



## Decoding category and familiarity information during visual imagery

Flavio Ragni<sup>a</sup>, Angelika Lingnau<sup>b</sup>, Luca Turella<sup>a,\*</sup>

<sup>a</sup> Center for Mind/Brain Sciences (CIMEC), University of Trento, Italy

<sup>b</sup> Institute of Psychology, University of Regensburg, Germany



### ARTICLE INFO

#### Keywords:

Visual imagery  
MVPA  
Faces  
Places  
Stimulus familiarity

### ABSTRACT

Visual imagery relies on a widespread network of brain regions, partly engaged during the perception of external stimuli. Beyond the recruitment of category-selective areas (FFA, PPA), perception of familiar faces and places has been reported to engage brain areas associated with semantic information, comprising the precuneus, temporo-parietal junction (TPJ), medial prefrontal cortex (mPFC) and posterior cingulate cortex (PCC). Here we used multivariate pattern analyzes (MVPA) to examine to which degree areas of the visual imagery network, category-selective and semantic areas contain information regarding the category and familiarity of imagined stimuli. Participants were instructed via auditory cues to imagine personally familiar and unfamiliar stimuli (i.e. faces and places). Using region-of-interest (ROI)-based MVPA, we were able to distinguish between imagined faces and places within nodes of the visual imagery network (V1, SPL, aIPS), within category-selective inferotemporal regions (FFA, PPA) and across all brain regions of the extended semantic network (i.e. precuneus, mPFC, IFG and TPJ). Moreover, we were able to decode familiarity of imagined stimuli in the SPL and aIPS, and in some regions of the extended semantic network (in particular, right precuneus, right TPJ), but not in V1. Our results suggest that posterior visual areas - including V1 - host categorical representations about imagined stimuli, and that stimulus familiarity might be an additional aspect that is shared between perception and visual imagery.

### 1. Introduction

If asked to locate a beloved person, for example a family member within a crowd, most of us would probably identify that person's face among many others in a considerably short amount of time. Due to their personal valence, personally familiar stimuli happen to be easier to recognize compared to unfamiliar stimuli (Visconti Di Oleggio Castello et al., 2017). This is also true when the quality of the visual input (i.e. picture) is degraded (Burton et al., 1999), or when attentional resources are reduced (Gobbini et al., 2013). This facilitation in recognizing familiar stimuli seems to be possible through the larger amount of associated features, i.e. subjective information and personal episodic knowledge, which are lacking in novel and unfamiliar stimuli (Gobbini et al., 2004; Gobbini and Haxby, 2007; Cloutier et al., 2011).

To identify the network of brain regions involved in the perception of personally familiar stimuli, previous studies mainly focused on two stimulus categories, i.e. faces and places. The network involved in processing personally familiar faces has been suggested to include a "core" network, comprising inferotemporal regions (e.g. the fusiform face area, FFA) responsible for extracting information about the visual appearance of faces, and an "extended" network involved in the retrieval of relevant semantic information, such as personal traits, intentions and attitudes of

that familiar person (Haxby et al., 2000). This "extended" semantic network for face perception has been proposed to include "Theory of Mind" areas, such as the medial prefrontal cortex (mPFC) and the temporo-parietal junction (TPJ), responsible for retrieving person knowledge (Cloutier et al., 2011), and anterior temporal cortices and precuneus, responsible for the retrieval of long-term episodic memories (Shah et al., 2001; Gobbini and Haxby, 2007; Haxby and Gobbini, 2012). Using multivariate pattern analyzes (MVPA), Visconti Di Oleggio Castello et al. (2017) showed that information about both the specific identity and the familiarity of perceived faces could be decoded within both the "core" (i.e. FFA, OFA) and the "extended" network for face perception (i.e. mPFC, TPJ, precuneus, IFG). Interestingly, familiarity of perceived faces was also encoded in the primary visual cortex (V1).

Regarding the perception of personally familiar places, Sugiura et al. (2005) identified a key role of two different brain regions: the precuneus, supporting episodic memory retrieval, and the posterior cingulate cortex (PCC), involved in the spatial representation and recognition of personally familiar places. In a more recent study, Silson et al. (2019) further extended these results, reporting the involvement of a similar portion of the medial parietal cortex during the recall of personally familiar places from memory. Moreover, a

\* Corresponding author.

E-mail address: [luca.turella@unitn.it](mailto:luca.turella@unitn.it) (L. Turella).

<https://doi.org/10.1016/j.neuroimage.2021.118428>.

Received 4 January 2021; Received in revised form 29 June 2021; Accepted 22 July 2021

Available online 24 July 2021.

1053-8119/© 2021 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

single case study exploring recognition mechanisms in a patient suffering from topographic agnosia highlighted a key role of the anterior parahippocampal cortex in mediating the recognition of personally familiar places (Van Assche et al., 2016).

To which degree are areas - sensitive to the familiarity of perceived stimuli - also involved in generating a mental image of a person or a place in the absence of a visual stimulus? Visual mental imagery has been interpreted as the reactivation of sensory representations stored in our memory by means of top-down mechanisms (Dentico et al., 2014; Dijkstra et al., 2017a; Mechelli et al., 2004; Pearson, 2019). In line with this idea, a growing number of studies showed that visual imagery recruits a similar set of brain regions as those involved in perception (for a meta-analysis, see Winlove et al., 2018). For example, several studies showed that visual mental imagery recruits retinotopically organized early visual cortices (Kosslyn and Thompson, 2003; Kosslyn, 2005; Winlove et al., 2018), but also parietal areas known to be recruited during visuospatial attention (Andersson et al., 2019; Formisano et al., 2002; Sack and Schuhmann, 2012; Slotnick et al., 2005). Moreover, visual mental imagery of specific stimulus categories (i.e. faces, places and objects) has been shown to recruit category selective regions within the inferior temporal cortex (i.e. FFA, PPA, lateral occipital complex - LOC; Ishai et al., 2000; O'Craven and Kanwisher, 2000). Using multivariate pattern analyses (MVPA), it has been demonstrated that it is possible to decode the identity of imagined stimuli in early visual cortices (Albers et al., 2013), parietal (Ragni et al., 2020) and inferior-temporal regions (Cichy et al., 2012; Reddy et al., 2011). Finally, several studies demonstrated that it is possible to predict imagined stimuli from activation patterns obtained during perception and vice versa (cross-decoding, Stokes et al., 2009; Albers et al., 2013; Ragni et al., 2020), indicating shared representations between perception and visual mental imagery.

Despite these similarities between imagery and perception, no study so far has explored to which degree it is possible to distinguish between visual mental imagery of personally familiar versus personally unfamiliar stimuli. To address this question, we instructed a group of participants to imagine different personally familiar and unfamiliar stimuli (i.e. faces and places) while measuring their brain activity using fMRI. Specifically, we asked in which brain regions it is possible to distinguish between (1) imagined stimulus categories (i.e., faces and places), and between (2) personally familiar and unfamiliar imagined faces and places. Using region-of-interest (ROI)-based MVPA, we focused on three sets of brain regions: (i) the visual imagery network, comprising key areas involved in the generation and maintenance of mental images (i.e. V1, SPL and aIPS); (ii) category-selective regions within inferior temporal cortex (i.e. FFA and PPA) - which are recruited during visual imagery of specific stimulus categories (i.e. faces and places); (iii) the extended semantic network for face perception, consisting in brain regions involved in the extraction of personal information from familiar stimuli (i.e. precuneus, mPFC, IFG and TPJ; Haxby et al., 2000; Visconti Di Oleggio Castello et al., 2017).

To anticipate our results, we were able to decode category information of imagined stimuli within occipital (V1) and parietal (SPL, aIPS) nodes of the visual imagery network. Similarly, we were able to distinguish between imagined stimulus categories in category-selective regions (FFA, PPA), and across the extended semantic network for face perception (i.e. precuneus, mPFC, IFG and TPJ). Familiarity of imagined stimuli, on the other hand, could be decoded (over and above vividness) in some regions of the visual imagery network (i.e. right SPL, right aIPS) and the extended semantic network for face perception (in particular, right precuneus and right TPJ), but not in V1 and category-selective areas. Our results suggest that posterior visual areas, including V1, contain category-specific information about imagined faces and places. Moreover, our results demonstrate that familiarity of imagined stimuli is encoded in higher-level semantic regions, extending previous findings on the perception of familiar and unfamiliar stimuli to visual mental imagery.

## 2. Materials & Methods

### 2.1. Participants

Twenty-five healthy volunteers participated in the study. All participants had normal or corrected-to-normal vision and had no history of neurological or psychiatric disease. Before taking part in the study, all participants gave their written informed consent. Due to excessive head movements during the scanning session (average relative mean displacement across all experimental runs > 1mm), two participants had to be excluded from the study, leading to a final sample of twenty-three participants (13 males, 10 females, mean age:  $26.6 \pm 3.8$ ). The study was approved by the Ethics Committee for research involving human participants at the University of Trento, Italy.

### 2.2. Setup

Visual stimuli were displayed on a liquid crystal monitor (Nordic NeuroLab, Norway; frame rate: 60 Hz; screen resolution:  $1920 \times 1200$  pixels). Participants lay horizontally in the scanner and viewed the screen ( $31.2^\circ \times 17.5^\circ$  of visual angle) binocularly via a rectangular mirror, positioned on the head coil. The auditory cue was delivered by means of MR-compatible earplugs (Siemens pneumatic earplugs). Button presses were collected via MR-compatible response buttons (Nordic NeuroLab Box, Nordic NeuroLab, Norway). Stimulus presentation, response collection and synchronization with the scanner were controlled using MATLAB 2017 (MathWorks, Natick, MA, U.S.A.) and the Psychtoolbox-3 for Windows (Brainard, 1997). Experiment presentation code is available on the Open Science Framework (<https://osf.io/ynp5z/>).

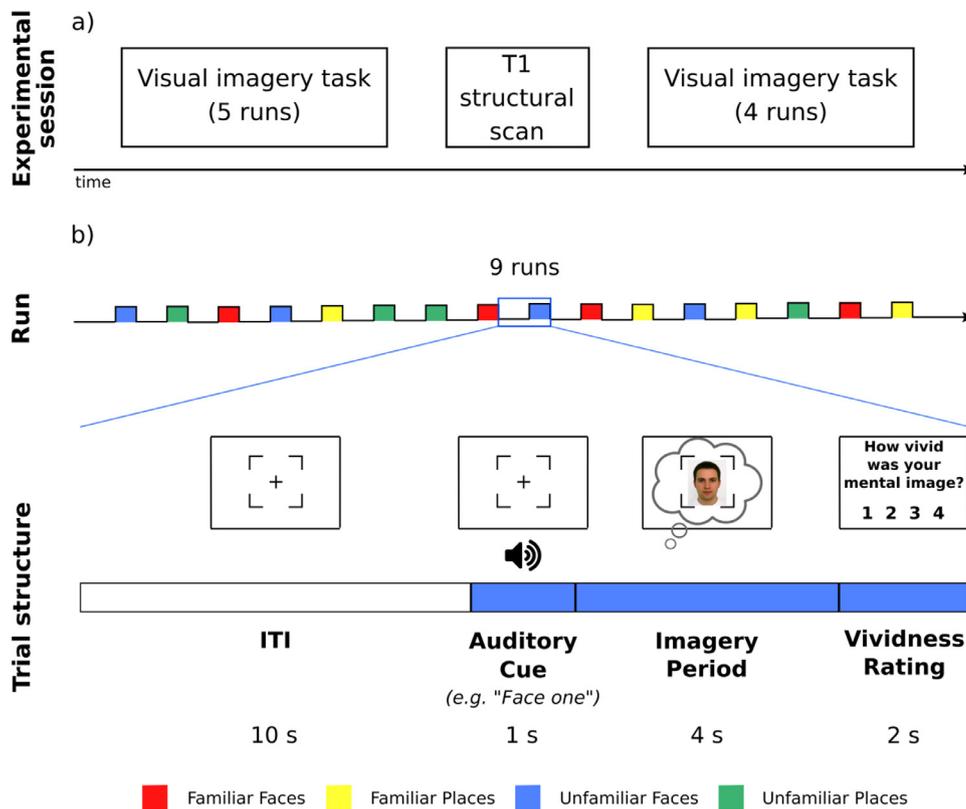
### 2.3. Stimuli

Stimuli consisted in pictures of different faces and places (4 exemplars each). We selected these two stimulus categories as their perception is known to preferentially recruit portions of high-level visual cortex (i.e. FFA and PPA; Epstein et al., 1999; Kanwisher and Yovel, 2006). Within each stimulus category, half of the stimuli (i.e. 2 exemplars) consisted of personally familiar stimuli, whereas the remaining half consisted of unfamiliar stimuli. Personally familiar faces were close-up photographs of two male or two female individuals with a close relationship with each participant (e.g. relatives, friends, partners, etc.). Personally familiar places, instead, were images of indoor spaces where participants felt comfortable (e.g. their own bedroom, living room, etc.). Participants provided images several days before the experimental session. A list of all personally familiar stimuli selected by participants is provided in the *Supplementary Materials (Table S1)*.

Each familiar stimulus was matched with an unfamiliar picture selected from two standard databases (Faces: Face Place Dataset, Tarr, 2008; Places: MIT Places-205 Dataset, Zhou et al., 2014). Faces were also matched for age, sex and ethnicity, whereas places were matched for the type of room. Pictures were scaled to  $155 \times 155$  pixels, corresponding approximately to  $4^\circ$  visual angle. Each stimulus was then associated with a name: unfamiliar stimuli were tagged with standard labels (i.e. "face one", "face two", "room one", "room two"), whereas familiar stimuli with names selected by each participant. These labels were used as auditory cues during the experimental task (see 2.4. *Experimental session and task* for more details).

### 2.4. Experimental session and task

Each participant completed a single experimental session, consisting of 9 functional runs ( $\sim 5$  min each; see Fig. 1a). Each functional run started and ended with a rest period (12 s at the beginning and 20 s at the end), and consisted of 16 trials, for a total of 144 trials per participant (36 trials for each condition). The conditions were embedded in a



**Fig. 1.** Layout of an experimental session, a single run and the trial structure. (a) Experimental session: participants performed a single experimental session, consisting in 9 experimental runs and a structural scan halfway through the experiment. (b) Each run (top panel) consisted of 16 trials, using an event-related design. Each trial (bottom panel) was preceded by an ITI (10 s) consisting of the presentation of a central fixation cross and a superimposed placeholder (10 s). Next, an auditory cue (1 s) instructed participants which stimulus to imagine (e.g. a face, identity '1') in the center of the placeholder, for 4 s. At the end of each trial, participants rated the vividness of the mental image they were able to generate on a 4-steps Likert scale, where 1 corresponded to low vividness and 4 to high vividness.

2 × 2 factorial design, with the factors: stimulus category (faces, places) and familiarity (personally familiar, unfamiliar). To examine patterns of brain activation during visual imagery of personally familiar and unfamiliar stimuli, we adopted an event-related design. The stimulus to be imagined was randomized across trials. Before entering the scanner, participants were requested to memorize the familiar and unfamiliar experimental stimuli and practiced the task for one experimental run.

Each trial of the visual imagery task started with a 10 s inter-trial interval (ITI in Fig. 1b), consisting in the presentation of a central fixation cross and a superimposed placeholder (Fig. 1b). The placeholder was represented by a square comprising 4° visual angle and served as a reference for the position and size of the mental image to be generated during the subsequent imagery period (Fig. 1b). Then, a verbal instruction (Auditory cue in Fig. 1b, duration: 1 s) instructed participants which stimulus to imagine in the current trial. After the cue, participants were instructed to imagine the corresponding stimulus as vividly as possible in the center of the placeholder (Imagery Period in Fig. 1b, duration: 4 s). Next, a question appeared on the screen (Vividness rating in Fig. 1, duration: 2 s), prompting participants to rate the vividness of the mental image they had previously generated on a 4-steps Likert scale (answer scale: 1. *Low vividness* – 4. *High vividness*). Participants were asked to indicate their response by button press. Across the entire trial, the only visual information present on the screen was the central fixation cross and the superimposed placeholder, and the question in the rating phase. At the end of the experimental session, participants filled out the Vividness of Visual Imagery Questionnaire (VVIQ; Marks, 1973). This questionnaire is aimed to assess individual differences in visual imagery abilities by asking the reader to imagine different scenarios (e.g. "Visualize a rising sun. Consider carefully the picture that comes before your mind's eye"), both with eyes open and eyes closed, and rate their vividness on a 5-steps Likert scale (answer alternatives: 1. *Perfectly clear and as vivid as normal vision*; 2. *Clear and reasonably vivid*; 3. *Moderately clear and vivid*; 4. *Vague and dim*; 5. *No image at all*).

## 2.5. Data acquisition

Magnetic resonance images were collected using a 3T Siemens MAGNETOM Prisma scanner equipped with a 64-channel head-coil. Functional data were acquired using a multiband EPI sequence (multi-band acceleration factor 3, TE/TR = 28/1000 ms, flip angle = 59°, matrix size = 64 × 64, 42 interleaved slices, in-slice resolution 3 mm × 3 mm). We acquired 294 volumes for each functional run, with axial slices slightly tilted to be approximately parallel to the calcarine sulcus in order to optimize brain coverage.

For each participant, we acquired a T1-weighted anatomical scan (MPRAGE; TR: 2140 ms; voxel resolution: 1 mm × 1 mm × 1 mm; TE: 2.9 ms; FA: 12°; FOV: 288 mm<sup>3</sup>; 208 slices; inversion time: 950 ms).

## 2.6. Behavioral data analyzes

For each participant, we computed average vividness rating for imagined faces and places, separately for familiar and unfamiliar stimuli. We performed a repeated measures ANOVA, with familiarity (2 levels) and stimulus category (2 levels) as factors. Post-hoc *t*-tests were adopted to compare specific conditions of interest.

## 2.7. fMRI data analyzes

### 2.7.1. Preprocessing

Data were preprocessed and analyzed using FSL 5.0.11 (FMRIB's Software Library, <https://fsl.fmrib.ox.ac.uk/fsl>) in combination with custom software written in MATLAB (MathWorks, Natick, MA, U.S.A.). Preprocessing included rigid-body motion correction to the mean image, followed by slice timing correction and high-pass temporal filtering (> 0.01 Hz). To prevent multivariate pattern (MVP) analyses from being affected by the removal of fine-scale information in activity patterns

(Kriegeskorte et al., 2006), we did not apply spatial smoothing to the functional data.

Each functional run was registered to the subject's anatomical high-resolution image with rigid body transformation (using the Boundary-Based Registration algorithm implemented in FSL 5.1; Greve and Fischl, 2009) and to the MNI152 2mm standard brain using linear transformation (FLIRT, 12 degrees of freedom; Jenkinson and Smith, 2001; Jenkinson et al., 2002). Visual inspection of the results of this preprocessing pipeline was performed for each participant. Due to excessive head motion, three functional runs of two different participants were excluded from the analyzes (average relative mean displacement > 1mm).

### 2.7.2. Univariate analyzes

To examine whether there are differences in univariate activation between familiar and unfamiliar stimuli, we performed a general linear model (GLM) analyzes. The first and second level analyzes (fixed effect) were performed in subject's individual space using FSL (fixed-effect model). To examine the amplitude of the blood-oxygen level-dependent (BOLD) response during visual imagery of familiar and unfamiliar stimuli, we created regressors for each factorial combination of stimulus category and familiarity, resulting in a total of 4 regressors (2 categories x 2 familiarity levels) for each experimental run. Regressors were time-locked to the start of the imagery delay (duration: 4 s). Moreover, regressors for the presentation of the auditory cue (time-locked to the onset of the auditory instruction; duration: 1 s) and response-phase (time-locked to the appearance of the question on the screen; duration: 2 s) were added to the model as nuisance regressors. Each regressor was convolved with a canonical hemodynamic response function (HRF). In addition, motion parameters (3 x rotation, 3 x translation) were added to the model as regressors of no interest.

Results from the second-level analyzes were then entered in a third-level analyzes across participants, performed using FSL's FLAME (mixed-effect model). For the resulting maps, we adopted a voxel-wise threshold of  $p < 0.001$  ( $z$  set at 3.1) and a (corrected) cluster-wise threshold ( $p = 0.001$ ) following Gaussian Random Field (GRF) theory (Worsley et al., 1996), which is embedded in the FSL's cluster routine.

### 2.7.3. ROI definition for multivariate pattern (MVP) analyzes

In this study, we aimed to investigate how neural representations of personally familiar stimuli are encoded within three different brain systems: the visual imagery network, category-selective areas of inferior temporal cortex and the extended semantic network for face perception. We extracted ROIs within these three networks by defining a sphere centered around a priori defined coordinates (radius 9 mm).

The visual imagery network is central for the internal generation and maintenance of mental images (Dijkstra et al., 2019a; Mechelli et al., 2004). This network encompasses the primary visual cortex (V1), and parietal areas (i.e. SPL and aIPS).

In addition, we identified ROIs in category-selective areas FFA and PPA in the inferior temporal cortex, which are known to be recruited when participants perceive or imagine specific stimulus categories (i.e. faces and places).

For ROIs of the extended semantic system for face perception, we included brain regions considered to be involved in the processing of personally familiar stimuli (Gobbini and Haxby, 2007; Haxby et al., 2000), such as the precuneus, the temporo-parietal junction (TPJ), the inferior frontal gyrus (IFG) and the medial prefrontal cortex (mPFC).

Specifically, we defined spherical ROIs centered around the coordinates of a previous study on the perception of personally familiar face stimuli (Visconti Di Oleggio Castello et al., 2017) for V1, FFA, precuneus, TPJ, mPFC and IFG. For SPL and aIPS spherical ROIs were centered on the peak probabilistic voxel of each region as defined by the Juelich Histological Atlas (SPL: area 7A; aIPS: area HIP3; Eickhoff et al., 2007). Since the Juelich Histological Atlas does not contain standard ROIs of category selective regions of inferior temporal cortex, the spherical ROIs for PPA were centered around peak values of a standard ROI selected

using an automated meta-analytic database (*Neurosynth*; Yarkoni et al., 2011). Considering that a more balanced comparison of effects might be achieved if ROIs are defined using the same criteria, we ensured FFA coordinates provided by Visconti Di Oleggio Castello et al. (2017) fell within the range of those provided by the same meta-analytic database used to define PPA (i.e. *Neurosynth*). All ROIs were visually inspected using FSLView to ensure there was no overlap between them (see Table S2 for the coordinates of the peaks of all our ROIs).

### 2.7.4. MVP analyzes

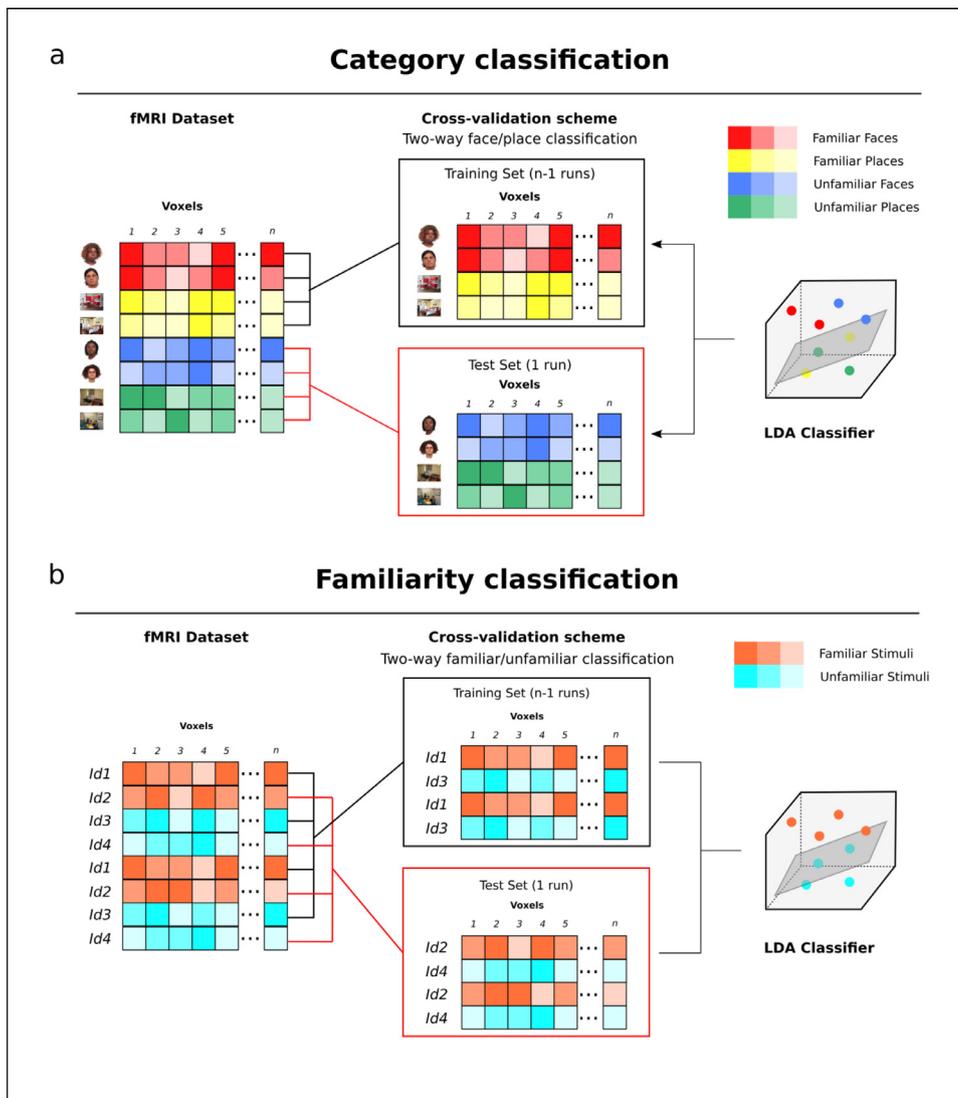
ROI-based MVP analyzes were implemented by means of a 2-class regularized linear discriminant analyzes (LDA) classifier using the CoS-MoMvpa Toolbox (Oosterhof et al., 2016). Estimates for the classification analyzes were generated using a general linear model (GLM) analyzes, performed in participant's individual anatomical space using FSL (fixed-effect model). We created regressors for each single trial within the 9 experimental runs. This led to a total of 144 estimates per participant (36 for each experimental condition). Regressors were time-locked to the start of the imagery delay (duration: 4 s). Moreover, regressors for the presentation of the auditory cue (time-locked to the onset of the auditory instruction; duration: 1 s) and the response-phase (time-locked to the appearance of the question on the screen; duration: 2 s) were added to the model as nuisance regressors. Each regressor was convolved with a canonical hemodynamic response function (HRF). Regressors were orthogonal to each other (maximum correlation between regressors within runs and participants:  $R^2 = 0.028$ ; average correlation:  $R^2 = 0.00016$ ). In addition, motion parameters (3 rotational and 3 translational parameters) resulting from 3D motion correction were added to the model as regressors of no interest, for a total of 24 regressors for each experimental run. Since  $t$ -values are computed dividing beta estimates by the estimate of the standard error, decoding on the basis of  $t$ -values has been shown to suppress the contribution of noisy voxels and therefore has been considered to be better suited for decoding (Misaki et al., 2010). We therefore performed MVP on  $t$ -values instead of beta values.

We performed two different MVP classification analyzes: (1) with the category classification analyzes, we aimed to identify brain regions representing the category (i.e. face or place) of imagined stimuli; (2) with the familiarity classification analyzes, we investigated which brain regions encode information relative to the familiarity of imagined stimuli, separately for faces and places.

*Cross-validation.* We adopted a leave-one-run-out cross-validation scheme.  $T$ -values were divided into training and test sets depending on the type of decoding. For *category classification* (i.e. faces vs places), we cross-validated across familiarity and runs (see Fig. 2a): the classifier was trained on  $t$ -values obtained during visual imagery of two familiar faces and two familiar places in  $n-1$  runs and tested on  $t$ -values corresponding to imagery of the left-out unfamiliar stimuli in the remaining run, and vice versa (18 cross-validation splits). This allowed us to control for familiarity information. For *familiarity classification*, we cross-validated across pairs of identities and runs (see Fig. 2b). Separately for faces and places, we trained the classifier on  $t$ -values corresponding to one familiar and one unfamiliar stimulus in  $n-1$  runs and tested on the  $t$ -values corresponding to two left-out identities (one personally familiar, one unfamiliar) in the remaining run, and vice versa (18 cross-validation splits). This allowed us to disregard identity information from familiarity classification.

### 2.7.5. ROI-based MVP

We performed classification analyzes across three separate networks: the visual imagery network (i.e. V1, SPL, aIPS), category-selective areas (i.e. FFA, PPA) and the extended semantic network for face processing (i.e. mPFC, precuneus, TPJ, IFG). For more information see 2.7.2. *ROI definition.* To assess significance of decoding accuracy, we performed permutation-based testing (Stelzer et al., 2013). For each ROI and classification analyzes, we created 100 random accuracy values for each



**Fig. 2.** Cross-validation schemes. Visual representation of the different cross-validation schemes adopted for (a) category classification (across familiarity and runs) and (b) familiarity classification (across pairs of identities and runs, separately for faces and places). In panel (b) the labels ID1, ID2, ID3 and ID4 represent the 4 different identities of face or place stimuli selected individually for each participant.

participant by permuting labels and performing classification. We then randomly selected one permuted accuracy value for each participant, computed the group average and repeated this operation 10000 times to create a null distribution of averaged accuracy values. Statistical significance of the decoding accuracy was computed with respect to this null distribution of group averaged accuracy values. Results were corrected for multiple comparisons (number of ROIs x number of tests) using a false discovery rate (FDR;  $q$ -value < 0.05; Benjamini and Yekutieli, 2001).

To replicate our ROI analyses and to identify additional brain regions representing category and familiarity of imagined faces and places, we also performed a whole brain searchlight-based MVPA (Kriegeskorte et al., 2006; Oosterhof et al., 2016). For details, see the *Searchlight-based MVPA* section in the *Supplementary Materials*.

Finally, we conducted three additional control analyses to test whether similarity between the auditory cues and/or vividness might partially explain the MVPA results.

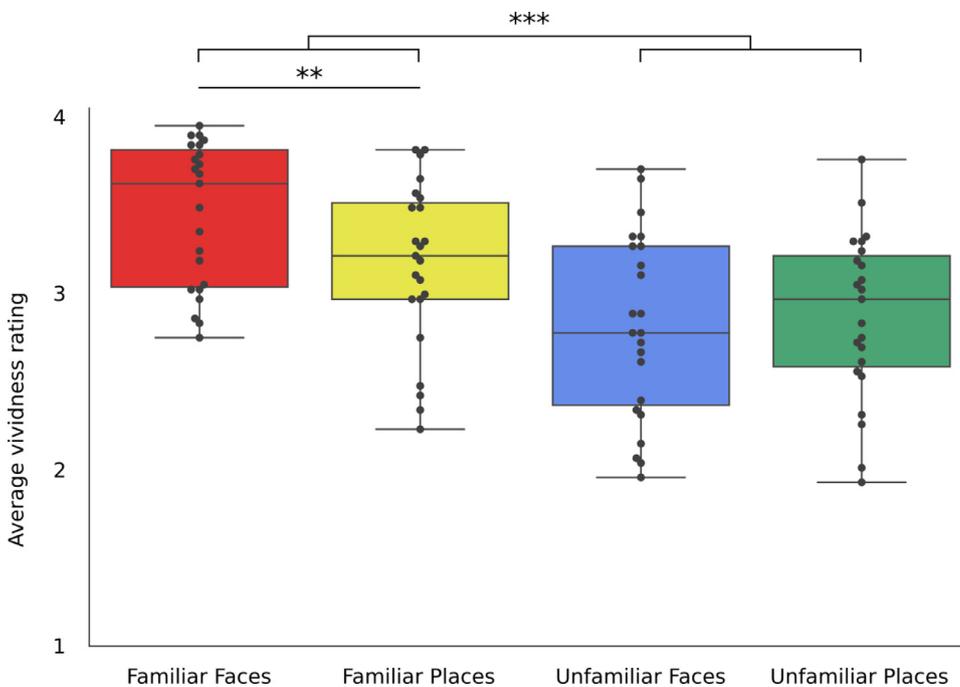
First, to exclude a potential influence of similarity between auditory cues for the different categories on the results of the decoding of familiarity, we performed a Representational Similarity Analyses (RSA), comparing the similarity between neural estimates of familiar and unfamiliar stimuli (for more detailed information see the section *The role of the auditory cue: Representational similarity analyses (RSA)* in the *Supplementary Materials*).

Second, to explore the potential impact of vividness on decoding analyses, we tried to directly decode vividness ratings provided by participants from patterns of activation using (a) a Support Vector Classification and (b) Support Vector Regression analyses (see *The role of vividness: Support Vector Classification (SVC)* and *The role of vividness: Support Vector Regression (SVR)* in the *Supplementary Materials*, for more information).

Third, we tested the significance of the differences between the control analyses (i.e., decoding of vividness) and the original analyses (i.e., decoding of familiarity). To this aim, we directly contrasted the classification accuracies for the SVC control analyses (decoding vividness information) with the classification accuracies of the two original decoding analyses (i.e., decoding familiarity information, separately for faces and places). For more detailed information please refer to section *Comparison between familiarity analyses and SVC analyses for vividness* in the *Supplementary Materials*.

#### 2.7.6. Data and code availability

Code and summary data for the ROI-based multivariate analyses are available on the Open Science Framework (<https://osf.io/ynp5z/>). The conditions of our ethics approval do not permit public archiving or peer-to-peer sharing of individual raw data. The raw functional and structural data of our study may be made available upon request after a confirmation from the ethical committee of our institution.



**Fig. 3.** Boxplot representing the distribution of vividness ratings ( $N=23$ ) as function of imagined stimulus category and familiarity. Data (black dots) represent average vividness ratings of individual participants, separately for each stimulus category. Each box encloses data comprised between the first (top edge) and third (bottom edge) quartile. The horizontal line within each box denotes the median. Whiskers indicate maximum (top) and minimum (bottom) values. Significance levels: one black asterisk,  $p < 0.05$ ; two black asterisks,  $p < 0.01$ ; three black asterisks,  $p < 0.001$ . Color scheme same as category decoding in Fig. 1.

### 3. Results

#### 3.1. Behavioral results

Vividness for each stimulus category (i.e. faces and places) was computed by averaging vividness ratings, separately for familiar and unfamiliar stimuli. Fig. 3 shows the distribution of vividness ratings as a function of stimulus category (i.e. face or place) and familiarity (i.e. familiar and unfamiliar stimuli). As expected, participants provided higher vividness ratings for visual mental imagery of familiar in comparison to unfamiliar stimuli [main effect of familiarity:  $F(1, 22) = 29.446$ ,  $p < 0.001$ ; average vividness familiar stimuli:  $3.34 \pm 0.47$ ; average vividness unfamiliar stimuli:  $2.87 \pm 0.5$ ]. We observed no difference in vividness ratings when comparing faces and places overall [main effect of stimulus category:  $F(1, 22) = 3.22$ ,  $p = 0.087$ ; average vividness faces:  $3.17 \pm 0.57$ ; average vividness places:  $3.05 \pm 0.49$ ]. However, visual imagery of familiar faces induced more vivid mental images compared to visual imagery of familiar places [interaction between stimulus category and familiarity:  $F(1, 22) = 6.503$ ,  $p = 0.018$ ;  $t(22) = 3.368$ ,  $p = 0.003$ ; average vividness familiar faces:  $3.49 \pm 0.41$ ; average vividness familiar places:  $3.19 \pm 0.48$ ]. By contrast, we obtained no difference in terms of vividness ratings between unfamiliar faces and places [ $t(22) = -0.542$ ,  $p = 0.593$ ; average vividness unfamiliar faces:  $2.84 \pm 0.53$ ; average vividness unfamiliar places:  $2.9 \pm 0.48$ ].

#### 3.2. Univariate analyses

We focused the univariate analyses on the familiarity effect, testing whether there was an increased activation associated with familiar compared to unfamiliar imagined stimuli (GLM contrast familiar > unfamiliar stimuli). Results from the mixed-effects analyses were projected on a segmented and inflated surface mesh of the two hemispheres in BrainVoyager QX 2.8.0 (Brain Innovation). As can be seen in Fig. 4, familiar stimuli elicited greater activation compared to unfamiliar stimuli within the bilateral mPFC, bilateral precuneus, left anterior temporal lobe and left angular gyrus. By contrast, we found no region where imagining unfamiliar stimuli was associated with an increase of activation compared to familiar stimuli.

#### 3.3. Multivariate pattern analyses

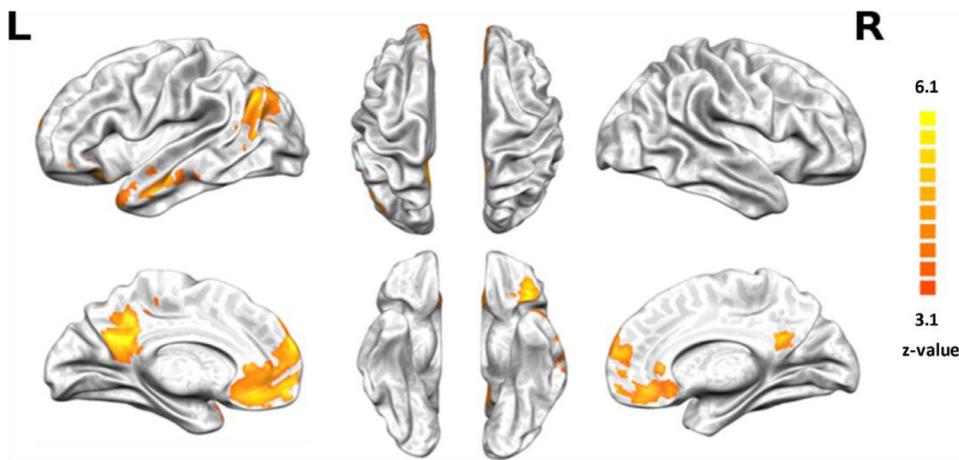
We performed a ROI-based multivariate pattern analyses to examine the representation of personally familiar stimuli (i.e. faces and places) within the visual imagery network, category-selective inferotemporal areas and the extended semantic network for face perception.

Regarding our first question, we were able to decode the category (faces, places) of imagined stimuli from patterns of activation in V1 and parietal (i.e. SPL, aIPS) regions of the visual imagery network (Fig. 5a), within category-selective areas (i.e. FFA and PPA; Fig. 5b) and the extended semantic network of face perception (i.e. precuneus, TPJ, mPFC, IFG; Fig. 5c). The corresponding statistics are shown in Table 1.

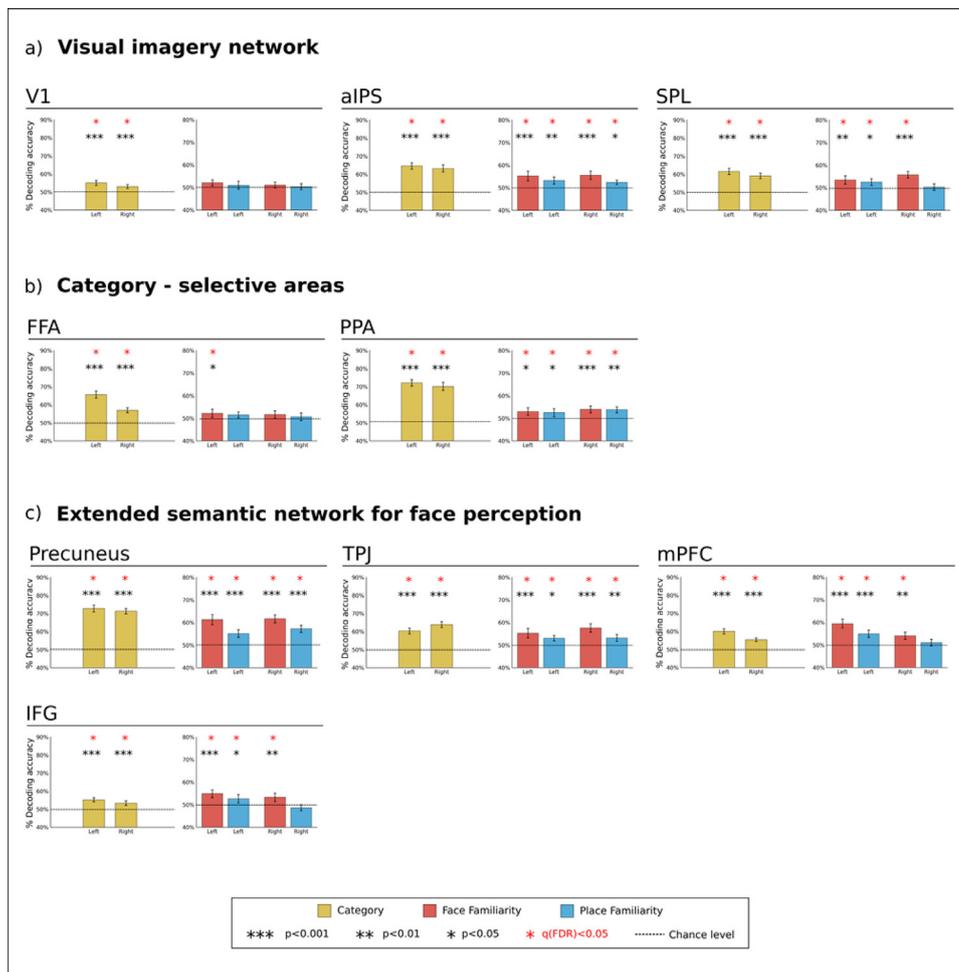
Regarding our second question, we were able to decode the familiarity of imagined faces and houses within parietal (i.e. SPL, aIPS) nodes of the visual imagery network, but not in V1 (Fig. 5a). Within category-selective areas, we were able to decode the familiarity of faces, but not places, in left FFA (Fig. 5b). By contrast, we were able to decode the familiarity of both places and faces in PPA. Finally, we were able to decode the familiarity of imagined faces and places within the extended semantic network (i.e. precuneus, TPJ, mPFC, IFG; Fig. 5c). The corresponding statistics are shown in Table 1. Moreover, the results of our ROI analyses are consistent with the results of the searchlight-based MVP analyses (see Supplementary Materials, Fig. S1-3).

To assess whether the familiarity decoding might be explained by differences in similarity of the adopted auditory cues and/or by the different levels of vividness of the imagined stimuli, we performed three additional control analyses. In the following sections, we will briefly describe these analyses (summarized in Fig. 6), whereas a more detailed description is provided in the Supplementary Materials.

First, we explored whether there was a possible influence of the similarity of the adopted auditory cues on familiarity decoding adopting RSA (see Supplementary Materials, page 5-9). The results of this analyses showed a possible effect of similarity between the auditory cues within several (i.e. left precuneus, left TPJ, left mPFC and right IFG), but not all of the examined ROIs (see Table S1), which might partially explain the decoding of familiarity in these regions. The result of the RSA analyses is summarized in Fig. 6: regions showing a significant effect in this analyses (and thus a potential role of the auditory cue) are highlighted with a dotted circle.



**Fig. 4.** Results of the univariate mixed-effects GLM contrast familiar > unfamiliar stimuli ( $N = 23$  participants). The z-map of group level activation was thresholded at  $p=0.001$  and cluster-correction was applied using GRF theory ( $p=0.001$ ). The resulting z-map was projected on a segmented and inflated surface mesh of the two hemispheres. L: Left hemisphere. R: Right hemisphere.



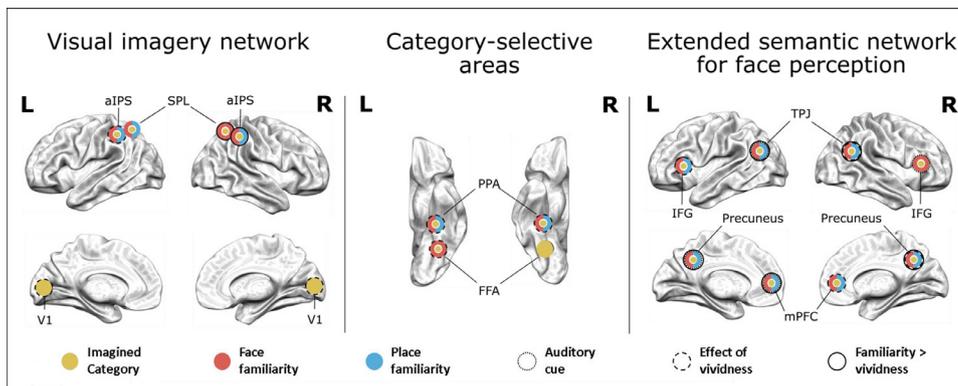
**Fig. 5.** Mean decoding accuracy of imagined stimuli, separately for the visual imagery network (i.e. V1, SPL and aIPS; panel (a)), category-selective areas (i.e. FFA and PPA; panel (b)) and the extended semantic network for face perception (i.e. precuneus, TPJ, mPFC and IFG; panel (c)). Statistical significance was assessed by means of permutation testing (10000 iterations) coupled with FDR-based correction for multiple comparisons (Benjamini and Yekutieli, 2001). Yellow: category decoding; blue: familiarity decoding of places; red: familiarity decoding of faces). Significance levels: one black asterisk,  $p < 0.05$ ; two black asterisks,  $p < 0.01$ ; three black asterisks,  $p < 0.001$ ; one red asterisk,  $q(\text{FDR}) < 0.05$ . Error bars: standard error of the mean (S.E.M.) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

The other two analyses, Support Vector Classification (SVC) and Support Vector Regression (SVR), aimed at assessing the possible role of vividness in driving the decoding of familiarity (see *Supplementary Materials, page 10–15*). These two analyses revealed a possible effect of vividness in some of the examined ROIs (i.e. bilateral V1, left aIPS, bilateral PPA, left FFA, left IFG, right mPFC, right precuneus and right TPJ; see *Figs. 6* and *S9,S10*). However, there were also regions in which we obtained an effect of vividness in the absence of a significant effect for the decoding of familiarity (e.g. V1). The results of the SVC and SVR analyses are summarized in *Fig. 6*: regions showing a significant

effect of vividness for one or both analyses are marked with a dashed circle.

Overall, the results of these control analyses seem to suggest that similarity between auditory cues and differences in vividness between categories might only partly explain the results of the familiarity decoding, as these effects were significant only within a subset of the examined ROIs, but not within all of them.

To complete our control analyses regarding the potential role of vividness, we directly compared the significance of the differences between the decoding accuracies for familiarity and vividness. This com-



**Fig. 6.** Visual representation of the results of the ROI-based classification analyses. **Left panel:** visual imagery network (i.e. V1, SPL and aIPS). **Middle panel:** category-selective areas (i.e. FFA and PPA). **Right panel:** extended semantic network for face perception (i.e. precuneus, TPJ, mPFC and IFG). Colored discs represent significant above-chance classification accuracy (color scheme similar to Fig. 5). Statistical significance was assessed by means of permutation testing (10,000 iterations) coupled with FDR-based correction for multiple comparisons (Benjamini and Yekutieli, 2001). Different line types around the disks represent a significant effect in one of the control analyses: dotted line (auditory cue effect, RSA ana-

lyzes), dashed line (vividness effect, SVR and/or SVC analyzes), continuous line (familiarity for face and/or place > vividness).

parison is possible only between the decoding accuracies for familiarity and the results of the SVC control analyzes for vividness, but it is not possible for the SVR analyzes for vividness (which provides normalized correlation coefficients as output). We thus compared the normalized decoding accuracies, separately for familiarity of faces and places, with the normalized classification accuracies for vividness (see *Supplementary Materials, pages 16-17*, for details). The result of this analyzes is summarized in Fig. 6: regions showing significant higher decoding of familiarity than for vividness are highlighted with a continuous circle.

We obtained a significant difference between the decoding accuracies for face familiarity and vividness within several ROIs (Fig. S11): 2 regions for the visual imagery network (Right SPL, Right aIPS) and 5 regions for the extended semantic network for face perception (Left mPFC, Left Precuneus, Right Precuneus, Left TPJ, Right TPJ). Decoding accuracy of place familiarity was significantly higher than decoding of vividness only in one region of the extended semantic network for face perception, i.e. right Precuneus (Fig. S12). In sum, the Right Precuneus was the only region where familiarity for both place and face were significantly higher than vividness.

#### 4. Discussion

Here we used multivariate pattern analyzes to investigate to which degree it is possible to decode: (1) the category and (2) the familiarity of imagined stimuli. We found that category information was encoded in regions of the visual imagery network, such as V1 and in parietal areas (SPL, aIPS), in category-selective areas of inferior temporal cortex (i.e. FFA and PPA) and the extended semantic network (precuneus, TPJ, mPFC, IFG). Familiarity of imagined faces and places, on the other hand, could be decoded more accurately than vividness from patterns of activation in parietal regions (left SPL, left aIPS) of the visual imagery network, in regions of the extended semantic network for face processing (precuneus, TPJ, mPFC), but not in V1 and within category-selective areas (i.e. FFA and PPA). In the following sections, we will discuss these results in more detail.

##### 4.1. Decoding category information of imagined stimuli

We were able to decode category (i.e. face vs. place) information about imagined stimuli from patterns of activation within V1. Previous studies reported encoding of imagined simple stimuli in V1, such as oriented gratings (Albers et al., 2013) or line drawings of individual letters, simple shapes and objects (Ragni et al., 2020). Regarding visual imagery of more complex stimuli (i.e. high-resolution pictures of different stimulus categories), Reddy et al. (2011) were able to decode the imagined stimulus category in category-selective areas (FFA, PPA), higher visual areas (LOC) and extrastriate visual areas (i.e. V2 – V4), but not in V1. Thus, to our knowledge the current study is the first showing decoding of the category of imagined stimuli within V1, expanding the knowledge

about the level of complexity of the stimuli that can be encoded in early visual cortices during visual imagery.

Beyond visual areas, we were able to decode category information of imagined stimuli within parietal nodes of the visual imagery network (i.e. aIPS, SPL), and within category-selective areas of the inferior temporal cortex (i.e. FFA, PPA). Parietal regions have repeatedly been shown to be recruited during visual imagery (Knauff et al., 2000; Formisano et al., 2002; Slotnick et al., 2005; Winlove et al., 2018), and SPL and aIPS have been reported to represent information about specific imagined stimulus exemplars (Ragni et al., 2020). Similarly, ventral temporal cortex has been shown to differentially respond during visual imagery of different stimulus categories (i.e. faces, houses and chairs: Ishai et al., 2000; faces and places: O’Craven and Kanwisher, 2000; faces: Ishai, 2002). Moreover, PPA has been shown to encode the identity of specific imagined places (Boccia et al., 2017; Johnson and Johnson, 2014). Our results confirm these findings, indicating that both FFA and PPA can host representations of category information during visual imagery.

Within the extended semantic network for face processing, we were able to decode the category of the imagined stimulus within all examined ROIs (i.e. precuneus, TPJ, mPFC and IFG). In a recent MVPA study, Visconti Di Oleggio Castello et al. (2017) presented participants with pictures of personally familiar and unfamiliar faces. The authors were able to decode identity information of perceived faces, independently of their familiarity, within high-level semantic regions, such as the precuneus, TPJ, mPFC and IFG. Our results indicate that the extended semantic network for face perception can represent familiarity-independent category information about a stimulus also during visual mental imagery.

##### 4.2. Decoding familiarity of imagined stimuli

Regarding familiarity decoding, given the univariate differences between imagery of familiar and unfamiliar stimuli in bilateral mPFC and precuneus, decoding results in these two areas are likely to be driven by these univariate differences rather than due to differences in the underlying patterns of activation. That said, we were able to decode familiarity of imagined faces and places from patterns of activation in parietal regions (SPL, aIPS) within the visual imagery network. As these regions have been frequently reported to be recruited during spatial tasks (for a meta-analyzes see Winlove et al., 2018), both SPL and aIPS have been considered to be part of a “dorsal” spatial network comprising parietal and premotor regions, involved in the representations of spatial configurations of imagined stimuli (Sack and Schuhmann, 2012). In addition, recent MVPA studies revealed representations of stimuli in parietal cortices during perception and imagery (Dijkstra et al., 2017a; Ragni et al., 2020), suggesting their role in detecting salient parts of imagined and perceived stimuli to support the orientation of top-down attention (Bogler et al., 2011; Dijkstra et al., 2019a). Our results suggest

**Table 1**  
Results of the ROI-based MVP analyses and corresponding statistics. Statistical significance was assessed by means of permutation testing (10000 iterations). Chance level: 50%. Red asterisks denote statistical tests surviving FDR-based correction for multiple comparisons (Benjamini and Yekutieli, 2001).

		Category classification			Familiarity classification - Faces			Familiarity classification - Places			
		Accuracy	Std Dev	p-value	Accuracy	Std Dev	p-value	Accuracy	Std Dev	p-value	
Visual Imagery Network	V1										
		Left	55.03%	6.58%	p<0.001*	52.08%	6.1%	p=0.055	50.98%	8.36%	p=0.256
		Right	53.07%	5.17%	p<0.001*	51.1%	6.34%	p=0.213	50.37%	6.43%	p=0.325
	SPL	Left	61.57%	8.22%	p<0.001*	53.47%	8.84%	p=0.0042*	52.52%	7.14%	p=0.02*
		Right	59.05%	7.19%	p<0.001*	55.7%	7.2%	p<0.001*	50.31%	7.01%	p=0.483
	aIPS	Left	64.61%	8.47%	p<0.001*	55.22%	7.2%	p<0.001*	53.24%	7.6%	p=0.0046*
Category - selective Areas		Right	63.19%	9.29%	p<0.001*	55.58%	8.75%	p<0.001*	52.37%	4.8%	p=0.0357*
	FFA	Left	65.63%	9.41%	p<0.001*	52.2	8.6%	p=0.038*	51.52%	6.42%	p=0.1327
		Right	56.95%	6.63%	p<0.001*	51.7%	7.8%	p=0.143	50.67%	8.11%	p=0.2888
	PPA	Left	72.08%	8.63%	p<0.001*	53.01%	7.94%	p=0.0114*	52.51%	8.48%	p=0.0226*
		Right	70.11%	10.63%	p<0.001*	53.95%	7.35%	p<0.001*	53.76%	6.47%	p=0.0032*
	mPFC	Left	60%	6.51%	p<0.001*	59.46%	9.36%	p<0.001*	54.95%	7.61%	p<0.001*
Extended Semantic Network for Face Perception		Right	55.37%	6.51%	p<0.001*	54.03%	7.59%	p=0.0017*	51.05%	7.13%	p=0.233
	Precuneus	Left	72.77%	9.48%	p<0.001*	61.29%	10.61%	p<0.001*	55.13%	8.16%	p<0.001*
		Right	71.3%	7.19%	p<0.001*	61.62%	8.44%	p<0.001*	57.18%	7.54%	p<0.001*
	TPJ	Left	60.32%	7.51%	p<0.001*	55.28%	10.03%	p<0.001*	53.08%	5.77%	p=0.0119*
		Right	63.82%	7.96%	p<0.001*	57.55%	9.04%	p<0.001*	53.17%	7.29%	p=0.008*
	IFG	Left	55.22%	5.31%	p<0.001*	54.89%	8.55%	p<0.001*	52.63%	8.77%	p=0.024*
	Right	55.32%	6.3%	p<0.001*	53.25%	8.98%	p=0.0069*	48.66%	6.32%	p=0.818	

that parietal regions also contribute to the distinction between imagined stimulus categories. Note that the control analyses showed that the effect of familiarity observed in parietal cortex might be partially (or fully) explained by vividness in the left hemisphere, but not in the right one (Figs. S9,S10): in the right SPL and aIPS, we were unable to decode vividness, and there was a significant difference between decoding familiarity and vividness (Fig. S11). Note that this effect was evident only for faces, suggesting a possible dissociation between imagined stimulus categories and/or hemisphere. Further studies will be required to examine the precise contribution of parietal regions to visual imagery of familiar stimuli.

We were unable to decode familiarity of imagined faces and places within V1. Using personally familiar face stimuli, [Visconti Di Oleggio Castello et al. \(2017\)](#) found distributed representations of familiarity information in early visual cortices (i.e. V1 - V3) during perception. Several fMRI studies suggested a feedback transfer of information to early visual cortex during the execution of different tasks: [Monaco et al. \(2020\)](#), for example, were able to decode different planned actions from patterns of activation within the primary visual cortex; [Vetter et al., \(2014, 2020\)](#) found encoding of real and imagined sounds within V1 in both normally sighted and congenitally blind individuals; [Bannert and Bartels \(2013\)](#) found representations of the prototypical colours of objects presented in black and white, and [Muckli et al., \(2015\)](#) were able to encode visual information in non-stimulated portions of V1, suggesting the existence of mechanisms of cortical feedback from higher-level visual areas ([Muckli and Petro, 2013](#)). Visual mental imagery has been argued to rely on a similar top-down mechanism, where information about imagined stimuli is transferred from prefrontal and parietal areas to occipital nodes of the visual imagery network ([Dijkstra et al., 2017a; Mechelli et al., 2004](#)). Our decoding results suggest that, as opposed to what has been observed during perception of personally familiar stimuli ([Visconti Di Oleggio Castello et al., 2017](#)), only category-specific, but not familiarity-related information, is possibly transferred to early visual cortex during imagery. Moreover, the results of the control analyses (Figs. 6 and S9,S10) point towards a specific role of V1 in representing information related to vividness of imagined stimuli, in line with previous research ([Dijkstra et al., 2017a](#)). In addition, V1 was the only region in which it was possible to decode vividness but not familiarity information, suggesting that vividness might be dissociated from familiarity in this region.

Regarding category-selective areas, we were able to decode the familiarity of imagined faces in left FFA and of imagined faces and places in PPA. The results in left FFA are in line with the observation that the familiarity of perceived faces can be decoded in bilateral FFA ([Visconti Di Oleggio Castello et al., 2017](#)). It is worth noting that we were able to decode familiarity of faces from the left but not from the right FFA. This asymmetry in the involvement of FFA during visual imagery of faces has been reported by previous studies, some indicating stronger recruitment within the right ([O'Craven and Kanwisher, 2000](#)) and others within the left hemisphere ([Ishai, 2002](#)). Note also that the left hemisphere has been reported to be more frequently activated by visual imagery tasks ([Winlove et al., 2018](#)).

Note that our control analyses revealed an effect of vividness in bilateral PPA and left FFA, and the effect of decoding of familiarity and vividness did not differ statistically in these regions (Figs. S9-13 and 6). Consequently, decoding of familiarity of imagined stimuli in category-selective regions might be driven by vividness, or a combination of vividness and familiarity. Further studies will be required to clarify this point.

A rich body of neuroimaging studies reported the involvement of regions of the extended semantic network of face perception in processing information about stimulus familiarity. Specifically, perceiving familiar faces has been reported to recruit, in addition to category selective regions within ventral temporal cortex (e.g. FFA and OFA. [Haxby et al., 2000; Natu and O'Toole, 2011](#)), a widespread network of brain regions ([Gobbini and Haxby, 2007; Haxby and Gobbini, 2012](#)). This network

includes different areas that allow us to recognize personally familiar individuals, such as mPFC and TPJ - possibly involved in the retrieval of personal knowledge (Cloutier et al., 2011) - and anterior temporal cortices and precuneus - engaged during the retrieval of long-term episodic memories (Shah et al., 2001). A similar engagement of this network has been reported during the perception of personally familiar places (Sugiura et al., 2005). These authors observed the recruitment of the precuneus, medial posterior cingulate cortex and retrosplenial cortex, in addition to PPA (Sugiura et al., 2005). Moreover, Visconti Di Oleggio Castello et al. (2017) were able to decode familiarity of perceived faces both within category-selective temporal regions (e.g. OFA, FFA) and semantic areas (i.e. mPFC, IFG, TPJ and STS).

Our control analyses further support the central role of the extended semantic network in processing the familiarity of imagined stimuli. Indeed, several regions within the network (left and right TPJ, left and right precuneus and the left mPFC) represented information about familiarity for face - over and above vividness information (Fig. S11). In addition, RSA showed that two regions of the network - right mPFC and right precuneus - represented familiarity of imagined stimuli irrespective of the information regarding the auditory cue. One of these two regions, right precuneus, represented familiarity - over and above vividness - both for faces and places. This ROI, which is part of the extended semantic network, was the only region across all examined ROIs showing all these properties. Overall, our findings further expand previous observations, indicating that visual imagery of personally familiar faces and places recruits a similar set of brain regions - comprising the extended semantic network - as those involved during perception.

The general functional overlap between bottom-up sensory stimulation and top-down mental imagery has been reported in several univariate (Winlove et al., 2018) and multivariate studies (Reddy et al., 2011; Lee et al., 2012; Albers et al., 2013; for a review see Dijkstra et al., 2019a), and across different domains (e.g. auditory: Zatorre and Halpern, 2005; motor: Monaco et al., 2020). Our results are in line with the existence of a shared neural substrate for perception and imagery. An open question remains whether visual imagery and perception of personally familiar stimuli rely on the same neural representations, especially in regions of the extended semantic network for face perception. Since a growing number of studies reported shared neural codes between visual imagery and perception in occipital (Albers et al., 2013), inferotemporal (Reddy et al., 2011; Stokes et al., 2011) and parietal (Dijkstra et al., 2017b; Ragni et al., 2020) nodes of the visual imagery network, a similar neural representation might underlie visual imagery and perception of personally familiar stimuli as well.

## 5. Conclusions

Our results revealed that during visual imagery of personally familiar faces and places, information about the stimulus category is distributed across the visual imagery network, category-selective areas and the extended semantic network of face perception. Familiarity information, instead, was represented within parietal regions of the visual imagery network and within some regions of the extended semantic network of face perception, but not in V1 - which seems to be mainly involved in processing vividness information. These results extend previous knowledge about the level of complexity of information that can be represented in early visual cortex during visual mental imagery. Moreover, our results suggest that stimulus familiarity might be an additional element that is shared across visual perception and visual imagery, extending a growing corpus of studies investigating the neural basis of visual mental imagery.

## Declaration of Competing Interest

The authors declare no competing financial interests.

## Credit authorship contribution statement

**Flavio Ragni:** Conceptualization, Methodology, Data curation, Formal analysis, Visualization, Investigation, Software, Writing - original draft, Writing - review & editing. **Angelika Lingnau:** Conceptualization, Methodology, Writing - original draft, Writing - review & editing, Funding acquisition. **Luca Turella:** Conceptualization, Methodology, Writing - original draft, Writing - review & editing, Funding acquisition.

## Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation programme of the European Research Council (ERC) grant agreement No. 339939 (Perceptual Awareness).

## Data and code availability statement

Code and summary data for the ROI-based multivariate analyses are available on the Open Science Framework (<https://osf.io/ynp5z/>). The conditions of our ethics approval do not permit public archiving or peer-to-peer sharing of individual raw data. The raw functional and structural data of our study may be made available upon request after a confirmation from the ethical committee of our institution.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.neuroimage.2021.118428](https://doi.org/10.1016/j.neuroimage.2021.118428).

## References

- Albers, A.M., Kok, P., Toni, I., Dijkerman, H.C., Lange, F.P. De, 2013. Report shared representations for working memory and mental imagery in early visual cortex 1-5. <https://doi.org/http://dx.doi.org/10.1016/j.cub.2013.05.065>
- Andersson, P., Ragni, F., Lingnau, A., 2019. Visual imagery during real-time fMRI neurofeedback from occipital and superior parietal cortex. *Neuroimage* 200, 332-343. [doi:10.1016/j.neuroimage.2019.06.057](https://doi.org/10.1016/j.neuroimage.2019.06.057).
- Bannert, M.M., Bartels, A., 2013. Decoding the yellow of a gray banana. *Curr. Biol.* 23, 2268-2272. [doi:10.1016/j.cub.2013.09.016](https://doi.org/10.1016/j.cub.2013.09.016).
- Benjamini, Y., Yekutieli, D., 2001. The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* 29, 1165-1188. [doi:10.1214/aos/1013699998](https://doi.org/10.1214/aos/1013699998).
- Boccia, M., Sulpizio, V., Palermo, L., Piccardi, L., Guariglia, C., Galati, G., 2017. I can see where you would be: Patterns of fMRI activity reveal imagined landmarks. *Neuroimage* 144, 174-182. [doi:10.1016/j.neuroimage.2016.08.034](https://doi.org/10.1016/j.neuroimage.2016.08.034).
- Bogler, C., Bode, S., Haynes, J.D., 2011. Decoding successive computational stages of saliency processing. *Curr. Biol.* 21, 1667-1671. [doi:10.1016/j.cub.2011.08.039](https://doi.org/10.1016/j.cub.2011.08.039).
- Brainard, D.H., 1997. The psychophysics toolbox. *Spat. Vis.* 10, 433-436. [doi:10.1163/156856897X00357](https://doi.org/10.1163/156856897X00357).
- Burton, A., Wilson, S., Cowan, M., Bruce, V., 1999. Face recognition in poor-quality video: evidence from security surveillance. *Psychol. Sci.* 10, 243-248.
- Cichy, R.M., Heinze, J., Haynes, J.D., 2012. Imagery and perception share cortical representations of content and location. *Cereb. Cortex* 22, 372-380. [doi:10.1093/cercor/bhr106](https://doi.org/10.1093/cercor/bhr106).
- Cloutier, J., Kelley, W.M., Heatherton, T.F., 2011. The influence of perceptual and knowledge-based familiarity on the neural substrates of face perception. *Soc. Neurosci.* 6, 63-75. [doi:10.1080/17470911003693622](https://doi.org/10.1080/17470911003693622).
- 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. [doi:10.1016/j.neuroimage.2014.05.081](https://doi.org/10.1016/j.neuroimage.2014.05.081).
- Dijkstra, Nadine, Bosch, S.E., van Gerven, M.A.J., 2019a. Shared neural mechanisms of visual perception and imagery. *Trends Cogn. Sci.* 23, 423-434. [doi:10.1016/j.tics.2019.02.004](https://doi.org/10.1016/j.tics.2019.02.004).
- Dijkstra, Nadine, Bosch, S.E., van Gerven, M.A.J., 2017a. Vividness of visual imagery depends on the neural overlap with perception in visual areas. *J. Neurosci.* 37, 1367-1373. [doi:10.1523/JNEUROSCI.3022-16.2016](https://doi.org/10.1523/JNEUROSCI.3022-16.2016).
- Dijkstra, N., Hinne, M., Bosch, S.E., van Gerven, M.A.J., 2019b. Between-subject variability in the influence of mental imagery on conscious perception. *Sci. Rep.* 9, 1-10. [doi:10.1038/s41598-019-52072-1](https://doi.org/10.1038/s41598-019-52072-1).
- Dijkstra, N., Zeidman, P., Ondobaka, S., Gerven, M.A.J., Van, Friston, K., 2017b. Distinct top-down and bottom-up brain connectivity during visual perception and imagery. *Sci. Rep.* 1-9. [doi:10.1038/s41598-017-05888-8](https://doi.org/10.1038/s41598-017-05888-8).
- Eickhoff, S.B., Paus, T., Caspers, S., Grosbras, M.H., Evans, A.C., Zilles, K., Amunts, K., 2007. Assignment of functional activations to probabilistic cytoarchitectonic areas revisited. *Neuroimage* 36, 511-521. [doi:10.1016/j.neuroimage.2007.03.060](https://doi.org/10.1016/j.neuroimage.2007.03.060).
- Epstein, R., Harris, A., Stanley, D., Kanwisher, N., 1999. The parahippocampal place area. *Neuron* 23, 115-125. [doi:10.1016/S0896-6273\(00\)80758-8](https://doi.org/10.1016/S0896-6273(00)80758-8).

- Formisano, E., Linden, D.E.J., Di Salle, F., Trojano, L., Esposito, F., Sack, A.T., Grossi, D., Zanella, F.E., Goebel, R., 2002. Tracking the mind's image in the brain I: Time-resolved fMRI during visuospatial mental imagery. *Neuron* 35, 185–194. doi:10.1016/S0896-6273(02)00747-X.
- Gobbini, M.I., Leibenluft, E., Santiago, N., Haxby, J.V., 2004. Social and emotional attachment in the neural representation of faces. *NeuroImage* 22 (4), 1628–1635. doi:10.1016/j.neuroimage.2004.03.049.
- Gobbini, M.I., Haxby, J.V., 2007. Neural systems for recognition of familiar faces. *Neuropsychologia* 45, 32–41. doi:10.1016/j.neuropsychologia.2006.04.015.
- Gobbini, M.I., Gors, J.D., Halchenko, Y.O., Rogers, C., Guntupalli, J.S., Hughes, H., Cipolli, C., 2013. Prioritized Detection of Personally Familiar Faces. *PLoS one* 8 (6), e66620. doi:10.1371/journal.pone.0066620.
- Greve, D.N., Fischl, B., 2009. Accurate and robust brain image alignment using boundary-based registration. *Neuroimage* 48, 63–72. doi:10.1016/j.neuroimage.2009.06.060.
- Haxby, J.V., Hoffman, E.A., Gobbini, M.I., 2000. The distributed human neural system for face perception. *Trends Cogn. Sci.* doi:10.1016/S1364-6613(00)01482-0.
- Haxby, J.V., Gobbini, M.I., 2012. Distributed Neural Systems for Face Perception. *Oxford Handbook of Face Perception* doi:10.1093/oxfordhb/9780199559053.013.0006.
- Ishai, A., 2002. Visual imagery of famous faces: effects of memory and attention revealed by fMRI. *Neuroimage* 17, 1729–1741. doi:10.1006/nimg.2002.1330.
- Ishai, A., Ungerleider, L.G., Haxby, J.V., 2000. Distributed neural systems for the generation of visual images. *Neuron* 28, 979–990. doi:10.1016/S0896-6273(00)00168-9.
- Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* 17, 825–841. doi:10.1006/nimg.2002.1132.
- Jenkinson, M., Smith, S., 2001. A global optimisation method for robust affine registration of brain images. *Med. Image Anal.* 5, 143–156.
- Johnson, M.R., Johnson, M.K., 2014. Decoding individual natural scene representations during perception and imagery. *Front. Hum. Neurosci.* 8, 59. doi:10.3389/fnhum.2014.00059.
- Kanwisher, N., Yovel, G., 2006. The fusiform face area: a cortical region specialized for the perception of faces. *Philos. Trans. R. Soc. B Biol. Sci.* doi:10.1098/rstb.2006.1934.
- Knauff, M., Kassubek, J., Mulack, T., Greenlee, M.W., 2000. Cortical activation evoked by visual mental imagery as measured by fMRI. *Neuroreport* 11, 3957–3962. doi:10.1097/00001756-200012180-00011.
- Kosslyn, S.M., 2005. Mental images and the Brain. *Cogn. Neuropsychol.* 22, 333–347. doi:10.1080/02643290442000130.
- Kosslyn, S.M., Thompson, W.L., 2003. When is early visual cortex activated during visual mental imagery? *Psychol. Bull.* 129, 723–746. doi:10.1037/0033-2909.129.5.723.
- Kriegeskorte, N., Goebel, R., Bandettini, P., 2006. Information-based functional brain mapping. *Proc. Natl. Acad. Sci. USA* 103, 3863–3868. doi:10.1073/pnas.0600244103.
- Lee, S., Kravitz, D.J., Baker, C.I., 2012. Disentangling visual imagery and perception of real-world objects. *Neuroimage* 59, 4064–4073. doi:10.1016/j.neuroimage.2011.10.055.
- Mechelli, A., Price, C.J., Friston, K.J., Ishai, A., 2004. Where bottom-up meets top-down: Neuronal interactions during perception and imagery. *Cereb. Cortex* 14, 1256–1265. doi:10.1093/cercor/bhh087.
- Misaki, M., Kim, Y., Bandettini, P., Kriegeskorte, N., 2010. Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. *Neuroimage* 53, 103–118. doi:10.1016/j.neuroimage.2010.05.051.
- Monaco, S., Malfatti, G., Culham, J.C., Cattaneo, L., Turella, L., 2020. Decoding motor imagery and action planning in the early visual cortex: Overlapping but distinct neural mechanisms. *Neuroimage* 218, 116981. doi:10.1016/j.neuroimage.2020.116981.
- Muckli, L., De Martino, F., Vizioli, L., Petro, L.S., Smith, F.W., Ugurbil, K., Goebel, R., Yacoub, E., 2015. Contextual feedback to superficial layers of V1. *Curr. Biol.* 25, 2690–2695. doi:10.1016/j.cub.2015.08.057.
- Muckli, L., Petro, L.S., 2013. Network interactions: Non-geniculate input to V1. *Curr. Opin. Neurobiol.* 23, 195–201. doi:10.1016/j.conb.2013.01.020.
- Natu, V., O'Toole, A.J., 2011. The neural processing of familiar and unfamiliar faces: a review and synopsis. *Br. J. Psychol.* 102, 726–747. doi:10.1111/j.2044-8295.2011.02053.x.
- O'Craven, K.M., Kanwisher, N., 2000. Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *J. Cogn. Neurosci.* 12, 1013–1023. doi:10.1162/08989290051137549.
- Oosterhof, N.N., Connolly, A.C., Haxby, J.V., 2016. CoSMoMVPA: multi-modal multivariate pattern analyzes of neuroimaging data in Matlab/GNU octave. *Front. Neuroinform.* 10, 27. doi:10.3389/fninf.2016.00027.
- Pearson, J., 2019. The human imagination: the cognitive neuroscience of visual mental imagery. *Nat. Rev. Neurosci.* doi:10.1038/s41583-019-0202-9.
- Ragni, F., Tucciarelli, R., Andersson, P., Lingnau, A., 2020. Decoding stimulus identity in occipital, parietal and inferotemporal cortices during visual mental imagery. *Cortex* 127, 371–387. doi:10.1016/j.cortex.2020.02.020.
- Reddy, L., Tsuchiya, T., Serre, T., 2011. Reading the mind's eye: decoding category information during mental imagery. *Neuroimage* 50, 818–825. doi:10.1016/j.neuroimage.2009.11.084.Reading.
- Sack, A.T., Schuhmann, T., 2012. Hemispheric differences within the fronto-parietal network dynamics underlying spatial imagery. *Front. Psychol.* 3, 1–10. doi:10.3389/fpsyg.2012.00214.
- Shah, N.J., Marshall, J.C., Zafiris, O., Schwab, A., Zilles, K., Markowitsch, H.J., Fink, G.R., 2001. The neural correlates of person familiarity. A functional magnetic resonance imaging study with clinical implications. *Brain* 124, 804–815. doi:10.1093/brain/124.4.804.
- Silson, E.H., Steel, A., Kidder, A., Gilmore, A.W., Baker, C.I., 2019. Distinct subdivisions of human medial parietal cortex are recruited differentially for memory recall of places and people. *bioRxiv* 1–25. https://doi.org/10.1101/554915
- Slotnick, S.D., Thompson, W.L., Kosslyn, S.M., 2005. Visual mental imagery induces retinotopically organized activation of early visual areas. *Cereb. Cortex* 15, 1570–1583. doi:10.1093/cercor/bhi035.
- Stelzer, J., Chen, Y., Turner, R., 2013. Statistical inference and multiple testing correction in classification-based multi-voxel pattern analyzes (MVPA): random permutations and cluster size control. *Neuroimage* 65, 69–82. doi:10.1016/j.neuroimage.2012.09.063.
- Stokes, M., Saraiva, A., Rohenkohl, G., Nobre, A.C., 2011. Imagery for shapes activates position-invariant representations in human visual cortex. *Neuroimage* 56, 1540–1545. doi:10.1016/j.neuroimage.2011.02.071.
- Stokes, M., Thompson, R., Cusack, R., Duncan, J., 2009. Top-down activation of shape-specific population codes in visual cortex during mental imagery. *J. Neurosci.* 29, 1565–1572. doi:10.1523/JNEUROSCI.4657-08.2009.
- Sugiura, M., Shah, N.J., Zilles, K., Fink, G.R., 2005. Cortical representations of personally familiar objects and places: Functional organization of the human posterior cingulate cortex. *J. Cogn. Neurosci.* 17, 183–198. doi:10.1162/0898929053124956.
- Tarr, M.J., 2008. Face - Place [WWW Document]. *Neural Basis of Cognitive*. Carnegie Mellon University.
- Van Assche, M., Kebets, V., Lopez, U., Saj, A., Goldstein, R., Bernasconi, F., Vuilleumier, P., Assal, F., 2016. Hurt but still alive: Residual activity in the parahippocampal cortex conditions the recognition of familiar places in a patient with topographic Agnosia. *NeuroImage Clin* 11, 73–80. doi:10.1016/j.nicl.2016.01.001.
- Vetter, P., Bola, L., Reich, L., Bennett, M., Muckli, L., Amedi, A., 2020. Decoding natural sounds in early "Visual" cortex of congenitally blind individuals. *Curr. Biol.* 30, 3039–3044. doi:10.1016/j.cub.2020.05.071, e2.
- Vetter, P., Smith, F.W., Muckli, L., 2014. Decoding sound and imagery content in early visual cortex. *Curr. Biol.* 24, 1256–1262. doi:10.1016/j.cub.2014.04.020.
- Visconti Di Oleggio Castello, M., Halchenko, Y.O., Guntupalli, J.S., Gors, J.D., Gobbini, M.I., 2017. The neural representation of personally familiar and unfamiliar faces in the distributed system for face perception. *Sci. Rep.* 7, 12237. doi:10.1038/s41598-017-12559-1.
- Winlove, C.I.P., Milton, F., Ranson, J., Fulford, J., MacKisack, M., Macpherson, F., Zeman, A., 2018. The neural correlates of visual imagery: a co-ordinate-based meta-analyzes. *Cortex* 105, 4–25. doi:10.1016/j.cortex.2017.12.014.
- Yarkoni, T., Poldrack, R.A., Nichols, T.E., Van Essen, D.C., Wager, T.D., 2011. Large-scale automated synthesis of human functional neuroimaging data. *Nat. Methods* 8, 665–670. doi:10.1038/nmeth.1635.
- Zatorre, R.J., Halpern, A.R., 2005. Mental concerts: musical imagery and auditory cortex. *Neuron* 47, 9–12. doi:10.1016/j.neuron.2005.06.013.
- Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., Oliva, A., 2014. Learning deep features for scene recognition using places database. In: *Proceedings of the Advances in Neural Information Processing Systems*, pp. 487–495.