

Decoding stimulus identity in occipital, parietal and inferotemporal cortices during visual mental imagery

Flavio Ragni¹, Raffaele Tucciarelli^{1,2}, Patrik Andersson^{1,3}, and Angelika Lingnau^{1,4,5*}

¹*Center for Mind/Brain Science (CIMEC), University of Trento, 38068 Rovereto (TN), Italy*

²*Department of Psychological Sciences, Birkbeck, University of London, Malet Street, WC1E 7HX*

³*Stockholm University Brain Imaging Centre (SUBIC), 106 91 Stockholm, Sweden*

⁴*Department of Psychology, Royal Holloway University of London, TW20 0EX, Egham (London)*

UK

⁵*Institute of Psychology, University of Regensburg, 93053 Regensburg, Germany*

Manuscript accepted for publication

*Corresponding author: Angelika Lingnau, University of Regensburg, 93053 Regensburg, Germany, e-mail: angelika.lingnau@ur.de

DOI: <https://doi.org/10.1016/j.cortex.2020.02.020>

Please cite this article as: Ragni F, Tucciarelli R, Andersson P, Lingnau A, Decoding stimulus identity in occipital, parietal and inferotemporal cortices during visual mental imagery, CORTEX, <https://doi.org/10.1016/j.cortex.2020.02.020>.

This article is distributed under the terms of the Creative Commons Attribution ([CC BY-NC-ND 4.0](https://creativecommons.org/licenses/by-nc-nd/4.0/)).

Abstract

In the absence of input from the external world, humans are still able to generate vivid mental images. This cognitive process, known as visual mental imagery, involves a network of prefrontal, parietal, inferotemporal, and occipital regions. Using multivariate pattern analysis (MVPA), previous studies were able to distinguish between the different orientations of imagined gratings, but not between more complex imagined stimuli, such as common objects, in early visual cortex (V1). Here asked whether letters, simple shapes, and objects can be decoded in early visual areas during visual mental imagery. In a delayed spatial judgment task, we asked participants to observe or imagine stimuli. To examine whether it is possible to discriminate between neural patterns during perception and visual mental imagery, we performed ROI-based and whole-brain searchlight-based MVPA. We were able to decode imagined stimuli in early visual (V1, V2), parietal (SPL, IPL, aIPS), inferotemporal (LOC) and prefrontal (PMd) areas. In a subset of these areas (i.e. V1, V2, LOC, SPL, IPL and aIPS), we also obtained significant cross-decoding across visual imagery and perception. Moreover, we observed a linear relationship between behavioral accuracy and the amplitude of the BOLD signal in parietal and inferotemporal cortices, but not in early visual cortex, in line with the view that these areas contribute to the ability to perform visual imagery. Together, our results suggest that in the absence of bottom-up visual inputs, patterns of functional activation in early visual cortex allow distinguishing between different imagined stimulus exemplars, most likely mediated by signals from parietal and inferotemporal areas.

Introduction

As human beings, we strongly rely on visual perception to process information coming from the external world. At the same time, we are able to create a mental representation of a percept in the absence of visual input, e.g. when we try to visualize the last place where we used our wallet when searching for it. This ability, often referred to as “seeing with the mind’s eye”, is called visual mental imagery.

During perception, information is processed along a pathway from the retina over the optic nerve, the optic chiasm and the lateral geniculate nucleus to the early visual cortex, responsible for processing low-level features of perceived stimuli (e.g. position, orientation, edges, contours), and to inferotemporal and parietal cortices, responsible for higher level visual processing, visual guidance and control of behavior, respectively (Ungerleider & Haxby, 1994). By contrast, during visual mental imagery, a top-down organization has been suggested in which sensory representations of external stimuli, e.g. objects in inferotemporal areas, are reenacted by means of signals coming from prefrontal areas (Mechelli et al., 2004; Dijkstra et al., 2017). The same top-down modulation has been proposed to exert an influence also over parietal areas, and to be involved in attentional mechanisms and in the representation of the spatial configuration of imagined stimuli (Sack et al., 2012). According to depictive theories of visual imagery (Kosslyn, 1981; 2005), this top-down modulation ultimately leads to the recruitment of primary visual cortex (V1), serving as a “dynamic blackboard” (Bullier, 2001) where mental images acquire their resemblance to real percepts.

There is a lack of a general consensus regarding the role of V1 during visual imagery. A number of positron emission tomography (PET; Kosslyn et al., 1993) and functional magnetic resonance imaging (fMRI) studies (Amedi et al., 2005; Klein et al., 2000; Ishai et al., 2002; Slotnick et al., 2005) demonstrated a recruitment of V1 during visual mental imagery tasks. By contrast, other studies failed to observe any reliable recruitment of V1 (Ishai et al., 2000; Formisano et al., 2002; Sack et al., 2002), or found a deactivation of V1 (Mellet et al., 2000; for a review, see Kosslyn & Thompson, 2003).

The advent of multivariate pattern analysis (MVPA; Haxby, 2001; Kriegeskorte et al., 2006) allowed exploring the representational content of V1 during visual imagery tasks. Albers et al. (2013) asked participants to imagine gratings with different orientations. Despite an overall low level of the blood-oxygen level dependent (BOLD) signal in early visual areas, decoding of imagined grating orientation based on the patterns of activation was possible in V1. Other studies investigated whether more complex imagined stimuli could be encoded in V1. Stokes et al. (2009) asked participants to imagine or observe two simple letters ('X' or 'O'). Results revealed that the identity of the imagined stimuli could be decoded from patterns of activation in high-level visual areas (i.e. LOC). Decoding was possible also in primary visual cortex, but only at a more liberal statistical threshold. Similar findings were later reported by Lee et al. (2012): in their study, participants imagined or observed complex pictures of 10 different real-world objects, differing widely in their orientation, shape and color. Classification analyses revealed significant encoding of imagined stimuli in visual areas anterior to V1 (i.e. V2, V3, V4, LO and pFS), but not in primary visual cortex.

Another debated issue within the literature on visual mental imagery is the degree to which patterns of activation are shared between bottom-up stimulation (i.e. perception) and top-down internal generation (i.e. imagery). Albers et al. (2013) demonstrated cross-decoding for the orientation of perceived and imagined gratings in V1. Likewise, Lee et al. (2012) found shared representations for imagery and perception of different complex objects in both striate and extrastriate visual areas. However, using different stimulus categories (e.g. faces, places, body parts), other studies found common neural patterns for imagery and perception in high-level visual areas (e.g. FFA, PPA and LOC; Stokes et al., 2009; 2011; Reddy et al., 2010; Cichy et al., 2011), but not in early visual cortices. To date, the degree of complexity of stimuli represented in primary visual cortex during imagery, and the extent to which these representations share similar neural codes with perception, is still unclear. Here, we aimed to examine whether stimuli that are more complex than gratings can be decoded in early visual cortex during visual imagery, and to which extent information about imagined stimuli is

shared with visual perception. To this aim, we asked participants to either imagine ('imagery' task) or observe ('perception' task) lowercase letters, simple shapes or objects.

To examine whether it is possible to distinguish between stimulus exemplars on the basis of neural patterns during the imagery and the perception condition in early visual cortex, we used region-of-interest (ROI)-based and a whole-brain searchlight-based MVPA (Kriegeskorte et al. 2006). Behavioral performance during the fMRI experiment was measured using a delayed spatial judgment task. We were able to distinguish between imagined stimulus exemplars pertaining to the three categories in early visual (V1 and V2), parietal (SPL, IPL and aIPS), inferotemporal (LOC) and prefrontal (PMd) areas. Moreover, in a subset of these regions (i.e. V1, V2, LOC, SPL, IPL and aIPS), we also obtained significant cross-decoding across visual imagery and perception. Our results are in line with the view that, during visual mental imagery, complex stimulus information is passed down to early visual cortex from higher-level brain regions even in the absence of retinal stimulation (Dentico et al., 2014; Vetter et al., 2014; Dijkstra et al., 2017).

Materials & Methods

We report how we determined our sample size, all data exclusions, all inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all manipulations, and all measures in the study.

Participants

Twenty-nine healthy volunteers participated in the study, with the sample size determined on the basis of previous studies using similar designs and analyses (Albers et al., 2013; Dijkstra et al., 2017). 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 artifacts in functional MR images, data from three participants had to be excluded from the study. Moreover, due to poor performance in the behavioral task (see section *Data analysis, behavioral data*), five additional participants were excluded, leading to a final sample of twenty-one participants (11 males, 10 females, mean age 26.1 ± 3.8). The study was approved by the Ethics Committee for research involving human participants at the University of Trento, Italy.

Setup

Visual stimuli were back-projected to a screen via a liquid crystal projector (OC EMP 7900, Epson Nagano, Japan; frame rate: 60 Hz; screen resolution: 1280x1024 pixels). Participants laid horizontally in the scanner and viewed the screen binocularly via a rectangular mirror ($17.8^\circ \times 13.4^\circ$ of visual angle), positioned on the head coil. The auditory cue was delivered by means of MR-compatible headphones (SereneSound, Resonance Technology, Inc.). Button presses were collected via MR-compatible response buttons (Lumina LP 400, Cambridge Research Systems). Stimulus presentation, response collection and synchronization with the scanner were controlled using “ASF” (Schwarzbach, 2011), based on MATLAB (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/nhv2r/>).

Stimuli

Stimuli consisted of six images pertaining to three stimulus categories: lowercase letters, simple shapes and objects (2 exemplars each; *Figure 1*) and are available on the Open Science Framework (<https://osf.io/nhv2r/>). These stimuli were selected based on the results of a behavioral experiment that made use of the eccentricity effect, i.e. faster generation of mental images near versus far from fixation (Marzi et al., 2006). This effect has been interpreted as a sign of the involvement of retinotopically organized visual cortex during visual imagery. We thus reasoned that imagined stimuli exhibiting a higher eccentricity effect might be more likely to recruit early visual cortex during imagery (see *Supplementary Material, Stimulus Selection and Validation* for additional information). To ensure that participants actively engaged in the visual mental imagery task during the fMRI experiment, we designed a delayed spatial judgment paradigm. Participants had to judge whether a black dot (0.5 degree visual angle, briefly presented on the screen) fell within or outside the stimulus they just had imagined (see *Experimental design, Imagery task*, for details). To this purpose, we defined a set of 20 non-overlapping dot positions for each individual stimulus exemplar: half of these positions were inside the outline of the stimulus (“matching” dot positions), and the other half outside the outline (“non-matching” dot positions; *Figure 1*).

To ensure that participants were able to perform the spatial judgment task for imagined stimuli, and that both “matching” and “non-matching” dots had comparable difficulty levels across stimulus categories, we performed a behavioral pilot experiment on an independent sample of N=6 participants. Results showed that all participants were able to perform the task above chance [mean accuracy: 72.5% ± 4.9%. One sample t-test: $t(5) = 11.2$, $p < 0.001$], both for “matching” and for “non-matching” dots [mean accuracy “matching” dots: 70.6% ± 10%; mean accuracy “non-matching” dots: 74.4% ± 8.3%. Paired samples t-test: $t(5) = 0.596$, $p = 0.577$]. The spatial judgment task proved to be

more difficult for visual imagery of lowercase letters [mean accuracy: $61.7\% \pm 8.7\%$] in comparison to visual imagery of objects [mean accuracy: $77.1\% \pm 8.8\%$] and simple shapes [mean accuracy: $78.8\% \pm 7.5\%$; repeated measures ANOVA; factors: stimulus category; main effect of stimulus category: $F(2, 10) = 7.766, p = 0.009$].

To prevent afterimages following stimulus presentation, we created visual masks, consisting of a phase-scrambled image of each stimulus exemplar by means of a Fourier transformation implemented in Matlab. To this purpose, original stimuli were imported in Matlab and a fast Fourier transform was applied. The fast Fourier transform converted images from the spatial to the frequency domain, which includes both amplitude and phase information. We then added a random phase to the phase angle of the transformed image, and applied an inverse Fourier transform to the modified image to revert it from frequency domain to spatial domain. With this procedure, we obtained a total of 6 masks, one for each individual stimulus exemplar (see *Figure 1*).

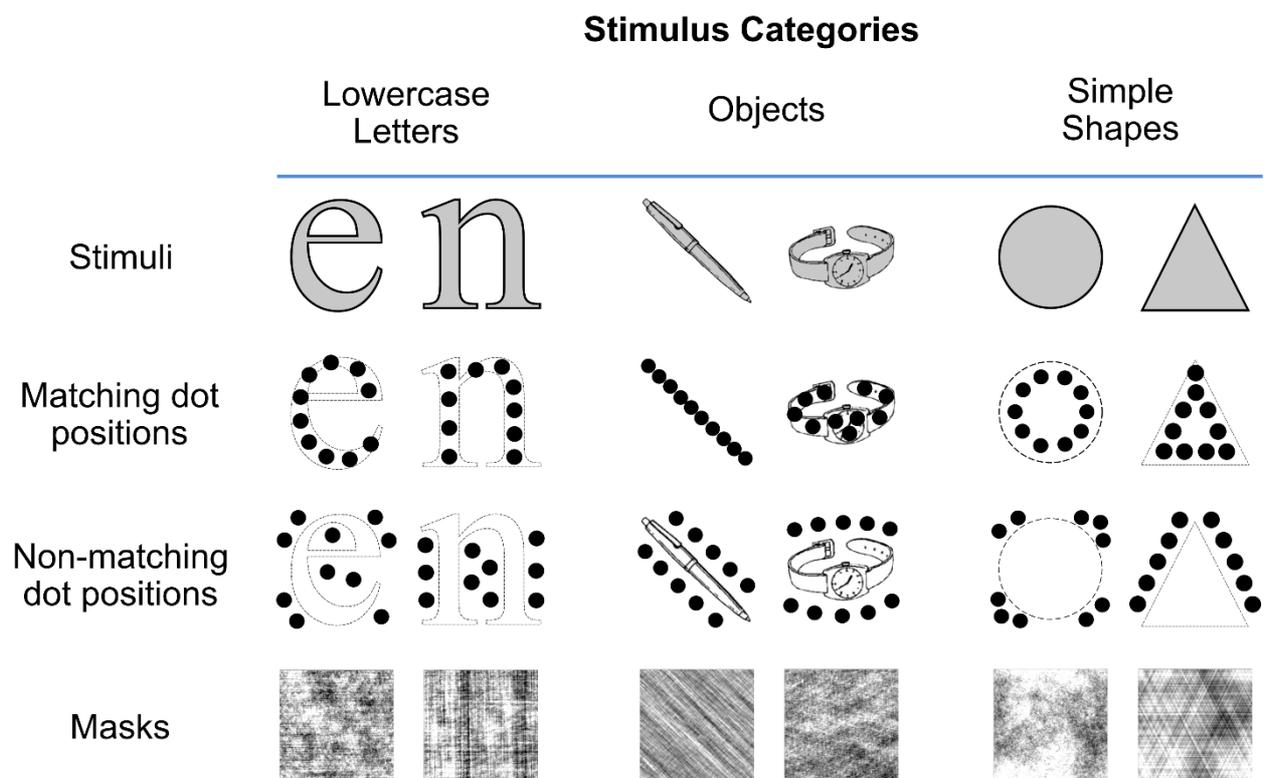


Figure 1. Stimulus exemplars. First row: there were two exemplars for each of the three stimulus categories (lowercase letters, objects, simple shapes), selected on the basis of the eccentricity effect (Marzi et al., 2006; see section Supplementary Material, Stimulus

Selection and Validation for details). Second row: “matching” dot positions for each stimulus exemplar. Third row: “non-matching” dot positions for each stimulus exemplar. Fourth row: Fourier-scrambled masks for each of the stimulus exemplars (see text for details).

Experimental design

To examine patterns of brain activation during visual mental imagery and the similarity with the patterns during visual perception, we used a mixed design, with stimulus category blocked, while the stimulus exemplar was randomized within each block.

To avoid the generation of visual mental images to be influenced by the repeated visual presentation of the stimuli to be imagined, participants performed the imagery condition (see section *Imagery condition* for details) in runs 1-5, and the perception condition (see section *Perception condition* for details) in runs 6-10. Each participant completed a single experimental session, consisting of a familiarization with the task outside the scanner (~20 minutes), a structural scan (~5 minutes) and 10 functional runs (~6 minutes each). Each functional run started and ended with 10 seconds rest, and contained 6 blocks (two for each stimulus category), interleaved by fixation blocks of 10 seconds each. Each block (43 seconds) consisted of 4 trials (2 “matching” and 2 “non-matching” dot positions), with a total of 240 trials per participant (20 trials for each factorial combination of task (2) and stimulus exemplar (6)). The order of blocks within each run and the exemplar to be imagined in each trial were randomized. The spatial position of the stimulus exemplars within each category in the instruction phase (i.e. left or right with respect to the fixation cross; see *Figure 2*) was counterbalanced across runs, following the order ABAB (where A is left and B is right) for half of the participants and the order BABA for the other half.

Imagery condition. Each block started with the presentation of the two exemplars of one of the three categories, one on the left and one on the right side of the screen (*Figure 2B*), for 1.5 seconds. To prevent after-images, stimulus presentation was followed by the appearance of two masks (1.5 s) at the same positions. Each trial was preceded by an inter-trial interval (ITI) of 1 s, consisting of the presentation of a central fixation cross and a superimposed placeholder. The placeholder comprised

6° of visual angle, and served as a reference for the position and size of the to-be-generated visual mental image. An auditory cue (i.e. “left” or “right”, 500 ms) instructed participants which of the two previously presented stimuli to imagine in the current trial. After the auditory cue, participants were instructed to imagine the corresponding stimulus in the central portion of the visual field delimited by the placeholder for a total of 7 seconds. During this time window, participants were instructed to generate the most vivid mental image they could, keeping the same size and position of the original stimuli inside the placeholder. Next, fixation cross and placeholder disappeared, and a black dot (0.5° visual angle) appeared for 2 seconds. Participants were asked to judge whether the dot fell within (“matching” trials) or outside (“non-matching” trials) the outline of the stimulus they had just imagined. Participants were instructed to provide the most accurate answer they could, favoring accuracy over speed. Participants were asked to indicate their response by button press with the index and middle finger of the right hand within 2 seconds during which the dot remained on the screen. The offset of the black dot was followed by the next trial.

Perception condition. The perception condition (*Figure 2C*) was similar to the imagery condition except for the following. After the presentation of the auditory cue, participants were presented with the stimulus corresponding to the auditory instruction within the placeholder for 0.5 seconds, followed by a mask (1.5 s), a fixation cross and the placeholder (4 s). This was followed by the presentation of a black dot (2 s). Participants had to judge whether the dot fell inside (“matching” trials) or outside (“non-matching” trials) the outline of the observed stimulus pressing the buttons with the index and middle finger of the right hand.

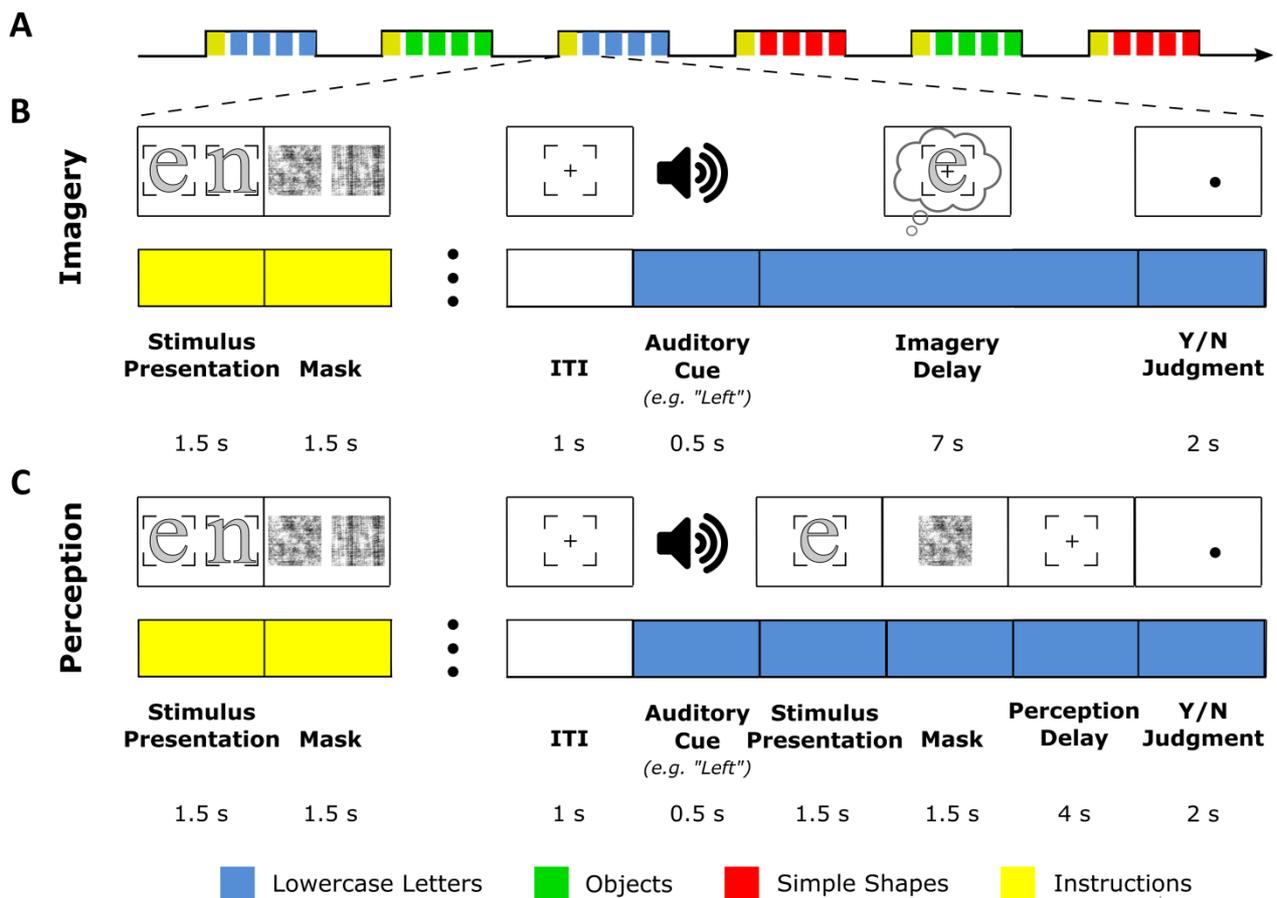


Figure 2. Task and experimental design. **A:** Experimental design. We used a mixed design: within each experimental block (black rectangles), only one stimulus category was tested (blue, lowercase letters; green, objects; red, simple shapes), whereas the stimulus exemplar was randomized within each block. **B:** Imagery condition. Participants were asked to imagine one of six stimulus exemplars pertaining to three different categories in a block design. Each block consisted of four runs and started with the presentation of the two stimulus exemplars pertaining to one of the three categories, one on the left and one on the right side of the screen, followed by a mask to prevent afterimages. Each trial started with a central fixation cross and a superimposed placeholder (1 s). Next, participants were presented with an auditory cue ('left', 'right') that indicated which of the two previously presented stimuli to imagine in the center of the placeholder for 7 seconds. Throughout the imagery delay, only the fixation cross and the placeholder were present on the screen. Next, placeholder and fixation cross disappeared, and a black dot appeared on the screen (2 s). Participants had to judge whether the dot fell within or outside the outline of the imagined stimulus. This was followed by the next trial. After a block of four trials, there was a fixation period of 10 seconds. **C:** Perception condition. The perception condition was identical to the imagery condition, except for the following. After the auditory cue, participants were presented with the visual stimulus corresponding to the instruction and the superimposed placeholder (1.5 s) and a mask (1.5 s). This was followed by the central fixation cross and the placeholder (4 s), and the presentation of the black dot (2 s). Participants had to judge whether the dot fell within or outside the outline of the observed stimulus.

To examine general visual imagery abilities, we asked each participant to fill out the Vividness of Visual Imagery Questionnaire (VVIQ; Marks, 1973) at the end of the session. This questionnaire

aims to assess individual variability in the strength and vividness of mental images by providing a set of scenarios to be imagined (e.g. “*think of some relative or friend whom you frequently see (but who is not with you at present), and consider carefully the picture that comes before your mind's eye*”). For each of the scenarios, participants had to rate the vividness of the mental image they are able to generate on a 5-point 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*), both with eyes open and eyes closed. We also collected vividness ratings for the six stimuli used in the current study using the same answer alternatives, and asked participants to rank the stimuli based on the difficulty they experienced in imagining them during the task (answer scale: 1. *Easiest stimulus to imagine* – 6. *Hardest stimulus to imagine*).

Data acquisition

MRI data were collected using a 4T Bruker MedSpec Biospin MRI scanner equipped with an eight-channel birdcage head coil. Functional data were acquired using an EPI sequence (TE/TR = 28.0/2000.0, flip angle = 73°, matrix size = 64x64, 30 interleaved slices, in-slice resolution 3 mm). Slices were axial, slightly tilted to be approximately parallel to the calcarine sulcus in order to optimize brain coverage. 171 volumes were acquired for each functional run.

To be able to coregister the low-resolution functional images to a high-resolution anatomical scan, we acquired a T1-weighted anatomical scan (MPRAGE; TR: 2700 ms; voxel resolution: 1 x 1 x 1 mm; TE: 4.18 ms; FA: 7°; FOV: 256 x 224 mm; 176 slices; GRAPPA with an acceleration factor of 2; inversion time: 1020 ms) for each participant.

Data analysis

Behavioral analyses

For each participant, we accessed accuracy in the delayed spatial judgment task by computing the percentage of correct answers separately for the imagery and perception condition, and for the six stimulus exemplars. Moreover, we compared the accuracy for “matching” and “non-matching” trials. Due to a technical fault, accuracy was not recorded for one run of the perception condition in one participant. We performed a repeated measure ANOVA, with condition (2 levels) and stimulus exemplars (6 levels) as factors.

fMRI data analysis

Preprocessing. Data were preprocessed and analyzed using FSL 5.1 (FMRIB’s Software Library, <https://fsl.fmrib.ox.ac.uk/fsl>) in combination with custom software written in Matlab. To assure to have reached steady-state magnetization, we discarded the first four volumes in each functional run. Preprocessing included motion correction to the mean image, followed by slice timing correction and high-pass temporal filtering (> 0.01 Hz). Spatial smoothing (Gaussian kernel FWHM = 6 mm) was applied for univariate analysis only, as it might remove fine structure information of activity patterns by averaging local signals within functional regions. These fine-scale patterns of neural activity may contain relevant information about experimental conditions, which can be detected by MVPA approaches (Kriegeskorte et al., 2006). Each functional run was registered to its corresponding coplanar high-resolution image with rigid body transformations (using the Boundary-Based Registration algorithm, as implemented in FSL 5.1; Greve & Fischl, 2009) and to the MNI152 2mm standard brain using linear transformation (FLIRT, 12 degrees of freedom; Jenkinson & Smith, 2001; Jenkinson et al., 2002).

Univariate RFX-GLM analysis. To estimate the amplitude of the BOLD response during the imagery and perception condition, we performed a random effects (RFX) general linear model (GLM) analysis

(N=21). Separately for the imagery and perception condition, we created regressors for each stimulus exemplar, resulting in a total of 6 (3 categories x 2 exemplars) regressors for each experimental run. Regressors for the imagery condition were time-locked to the onset of the imagery delay (duration: 6 s). Regressors for the perception condition were time-locked to the appearance of the stimulus, and included also the appearance of the visual mask (duration: 1.5 s + 1.5 s = 3 s). For both conditions, we added nuisance regressors for the presentation of the auditory cue (time-locked to the onset of the instructing cue), response phase (time-locked to the appearance of the black dot), presentation of the instruction (consisting of two stimulus exemplars and the subsequent mask) and 3D head motion estimates (3 rotation and 3 translation parameters) to the model. This leads to a total of 15 regressors for each experimental run. Each regressor was convolved with a canonical hemodynamic response function (HRF). The baseline was defined as all time-points not modeled in the design matrix (i.e. inter-trial interval, fixation intervals between blocks, and resting baseline at the beginning and the end of the runs). Results from the univariate RFX-GLM analysis were family-wise error (FWE) cluster-corrected using Gaussian Random Field (GRF) theory (Worsley et al., 1996) embedded in the FSL *cluster* routine. In the FSL *cluster* routine, statistical maps were thresholded at a $z = 2.3$ to define contiguous clusters; z -scores were computed applying an inverse cumulative distribution function (CDF) on p -values obtained from t statistical maps. The significance level of each cluster (estimated using GRF) was then compared with a probability threshold set at $p = 0.01$ to detect significant clusters. Results from the mixed-effect analysis were then projected onto an inflated brain in BrainVoyager QX 2.8.0 (BrainInnovation). Brain areas were labeled by means of the Juelich Histological Atlas, as implemented in FSL 5.1 except for area LOC, which is not contained in that atlas. We verified that the coordinates of the area we labelled as LOC are within the range of those indicated in previous studies (Grill-Spector et al., 1999; 2001; Pourtois et al., 2009).

ROI definition. V1 and V2 ROIs were defined using standard masks provided by the Juelich Histological Atlas (Eickhoff et al., 2007). To obtain non-overlapping ROIs, we applied a threshold

of .5 to the probabilistic maps and binarized these maps using FSLmaths. For each participant, we selected the 250 voxels within each standard ROI showing the highest t values during both the perception and the imagery condition. This selection consisted in four steps: first, for each participant, we averaged the t -values across the five experimental runs for each voxel, separately for the imagery and perception condition.

Next, we normalized the average t -value of each voxel in the imagery and perception condition separately, using the following formula (feature scaling):

$$x_{norm} = \frac{(x - x_{min})}{(x_{max} - x_{min})}$$

where x represents the t -value of the n^{th} voxel averaged across the five experimental runs of the selected condition (perception or imagery), x_{min} the lowest t -value across all voxels in the selected condition, and x_{max} the highest t -value across all voxels in the selected condition.

In this second step, the t -value expressed in each voxel is rescaled with respect to the maximum and minimum value in the imagery and perception condition separately. Separately for the imagery and perception condition, the output of this normalization is a vector with values for each voxel ranging between 0 and 1. This allowed us to refer the recruitment of each voxel in the two conditions to a common scale, despite the overall difference in magnitude of the strength of the activation between imagery and perception in visual areas.

Third, in order to select voxels showing the highest activation both in the imagery and the perception condition, we transformed our data using the following index:

$$I = (x_{norm, perception} \times x_{max, perception}) + (x_{norm, imagery} \times x_{max, imagery})$$

where $x_{norm, perception}$ and $x_{norm, imagery}$ represent the normalized t -values (averaged across the five runs) of a single voxel in the perception and imagery condition, respectively, whereas $x_{max, perception}$ and $x_{max, imagery}$ represent the maximum normalized t -values of a voxel of the perception and imagery

condition, respectively. Finally, we selected the 250 voxels showing the highest I values (see *Figure 3*).

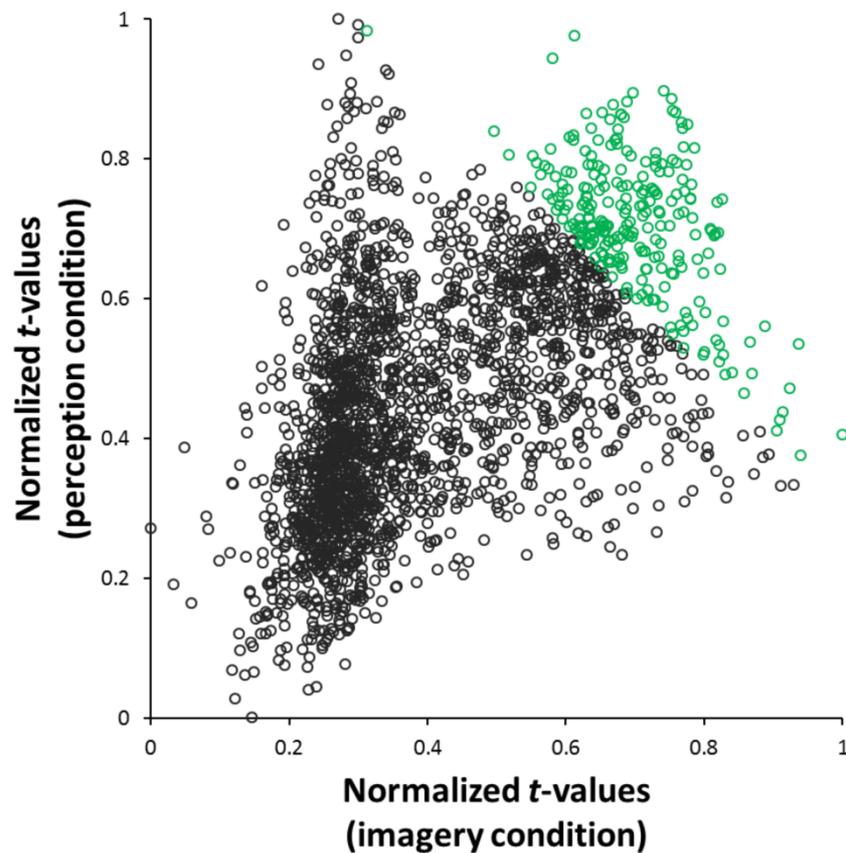


Figure 3. Example voxel selection for the VI ROI in one representative participant. The 250 voxels showing the highest normalized t -value (see text for details) both in the perception and the imagery condition are highlighted in green.

To examine classification accuracy in areas not expected to be involved in visual mental imagery, we included a control ROI encompassing the ventral bilateral striatum (see also Andersson, Ragni, & Lingnau, 2019), defined using the Oxford-GSK-Imanova Striatal Structural Atlas (Tziortzi et al., 2011). Within the ventral bilateral striatum, we selected 250 voxels repeating the steps described above.

Multivariate pattern analysis (MVPA). We performed both an ROI- and a searchlight-based MVPA using a multi-class regularized linear discriminant analysis (LDA) classifier, as implemented in CoSMoMVPA (Oosterhof et al., 2016). The purpose of the ROI analysis was to directly test whether

it is possible to decode imagined and perceived stimulus exemplars in early visual areas (i.e. V1 and V2). The whole-brain searchlight analysis was performed to explore which additional areas potentially represent imagined and observed stimuli.

ROI-based MVPA. For the multivariate pattern analysis, separately for each participant, we conducted an additional GLM analysis including regressors for the two trials within each block pertaining to each stimulus exemplar in the perception and imagery condition. Nuisance regressors were included, and were identical to those used for the *Univariate RFX-GLM analysis*, resulting in a total of 15 regressors for each experimental run. We estimated t -values for each block separately for the perception and imagery condition, resulting in 2 estimates for each stimulus exemplar and run (for a total of 60 estimates for the imagery condition, and 60 estimates for the perception condition per participant). We used t -values as input for the classifier as they are considered to be better suited for decoding (Misaki et al., 2010). Specifically, t -values are computed by dividing the beta estimates by its standard error estimate, allowing to suppress the contribution of noisy voxels.

Classification accuracies were computed using a leave-one-run-out cross-validation method (i.e., patterns from $N-1$ runs served as the training set, whereas the pattern from the remaining run served as the testing dataset). Moreover, to examine the similarity between the neural patterns obtained during imagery and perception, we performed cross-condition decoding. Specifically, the classifier was trained to discriminate between the six stimulus exemplars in one condition (e.g. perception), and tested on its ability to discriminate between the six exemplars in the other condition (e.g. imagery), and vice versa. Results from the two cross-condition classifications were averaged, resulting in one accuracy score for each ROI. To assess the significance of the decoding accuracy, we performed permutation testing coupled with Threshold-Free Cluster Enhancement (TFCE), as implemented in CoSMoMVPA (Oosterhof et al., 2016). For each classification analysis, we computed a null distribution by randomly permuting targets and performing classification. As suggested by Stelzer et al. (2013), we repeated this operation 100 times, resulting in 100 accuracy

values for each participant. We then randomly selected permutation accuracy values and averaged across participants (10000 iterations) to create a null-distribution of averaged accuracy values.

Searchlight-based MVPA. To identify any additional regions representing imagined and observed stimulus exemplars belonging to three different categories, we performed a whole brain searchlight-based MVPA (Kriegeskorte et al. 2006; Oosterhof et al., 2016). Decoding parameters and procedures were very similar to the ROI-based MVPA, except that a searchlight approach was applied, using a spherical ROI (100 voxels) centered around each voxel in the brain. Decoding accuracies from each searchlight were assigned to the central voxel. To identify voxels where classification accuracy was greater than chance (16.6%), we performed permutation testing coupled with Threshold-Free Cluster Enhancement (TFCE), as implemented in CoSMoMVPA (Oosterhof et al., 2016), similarly to the procedure we used for the ROI-based analysis. For each classification analysis, we computed a null distribution by randomly permuting targets and performing classification. We repeated this operation 100 times for each participant, as suggested by Stelzer et al. (2013). We then randomly selected permutation maps and averaged across participants (10000 iterations) to create a null-distribution of averaged accuracy maps. For visualization purposes, we projected group maps on a segmented and inflated MNI aligned brain (Colin Holmes' 27-scan average brain image, as implemented in NeuroElf, v 1.1) in BrainVoyager QX 2.8.0 (BrainInnovation), separately for the decoding of stimulus exemplars during perception, imagery and across the two conditions (cross-condition decoding). Code and summary data for the uni- and multivariate analyses are available on the Open Science Framework (<https://osf.io/nhv2r/>). The conditions of our ethical approval do not permit public archiving or peer-to-peer sharing of individual raw data. The data supporting the conclusions of this article are therefore not available to any individual outside the author team under any circumstances.

Correlation between the BOLD signal and behavioral performance. To explore the contribution of different brain areas to behavioral performance during visual mental imagery, we determined the relation between the amplitude of the BOLD signal and behavioral accuracy in the imagery task, as

well as the relation between the amplitude of the BOLD signal and individual vividness ratings. To this aim, we selected four ROIs as highlighted by the univariate and/or multivariate analyses, namely, primary visual cortex (V1), left superior parietal lobe (SPL), left anterior intraparietal sulcus (aIPS) and left lateral occipital complex (LOC). For V1, we used the same ROIs as defined for the ROI-based MVPA (see *ROI definition* for more details). For SPL and aIPS, ROIs were defined based on the univariate contrast [imagery > baseline] (see *Univariate RFX-GLM analysis* for more details). LOC, instead, was defined based on the group accuracy map resulting from the searchlight-based MVPA of the imagery condition (see *Multivariate pattern analysis (MVPA)* for more details). For all the considered regions, we extracted the average BOLD amplitude expressed as % BOLD signal change from the univariate contrast [imagery > baseline] (see *Univariate RFX-GLM analysis* for more details), using the Featquery tool in FSL 5.1. We then computed the correlation (Spearman's rank-order correlation) between the average activation extracted from the four ROIs and two behavioral indices: the accuracy in the imagery condition, expressed as the percentage of correct answers, and individual ratings of the ability to generate vivid mental images as assessed by the VVIQ questionnaire. The VVIQ score is expressed as the average between the score obtained during visual mental imagery performed with eyes open and eyes closed, as done in previous studies (Amedi et al., 2005). Significance was assessed by means of permutation testing (10000 iterations), performed using the *multi_comp_perm_corr* function (Groppe et al., 2011) in Matlab. No part of the study procedures and analyses was pre-registered prior to the research being conducted.

Results

Behavioral results. Accuracy in the behavioral task was computed individually for each participant as the percentage of correct answers (i.e. correct spatial localization of a “matching” dot as positioned on the imagined/observed stimulus; correct spatial localization of a “non-matching” dot as positioned outside the imagined/observed stimulus), separately for the perception and imagery condition. Since we instructed participants to favor accuracy over speed in the completion of the task, reaction times were not included in this analysis.

Figure 4 shows the accuracy in the spatial judgement task as a function of condition (imagery, perception) and stimulus exemplar (letter ‘e’, letter ‘n’, pen, watch, circle and triangle). Not surprisingly, we found a higher accuracy in the spatial judgement task for the perception condition [mean: 76,31% ± 10.15%] in comparison to the imagery condition [mean: 67.9% ± 10.22%] in all six exemplars.

These observations are supported by the corresponding statistics, using a repeated measures ANOVA with the factors task (2) and stimulus exemplars (6). Accuracy differed between the two tasks [main effect of task: $F(1, 20) = 26.425, p < 0.001$] and between stimulus exemplars [main effect of stimulus exemplars: $F(5, 100) = 23.77, p < 0.001$]. In both the imagery and perception task, accuracy for lowercase letters (i.e. ‘e’ and ‘n’) was lower than for objects (i.e. pen and watch) and simple shapes (i.e. circle and triangle) [imagery task, paired samples t-tests: lowercase letters (mean accuracy: 60.5% ± 13%) vs objects (mean accuracy: 73.1% ± 12.8%); $t(20) = -3.051, p = 0.006$. Lowercase letters (mean accuracy: 60.5% ± 13%) vs simple shapes (mean accuracy: 70.1% ± 10%); $t(20) = -4.041, p = 0.001$. Objects (mean accuracy: 73.1% ± 12.8%) vs simple shapes (mean accuracy: 70.1% ± 10%); $t(20) = -0.687, p = 0.506$. Perception task, paired samples t-tests: lowercase letters (mean accuracy: 63.9% ± 13.3%) vs objects (mean accuracy: 82.5% ± 10.6%); $t(20) = -9.314, p < 0.001$. Lowercase letters (mean accuracy: 63.9% ± 13.3%) vs simple shapes (mean accuracy: 82.5% ± 10.24%); $t(20) = -9.314, p < 0.001$. Objects (mean accuracy: 82.5% ± 10.6%) vs simple shapes (mean

accuracy: $82.5\% \pm 10.24\%$; $t(20) = 0$, $p = 0.999$]. The effect of stimulus exemplars on accuracy did not interact with the task [interaction stimulus exemplars*task, $F(5, 100) = 1.949$, $p = 0.093$].

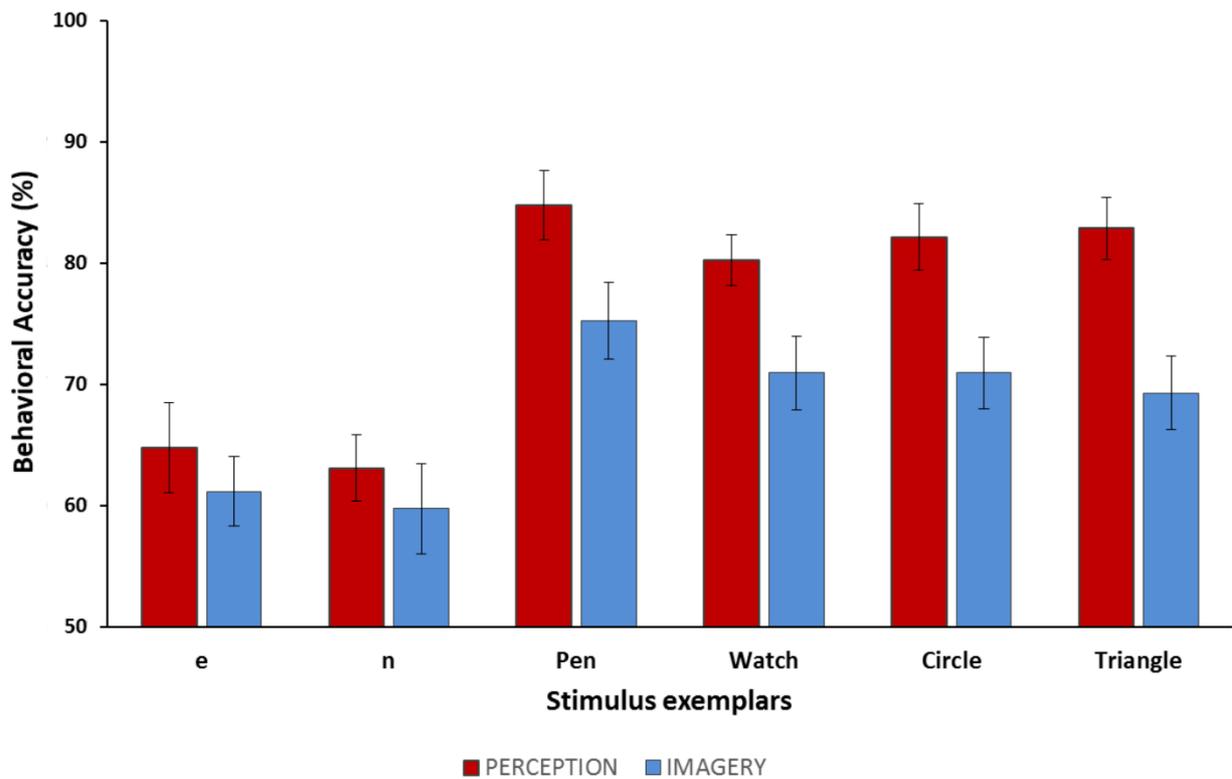
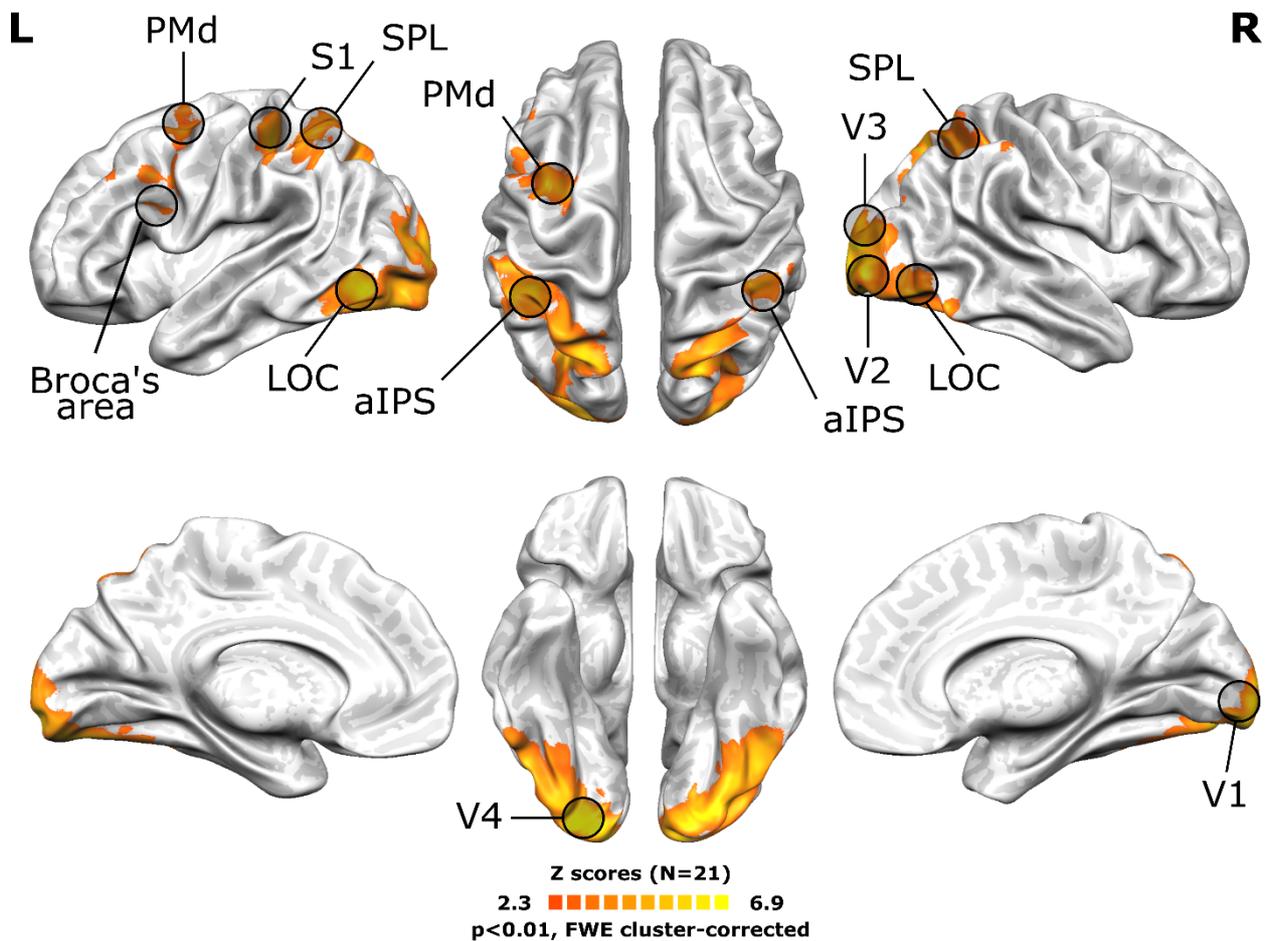


Figure 4. Mean behavioral accuracy ($N=21$) in the delayed spatial judgment task during the perception and imagery condition as a function of stimulus exemplars. Chance level: 50%. Error bars: standard error of the mean (S.E.M.).

Univariate analysis

Perception condition. As can be seen in *Figure 5* showing the RFX GLM contrast [perception > baseline], the perception condition lead to a widespread recruitment of striate and extrastriate visual areas. In addition to left and right V1, we found bilateral recruitment of areas V2, V3, and V4. Moreover, we also found a bilateral recruitment of LOC, known to be involved in the processing of shape and objects (Kourtzi and Kanwisher, 2000; Grill-Spector et al., 2001).

Additionally, this contrast recruited a network of parietal - bilateral SPL, bilateral aIPS, left primary somatosensory cortex (S1) - and frontal - left dorsal premotor cortex (PMd) and Broca's area -



regions.

Figure 5. Results of the univariate RFX GLM contrast ($N = 21$ participants) [perception > baseline]. The group activation map was FWE cluster-corrected using GRF theory and projected on an inflated surface mesh. L: Left hemisphere. R: Right hemisphere.

Imagery condition. Figure 6 shows the RFX GLM contrast [imagery > baseline]. We found a selective recruitment of the left hemisphere, involving SPL, aIPS, IPL, S1 and PMd.

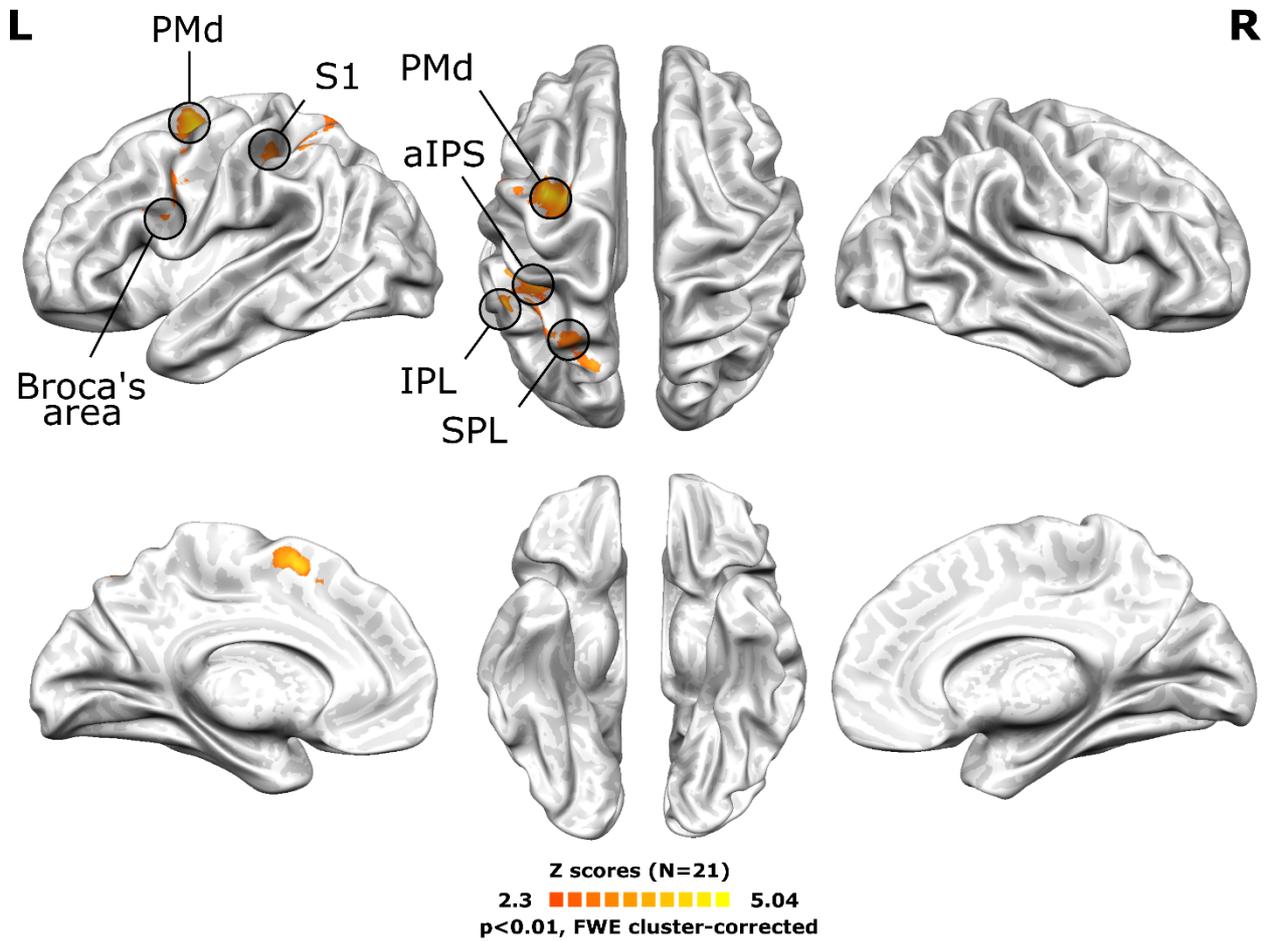


Figure 6. Results of the univariate RFX GLM contrast ($N = 21$ participants) [imagery > baseline]. The group activation map was FWE cluster-corrected using GRF theory and projected on an inflated surface mesh. Abbreviations same as in Figure 5.

Multivariate pattern analysis

ROI-based MVPA. In the ROI-based MVPA we tested whether the six stimulus exemplars could be decoded on the basis of patterns of brain activity obtained during perception and visual imagery of the same stimuli. We found significant above chance classification accuracy for observed stimuli in both V1 (mean accuracy: $23.10\% \pm 5.4\%$; $p < 0.001$) and V2 (mean accuracy: $21.67\% \pm 4.8\%$; $p < 0.001$; see *Figure 7*). Within the same ROIs, however, we were not able to decode imagined stimulus exemplars (V1: mean accuracy $18.81\% \pm 3.9\%$; $p = 0.074$. V2: mean accuracy $20.71\% \pm 6.9\%$; $p = 0.055$). As expected, classification accuracy did not reach significance in the control region (ventral bilateral striatum), neither in the perception condition (mean accuracy $16.67\% \pm 5.5\%$; $p = 0.5$) nor in the imagery condition (mean accuracy $17.3\% \pm 5.5\%$; $p = 0.5$).

Cross-condition MVPA. In the cross-condition MVPA we aimed to investigate the similarity between the representation of imagined and observed stimuli in early visual areas. In particular, we examined whether it is possible to train a classifier to successfully distinguish between the six stimulus exemplars based on the patterns of activation elicited by visual stimuli (perception condition), and then test the classifier on patterns of activation elicited by visual mental imagery (imagery condition) of the same stimuli (and vice versa). As shown in *Figure 7*, the results indicate above chance cross-classification accuracy in V1 (mean accuracy $20.24\% \pm 4.1\%$; $p = 0.002$), and in V2 (mean accuracy $20.95\% \pm 4.9\%$; $p = 0.002$). Classification accuracy did not reach significance in the ventral bilateral striatum (mean accuracy $16.19\% \pm 3.1\%$; $p = 0.5$).

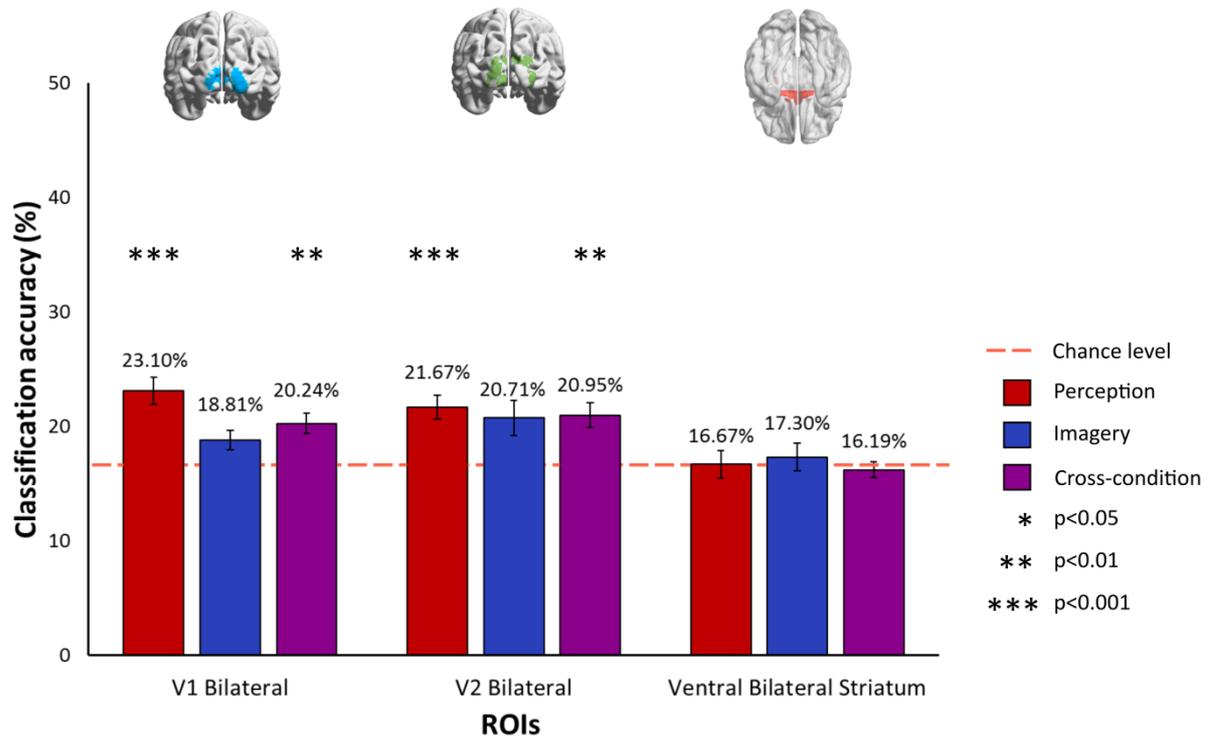


Figure 7. Mean decoding accuracy of stimulus exemplars, separately for V1, V2, and a control ROI (ventral bilateral striatum). Red bars, perception condition. Blue bars, imagery condition. Purple bars, cross-condition decoding. Statistical significance was assessed by means of permutation testing coupled with Threshold-Free Cluster Enhancement (10000 iterations). Significance levels: one black asterisk, $p < 0.05$; two black asterisks, $p < 0.01$; three black asterisks, $p < 0.001$. Error bars: standard error of the mean (S.E.M.).

Searchlight-based MVPA

Perception condition. To identify additional areas that can distinguish between the six stimulus exemplars, we performed a whole-brain searchlight-based MVPA. *Figure 8* shows the group t map for the decoding of stimulus exemplars in the perception condition. As can be seen, this analysis revealed significant decoding in early visual areas (V1, V2 bilaterally), in V4 (bilaterally) and in hMT (bilaterally).

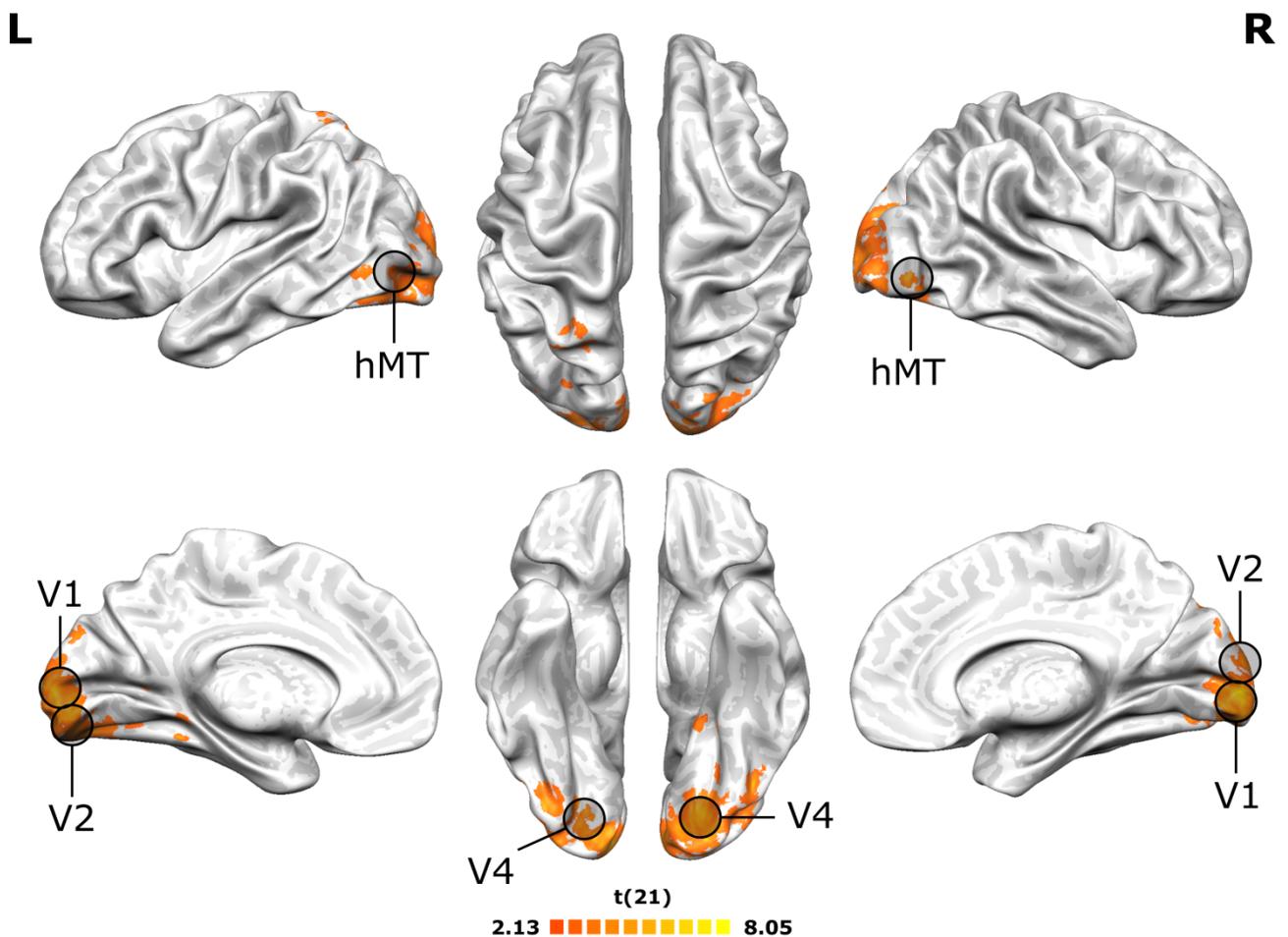


Figure 8. Results of the searchlight-based MVPA for the perception condition (see text for details). The spherical searchlight comprised 100 voxels. The group t map was thresholded at a z -score of 1.65, corresponding to $p < 0.05$ one-tailed (corrected for multiple comparisons) and projected on an inflated surface. Abbreviations same as in Figure 5.

Imagery condition. Figure 9 illustrates the group t map for the searchlight-based MVPA of the imagined stimulus exemplars. We found significant clusters in early visual areas (left V1 and bilateral V2), in the bilateral LOC and bilateral SPL. Moreover, we also found significant above-chance classification accuracy in the left aIPS, left Broca's area, bilateral S1, and in the left PMd.

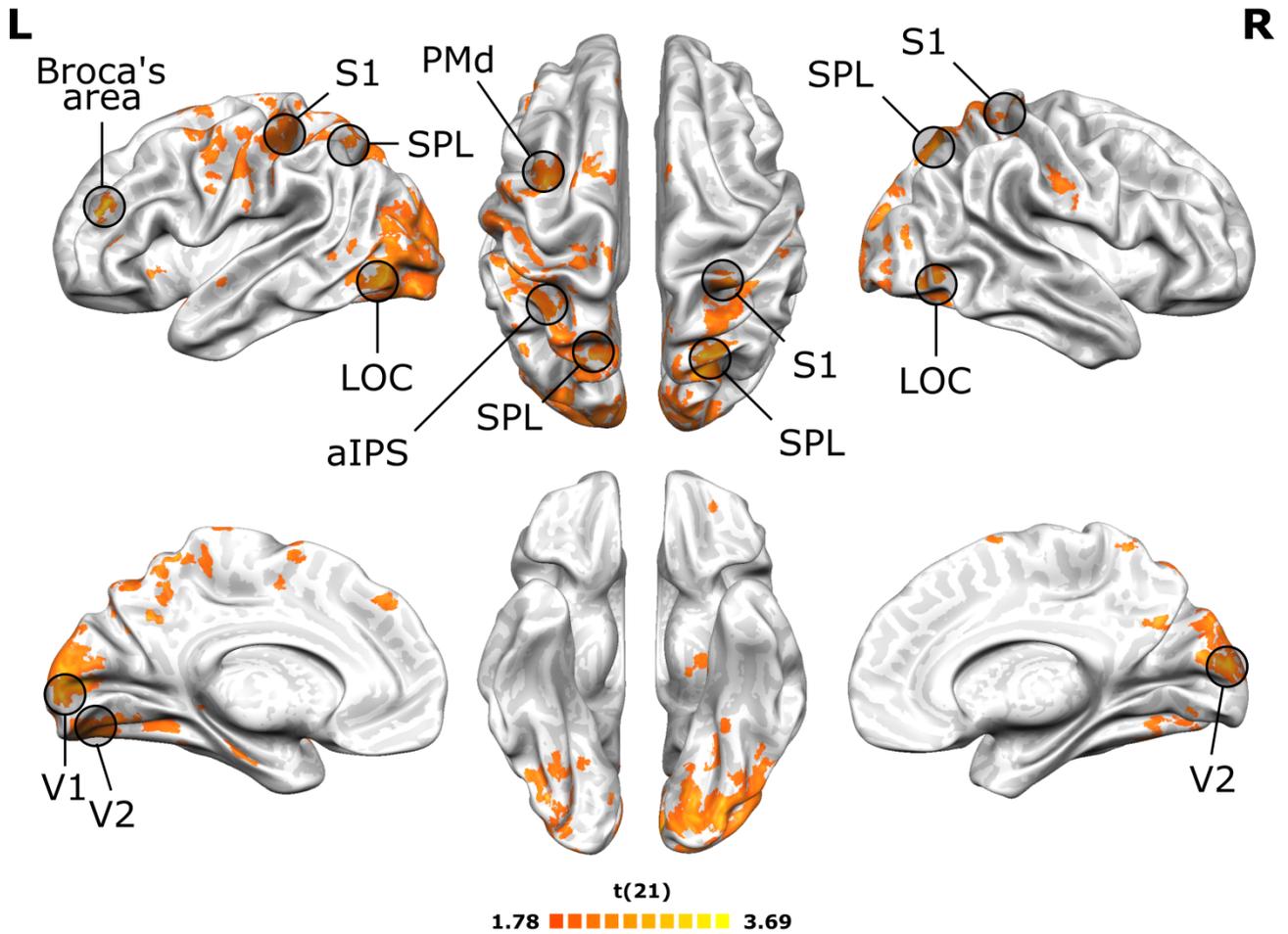


Figure 9. Results of the searchlight-based MVPA for the imagery condition (see text for details). The spherical searchlight comprised 100 voxels. The group t map was thresholded at a z -score of 1.65, corresponding to $p < 0.05$ one-tailed (corrected for multiple comparisons) and projected on an inflated surface. Abbreviations same as in Figure 5.

Cross-decoding. Figure 10 illustrates the group t map for the cross-condition decoding. We obtained significant clusters in bilateral V1 and V2, bilateral LOC, left IPL, and bilateral aIPS, SPL and S1.

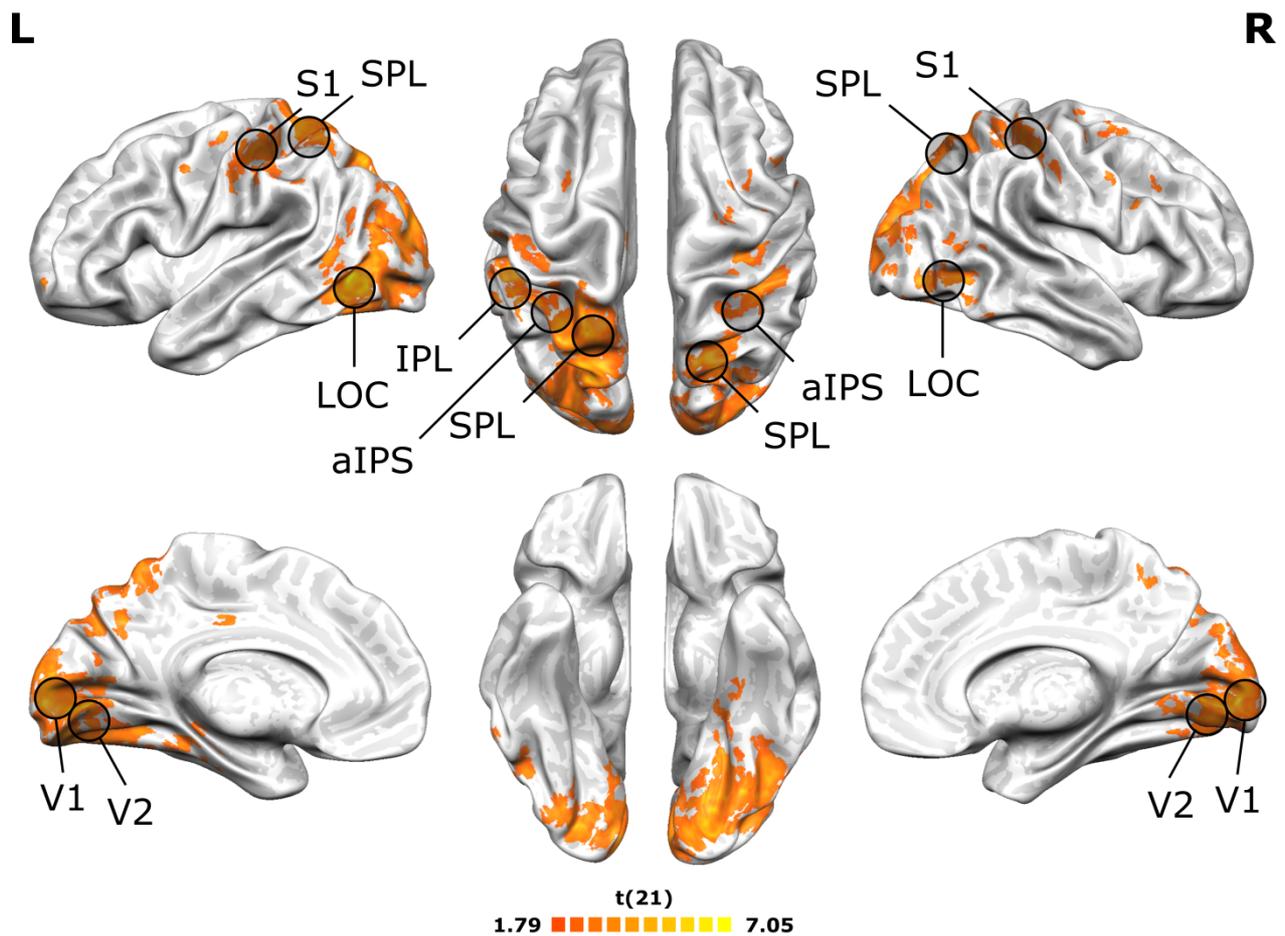


Figure 10. Results of the searchlight-based MVPA for cross-condition decoding (see text for details). The spherical searchlight comprised 100 voxels. The group t map was thresholded at a z -score of 1.65, corresponding to $p < 0.05$ one-tailed (corrected for multiple comparisons) and projected on an inflated surface. Abbreviations same as in Figure 5.

Correlation between bold signal and behavioral measures

We obtained a positive correlation between the amplitude of the BOLD signal and behavioral accuracy in the imagery condition in the left SPL ($r_s=0.614$, $p=0.003$), in the left aIPS ($r_s=0.572$, $p=0.009$) and in the left LOC ($r_s=0.532$, $p=0.016$; *Figure 11, left column*). By contrast, we obtained no systematic relationship between the amplitude of the BOLD signal and behavioral performance in V1 ($r_s=0.205$, $p=0.961$).

Regarding the correlation between cortical activity during the imagery condition, expressed as % of BOLD signal change, and the subjective vividness ratings for visual mental imagery, the analysis did not reveal any significant correlation in any of the examined ROIs (V1, $r_s=-0.012$, $p=0.961$; left SPL, $r_s=-0.286$, $p=0.218$; left aIPS, $r_s=-0.297$, $p=0.19$; left LOC, $r_s=0.085$, $p=0.697$; *Figure 11, right column*).

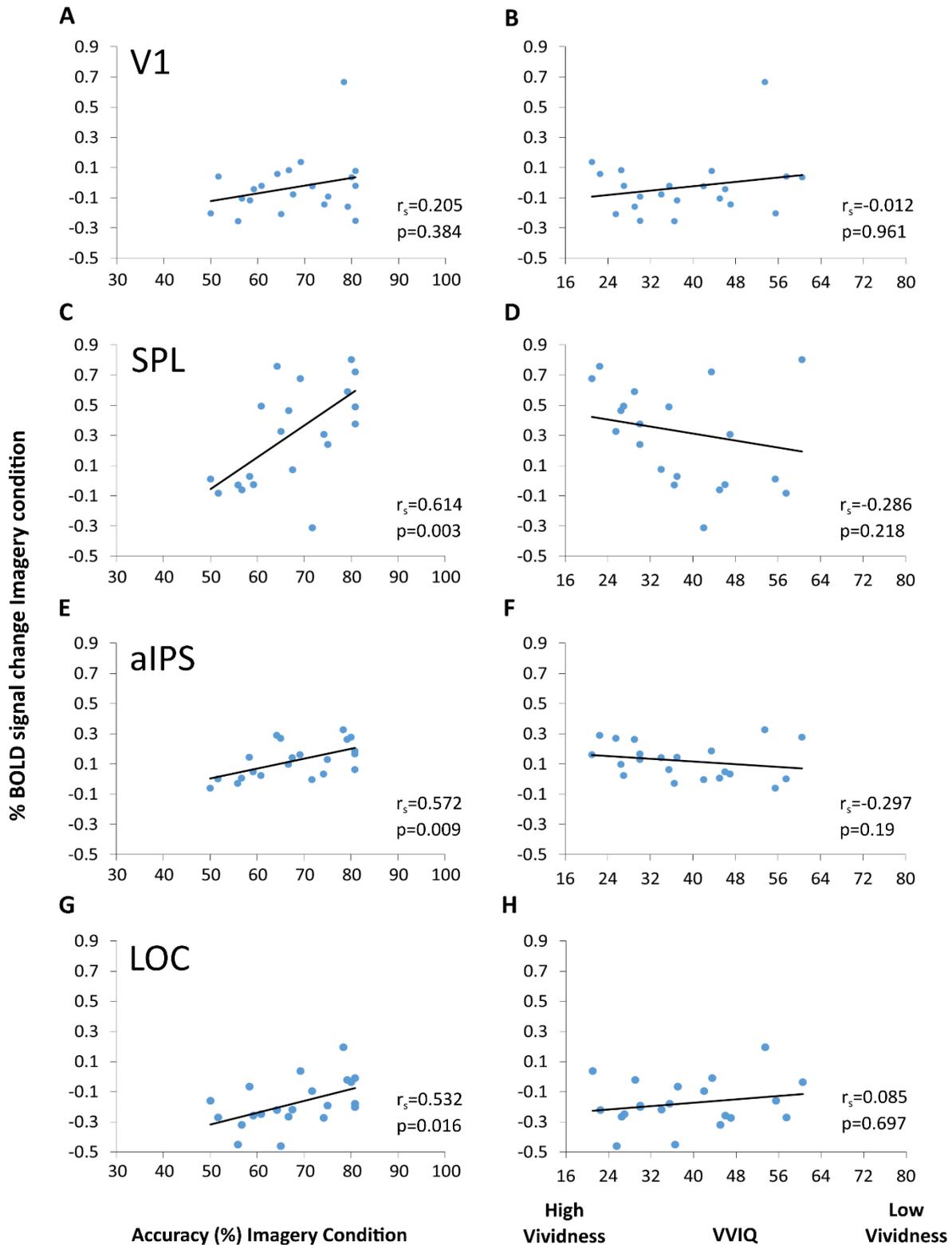


Figure 11. Correlation between the amplitude of the BOLD signal and behavioral measures. Correlation between mean activation in V1 (A, B), SPL (C, D), aIPS (E, F) and LOC (G, H) expressed as % of BOLD signal change and accuracy in the imagery condition (left column), and vividness of visual mental imagery as assessed by the VVIQ (right column).

Discussion

To explore whether stimulus exemplars pertaining to different categories (letters, simple shapes, objects) can be decoded in early visual cortex during visual mental imagery, we used a multivariate pattern analysis approach. We found that neural activity patterns in early visual (V1, V2), parietal (SPL, IPL and aIPS), infero-temporal (LOC) and prefrontal (PMd) areas can distinguish between the six imagined stimulus exemplars. Moreover, in parietal, inferotemporal and early visual cortices, we found shared representations across visual mental imagery and visual perception of the same stimuli. In the following, we will discuss these results in more detail.

Encoding of imagined stimuli in early visual cortex

The absence of a reliable activation of early visual cortex during visual imagery is a well-known result within the emerging literature suggesting a high variability in the recruitment of V1 (e.g. Ishai et al., 2000; Formisano et al., 2002; Yomogida et al., 2004). Despite the overall weak signal in early visual cortex, we were able to decode imagined stimulus exemplars on the basis of patterns of activation in V1 and V2. This result was stronger in the searchlight-based compared to ROI-based MVPA, possibly due to the fact that this effect was lateralized, with stronger results in the left compared to the right hemisphere. This finding is compatible with the results of our behavioral experiment (see *Supplementary Material - Stimulus selection and validation*), which highlighted a stronger eccentricity effect in the right compared to the left hemifield for the three selected stimulus categories. Previous studies adopted a similar approach and tried to decode specific exemplars within a category during imagery tasks. As an example, Albers et al. (2013) demonstrated decoding of relatively simple stimuli (i.e. gratings with different orientations) in early visual cortex (V1 to V4). Stokes et al. (2009) were able to decode simple letters ('X' and 'O') from patterns of activation in lateral occipitotemporal (LOC) cortex. Decoding was possible also in primary visual cortex, but only at a more liberal statistical threshold. By contrast, Lee et al. (2012) used images of common objects

(i.e. colored and highly detailed pictures of different real-world objects). Their results revealed a representation of imagined stimuli in extrastriate cortices (i.e. V2, V3, V4, LO and pFS), but not in V1.

Together, these results suggest that the degree to which it is possible to decode the content of imagined stimuli in early visual cortex might depend on the type of stimulus adopted, with a preference for low-level stimuli such as the ones used by Albers et al. (2013) (i.e. gratings) and the ones used in the current study (i.e., line drawings of lowercase letters, common objects and simple shapes).

We hypothesize that in the absence of bottom-up visual stimulation, early visual cortex receives information about imagined stimuli from inferotemporal and parietal brain regions via top-down feedback, at least for relatively simple stimulus categories. In line with this view, several fMRI studies showed the existence of nonretinal influences on early visual cortex (Muckli et al., 2005; Bannert & Bartels, 2013; Muckli & Petro, 2013), supporting the hypothesis of a top-down modulation from higher-level areas during visual imagery. Vetter et al. (2014), for example, were able to decode real and imagined sound categories from patterns of activity in V1 which, according to the authors, is likely to have been mediated by top-down feedback from multisensory brain areas such as pSTS and the precuneus.

The involvement of parietal and premotor cortex during visual imagery

Our analyses revealed a recruitment of parietal (SPL, aIPS, IPL) and premotor (PMd) cortices of the left hemisphere during the imagery condition. Likewise, the searchlight-based MVPA showed significant above-chance decoding of imagined stimulus exemplars in SPL, IPL, S1, and PMd. Both in SPL and aIPS, we obtained a positive correlation between neural activity and behavioral performance, suggesting a critical role of both regions in participants' ability to determine the position of a dot with respect to an imagined stimulus. However, despite this type of analysis being common practice in the visual imagery literature (e.g. Amedi et al., 2005; Cui et al., 2007; Andersson et al., 2019), we acknowledge that the general noisiness of brain-behavior correlations together with the

low-sample size is likely to have limited the power of our results, which therefore should be interpreted with caution.

We did not find any significant correlation between cortical activity and subjective vividness of mental imagery as assessed by the VVIQ. Considering that the questionnaire is aimed to assess the general experience of visual imagery, it provided a measure of more general abilities in visual mental imagery which might not directly correspond to the neural mechanisms related to the present task.

The recruitment of parietal regions during visual mental imagery is in line with previous studies (Knauff et al., 2000; Formisano et al., 2002; Ganis et al., 2004) which also highlighted a prominent role of the left hemisphere in imagery tasks (see Winlove et al., 2018). Recent studies showed parietal regions (i.e. SPL and aIPS) to be involved during mental imagery of different hand actions (Oosterhof et al., 2012) and in the encoding of the identity of artificial stimuli during visual working memory (Christophel et al., 2012; 2014). According to Naughtin et al. (2014), premotor cortex can host distinct representations of both identity and spatial position of stimuli in a visual working memory task during the retention delay. Our results suggest that PMd might have a role in encoding the identity of stimuli pertaining to different categories during visual mental imagery as well.

S1 has been shown to be involved in tasks requiring the use of tactile imagery, both at the univariate (Savini et al., 2012; Schmidt et al., 2014) and multivariate (de Borst & de Gelder, 2017) level. Christophel & Haynes (2014) were able to decode the identity of previously memorized visual stimuli in primary somatosensory cortex. The authors hypothesized that visual features of memorized stimuli might be encoded using cross-modal mapping with tactile features, which in turn would improve storage of visual information in working memory throughout delayed periods of time. It is possible that a similar mechanism occurred in our delayed spatial judgment task as well.

To help understanding how different types of information are processed within the visual mental imagery network, some studies examined the patterns of connectivity elicited when imaging and perceiving real stimuli. What these studies revealed was a reversed flow of information during visual imagery and perception. In particular, they suggested that during imagery information is transmitted

from prefrontal (Mechelli et al., 2004; Dijkstra et al., 2017) and parietal (Dentico et al., 2014) to occipital nodes of the visual imagery network, opposite to the bottom-up flow of information arising during early stages of visual perception. We hypothesize that a similar top-down mechanism might underlie the results of the current study as well.

The role of the LOC in visual imagery

In the absence of a significant recruitment, we were able to decode imagined stimulus exemplars in LOC, a brain area known to be involved in object recognition (Grill-Spector et al., 2001; Pourtois et al., 2009). Its involvement in visual mental imagery of different stimulus categories, such as letters and common objects, has been reported in previous studies (Stokes et al., 2009; Reddy et al., 2010; Lee et al., 2012). Similarly to our observations in parietal regions, we found a positive correlation between neural activity and behavioral performance in the delayed spatial judgment task in this region. Taken together, these results indicate a pivotal role of parietal and inferotemporal regions in the execution of tasks requiring to perform mental operations on imagined stimuli (e.g. the delayed spatial-judgment task in the current study).

Shared neural representation for imagined and observed stimuli

The ROI- and searchlight-based MVPA revealed shared representations for imagery and perception in bilateral V1, V2 and left LOC. Note that we cannot fully exclude that the cross-decoding results have been influenced by spontaneous forms of visual mental imagery occurring during the perception condition. Likewise, we observed similar differences between stimulus exemplars in terms of behavioral accuracy that might reflect difficulty and/ or complexity of the stimuli, which may have contributed to cross-decoding (albeit in specific brain areas). That said, our observations are in line with the results of previous studies, which demonstrated cross-decoding in V1 for the orientation of perceived and imagined gratings (Albers et al., 2013) and complex objects (Lee et al., 2012). Likewise, several studies revealed shared representations for imagery and perception of different

stimulus categories, such as faces, body parts, objects and scenes in higher-level visual areas (e.g. FFA, PPA, EBA and LOC; Stokes et al., 2009; 2011; Reddy et al., 2010; Cichy et al., 2011; Lee et al., 2012).

Moreover, we found shared representations for imagined and observed stimuli in a network of parietal areas, comprising bilateral SPL and aIPS, and left IPL. These areas are known to be recruited in tasks involving a strong spatial component (Mellet et al., 1996; Trojano et al., 2000; Sack et al., 2002). Sack et al. (2012) proposed the existence of two distinct pathways engaged during visual mental imagery. In analogy with perception, a ventral occipito-temporal imagery network involving category-selective regions in inferior temporal cortex is assumed to represent the content of mental images, whereas a dorsal “spatial” network, encompassing parietal and premotor cortices, is assumed to be involved in the encoding of the spatial configuration of imagined stimuli. Following this line of reasoning, considering that our task required participants to judge the position of a dot with respect to a visual or imagined stimulus, the cross-decoding within parietal areas might reflect the processing of the spatial configuration of the observed and imagined stimulus. Different tasks, requiring to judge different aspects of the stimulus such as its shape, its weight or size in the real world, might reveal a different set of areas in comparison to those found in the current study.

Conclusions

We demonstrated that it is possible to decode between imagined stimulus exemplars belonging to different categories not only in parietal, inferotemporal and frontal cortex, but also in early visual areas. Moreover, in a subset of these areas, we also obtained significant cross-decoding, indicating shared representations for imagined and observed stimuli in early visual, parietal and inferotemporal cortices. The correlation between behavioral accuracy and the amplitude of the BOLD signal in parietal and inferotemporal cortices, but not in early visual cortex, are in line with the view that these areas contribute to the ability to perform visual imagery, and that in the absence of bottom-up

information, early visual cortex has access to information about imagined stimuli via feedback connections.

Acknowledgements

This work was supported by the European Research Council (ERC) Grant number 339939 “Perceptual Awareness”.

References

- Albers, A. M., Kok, P., Toni, I., Dijkerman, H. C., and de Lange, F. P. (2013). Shared representations for working memory and mental imagery in early visual cortex. *Current Biology*, 23(15), 1427–31. <http://doi.org/10.1016/j.cub.2013.05.065>
- Amedi, A., Malach, R., and Pascual-Leone, A. (2005). Negative BOLD differentiates visual imagery and perception. *Neuron*, 48, 859–872. <https://doi.org/10.1016/j.neuron.2005.10.032>
- Andersson, P., Ragni, F., and Lingnau, A. (2019). Visual imagery during real-time fMRI neurofeedback from occipital and superior parietal cortex. *Neuroimage*, 200, 332-343. <http://doi.org/10.1016/j.neuroimage.2019.06.057>
- Bannert, M. M., and Bartels, A. (2013). Decoding the yellow of a gray banana. *Current Biology*, 23, 2268–2272. <http://doi.org/10.1016/j.cub.2013.09.016>
- Brainard, D. H. (1997) The Psychophysics Toolbox. *Spatial Vision*, 10, 433-436.
- Bullier, J. (2001). Integrated model of visual processing. *Brain Research Reviews*, 36, 96–107.
- Christophel, T. B., and Haynes, J. D. (2014). Decoding complex flow-field patterns in visual working memory. *NeuroImage*, 91, 43–51. <http://doi.org/10.1016/j.neuroimage.2014.01.025>
- Christophel, T. B., Hebart, M. N., and Haynes, J. D. (2012). Decoding the contents of visual short-term memory from human visual and parietal cortex. *Journal of Neuroscience*, 32(38), 12983–12989. <http://doi.org/10.1523/JNEUROSCI.0184-12.2012>
- Cichy, R. M., Heinze, J., and Haynes, J. D. (2012). Imagery and perception share cortical representations of content and location. *Cerebral Cortex*, 22(2), 372–380. <http://doi.org/10.1093/cercor/bhr106>
- Cui, X., Jeter, C.B., Yang, D., Montague, P.R. and Eagleman, D.M. (2007). Vividness of mental imagery: individual variability can be measured objectively. *Vision Research*, 47(4), 474-478. <http://doi.org/10.1016/j.visres.2006.11.013>
- de Borst, A. W., and de Gelder, B. (2017). fMRI-based Multivariate Pattern Analyses reveal imagery modality and imagery content specific representations in primary somatosensory, motor and

- auditory cortices. *Cerebral Cortex*, 27(8), 3994–4009. <https://doi.org/10.1093/cercor/bhw211>
- Dentico, D., Cheung, B. L., Chang, J. Y., Guokas, J., Boly, M., Tononi, G., and Van Veen, B. (2014). Reversal of cortical information flow during visual imagery as compared to visual perception. *NeuroImage*, 100, 237–243. <http://doi.org/10.1016/j.neuroimage.2014.05.081>
- Dijkstra, N., Zeidman, P., Ondobaka, S., Van Gerven, M. A. J., and Friston, K. (2017). Distinct top-down and bottom-up brain connectivity during visual perception and imagery. *Scientific Reports*, 7(1), 1–9. <http://doi.org/10.1038/s41598-017-05888-8>
- Eickhoff, S. B., Paus, T., Caspers, S., Grosbras, M.-H., Evans, A. C., Zilles, K., and Amunts, K. (2007). Assignment of functional activations to probabilistic cytoarchitectonic areas revisited. *NeuroImage*, 36(3), 511–521. <http://doi.org/10.1016/j.neuroimage.2007.03.060>
- Formisano, E., Linden, D. E. J., Di Salle, F., Trojano, L., Esposito, F., Sack, A. T., Grossi, D., Zanella, F.E., and Goebel, R. (2002). Tracking the mind's image in the brain I: Time-resolved fMRI during visuospatial mental imagery. *Neuron*, 35(1), 185–194. [http://doi.org/10.1016/S0896-6273\(02\)00747-X](http://doi.org/10.1016/S0896-6273(02)00747-X)
- Ganis, G., Thompson, W. L., and Kosslyn, S. M. (2004). Brain areas underlying visual mental imagery and visual perception: An fMRI study. *Cognitive Brain Research*, 20(2), 226–241. <http://doi.org/10.1016/j.cogbrainres.2004.02.012>
- Greve, D.N., and Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, 48(1), 63-72. <http://doi.org/10.1016/j.neuroimage.2009.06.060>
- Grill-Spector, K., Kourtzi, Z., and Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research*, 41(10–11), 1409–1422. [http://doi.org/10.1016/S0042-6989\(01\)00073-6](http://doi.org/10.1016/S0042-6989(01)00073-6)

- Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzhak, Y., and Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron*, 24(1), 187–203. [http://doi.org/10.1016/S0896-6273\(00\)80832-6](http://doi.org/10.1016/S0896-6273(00)80832-6)
- Groppe, D. M., Urbach, T. P. and Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiology*, 48(12), 1711–1725. <http://doi.org/10.1111/j.1469-8986.2011.01273.x>.
- Haxby, J. V. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293(5539), 2425–2430. <http://doi.org/10.1126/science.1063736>
- Ishai, A. (2002). Visual imagery of famous faces: effects of memory and attention revealed by fMRI. *NeuroImage*, 17(4), 1729–1741. <http://doi.org/10.1006/nimg.2002.1330>
- Ishai, A., Ungerleider, L. G., and Haxby, J. V. (2000). Distributed neural systems for the generation of visual images. *Neuron*, 28(3), 979–990. [http://doi.org/10.1016/s0896-6273\(00\)00168-9](http://doi.org/10.1016/s0896-6273(00)00168-9)
- Jenkinson, M., Bannister, P. R., Brady, J. M., and Smith, S. M. (2002). Improved optimisation for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, 17(2):825-841.
- Jenkinson, M., and Smith, S. M. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, 5(2), 143-156.
- Klein, I., Paradis, A., Poline, J., Kosslyn, S. M., Le Bihan, D. (2000). Transient activity in the human calcarine cortex during visual-mental imagery: an event-related fMRI study. *Journal of Cognitive Neuroscience*, 12(2), 15-23.
- Knauff, M., Kassubek, J., Mulack, T., and Greenlee, M. W. (2000). Cortical activation evoked by visual mental imagery as measured by fMRI. *Neuroreport*, 11(18), 3957–3962. <http://doi.org/10.1097/00001756-200012180-00011>
- Kosslyn, S. M. (1981). The medium and the message in mental imagery: A theory. *Psychological Review*, 88(1), 46–66. <http://doi.org/10.1037/0033-295X.88.1.46>

- Kosslyn, S. M. (2005). Mental images and the Brain. *Cognitive Neuropsychology*, 22(3–4), 333–347.
<http://doi.org/10.1080/02643290442000130>
- Kosslyn, S. M., Alpert, N. M., Thompson, W. L., Maljkovic, V., Weise, S. B., Chabris, C. F., Hamilton, S.E., Rauch, S.L., and Buonanno, F. S. (1993). Visual mental imagery activates topographically organized visual cortex: PET investigations. *Journal of Cognitive Neuroscience*, 5, 263–287. <http://doi.org/10.1162/jocn.1993.5.3.263>
- Kosslyn, S. M., and Thompson, W. L. (2003). When is early visual cortex activated during visual mental imagery? *Psychological Bulletin*, 129(5), 723–746. <http://doi.org/10.1037/0033-2909.129.5.723>
- Kourtzi, Z., and Kanwisher, N. (2000). Cortical regions involved in perceiving object shape. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 20(9), 3310–8.
- Kriegeskorte, N., Goebel, R., and Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 3863–3868. <http://doi.org/10.1073/pnas.0600244103>
- Lee, S., Kravitz, D. J., and Baker, C. I. (2012). Disentangling visual imagery and perception of real-world objects. *NeuroImage*, 59(4), 4064–4073.
<http://doi.org/10.1016/j.neuroimage.2011.10.055>
- Marks, D.F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology*, 64, 17-24.
- Marzi, C., Mancini, F., Metitieri, T., and Savazzi, S. (2006). Retinal eccentricity effects on reaction time to imagined stimuli. *Neuropsychologia*, 44(8), 1489–1495.
<http://doi.org/10.1016/j.neuropsychologia.2005.11.012>
- Mechelli, A., Price, C.J., Friston, K.J., and Ishai, A. (2004). Where bottom-up meets top-down: Neuronal interactions during perception and imagery. *Cerebral Cortex*, 14(11), 1256–1265.
<http://doi.org/10.1093/cercor/bhh087>

- Mellet, E., Briscoigne, S., Tzourio-Mazoyer, N., Ghaëm, O., Petit, L., Zago, L., Etard, O., Berthoz, A., Mazoyer, B., and Denis, M. (2000). Neural correlates of topographic mental exploration: the impact of route versus survey perspective learning. *NeuroImage*, 12(5), 588–600. <http://doi.org/10.1006/nimg.2000.0648>
- Misaki, M., Kim, Y., Bandettini, P. A., and Kriegeskorte, N. (2010). Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. *NeuroImage*, 53(1), 103–118. <http://doi.org/10.1016/j.neuroimage.2010.05.051>
- Muckli, L., Kohler, A., Kriegeskorte, N., and Singer, W. (2005). Primary visual cortex activity along the apparent-motion trace reflects illusory perception. *PLoS Biology*, 3(8). <http://doi.org/10.1371/journal.pbio.0030265>
- Muckli, L., and Petro, L. S. (2013). Network interactions: Non-geniculate input to V1. *Current Opinion in Neurobiology*, 23(2), 195-201. <http://doi.org/10.1016/j.conb.2013.01.020>
- Naughtin, C. K., Mattingley, J. B., and Dux, P. E. (2014). Distributed and overlapping neural substrates for object individuation and identification in visual short-term memory. *Cerebral Cortex*, 26(2), 566–575. <http://doi.org/10.1093/cercor/bhu212>
- Oosterhof, N. N., Connolly, A. C., and Haxby, J. V. (2016). CoSMoMVPA: Multi-Modal Multivariate Pattern Analysis of neuroimaging data in Matlab/GNU Octave. *Frontiers in Neuroinformatics*, 10, 27. <http://doi.org/10.3389/fninf.2016.00027>
- Oosterhof, N. N., Tipper, S. P., and Downing, P. E. (2012). Visuo-motor imagery of specific manual actions: A multi-variate pattern analysis fMRI study. *NeuroImage*, 63(1), 262-271. <http://doi.org/10.1016/j.neuroimage.2012.06.045>
- Pourtois, G., Schwartz, S., Spiridon, M., Martuzzi, R., and Vuilleumier, P. (2009). Object representations for multiple visual categories overlap in lateral occipital and medial fusiform cortex. *Cerebral Cortex*, 19(8), 1806–1819. <http://doi.org/10.1093/cercor/bhn210>
- Reddy, L., Naotsugu, T., and Serre, T. (2011). Reading the mind's eye: Decoding category information during mental imagery. *Neuroimage*, 50(2), 818–825.

<http://doi.org/10.1016/j.neuroimage.2009.11.084>

Sack, A. T., and Schuhmann, T. (2012). Hemispheric differences within the fronto-parietal network dynamics underlying spatial imagery. *Frontiers in Psychology*, 3(JUN), 1–10.

<http://doi.org/10.3389/fpsyg.2012.00214>

Sack, A. T., Sperling, J. M., Prvulovic, D., Formisano, E., Goebel, R., Di Salle, F., Dierks, T., and Linden, D. E. J. (2002). Tracking the mind's image in the brain II: Transcranial magnetic stimulation reveals parietal asymmetry in visuospatial imagery. *Neuron*, 35(1), 195–204.

[http://doi.org/10.1016/S0896-6273\(02\)00745-6](http://doi.org/10.1016/S0896-6273(02)00745-6)

Savini, N., Brunetti, M., Babiloni, C., and Ferretti, A. (2012). Working memory of somatosensory stimuli: an fMRI study. *International Journal of Psychophysiology*, 86(3), 220-228.

<https://doi.org/10.1016/j.ijpsycho.2012.09.007>

Schwarzbach, J. (2011). A simple framework (ASF) for behavioral and neuroimaging experiments based on the psychophysics toolbox for MATLAB. *Behavior Research Methods*, 43(4), 1194-201. <https://doi.org/10.3758/s13428-011-0106-8>

Schmidt, T. T., Ostwald, D., and Blankenburg, F. (2014). Imaging tactile imagery: Changes in brain connectivity support perceptual grounding of mental images in primary sensory cortices.

NeuroImage, 98, 216–224. <http://dx.doi.org/10.1016/j.neuroimage.2014.05.014>

Stelzer, J., Chen, Y. and Turner, R. (2013). Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): random permutations and cluster size control. *NeuroImage*, 65, 69–82. <http://dx.doi.org/10.1016/j.neuroimage.2012.09.063>

Stokes, M., Saraiva, A., Rohenkohl, G., and Nobre, A. C. (2011). Imagery for shapes activates position-invariant representations in human visual cortex. *NeuroImage*, 56(3), 1540-1545.

<http://doi.org/10.1016/j.neuroimage.2011.02.071>

- Stokes, M., Thompson, R., Cusack, R., and Duncan, J. (2009). Top-down activation of shape-specific population codes in visual cortex during mental imagery. *Journal of Neuroscience*, *29*(5), 1565–1572. <http://doi.org/10.1523/JNEUROSCI.4657-08.2009>
- Trojano, L. (2000). Matching two imagined clocks: The functional anatomy of spatial analysis in the absence of visual stimulation. *Cerebral Cortex*, *10*(5), 473–481. <http://doi.org/10.1093/cercor/10.5.473>
- Tziortzi, A.C., Searle, G.E., Tzimopoulou, S., Salinas, C., Beaver, J.D., Jenkinson, M., Laruelle, M., Rabiner, E.A., and Gunn, R.N. (2011). Imaging dopamine receptors in humans with [11C]-(+)-PHNO: dissection of D3 signal and anatomy. *NeuroImage*, *54*(1), 264–77. <http://doi.org/10.1016/j.neuroimage.2010.06.044>
- Ungerleider, L. G., and Haxby, J. V. (1994). “What” and “where” in the human brain. *Current Opinion in Neurobiology*, *4*(2), 157–165.
- Vetter, P., Smith, F. W., and Muckli, L. (2014). Decoding sound and imagery content in early visual cortex. *Current Biology*, *24*(11), 1256–1262. <http://doi.org/10.1016/j.cub.2014.04.020>
- Winlove, C. I. P., Milton, F., Ranson, J., Fulford, J., MacKisack, M., Macpherson, F., and Zeman, A. (2018). The neural correlates of visual imagery: A co-ordinate-based meta-analysis. *Cortex*. <http://doi.org/10.1016/j.cortex.2017.12.014>
- Worsley, K. J., Marrett, S., Neelin, P., Vandal, A. C., Friston, K. J., and Evans, A. C. (1996). A unified statistical approach for determining significant signals in images of cerebral activation. *Human Brain Mapping*, *4*(1), 58–73. [http://doi.org/10.1002/\(SICI\)1097-0193\(1996\)4:1<58::AID-HBM4>3.0.CO;2-O](http://doi.org/10.1002/(SICI)1097-0193(1996)4:1<58::AID-HBM4>3.0.CO;2-O)
- Yomogida, Y., Sugiura, M., Watanabe, J., Akitsuki, Y., Sassa, Y., Sato, T., Matsue, Y., and Kawashima, R. (2004). Mental visual synthesis is originated in the fronto-temporal network of the left hemisphere. *Cerebral Cortex*, *14*(12), 1376–1383. <http://doi.org/10.1093/cercor/bhh098>

Supplementary Material

Stimulus selection and validation

To select the stimuli to be used in the fMRI task, we performed a behavioral study outside the scanner. This study was based on the eccentricity effect (Chelazzi et al., 1988): participants are faster to detect stimuli presented closer to fixation in comparison to stimuli presented further away from fixation. This effect is assumed to be based on cortical magnification, i.e. the reduction of the number of neurons in the visual cortex responsible for processing visual stimuli as function of visual field location (Kitterle, 1986; Marzi & Di Stefano, 1981). In particular, visual stimuli presented in the center of the visual field (i.e. fovea) are processed by a larger number of neurons compared to more peripheral locations. Marzi et al. (2006) demonstrated that a similar eccentricity effect can be obtained for imagined stimuli and interpreted this observation as a sign of the involvement of retinotopically organized early visual cortex during visual mental imagery. Following this logic, we reasoned that stimuli that show a reliable eccentricity effect should be good candidates to be chosen for our fMRI study.

Participants. Fourteen 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 difficulties in maintaining their gaze at fixation during the task in at least two out of four experimental runs (see section *Data analysis, eye-tracking data*), three participants were excluded. Two additional participants were excluded due to a high number of missed responses (15% and 69%), leading to a final sample of nine participants (5 females, 4 males, mean age 25.2 ± 4.7). The study was approved by the Ethics Committee for research involving human participants at the University of Trento, Italy.

Stimuli. We asked participants to imagine six different stimulus categories (checkerboards, gratings, simple shapes, lowercase letters, objects and invented shapes), each one composed by four different stimulus exemplars (*Figure S1*). Checkerboards, gratings, invented shapes, lowercase letters and simple shapes were created in Inkscape (<http://www.inkscape.org>). Objects were selected from the 260 standardized pictures composing the Snodgrass database (Snodgrass & Vanderwart, 1980), based on the criteria that they should differ widely for low-level visual features (i.e. orientation and shape) and the amount of visual details. Stimuli were then edited in Inkscape (<http://www.inkscape.org>), to match their background color across categories.

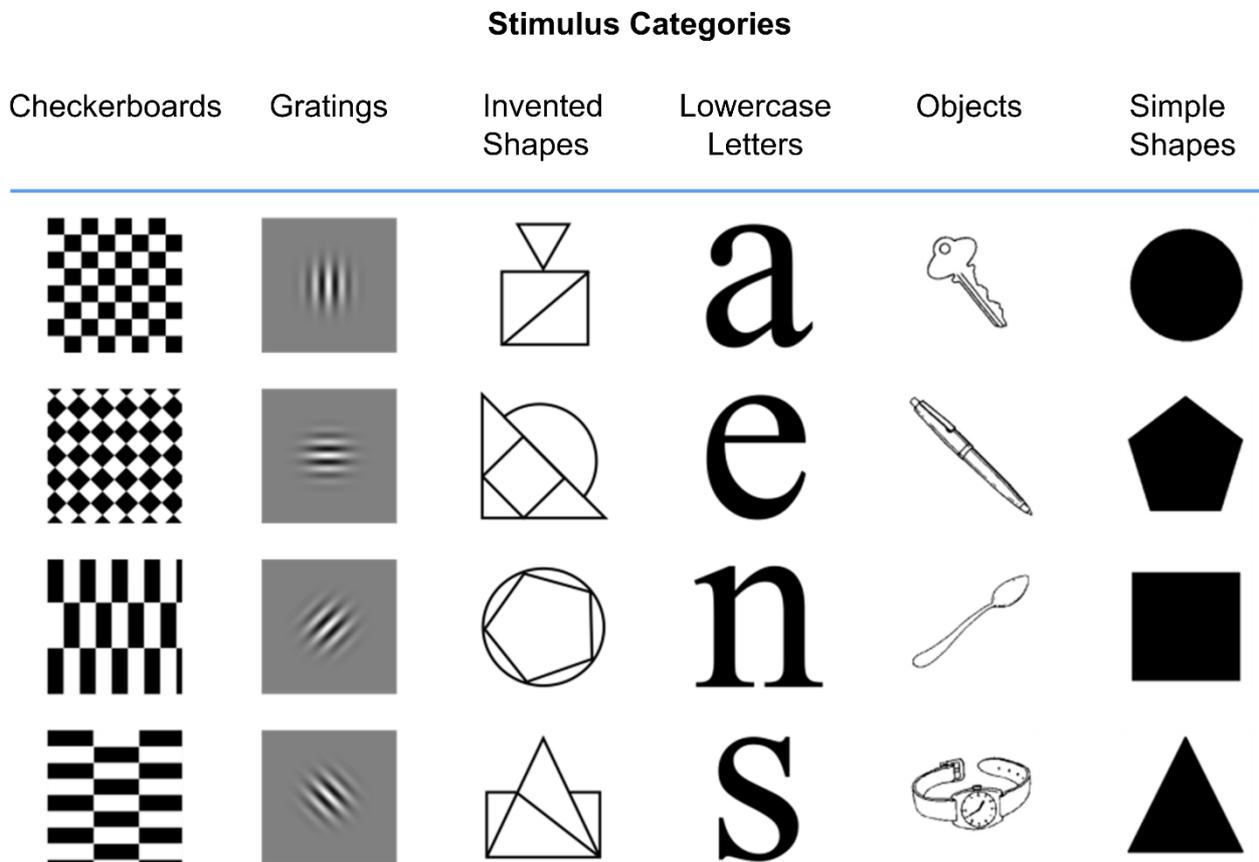


Figure S1. Full set of stimulus exemplars used in the behavioral pilot study. For each stimulus category (checkerboards, gratings, invented shapes, lowercase letters, objects and simple shapes), four stimulus exemplars were used.

To prevent afterimages during visual presentation of the stimuli, we created a visual mask, common for all stimulus exemplars. The mask was created using a custom-written MATLAB script. Each stimulus was converted to grayscale, eliminating hue and saturation information. Then, values for

each pixel were averaged across all stimuli, resulting in a sample picture. As the last step, the spatial position of pixels of this sample picture was shuffled, resulting in a black and white visual mask (*Figure S2*).

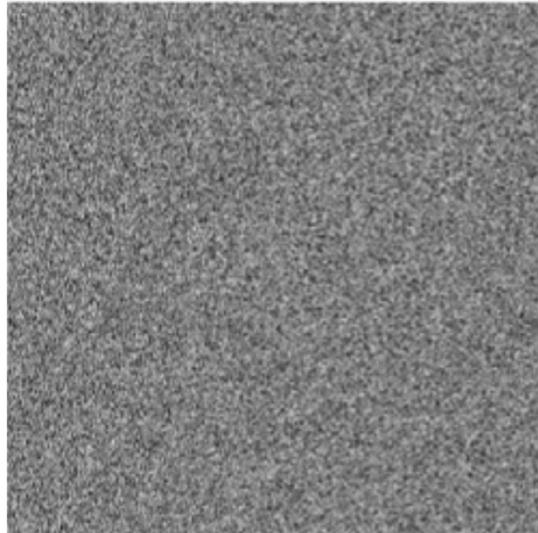


Figure S2. Visual mask used in the behavioral pilot study. To create the visual mask, each stimulus was converted to grayscale, eliminating hue and saturation information. Then, values for each pixel were averaged across all stimuli. In the last step, the spatial position of pixels was shuffled, resulting in a black and white visual mask (see text for more details).

Experimental design. The experiment was divided in two sessions, to be performed on two consecutive days. Each session comprised two runs (24 blocks each), each one lasting approximately 20 minutes. The entire experiment comprised a total of 960 trials (40 trials for each combination of spatial position (4) and stimulus exemplar (6)). On each day, participants saw all the stimuli twice, in a randomized order. Participants were instructed to perform visual imagery at four different spatial locations, either 2° or 8° of visual angle (relatively to the central fixation point) to the left or right side of the screen. To keep the size of the imagined stimuli constant, participants were asked to performed visual imagery within placeholders positioned at the aforementioned locations and comprising 2° of visual angle (*Figure S3*). All four placeholders were presented on the screen throughout the trial. Each block started with the presentation of the stimulus exemplar to be imagined (1 s) at the center of the screen, followed by a mask (1 s). An auditory cue (“far left”, “far right”,

“near left”, “near right”; 400 ms) instructed participants at which location they had to imagine the stimulus. Participants were asked to indicate by button press when they reached a vivid mental image. If no response was provided, the program automatically passed to the next trial after 6 seconds. After 10 trials, a new block started, and the stimulus to be imagined changed. At the end of the experiment, we obtained subjective ratings of the vividness of visual imagery and the perceived difficulty in generating the imagined stimuli from each participant. We recorded reaction times, and monitored eye-movements using a video-based eye tracking system (Eyelink 1000, SR Research).

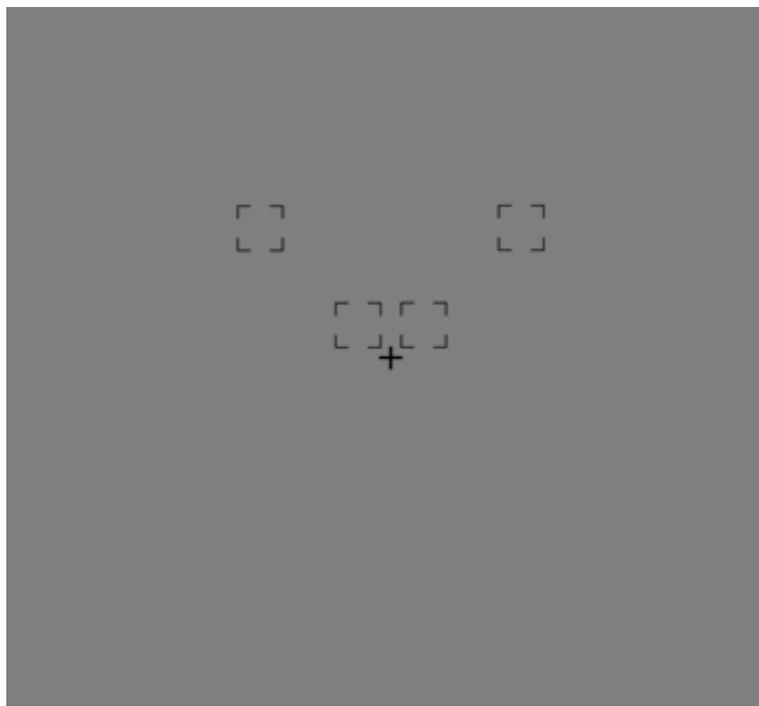


Figure S3. Behavioral pilot study. During each block, participants had to imagine the selected stimulus at four different spatial locations, either 2° or 8° from central fixation, both on the left and on the right side of the screen. To keep the size and position of the mental image constant, participants were instructed to perform visual mental imagery within one of four placeholders positioned at the aforementioned locations, comprising 2° visual angle.

Data analysis.

To assess the ability of our participants in maintaining their gaze at fixation, we examined their gaze position throughout each run with respect to the central fixation cross. For each participant, we analyzed the 6 seconds imagery delay starting after the cue onset. We averaged reaction times (RTs)

for each eccentricity (2° and 8°) and side (left or right), separately for each stimulus category. We then performed a repeated-measures ANOVA, with eccentricity (2 levels), side (2 levels), and stimulus category (6 levels) as factors. Moreover, to assess differences in vividness and difficulty in generating mental images between the six selected stimulus categories, we performed two additional repeated-measures ANOVAs (factors: vividness rating and stimulus category (6 levels) for vividness ratings; difficulty ratings and stimulus categories (6 levels) for difficulty ratings). Degrees of freedom were adjusted by the Greenhouse-Geisser procedure where appropriate (corresponding p-values denoted as p_{GG}).

Results.

Participants were faster to indicate that they experienced a vivid mental image when they were asked to perform visual imagery near (2°; mean RT = 2669.3 ± 333 ms) in comparison to further away (8°; mean RT = 2973.5 ± 417.8 ms) from fixation, in line with previous studies (Marzi et al., 2006). This observation is supported by the corresponding repeated-measures ANOVA [main effect eccentricity: $F(1, 8) = 16.8$, $p = 0.003$]. Reaction times differed between stimulus categories [main effect of category: $F(2.3, 26.1) = 10.46$, $p_{GG} = 0.001$; mean simple shapes: 2575.5 ± 114.6 ms; mean objects: 2774.2 ± 125.7 ms; mean lowercase letters: 2827.3 ± 112.9 ms; mean gratings: 2828.7 ± 133 ms; mean checkerboards: 3005.6 ± 133.5 ms; mean invented shapes: 3020.5 ± 154.8 ms;]. The eccentricity effect was not modulated by stimulus category [interaction categories*eccentricity: $F(2.42, 19.40) = 0.553$, $p_{GG} = 0.616$], suggesting that the eccentricity effect was present for all six examined stimulus categories.

For two categories (i.e. checkerboards and invented shapes) we found a stronger eccentricity effect when imagery was performed on the left side of the screen, whereas for the remaining ones (i.e. gratings, lowercase letters, simple shapes and objects), the pattern was the opposite [interaction side*eccentricity*category: $F(5.55) = 2.63$, $p = 0.034$. See *Figure S4*]. None of the other interactions

were significant [interaction category*side: $F(2.29, 18.38) = 1.2, p = 0.328$; interaction side*eccentricity: $F(1, 8) = 0.340, p = 0.576$].

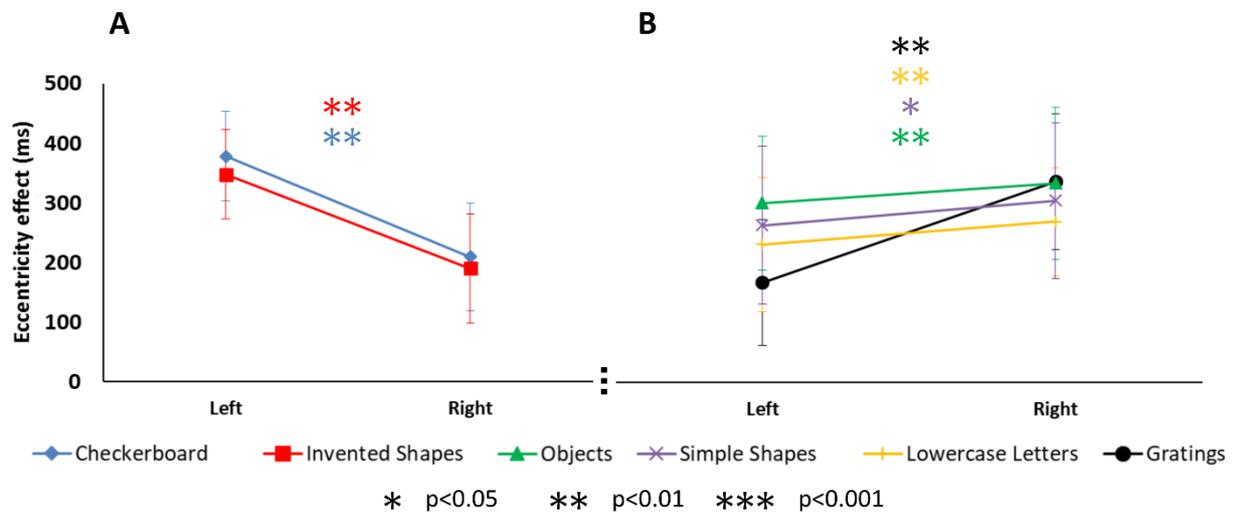


Figure S4. Magnitude of the eccentricity effect (RTs for stimuli presented at 8° - RTs for stimuli presented at 2° visual angle) as a function of stimulus position (left, right), separately for the six stimulus categories. As can be seen, the eccentricity effect was larger for stimuli presented on the left side of the screen for checkerboards and invented shapes (A), whereas the opposite pattern was obtained for objects, simple shapes, lower case letters, and gratings (B). Error bars: standard error of the mean (S.E.M.).

A repeated-measures ANOVA of the vividness and difficulty ratings revealed that both ratings were modulated by stimulus category [main effect of stimulus category for vividness rating: $F(5, 40) = 6.45, p = 0.003$; main effect of stimulus category for difficulty ratings: $F(5, 40) = 8.157, p < 0.001$; see *Figure S5a*].

We thus decided to select the three categories associated with the lowest vividness ratings (indicating the most vivid experience) and the lowest difficulty ratings (i.e. lowercase letters, objects and simple shapes; see *Figure S5a*) for the fMRI study.

To test whether the eccentricity effect differed between the four stimulus exemplars constituting each category, we performed an additional ANOVA, separately for each stimulus category [repeated-measures ANOVA, factors: eccentricity, side, stimulus exemplars. Lowercase letters. Interaction eccentricity*stimulus exemplar: $F(3, 24) = 1.69, p = 0.195$; Objects. Interaction stimulus exemplars*eccentricity: $F(3, 24) = 1.137, p = 0.354$. Simple Shapes. Interaction stimulus

exemplars*eccentricity: $F(3, 24) = 0.846, p = 0.482$]. Due to the lack of a clear difference in the magnitude of the eccentricity effects between different stimulus exemplars, we selected two exemplars within each category that differed widely for low-level visual features, such as orientation or shape (i.e. circle and triangle for simple shapes; pen and watch for objects; letter ‘e’ and ‘n’ for lowercase letters. *Figure S5b*).

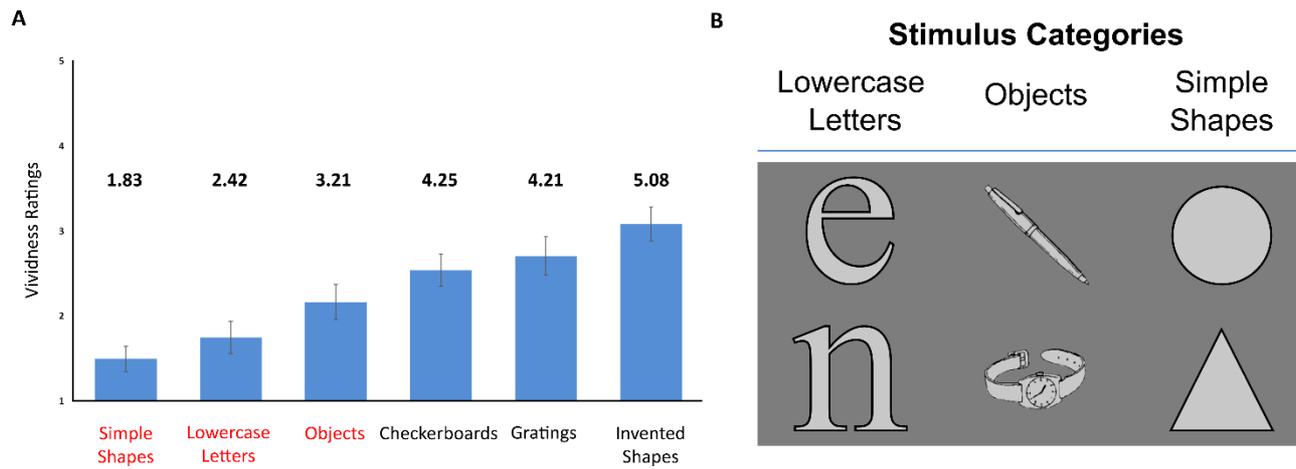


Figure S5. (A) Individual ratings of difficulty (black numbers) and vividness (blue bars) of visual mental imagery for each stimulus category used in the behavioral experiment carried out to select stimuli to be used in the fMRI experiment. Stimulus categories highlighted in red are those selected for the fMRI study. Vividness scale: 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. Difficulty scale: 1. Easiest category to imagine – 6. Hardest category to imagine. (B) Stimulus categories and exemplars used for the fMRI experiment. For each stimulus category (i.e. lowercase letters, objects and simple shapes), we selected two stimulus exemplars on the basis of the behavioral pilot experiment described above.

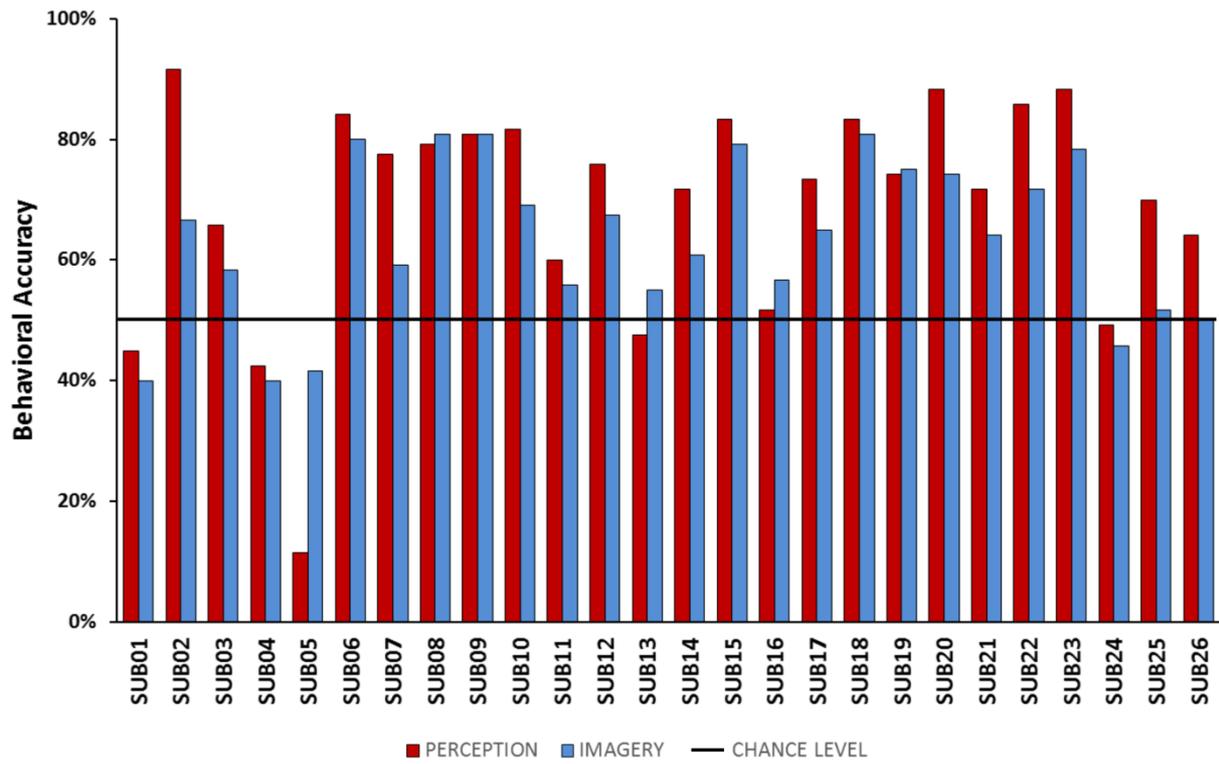


Figure S6. Individual accuracies for $N=26$ participants that took part in the fMRI experiment, separately for the perception (red) and imagery (blue) condition. The black line indicates chance level performance for the delayed spatial judgment task. Participants 1, 4, 5, 13, 24 showed a performance below chance level either in the imagery or in the perception condition, and were therefore excluded from further analyses.

References

- Chelazzi, L. (1988). Hemiretinal difference in speed of light detection in esotropic amblyopes. *Vision Research*, 28(1), 95–104.
- Kitterle, F. L. (1986). Psychophysics of lateral tachistoscopic presentation. *Brain and Cognition*, 5(2), 131–162. [http://doi.org/10.1016/0278-2626\(86\)90052-7](http://doi.org/10.1016/0278-2626(86)90052-7)
- Marzi, C. A., and Di Stefano, M. (1981). Hemiretinal differences in visual perception. *Documenta Ophthalmologica, Proceeding Series*, 30, 273–278. <http://doi.org/10.1371/journal.pone.0094539>
- Snodgrass, J. G., and Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning & Memory*, 6(2), 174–215. <http://doi.org/10.1037/0278-7393.6.2.174>

Tables

Univariate Analysis

Analysis	Cluster size	Cluster regions (Juelich Histological Atlas)	Hemisphere	Max Z	Max X (mm)	Max Y (mm)	Max Z (mm)
Perception > Baseline	24416	V3	Right	6.94	30	-94	-2
		V2	Right	6.8	22	-94	-12
				6.41	26	-92	8
		V4	Left	6.65	-22	-84	-18
		V1	Right	6.51	16	-102	6
				6.43	20	-100	6
	2147	Dorsal Premotor Cortex (PMd)	Left	5.42	-40	4	62
				4.98	-28	0	52
		Broca's Area	Left	4.76	-46	10	34
				4.43	-42	4	34
				3.58	-30	8	70
				3.54	-44	32	40
Imagery > Baseline	1794	Primary Somatosensory Cortex (S1)	Left	4.34	-48	-42	60
				3.43	-46	-32	44
	Anterior Intraparietal Sulcus (aIPS)	Left	4.21	-34	-44	42	
			3.69	-34	-34	36	
	Inferior Parietal Lobe (IPL)	Left	3.58	-54	-46	52	
	Superior Parietal Lobe	Left	3.54	-14	-70	58	

		(SPL)					
Analysis	Cluster size	Cluster regions (Juelich Histological Atlas)	Hemisphere	Max Z	Max X (mm)	Max Y (mm)	Max Z (mm)
Imagery > Baseline	1697	PMd	Left	4.68	-24	-4	52
				3.97	-6	6	56
				3.94	-42	4	62
				3.16	-48	2	40
		Broca's Area	Left	3.6	-50	8	30
				3.23	-30	2	34

Searchlight-based MVPA

Analysis	Cluster size	Cluster regions (Juelich Histological Atlas)	Hemisphere	Max t	Max X (mm)	Max Y (mm)	Max Z (mm)
Perception Condition	2820	V1	Right	8.05	12	-93	3
				7.33	15	-99	3
				7.25	15	-78	3
				6.73	15	-96	-6
				6.57	9	-96	-6
		V2	Left	7.16	-3	-93	-9
	160	SPL	Left	4.66	-18	-72	45
				4.46	-9	-57	72
				4.38	-27	-66	54
				4.18	-24	-57	69
aIPS		Left	4.64	-21	-60	45	
			4.53	-21	-54	51	
Imagery Condition	8107	V1	Left	6.93	-3	-93	9
				6.38	-9	-87	12
				5.69	-27	-66	3
		V4	Left	6.42	-39	-87	-15
				5.52	-51	-81	0

Analysis	Cluster size	Cluster regions (Juelich Histological Atlas)	Hemisphere	Max Z	Max X (mm)	Max Y (mm)	Max Z (mm)
Imagery condition	137	Broca's Area	Left	5.16	-45	39	24
Cross-decoding condition	8609	LOC (Grill-Spector et al., 1999; Pourtois et al., 2008)	Left	7.05	-42	-60	-3
				6.74	-54	-60	-6
		hMT	Left	6.78	-42	-75	9
		V2	Left	6.29	-6	-75	3
		SPL	Left	6.21	-21	-72	48
		aIPS	Left	6.09	-24	-66	54