shorter sliding window because the temporal misalignment problem is then more pronounced.

There are methodological challenges associated with spatial covariance estimation using neuroimaging data (Pesaran et al., 2018). These criticalities must be taken into account to support the assumption that

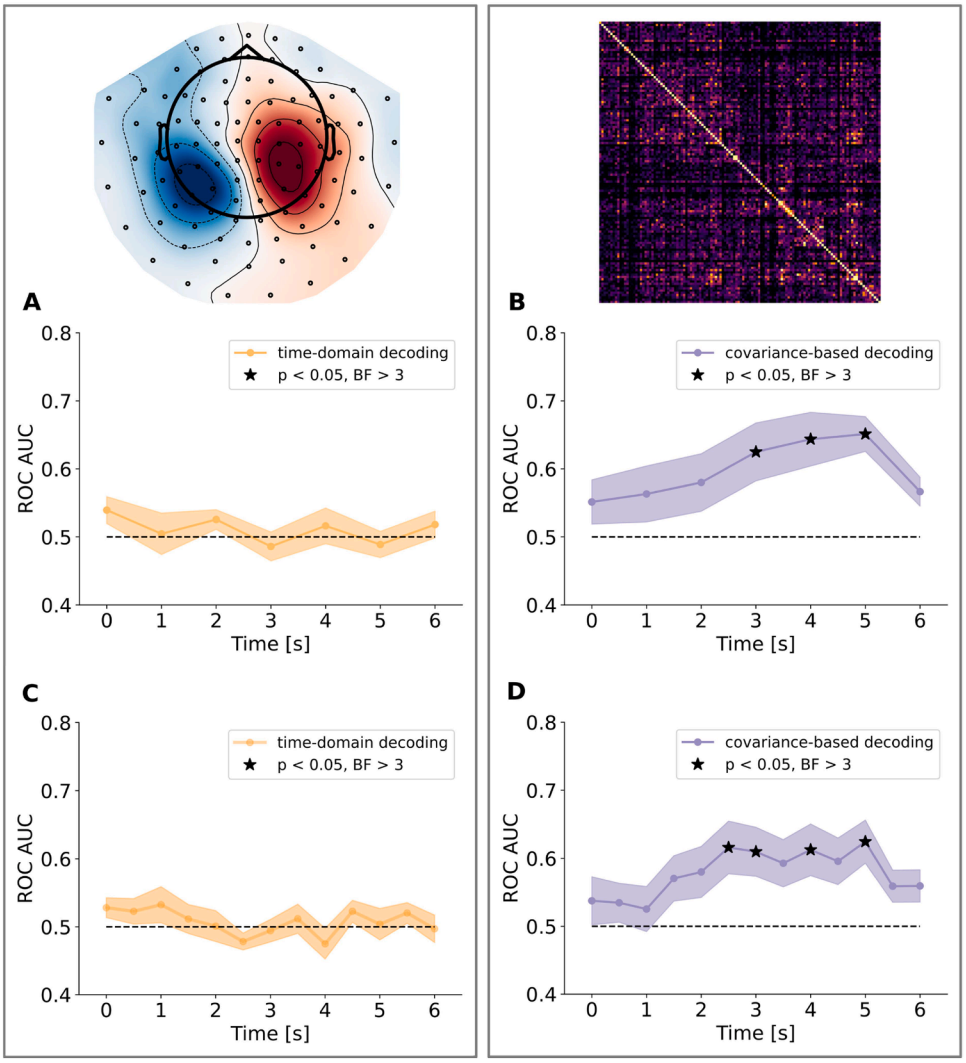


Fig. 2. Time-domain decoding and covariance-based decoding applied to visual imagery MEG data. Left panels **(A and C)** show group-level decoding results in sensor space (including gradiometers and magnetometers) obtained from the classic time-domain decoding method, using 1 s **(A)** or 500 ms **(C)** sliding windows, respectively. Classification score is at chance level and not statistically significant in any section of the trial epoch. Right panels **(B, D)** show group-level decoding results in sensor space (including gradiometers and magnetometers) obtained from the covariance-based decoding method, using 1 s **(B)** and 500 ms **(D)** sliding windows, respectively. Classification score is above chance and statistically significant in three (or four) time windows spanning from 2.5 to 5.5 s. The solid line indicates the mean, while the shaded area indicates the standard error of the mean (s.e.m.).

functional connectivity estimates reflect variations associated with imagery content. On the one hand, there are signal-to-noise ratio (SNR) potential confounds. In fact, random variations in amplitude across trials (e.g., due to fatigue, alertness, etc.…) can lead to task-irrelevant changes in covariance estimation. To minimize spurious variations, we applied a baseline correction. Specifically, we applied a whitening transformation to remove the covariance associated with task-irrelevant variations (as estimated from the baseline time window) and retain the covariance associated with task-relevant variations (as estimated from the trial time window). On the other hand, covariance estimates can be biased by volume conduction, i.e., field spread across sensors and sources (Bastos and Schoffelen, 2016). However, while volume conduction affects instantaneous phase relationships, we estimated covariance over 1 s time-windows capturing only variations which are more sustained over time. Moreover, volume conduction should not influence neural decoding performance. In fact, the variations associated with volume conduction are independent from experimental conditions. In other words, volume conduction equally affects both face and place imagery trials because it is due to measurement error. In contrast, the classification algorithm detects similarities and differences between conditions remaining unaffected by random variations that are common to both conditions.

3.3. Relationship between decoding performance and vividness ratings

Next, we tested whether covariance-based decoding performance correlated with participants' subjective evaluation of the vividness of visual imagery. We tested the relation between decoding performance and subjective ratings both across and within subjects (Fig. 3). Across subjects, significant decoding performance (i.e., averaged ROC AUC scores in the time windows ranging from 2.5 to 5.5 s) correlated with self-reported individual differences in the general vividness of visual imagery ($r(10)=0.56$, $p < 0.05$, Fig. 4A), as assessed by the Vividness of Visual Imagery Questionnaire (VVIQ). Moreover, within subjects, we split the MEG dataset into trials associated with high vividness reports (scores ≥ 3 in the vividness rating provided at the end of the trial, see Fig. 1A) and trials associated with low vividness reports (scores≤ 2), to contrast the decoding performance associated with different subjective vividness ratings. We reasoned that if covariance-based decoding relies on information that contributes to the perceived vividness of visual

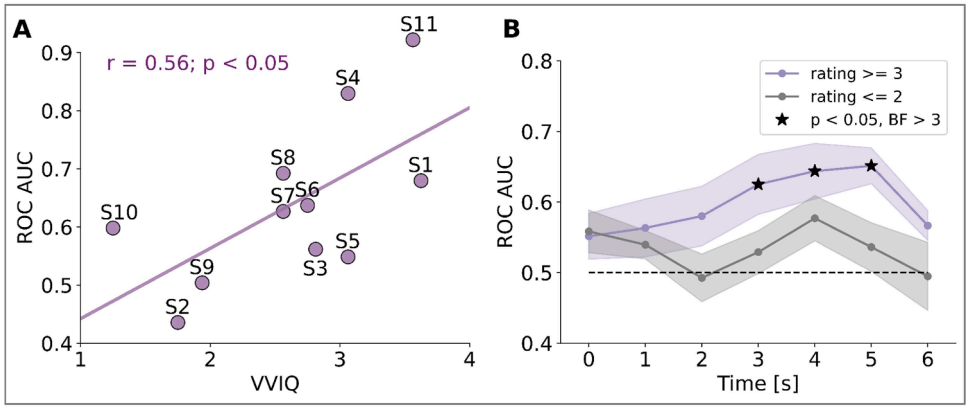


Fig. 3. Relation between covariance-based decoding performance and subjective vividness ratings. (A) Correlation between vividness of visual imagery questionnaire (VVIQ) scores and averaged ROC AUC scores in the time windows ranging from 2.5 to 5.5 s obtained using covariance-based decoding. **(B)** We obtained higher and significant covariance-based decoding performance when using only trials associated with high vividness ratings (≥3, purple line), compared to lower and not significant covariance-based decoding performance when using only trials with lower vividness ratings (≤2, gray line). The solid line indicates the mean, while the shaded area indicates the standard error of the mean (s.e.m.).

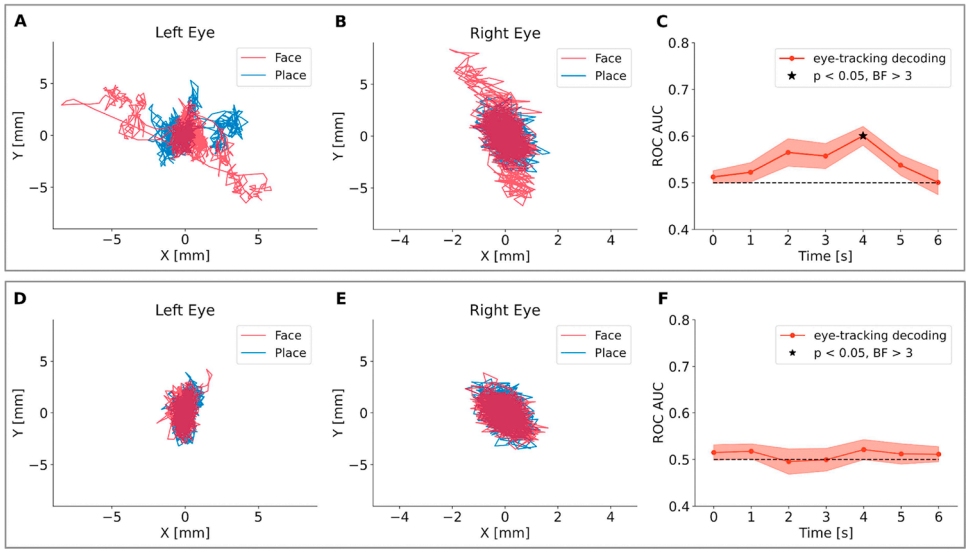


Fig. 4. Detection and removal of trials containing predictive microsaccades. (A, B) Examples of left and right eye movements in a trial in which the micro-saccade activity contained information about the imagination target (faces, red line, versus places, blue line). **(C)** Group-level decoding based on eye-tracking data only before microsaccade removal (mean and s.e.m). At this level, only trials containing overt saccades exceeding a rejection threshold were excluded. Trials containing sub-threshold, but predictive micro-saccade activation still contributed to the decodability of the imagination target. **(D, F)** After removing all trials with increased decoding probabilities, only trials in which eye-movement traces did not contain information about the imagination target remained in the sample. **(F)** Resulting group-level decoding after microsaccade removal.

imagery then high vividness ratings will be associated with higher ROC_AUC scores. Indeed, decoding scores were higher and significant in the time windows from 2.5 to 5.5 s when using only trials with high vividness ratings (≥3, Fig. 4B) while the decoding scores were lower and not significantly above chance level (at any time) when using only trials with low vividness ratings (≤2). Importantly, there were no significant differences in vividness ratings between imagination categories ($t(10) = -0.48$, $p = n.s.$) suggesting that participants' imagination was equally vivid in faces and places trials. Additionally, behavioral performance showed no evidence of learning effects, and decoding performance exhibited no habituation effects across experimental blocks (see Figure S3).

3.4. Detection and elimination of predictive saccades and microsaccades

In addition, we ran a control analysis to rule out the possibility that the decoding was driven by systematic differences in eye movements associated with face and place imagery. This is a general concern for any neural decoding study, and the covariance-based decoding approach offers a straightforward and clean solution to rule out this potential confound. Although participants were instructed to keep their eyes fixated at the center of the screen during the imagination task, co-registered eye-tracking revealed some residual but systematic micro-saccadic activity that was related to the imagination targets (Fig. 4A and B), even after excluding trials with supra-threshold eye-movements (saccades). Therefore, we used covariance-based decoding to read out imagination categories from eye-tracking data that survived the threshold-based exclusion. By visual inspection of eye-tracking data, we observed systematic differences in the eye movement position covariance that may contribute to the classification of face versus place imagery, at least partially for some participants in some time windows (Fig. 4A-B). At the group level, eye movement decoding was statistically significant ($p < 0.05$, BF=3) from 3.5–4.5 s (Fig. 4C). We observed no significant correlation between eye-movement covariance-based

decoding scores in the significant time window and self-reported vividness as assessed by the Vividness of Visual Imagery Questionnaire (VVIQ) (Fig. S4). In order to correct for that, we cleaned the MEG dataset from all trials containing any such predictive microsaccades, by training a linear classifier on the eye tracking dataset and estimating the predictive probabilities for each trial. All trials with increased classification probability were excluded from further analyses on the MEG dataset. After predictive microsaccade removal, eye tracking decoding was no longer significant (Fig. 4D-F). A comparison between MEG covariance-based decoding scores before and after predictive microsaccade removal (Fig. S5) shows a small difference in performance reflecting the portion of neural decoding explained by subthreshold (microsaccadic) eye movements.

3.5. The functional connectivity network distinguishing face vs. place imagery

One major advantage of the covariance-based decoding approach for the purpose of this study is that it is inherently based on a functional connectivity measure, i.e. the degree of interaction between node pairs. This allows us to map the most informative connectivity patterns underlying covariance-based decoding. In the following, we use two different types of visualization: edge maps and hub maps (see Fig. 5). While the edge map allows us to visualize what sensor pairs are more informative to distinguish face and place imagery, the hub map allows us to visualize what individual nodes are most informative to distinguish face and place imagery. Edge maps are based on the normalized absolute difference in covariance, averaged across trials. Hub maps are based on a cluster-based permutation test between covariance matrices collapsed

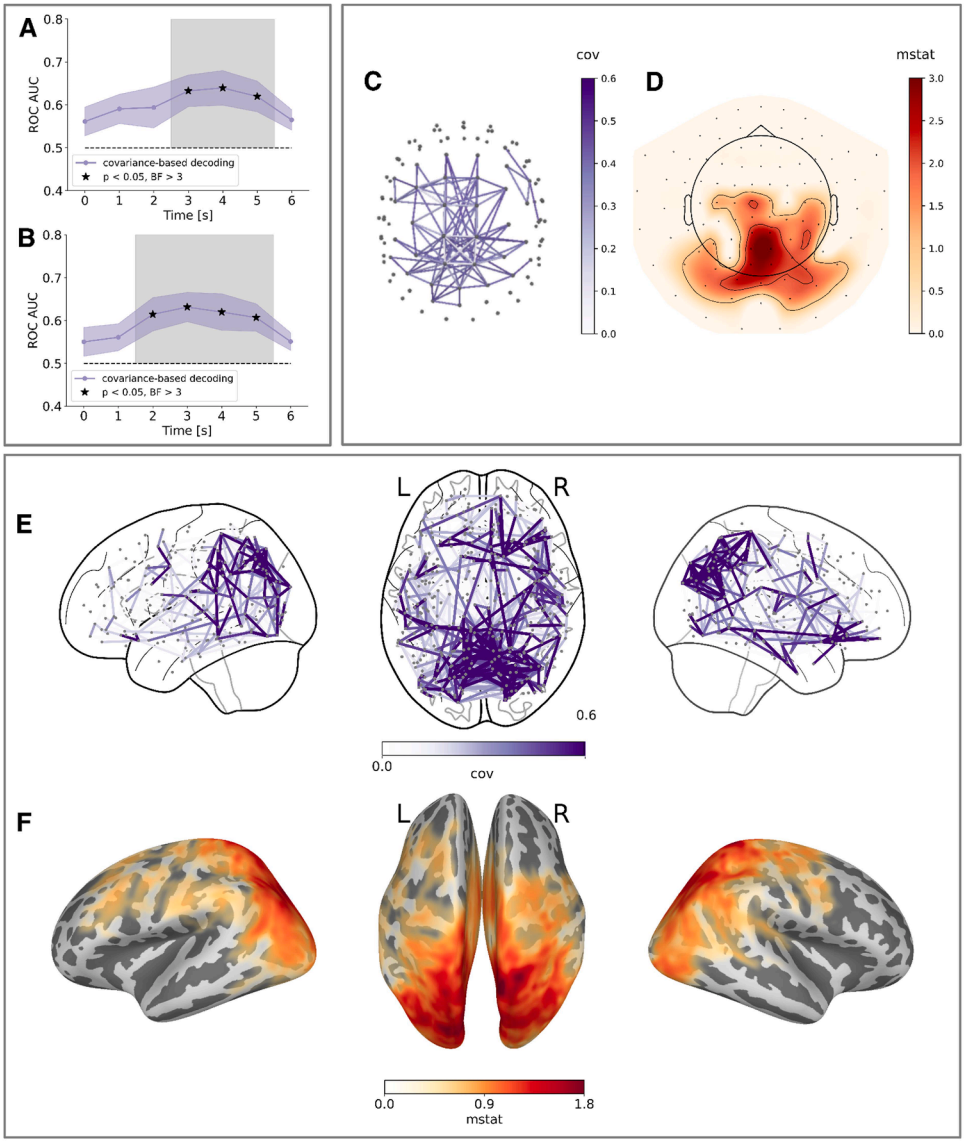


Fig. 5. Whole-brain visualization of the connectivity patterns underlying covariance-based decoding in sensor and source space. Decoding results obtained using all sensors (i.e., both gradiometers and magnetometers) **(A)** and all reconstructed sources parcellated using the Glasser atlas **(B),** after removing trials with predictive microsaccades. Edge maps represent the normalized absolute difference in covariance (purple color map). Hub maps represent the output of a cluster-based permutation test between face and place covariance matrices collapsed along one dimension (red color map). **(C)** Edge map showing the most informative connections between sensors distinguishing face and place trials (highest 5 percentiles). Each gray dot represents a sensor, and each purple line represents the covariance between two sensors. **(D)** Hub map showing the most informative individual sensors distinguishing face and place trials. **(E)** Edge map showing most informative connections between parcellated areas distinguishing face and place trials (highest 5 percentiles). Each gray dot represents a parcellated area and each purple line represents the covariance between two parcellated areas. **(F)** Hub map showing most informative individual sources distinguishing face and place trials.

along one dimension (we refer to this metric as *mstat,* i.e., matrix statistics, for details see Methods).

To allow for a better localization of these connectivity patterns we applied covariance-based decoding both in sensor space and in source space. In sensor space, we obtained significant covariance-based decoding ($p < 0.05$, BF>3, Fig. 5A) from 2.5 to 5.5 s using all sensors (i.e., both gradiometers and magnetometers) also after removing trials with predictive microsaccades. The hub map estimated for this time window showed that most informative connectivity hubs distinguishing face and place trials are in the posterior sensors (Fig. 5D). The edge map estimated for this time window - including all connections within the highest 5 percentiles of normalized absolute differences in covariance - showed that the most informative connections include not only short-range connections within both anterior and posterior sensors but also long-range connections between anterior and posterior sensors (Fig. 5C). In source space, we obtained significant covariance-based decoding ($p < 0.05$, BF>3, Fig. 5B) from 1.5 to 5.5 s using all parcellated sources also after removing trials with predictive microsaccades. The hub map estimated for this time window showed that the most informative connectivity hubs distinguishing face and place trials are in the occipital and parietal cortices but also in temporal and frontal regions, albeit weaker (Fig. 5F). The edge map estimated for this time window - including all connections within the highest 5 percentiles of normalized absolute differences in covariance - showed that the most informative connections include not only short-range connections within both occipital and parietal areas but also long-range connections between occipital, parietal, temporal and frontal areas (Fig. 5E).

Additionally, we investigated the most informative connections associated with either face or place imagery decoding by looking at the feature coefficient weights of the logistic regression classifier (Haufe et al., 2014). The edge map of face-selective (red) and place-selective (blue) connections reveals a largely overlapping network spanning across visual, parietal, inferior temporal and inferior frontal regions (Fig. 6).

Overall, these results suggest that imagined faces and places involve differences in functional connectivity spanning a broad network of brain areas including not only short-range connections within posterior and anterior areas but also long-range connections between posterior and anterior areas.

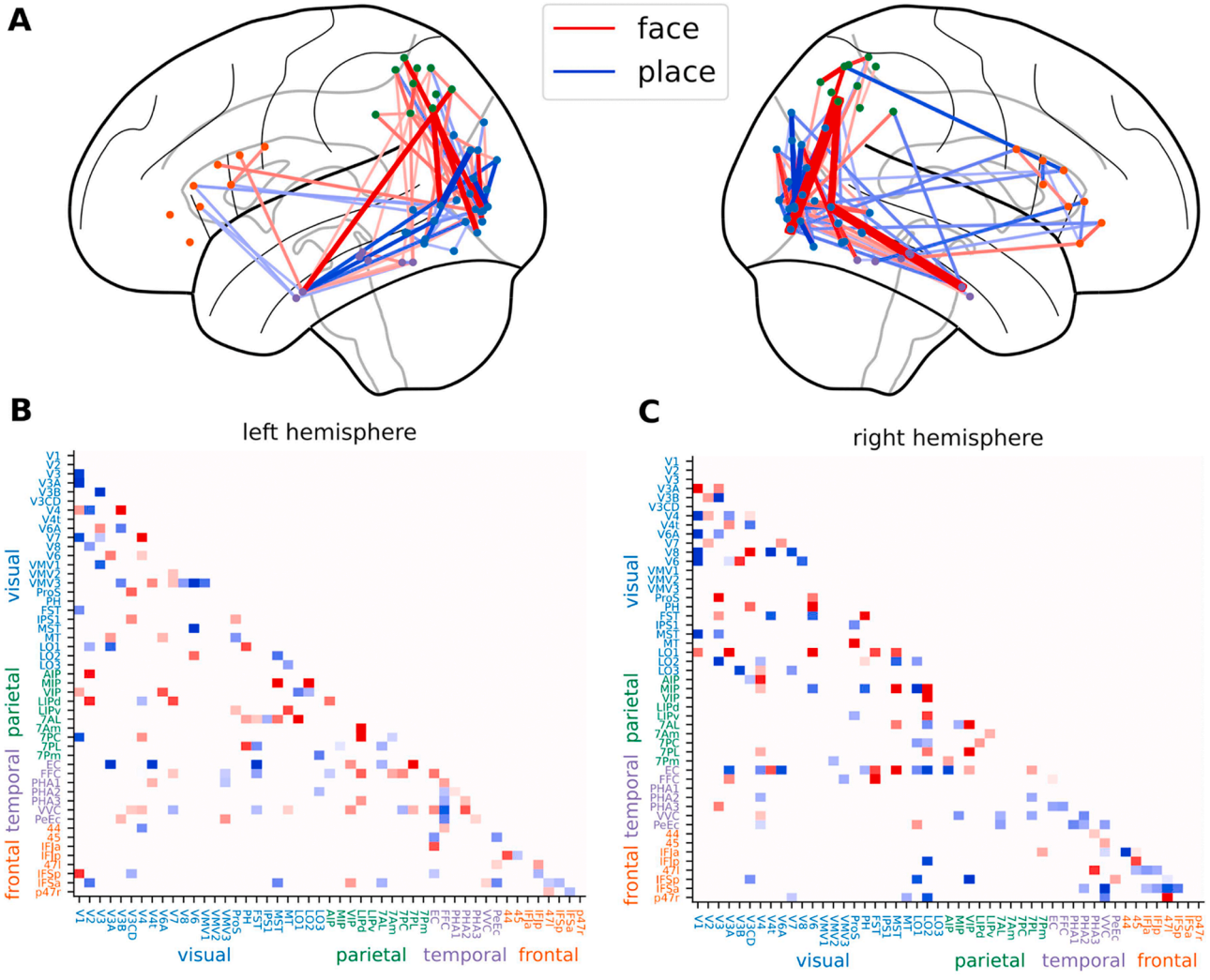


Fig. 6. Model inspection of the most informative connections for faces and places imagery decoding. The colored dots on the lateral view of the glass brain plot indicate the spatial locations of ROIs in visual (blue), parietal (green), inferior temporal (purple), and inferior frontal (orange) regions for each hemisphere. **(A)** Group-level edge map showing the most informative connections for covariance-based decoding of imagined faces (red) versus places (blue). **(B, C)** Group-level connectivity matrices displaying the names of ROI pairs for each hemisphere. ROI names are sorted and color-coded by their spatial location (posterior to anterior) and brain region (occipital to frontal). Only decoding coefficients with feature importance above the 95th percentile, determined via permutation importance, are displayed.

3.6. Sub-networks contribution to overall decoding performance

Finally, we tested the contribution from task-relevant sub-networks - including specific regions of interest (ROIs) - to covariance-based decoding. ROIs were selected based on previous literature and included: occipital areas (i.e., dorsal and ventral streams), parietal areas (i.e., inferior and superior parietal), temporal areas (i.e., inferior and medial temporal) and frontal areas (i.e., inferior frontal) (see Supplementary Methods for the complete list, and Fig. S6 for visualization, Glasser et al., 2016). In particular, the aim of this sub-network analysis was to further disentangle the contribution of short-range and long-range connections to overall decoding performance. By restricting the decoding analysis to subsets of the covariance matrices, we tested the relative contributions of short-range connections (e.g., including distributed nodes within the visual areas) and long-range connections (e.g., including distributed nodes between parietal and visual areas, temporal and visual areas, frontal and visual areas). In this case, since we tested multiple sub-networks at the same time we applied multiple comparisons correction. When testing the contribution of short-range connections (Fig. 7A), we obtained significant decoding results ($p < 0.01$, BF>8) using connections within the visual areas (from 2.5 to 5.5 s, light blue line) and within the parietal areas (from 2.5 to 3.5 s, light green line). Decoding within the temporal areas (purple line) and within the frontal areas (orange lines) was not significant. When testing the contribution of long-range connections (Fig. 7B), we obtained significant decoding results ($p < 0.01$, BF>8) between parietal and visual areas (from 1.5 to 3.5 s, cyan line), between temporal and visual areas (from 1.5 to 3.5 s, yellow line), and between frontal and visual areas (from 3.5 to 5.5 s, fuchsia line). Decoding between temporal and frontal areas was not significant. We also observed that the decoding was significant when using short-range connections within posterior cingulate areas and long-range connections between posterior cingulate and visual areas (see Fig. S7). This sub-network analysis revealed that both short-range and long-range connections are incremental to overall decoding performance.

To test how spatially specific these contributions from different sub-networks were, we ran a control analysis (Fig. S8) consisting in selecting task-irrelevant sub-networks that were little or not at all involved in the current visual imagery tasks, such as motor areas (i.e., premotor and motor) and auditory areas (i.e., primary and secondary auditory). We observed no significant decoding results for any of the short-range connections within these areas (Fig. S8 A) as well as the long-range connections between these areas (Fig. S8 B). This control analysis revealed that covariance-based decoding relies on spatially specific connectivity patterns associated with task-relevant sub-networks only.

4. Discussion

We investigated whether imagined faces and places are associated with distinct functional connectivity patterns reflecting category-specific inter-area communication associated with internal image generation. To do so, we used an experimental paradigm wherein vivid and detailed visual representations were generated internally, in the absence of any external stimulus. Such endogenous neural processes are challenging to decode due to their temporal misalignment across trials and because they involve the cooperation of different brain areas. To address these theoretical and methodological issues we introduced a covariance-based connectivity decoding method, originally designed for brain computer interface (BCI) applications. In particular, we used covariance-based decoding to read out endogenous functional connectivity variations associated with the mental imagery of faces and places. Our results reveal a category-specific network in which short-range and long-range interactions between brain regions can distinguish between face and place imagery. Moreover, we demonstrate the potential and suitability of the covariance-based decoding approach to answer key questions about the neural mechanisms underlying endogenous brain

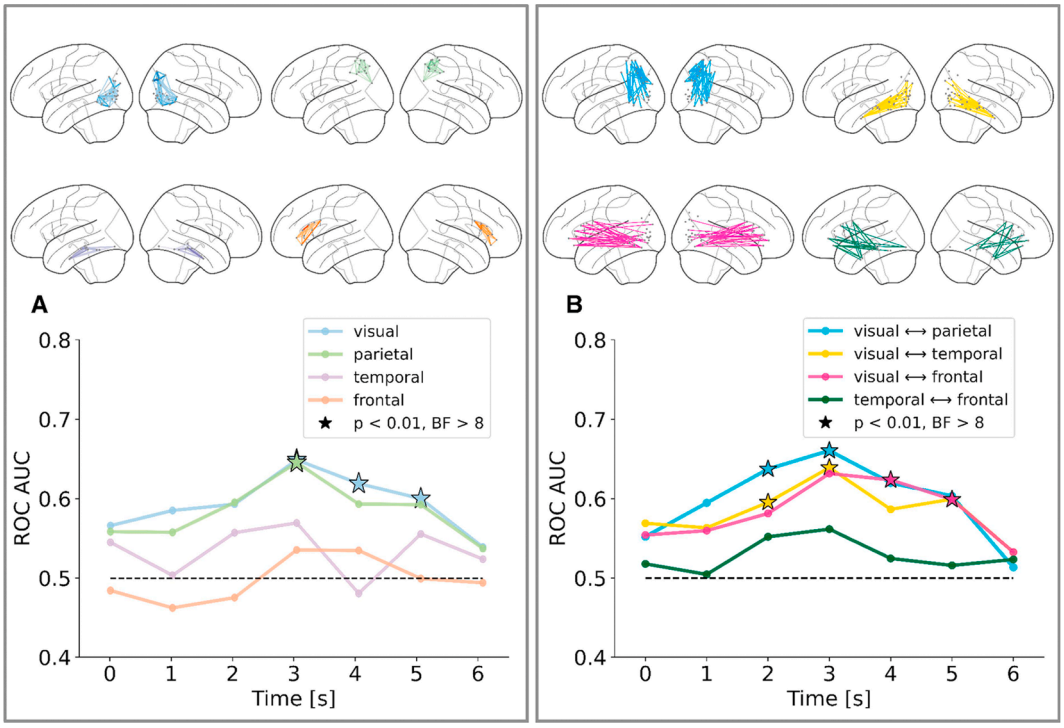


Fig. 7. Task-relevant sub-networks contribution to covariance-based decoding. (A-B) Task-relevant sub-networks. **(A)** Decoding results obtained using short-range connections within visual (light blue line), parietal (light green line), temporal (purple line) and frontal (orange line) areas. **(B)** Decoding results obtained using long-range connections between visual and parietal areas (cyan line), between visual and temporal areas (yellow line), between visual and frontal areas (fuchsia line) and between temporal and frontal areas (dark green line). For each sub-network, representative connections are shown on a lateral brain view using the same color-coding scheme.

signals in visual imagery.

One novelty of this study is the use of a minimally constrained experimental paradigm. Previous imagery studies often used a retro-cue paradigm, in which a pictorial cue is displayed on the screen, and participants are instructed to internally recreate that exact mental image, usually in a short amount of time. This paradigm assumes that imagery is similar to visual working memory (Tong, 2013), namely the internal replay and maintenance of a recently encountered image. Even though imagery events are better time-locked across trials when triggered by a pictorial cue, there are some methodological issues associated with the retro-cue paradigm. For instance, merely retrieving a recently presented visual stimulus is more constrained and arguably easier as it induces low–level visual features which a decoder can rely on. In contrast, we conceived of mental imagery as a constructive process based on the internal generation of images (Kosslyn, 1988) as opposed to a reproductive process based on a replay of recently seen images. Therefore, we did not use any pictorial aid as external reference but word cues only. To preclude the possibility that our decoder could rely on visual signals evoked by visual word cue presentation, the cueing period was separated in time from the subsequent imagination period by a jittered interval of about one second, and furthermore any remaining afterimages were eliminated by a dynamic phase scrambled mask. Moreover, we provided participants with a long imagery time window (6 s), assuming that truly internally re-constructing an image requires time. We cannot exclude the possibility that, among other things, we also decoded the memory trace of the semantic content associated with the word cues, even though this is unlikely, or at least not predominant, given the high loading on visual areas as well as the relationship between decoding performance and subjective vividness ratings. All these design choices were made to emphasize the internally generated aspects and to minimize potential stimulus-driven aspects. However, the costs of these experimental choices are high in terms of the methodological challenges thereby introduced. In particular, the temporal misalignment of endogenous signals across trials often renders time–domain decoding largely ineffective.

We demonstrate a novel methodological approach that can effectively read out purely endogenous signals which, until now, have been challenging to decode from neuroimaging data. Previous studies have investigated the temporal dynamics of visual imagery using time-domain decoding methods that were optimally tuned to decode fast transient changes in brain activity driven by external stimuli (Dijkstra et al., 2018). However, classic time-domain decoding methods are dramatically impeded by temporal misalignment across trials. The covariance-based decoding approach that we used here is well-suited to decode temporally misaligned signals in both experimental and simulated data. Another advantage of the covariance-based decoding method is its emphasis on functional connectivity dynamics unfolding over time. Statistical interdependence measures can be used to investigate large-scale brain network dynamics associated with different cognitive tasks (Bassett and Sporns, 2017; Bressler and Menon, 2010). For example, previous fMRI studies distinguished cognitive states (e.g., free recall, mathematical calculation) based on whole-brain functional connectivity (Shirer et al., 2012). However, accurate covariance estimation requires a substantial number of timepoints. Due to the low temporal resolution of fMRI, reliable covariance estimation is impractical, requiring long recording sessions (i.e., minutes) that fail to capture fast cognitive processes like memory, attention, or imagery. As an alternative, functional connectivity can be estimated over longer time windows to inform voxel selection for multi-voxel pattern analysis (MVPA) decoding within shorter time windows (Walbrin et al., 2024). However, this method does not explicitly use statistical interdependence measures as classifier inputs. In contrast, MEG provides sufficient temporal resolution for reliable covariance estimation over short time windows. Therefore, functional connectivity dynamics captured by spatial covariance matrices can be directly used as input features for classification. In this study, significant decoding results were obtained using a sliding window of just 500 ms.

The covariance-based decoding method offers several advantages, including simplicity, computational efficiency, and interpretability. The method is mathematically simple because it doesn't require specific assumptions about frequency and phase, unlike other functional connectivity metrics (e.g., coherence). It is interpretable because it provides information about the spatiotemporal connectivity patterns that allow discrimination between two classes. Information about reciprocal interconnections over time can also be captured by more complex decoding models, for instance graph neural networks. However, deep learning methods require extensive parameter tuning, which can be challenging to interpret, and a large number of trials, which is often impractical for neuroscience experiments. In contrast, covariance-based decoding offers a computationally efficient alternative, as it remains effective even with a limited number of trials. Importantly, we extend the covariance-based decoding approach to generate both sensor space and source space visualizations of the most informative functional connectivity patterns, which allows for a meaningful interpretation of contributing hubs and edges all over the cortical fold. In other words, because covariance-based decoding directly captures variations in functional connectivity over time, it allows us to identify the specific functional connections that provide relevant information for solving the cognitive task at hand.

To test the validity of the covariance-based decoding method for cognitive neuroscience applications, we performed a series of control analyses to ensure that decoding performance was related to the imagery task and was not driven by potential confounds (e.g. eye movements). A potential confound for every neural decoding study involving a visual task is that classification might be (at least partially) driven by eye movements. There is evidence for the fact that visual imagery, too, may be accompanied by oculomotor activation. For instance, when participants are instructed to imagine a recently seen grid pattern while looking at a blank screen, they produce oculomotor patterns which resemble the oculomotor patterns observed during perception of the actual grid pattern (Brandt and Stark, 1997). Systematic differences in eye movements associated with different imagery categories may affect neural decoding. Previous studies investigating visual perception, visual imagery and visual working memory used co-registered eye-tracking data to rule out potential eye movements confounds (Dijkstra et al., 2018; Mostert et al., 2018). In contrast, when eye movements are not controlled, there is evidence that they can partially explain neural decoding performance (Quax et al., 2019). The effect of eye movements on neural decoding performance can be explained by the overlap of their underlying neural generators with visual processing. For instance, there is evidence for an extensive overlap between brain areas activated by peripheral oculomotor activity and visual attention (Corbetta et al., 1998). In order to rule out potential eye movements confounds, we ran a control analysis using co-registered eye-tracking data. We first detected all trials associated with a high predictive probability based on the eye tracking data and then we removed all these trials with predictive eye movements from the MEG dataset. This control analysis is also important for the interpretation of the most informative functional connectivity patterns associated with covariance-based decoding since it ensures that these connections cannot be explained by systematic differences in eye movements.

Another important step was to show that classification accuracy correlates with participants' behavioral performance in the imagery task. Since there was no direct behavioral measure to assess whether participants performed well or not, we collected subjective vividness ratings for each trial. We expected to obtain better decoding results for those trials in which participants reported a highly vivid mental image. In line with this conjecture, we observed that higher vividness ratings were associated with higher decoding performance. Moreover, we assessed how well participants considered themselves able to internally generate vivid images in their mind's eye using the VVIQ. There is evidence for the fact that there are important individual differences in

visual imagery (Bainbridge et al., 2021; Keogh, and Pearson, 2018). In particular, the ability to generate mental images ranges from poorly vivid, almost absent imagery (i.e., aphantasia) to highly vivid, almost realistic imagery (i.e., hyperphantasia). We expected to obtain better decoding results for participants who considered themselves able to internally generate vivid mental images. In line with this prediction, we observed a significant positive correlation between decoding performance and VVIQ scores (Fig. 3).

The results align well with the existing literature on the neural mechanisms underlying visual imagery. We asked whether inter-area communication within visual areas can distinguish between object categories during visual imagery similarly to visual perception, regardless of whether an external stimulus is presented or not. Since faces and places involve different visual features (e.g., edginess, roundness, size) and different neural generators, we hypothesized that imagining these two different object categories will be associated with distinct functional connectivity patterns. Specifically, we expected that short-range connections within dorsal and ventral streams - including different brain areas representing specific visual features - will be associated with category-specific information integration associated with mental image generation. In line with our predictions, we obtained significant decoding results when using short-range connections within visual areas. Our results are consistent with previous studies suggesting that face and scene perception are associated with distinct brain networks (Epstein, and Baker, 2019; Ishai et al., 2005). Indeed, our sub-network analysis included brain areas that are considered to be part of both the face perception network (e.g., the fusiform or occipital face areas) and the scene perception network (e.g., the parahippocampal place area, the retrosplenial cortex, or the occipital place area). We also obtained significant decoding results when using short-range connections within parietal areas. Previous studies have shown that parietal areas are involved in feature integration during visual perception (Shafritz et al., 2002). There is also evidence from patients with parietal lesions who experience the clinical condition 'hemispatial neglect' that is associated with the incapability to visualize a visual hemifield both during perception and imagery (Bisiach, and Luzzatti, 1978). In line with this literature, we interpret category-specific short-range connections within the parietal areas as reflecting manipulation of spatial information that is necessary to achieve feature integration. Taken together, these findings suggest that category-specific information integration does not necessarily require constant monitoring of an external stimulus. In contrast, coordination between different brain areas associated with category-specific information integration is a basic computational mechanism that can be deployed even in the absence of an external stimulus.

A second question we aimed to address was whether top-down processing exerted from cognitive control mechanisms can be specific for different imagination categories. Previous studies have shown that imagery-related local activation occurs in temporal regions linked to memory retrieval (Kreiman et al., 2000; Slotnick et al., 2012), parietal regions associated with visuospatial attention (Andersson et al., 2019; Formisano et al., 2002), and frontal regions related to focal attention (Ishai et al., 2000; Yomogida et al., 2004). Rather than operating in isolation, these control areas coordinate within broader networks, such as the default mode network (Fox et al., Van Essen & Raichle, 2005) and the multiple-demand network (Duncan and Owen, 2000), during cognitive tasks like memory recall, daydreaming and attentional control. Moreover, there is evidence that, according to the global workspace theory (Dehaene et al., 1998), control areas constantly exchange information with sensory areas during effortful cognitive tasks (Schlegel et al., 2013). In line with this, we expected long-range connections between temporal and visual areas, parietal and visual areas, frontal and visual areas to be involved in visual imagery. However, it remained uncertain whether control areas provide distinct contributions depending on the specific content of visual imagery. To investigate this, we tested whether information contained in long-range connections can be used to distinguish the imagined categories (e.g., face versus place). The individual sub-networks results indicate that long-range connections between temporal and visual areas, parietal and visual areas, and frontal and visual areas contain category-specific information that is captured by covariance-based decoding to distinguish face and place imagery. Additionally, the most informative connections for face vs. place imagery decoding suggest that connections between temporal and frontal regions also play a crucial role. This finding aligns with previous studies highlighting their involvement in non-spatial attention tasks (Bedini, and Baldauf, 2021, 2023; Soyuhos, and Baldauf, 2023). The specificity of these long-range connections suggests that cognitive control mechanisms do not provide generalized support to visual cortex. Instead, they provide content-specific information that is tailored to different imagined categories (i.e., imagery/attentional templates, see Baldauf and Desimone, 2014).

It is important to point out that we cannot tell apart top-down from bottom-up information flow associated with long-range connections because covariance is an undirected functional connectivity measure. There is evidence suggesting that information predominantly flows top-down during imagery and bottom-up during perception (Dentico et al., 2014; Dijkstra et al., 2020; Mechelli et al., 2004; Dijkstra et al., 2017). During visual perception, all sensory information is available simultaneously, leading to the sequential activation of different cognitive processes (e.g., perception, attention, memory) in a processing cascade. In contrast, during visual imagery, a reverse processing cascade can be observed when brain activity is time-locked to an external stimulus, regardless of whether the imagery content is cue-induced or self-generated (Hu, and Yu, 2023). However, in minimally constrained experimental paradigms, different cognitive processes (e.g., memory, attention, perception) are activated asynchronously and often in varying orders across trials. In this study, the analysis of the most relevant connections for imagined category decoding does not indicate a clear dominance of specific connection types (e.g., control vs. sensory) at different time points. Instead, various connection types appear to contribute equally over time. In line with recent theoretical proposals based on converging experimental and neuropsychological evidence (Spagna et al., 2023), another possible explanation for the success of covariance-based decoding and the failure of time-domain decoding is that the category-selective network underlying visual imagery operates within a heterarchical rather than a hierarchical organization (Hochstein, and Ahissar, 2002). Within this framework, no single system (e.g., control vs. sensory) holds primary influence over the other (e.g., top-down vs. bottom-up); rather, both are equally essential for re-instantiating visual images, especially in minimally constrained experimental settings. Future studies will determine which proposed organization most accurately reflects the visual system's capacity to re-instantiate visual information, independent of the experimental paradigm.

This study primarily aimed to demonstrate the effectiveness of the covariance-based decoding method in reading out temporally misaligned brain signals, such as those associated with visual imagery. Additionally, we assessed the interpretability of the decoding results to confirm that the decoding model captures task-related brain activity. To support these findings, we demonstrated that decoding performance correlates with vividness self-reports and is driven by a category-selective functional connectivity network consistent with existing literature. However, a limitation of this study is the small sample size, which restricted a comprehensive analysis of the individual differences in visual imagery vividness as well as the specific organization of the category-selective functional connectivity network. Despite this limitation, the successful application of covariance-based decoding in our dataset suggests that future research with larger sample sizes can further explore these aspects using the same methodology.

The theoretical and methodological implications of this study extend well-beyond visual imagery. Mental imagery is one example of a purely internally driven cognitive process but there are many others: for

instance, endogenous visual attention and visual working memory. All such internally driven cognitive processes are not time-locked to an external stimulus and the endogenous brain activity associated with them is hard to detect because there is no overt behavior. Moreover, internally driven cognitive processes are typically associated with reciprocal interconnections between control areas and perceptual areas. We showed the suitability of the covariance-based decoding approach to answer questions about visual imagery. In addition, we suggest that the same method would be suitable to also answer research questions about visual attention and visual working memory. Our findings also have implications for visual prediction. Indeed, endogenous signals involving reciprocal interconnections across multiple areas play an important role in visual prediction, as suggested by analysis-by-synthesis (Yuille, and Kersten, 2006) and predictive coding (Rao and Ballard, 1999) theories. For instance, there is evidence from psychophysics that cueing upcoming images with visual and auditory word cues enhances subsequent visual detection (Lupyan, and Spivey, 2010). However, it is difficult to detect endogenous signals associated with visual prediction. We suggest that the covariance-based decoding approach also offers promising applications in that direction.

Taken together, we showed that imagined faces and places can be decoded from MEG signals using spatial covariance as a measure of functional connectivity. This finding has two implications for the understanding of visual imagery. On the one hand, we show that category-specific information integration in the visual cortex also occurs when there is no external stimulus. On the other hand, we show that reciprocal interconnections between cognitive control areas and perceptual areas are content-specific, i.e., different for imagined faces and places. To arrive at these conclusions, we used a minimally constrained experimental design that was structured to emphasize the internal generation of mental images. We proposed the application of a covariance-based decoding method originally designed for brain computer interface. We suggest that our successful application of covariance-based decoding to capture category specific functional connectivity associated with visual imagery paves the way for future applications to other internally driven cognitive processes, such as visual attention, visual working memory, and visual prediction.

Data availability

The data for the paper *Covariance-based decoding reveals a category-specific functional connectivity network for imagined visual objects* by Francesco Mantegna, Emanuele Olivetti, Philipp Schwedhelm and Daniel Baldauf is available from the corresponding author upon request (fmantegna93@gmail.com)

The code used to run the analyses can be found at DOI: 10.5281/zenodo.14536753 and will be publicly available as of the date of publication.

CRediT authorship contribution statement

Francesco Mantegna: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Emanuele Olivetti:** Writing – review & editing, Supervision, Software, Resources, Project administration, Methodology, Formal analysis, Data curation, Conceptualization. **Philipp Schwedhelm:** Writing – review & editing, Methodology, Investigation, Conceptualization. **Daniel Baldauf:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Fondazione Cassa di Risparmio di Trento e Rovereto. The authors would like to thank Gianpiero Manitolla and Davide Tabarelli for their technical support during data acquisition. The authors also thank Omri Raccah, Arianna Zuanazzi, Joan Orpella and David Poeppel for insightful comments on the manuscript.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2025.121171.

Data availability

Data will be made available on request.

References

Andersson, P., Ragni, F., Lingnau, A., 2019. Visual imagery during real-time fMRI neurofeedback from occipital and superior parietal cortex. Neuroimage 200, 332–343.
Bainbridge, W.A., Hall, E.H., Baker, C.I., 2021a. Distinct representational structure and localization for visual encoding and recall during visual imagery. Cerebral Cortex 31 (4), 1898–1913.
Bainbridge, W.A., Pounder, Z., Eardley, A.F., Baker, C.I., 2021b. Quantifying aphantasia through drawing: those without visual imagery show deficits in object but not spatial memory. Cortex 135, 159–172.
Baldauf, D., Desimone, R., 2014. Neural mechanisms of object-based attention. Science (1979) 344 (6182), 424–427.
Bannert, M.M., Bartels, A., 2018. Human V4 activity patterns predict behavioral performance in imagery of object color. J. Neurosci. 38 (15), 3657–3668.
Barachant, A. (2023). pyriemann 0.6.
Barachant, A., Bonnet, S., Congedo, M., Jutten, C., 2013. Classification of covariance matrices using a Riemannian-based kernel for BCI applications. Neurocomputing 112, 172–178.
Bartolomeo, P., Hajhajate, D., Liu, J., Spagna, A., 2020. Assessing the causal role of early visual areas in visual mental imagery. Nature Rev. Neurosci. 21 (9), 517. -517.
Bartolomeo, P., Liu, J., Spagna, A., 2024. Colors in the mind's eye. Cortex 170, 26–31.
Bassett, D.S., Sporns, O., 2017. Network neuroscience. Nat. Neurosci. 20 (3), 353–364.
Bastos, A.M., Schoffelen, J.M., 2016. A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. Front. Syst. Neurosci. 9, 175.
Bedini, M., Baldauf, D., 2021. Structure, function and connectivity fingerprints of the frontal eye field versus the inferior frontal junction: A comprehensive comparison. Eur. J. Neurosci. 54 (4), 5462–5506.
Bedini, M., Olivetti, E., Avesani, P., Baldauf, D., 2023. Connectomic investigation of the frontal eye field and inferior frontal junction. J. Vis. 23 (9), 5884. -5884.
Bisiach, E., Luzzatti, C., 1978. Unilateral neglect of representational space. Cortex 14 (1), 129–133.
Brandman, T., Peelen, M.V., 2017. Interaction between scene and object processing revealed by human fMRI and MEG decoding. J. Neurosci. 37 (32), 7700–7710.
Brandt, S.A., Stark, L.W., 1997. Spontaneous eye movements during visual imagery reflect the content of the visual scene. J. Cogn. Neurosci. 9 (1), 27–38.
Bressler, S.L., Menon, V., 2010. Large-scale brain networks in cognition: emerging methods and principles. Trends Cogn. Sci. (Regul. Ed.) 14 (6), 277–290.
Cai, T., Liu, W., Xia, Y., 2013. Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. J. Am. Stat. Assoc. 108 (501), 265–277.
Chen, Q., Garcea, F.E., Almeida, J., Mahon, B.Z., 2017. Connectivity-based constraints on category-specificity in the ventral object processing pathway. Neuropsychologia 105, 184–196.
Chen, Y., Wiesel, A., Eldar, Y.C., Hero, A.O., 2010. Shrinkage algorithms for MMSE covariance estimation. IEEE Trans. Signal Process. 58 (10), 5016–5029.
Cichy, R.M., Heinzle, J., Haynes, J.D., 2012. Imagery and perception share cortical representations of content and location. Cerebral Cortex 22 (2), 372–380.
Cichy, R.M., Pantazis, D., Oliva, A., 2014. Resolving human object recognition in space and time. Nat. Neurosci. 17 (3), 455–462.
Corbetta, M., Akbudak, E., Conturo, T.E., Snyder, A.Z., Ollinger, J.M., Drury, H.A., Shulman, G.L., 1998. A common network of functional areas for attention and eye movements. Neuron 21 (4), 761–773.
Dehaene, S., Kerszberg, M., Changeux, J.P., 1998. A neuronal model of a global workspace in effortful cognitive tasks. Proc. Natl. Acad. Sci. 95 (24), 14529–14534.
Dentico, D., Cheung, B.L., Chang, J.Y., Guokas, J., Boly, M., Tononi, G., Van Veen, B., 2014. Reversal of cortical information flow during visual imagery as compared to visual perception. Neuroimage 100, 237–243.
Dijkstra, N., Ambrogioni, L., Vidaurre, D., van Gerven, M., 2020. Neural dynamics of perceptual inference and its reversal during imagery. Elife 9, e53588.
Dijkstra, N., Mostert, P., de Lange, F.P., Bosch, S., van Gerven, M.A., 2018. Differential temporal dynamics during visual imagery and perception. Elife 7, e33904.

Dijkstra, N., Zeidman, P., Ondobaka, S., van Gerven, M.A., Friston, K., 2017a. Distinct top-down and bottom-up brain connectivity during visual perception and imagery. Sci. Rep. 7 (1), 1–9.
Dijkstra, N., Zeidman, P., Ondobaka, S., van Gerven, M.A., Friston, K., 2017b. Distinct top-down and bottom-up brain connectivity during visual perception and imagery. Sci. Rep. 7 (1), 1–9.
Douw, L., Nieboer, D., Stam, C.J., Tewarie, P., Hillebrand, A., 2018. Consistency of magnetoencephalographic functional connectivity and network reconstruction using a template versus native MRI for co-registration. Hum. Brain Mapp. 39 (1), 104–119.
Duncan, J., Owen, A.M., 2000. Common regions of the human frontal lobe recruited by diverse cognitive demands. Trends. Neurosci. 23 (10), 475–483.
Epstein, R.A., Baker, C.I., 2019. Scene perception in the human brain. Annu Rev. Vis. Sci. 5, 373.
Fischl, B., Sereno, M.I., Dale, A.M., 1999. Cortical surface-based analysis: II: inflation, flattening, and a surface-based coordinate system. Neuroimage 9 (2), 195–207.
Formisano, E., Linden, D.E., Di Salle, F., Trojano, L., Esposito, F., Sack, A.T., Goebel, R., 2002. Tracking the mind's image in the brain I: time-resolved fMRI during visuospatial mental imagery. Neuron 35 (1), 185–194.
Fox, M.D., Snyder, A.Z., Vincent, J.L., Corbetta, M., Van Essen, D.C., Raichle, M.E., 2005. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. Proc. Natl. Acad. Sci. 102 (27), 9673–9678.
Glasser, M.F., Coalson, T.S., Robinson, E.C., Hacker, C.D., Harwell, J., Yacoub, E., Van Essen, D.C., 2016. A multi-modal parcellation of human cerebral cortex. Nature 536 (7615), 171–178.
Goebel, R., Khorram-Sefat, D., Muckli, L., Hacker, H., Singer, W., 1998. The constructive nature of vision: Direct evidence from functional magnetic resonance imaging studies of apparent motion and motion imagery. Eur. J. Neurosci. 10 (5), 1563–1573.
Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., Hämäläinen, M., 2013. MEG and EEG data analysis with MNE-Python. Front. Neurosci. 267.
Guariglia, C., Padovani, A., Pantano, P., Pizzamiglio, L., 1993. Unilateral neglect restricted to visual imagery. Nature 364 (6434), 235–237.
Hämäläinen, M.S., Ilmoniemi, R.J., 1994. Interpreting magnetic fields of the brain: minimum norm estimates. Med. Biol. Eng. Comput. 32 (1), 35–42.
Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.D., Blankertz, B., Bießmann, F., 2014. On the interpretation of weight vectors of linear models in multivariate neuroimaging. Neuroimage 87, 96–110.
Haynes, J.D., Rees, G., 2006. Decoding mental states from brain activity in humans. Nature Rev. Neurosci. 7 (7), 523–534.
Hochstein, S., Ahissar, M., 2002. View from the top: hierarchies and reverse hierarchies in the visual system. Neuron 36 (5), 791–804.
Hu, Y., Yu, Q., 2023. Spatiotemporal dynamics of self-generated imagery reveal a reverse cortical hierarchy from cue-induced imagery. Cell Rep. 42 (10).
Hutchison, R.M., Culham, J.C., Everling, S., Flanagan, J.R., Gallivan, J.P., 2014. Distinct and distributed functional connectivity patterns across cortex reflect the domain-specific constraints of object, face, scene, body, and tool category-selective modules in the ventral visual pathway. Neuroimage 96, 216–236.
Ishai, A., Schmidt, C.F., Boesiger, P., 2005. Face perception is mediated by a distributed cortical network. Brain Res. Bull. 67 (1–2), 87–93.
Ishai, A., Ungerleider, L.G., Haxby, J.V., 2000. Distributed neural systems for the generation of visual images. Neuron 28 (3), 979–990.
Isik, L., Meyers, E.M., Leibo, J.Z., Poggio, T., 2014. The dynamics of invariant object recognition in the human visual system. J. Neurophysiol. 111 (1), 91–102.
Keogh, R., Pearson, J., 2018. The blind mind: no sensory visual imagery in aphantasia. Cortex 105, 53–60.
Kleiner, M., Brainard, D., Pelli, D., 2007. What's new in Psychtoolbox-3?.
Kosslyn, S.M., 1988. Aspects of a cognitive neuroscience of mental imagery. Science (1979) 240 (4859), 1621–1626.
Kosslyn, S.M., Thompson, W.L., Klm, I.J., Alpert, N.M., 1995. Topographical representations of mental images in primary visual cortex. Nature 378 (6556), 496–498.
Kreiman, G., Koch, C., Fried, I., 2000. Imagery neurons in the human brain. Nature 408 (6810), 357–361.
Linde-Domingo, J., Treder, M.S., Kerrén, C., Wimber, M., 2019. Evidence that neural information flow is reversed between object perception and object reconstruction from memory. Nat. Commun. 10 (1), 1–13.
Lupyan, G., Spivey, M.J., 2010. Making the invisible visible: verbal but not visual cues enhance visual detection. PLoS. One 5 (7), e11452.
Mahon, B.Z., Almeida, J., 2024. Reciprocal interactions among parietal and occipito-temporal representations support everyday object-directed actions. Neuropsychologia 198, 108841.
Makeig, S., Bell, A., Jung, T.P., Sejnowski, T.J., 1995. Independent component analysis of electroencephalographic data. Adv. Neural Inf. Process. Syst. 8.
Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG-and MEG-data. J. Neurosci. Methods 164 (1), 177–190.
Marks, D.F., 1973. Visual imagery differences in the recall of pictures. British J. Psychol. 64 (1), 17–24.
Mechelli, A., Price, C.J., Friston, K.J., Ishai, A., 2004. Where bottom-up meets top-down: Neuronal interactions during perception and imagery. Cerebral Cortex 14 (11), 1256–1265.
Moro, V., Berlucchi, G., Lerch, J., Tomaiuolo, F., Aglioti, S.M., 2008. Selective deficit of mental visual imagery with intact primary visual cortex and visual perception. Cortex 44 (2), 109–118.
Mostert, P., Albers, A.M., Brinkman, L., Todorova, L., Kok, P., De Lange, F.P., 2018. Eye movement-related confounds in neural decoding of visual working memory representations. eNeuro 5 (4).
Norman-Haignere, S.V., McCarthy, G., Chun, M.M., Turk-Browne, N.B., 2012. Category-selective background connectivity in ventral visual cortex. Cerebral Cortex 22 (2), 391–402.
O'Craven, K.M., Kanwisher, N., 2000. Mental imagery of faces and places activates corresponding stimulus-specific brain regions. J. Cogn. Neurosci. 12 (6), 1013–1023.
Olivetti, E., Veeramachaneni, S., Nowakowska, E., 2012. Bayesian hypothesis testing for pattern discrimination in brain decoding. Pattern. Recognit. 45 (6), 2075–2084.
Pearson, J., 2019. The human imagination: The cognitive neuroscience of visual mental imagery. Nature Rev. Neurosci. 20 (10), 624–634.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Duchesnay, E., 2011. Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.
Pesaran, B., Vinck, M., Einevoll, G.T., Sirota, A., Fries, P., Siegel, M., Srinivasan, R., 2018. Investigating large-scale brain dynamics using field potential recordings: Analysis and interpretation. Nat. Neurosci. 21 (7), 903–919.
Quax, S.C., Dijkstra, N., van Staveren, M.J., Bosch, S.E., van Gerven, M.A., 2019. Eye movements explain decodability during perception and cued attention in MEG. Neuroimage 195, 444–453.
Ragni, F., Lingnau, A., Turella, L., 2021. Decoding category and familiarity information during visual imagery. Neuroimage 241, 118428.
Rao, R.P., Ballard, D.H., 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat. Neurosci. 2 (1), 79–87.
Reddy, L., Tsuchiya, N., Serre, T., 2010. Reading the mind's eye: decoding category information during mental imagery. Neuroimage 50 (2), 818–825.
Sakai, K., Miyashita, Y., 1994. Visual imagery: an interaction between memory retrieval and focal attention. Trends. Neurosci. 17 (7), 287–289.
Salvucci, D.D., Goldberg, J.H., 2000. Identifying fixations and saccades in eye-tracking protocols. In: Proceedings of the 2000 symposium on Eye tracking research & applications, pp. 71–78.
Saygin, Z.M., Osher, D.E., Koldewyn, K., Reynolds, G., Gabrieli, J.D., Saxe, R.R., 2012. Anatomical connectivity patterns predict face selectivity in the fusiform gyrus. Nat. Neurosci. 15 (2), 321–327.
Schlegel, A., Kohler, P.J., Fogelson, S.V., Alexander, P., Konuthula, D., Tse, P.U., 2013. Network structure and dynamics of the mental workspace. Proc. Natl. Acad. Sci. 110 (40), 16277–16282.
Shafritz, K.M., Gore, J.C., Marois, R., 2002. The role of the parietal cortex in visual feature binding. Proc. Natl. Acad. Sci. 99 (16), 10917–10922.
Shirer, W.R., Ryali, S., Rykhlevskaia, E., Menon, V., Greicius, M.D., 2012. Decoding subject-driven cognitive states with whole-brain connectivity patterns. Cerebral Cortex 22 (1), 158–165.
Slotnick, S.D., Thompson, W.L., Kosslyn, S.M., 2012. Visual memory and visual mental imagery recruit common control and sensory regions of the brain. Cogn. Neurosci. 3 (1), 14–20.
Soyuhos, O., Baldauf, D., 2023. Functional connectivity fingerprints of the frontal eye field and inferior frontal junction suggest spatial versus nonspatial processing in the prefrontal cortex. Eur. J. Neurosci. 57 (7), 1114–1140.
Spagna, A., Heidenry, Z., Miselevich, M., Lambert, C., Eisenstadt, B.E., Tremblay, L., Bartolomeo, P., 2023. Visual mental imagery: evidence for a heterarchical neural architecture. Phys. Life Rev.
Sporns, O., 2002. Network analysis, complexity, and brain function. Complexity. 8 (1), 56–60.
Taulu, S., Simola, J., 2006. Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. Phys. Med. Biol. 51 (7), 1759.
Tong, F., 2013. Imagery and visual working memory: One and the same? Trends Cogn. Sci. (Regul. Ed.) 17 (10), 489–490.
Walbrin, J., Almeida, J., 2021. High-level representations in human occipito-temporal cortex are indexed by distal connectivity. J. Neurosci. 41 (21), 4678–4685.
Walbrin, J., Downing, P.E., Sotero, F.D., Almeida, J., 2024. Characterizing the discriminability of visual categorical information in strongly connected voxels. Neuropsychologia 195, 108815.
Yomogida, Y., Sugiura, M., Watanabe, J., Akitsuki, Y., Sassa, Y., Sato, T., Kawashima, R., 2004. Mental visual synthesis is originated in the fronto-temporal network of the left hemisphere. Cerebral Cortex 14 (12), 1376–1383.
You, K., Park, H.J., 2021. *Re*-visiting Riemannian geometry of symmetric positive definite matrices for the analysis of functional connectivity. Neuroimage 225, 117464.
Yuille, A., Kersten, D., 2006. Vision as Bayesian inference: nalysis by synthesis? Trends Cogn. Sci. (Regul. Ed.) 10 (7), 301–308.