Face-, color-, and word-specific patches in the human orbitofrontal cortex

Jianghao Liu[1,2*], Laurent Cohen[1,3], Minye Zhan[1,4,†], Paolo Bartolomeo[1,†]

1. Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, Inserm, CNRS, AP-HP, Hôpital de la Pitié-Salpêtrière, F-75013 Paris, France

2. Corporate Research, Dassault Systèmes, 78140 Vélizy-Villacoublay, France

3. AP-HP, Hôpital de la Pitié Salpêtrière, Fédération de Neurologie, Paris, France

4. Cognitive Neuroimaging Unit, Université Paris-Saclay,CEA, INSERM, CNRS ELR9003, NeuroSpin center, 91191 Gif/Yvette, France

† Shared senior authorship

* Correspondence: jianghaolouisliu@gmail.com (J.L)

Abstract

The human ventral occipitotemporal cortex (VOTC) contains multiple category-specific areas, organized along posterior-to-anterior and medial-to-lateral axes. However, the role of regions beyond the VOTC in category-specific processing remains less explored. Here, we report the presence of face-, color- and word-specific patches in the human orbitofrontal cortex (OFC) and systematically describe their location, activity amplitude, category selectivity, representational content, and functional connectivity. We compare these features with those of corresponding VOTC category-specific patches. Our findings reveal that face- and color-specific patches in the OFC follow a similar medial-lateral organization to those in the VOTC and form continuous functional gradients with VOTC patches. These results suggest that the OFC contains a topographic organization similar to, but at a higher hierarchical level than, the VOTC object categorization system.

One Sentence Summary: Face-, color-, and word-specific patches are present in the human orbitofrontal cortex and form a continuous functional hierarchy with corresponding patches in the ventral occipitotemporal cortex.

Introduction

The ability of object recognition in humans is supported by the ventral occipitotemporal cortex (VOTC) (see Ref.Grill-Spector & Weiner, 2014 for a review). Within VOTC, multiple sets of areas were found to be sensitive to specific object categories or visual features, including faces[2], bodies[3], words[4], colors[5], places[6]. In this article, we use "categories" to denote these domains or features, and use "specificity" interchangeably with preference/selectivity/sensitivity.

Spatially, these VOTC category-specific areas are hierarchically organized, forming topographic maps that follow two major organizational rules across the cortical space. In the posterior-to-anterior axis from early visual cortices to the anterior temporal cortex, each set of category-specific areas exhibited increasing specificity and invariance, implementing distinct levels of categorical abstraction[1]. In the medial-to-lateral axis, regions specific for places, colors, faces, words were found laying out in this order, which is thought to follow a larger organizational rule of eccentricity[1,7], where preference for more foveal stimuli would lie in more lateral regions. Across multiple category-specificity, patches often localize closely, and repeat along the posterior-to-anterior axis, for example, parallel face- and color-specific patches in both humans and monkeys[8,9], and multiple repetitions of patches sensitive to faces, bodies, spiky, stubby objects in monkeys[10].

Most previous studies focused on VOTC regions, particularly around the fusiform gyrus and nearby areas, with the anterior temporal cortex considered the apex of these category-specific processes[11]. Category-specificity beyond the vicinity of the VOTC[12,13] is much less explored. Few studies reported prefrontal face-specific areas(see Ref.Duchaine & Yovel, 2015 for a review), but it remains uncertain whether these areas contain patches for multiple categories with a topographical organization.

We found category-specific patches for faces and colors in the lateral orbitofrontal cortex (OFC) in a recent high-resolution (1.2 mm isotropic) 7T fMRI study on perception and mental imagery of the same visual items across five categories (faces, colors, words, shapes, maps)[15]. The majority of participants exhibited activity in such patches during both perception and imagery, even those who reported a lifetime absence of voluntary imagery, or aphantasia[16]. Interestingly, the OFC face and color patches seemed to follow the same medial-to-lateral layout as in the VOTC.

These OFC face- and color-specific patches cannot be fully accounted for by the major theories on the functional architecture of the OFC. One theory posits that the OFC

acts as a “top-down predictor” to facilitate visual recognition[17,18]. In this view, the OFC receives visual information carried by the low spatial frequencies, predicts ambiguous visual inputs in a top-down manner, as indicated by its shorter response latency (~50 ms earlier than in the VOTC) and its functional connectivity with the VOTC. Another theory posits that the OFC acts as a highly abstract “cognitive map”[19], which integrates relational structures of stimuli to guide decision-making(Wikenheiser & Schoenbaum, 2016) and organize abstract knowledge for behavior(Behrens et al., 2018). Monkey and rodent studies supporting this theory showed that the OFC represents abstract information pertinent to task states and integrates perceptual inputs from other brain areas[20]. Both theories suggested that the OFC constructs perception and integrates it into broader cognitive contexts, with a predictive role. However, none of the theories predicted the multiple category-specificity found in our study. A small number of monkey studies reported face-specificity or other stimulus-specificity in the OFC[21–23], which haven’t been well integrated into the major OFC theories.

In the current study, we use two independent 7T fMRI datasets to specify the OFC category selectivity for faces, colors, and words. In one dataset 1[15], we perform an in-depth study on the OFC face- and color-specific patches, focusing on their location, category-selectivity, activity amplitude, representational content, and functional connectivity. This unique ultra-high-resolution dataset of perception and imagery offers us the opportunity to examine OFC functions both with and without direct visual input, and to compare the OFC functions of typical imagers to that of aphantasic individuals. The inability to visualize in aphantasia was associated with difficulties in face recognition[16], potentially related to deficits of top-down prediction for faces. Our results show that the OFC face- and color-specific patches had very high category specificity, forming multiple functional trends integrated with VOTC face- and color-specific patches. In searching for potential OFC selectivity for other categories, in an independent 7T fMRI dataset of word reading (dataset 2[24]), we found word-specific patches in the OFC in 13 of 21 participants. Our results suggest that the OFC contains a topographic organization similar to, but at a higher hierarchical level than, the VOTC object categorization system.

Results

We report 7-Tesla fMRI data, focusing on OFC face- and color-specific patches (Dataset 1) and OFC word-specific patches (Dataset 2). No spatial smoothing was applied to any data

in this study. We describe the functional anatomy and activity profiles of the OFC patches, as well as their relationship with the VOTC patches. In Dataset 1, we further examine in depth for their functional properties, and compare imagery with perception and aphantasia with typical imagery in each section.

Dataset 1: Adjacent face- and color-specific patches in the OFC

Twenty participants (10 typical imagers and 10 aphantasic individuals) compared pairs of stimuli during visual perception and visual mental imagery in five different categories: object shapes, object colors, written words, famous faces, and a map of France[15]. The imagery task was carried out before the perception task. For the face imagery task, participants heard a word indicating the category (e.g., “face”), followed by two names (e.g., “E. Macron”, “F. Hollande”). They had to imagine and maintain the faces for eight seconds, then decide which face matched an attribute word (e.g., “round”) by pressing a button. For the color imagery task, they decided which food item had a darker or lighter color. In the corresponding perception tasks, the same stimuli were presented as auditory words with visual pictures. An abstract words task was also included to minimize mental imagery. Behavioral results showed that aphantasics performed the imagery tasks as accurately as typical controls but with a slower response time[15], consistent with the results from a larger sample using the same tasks[25].

Along the VOTC, we identified individual patches activated by the perception of faces or colors relative to stimuli from the other categories (functional voxel of 1.2 mm, no smoothing; voxelwise $p<0.001$ uncorrected, cluster size threshold = 12 voxels). For face-specific patches, which are generally bilateral, the corresponding left- and right-hemispheric patches were merged in subsequent ROI analyses. We identified the following areas: occipital face area (OFA; group-average Talairach coordinates, x = 38, y = -79, z = -6), fusiform face area 1 (FFA1; x = 37, y = -63, z = -11), fusiform face area 2 (FFA2; x = 37, y = -47, z = -15), anterior face patch (AFP; x = 36, y = -23, z = -16). We similarly identified bilateral color-specific patches: posterior color patch (PC; x = 29, y = -69, z = -10), central color patch (CC; x = 30, y = -51, z = -12), anterior color patch (AC; x = 29, y = -34, z = -15). The medial-lateral layout between face and color patches (Bayesian paired t-test of the x-coordinates, Bayes factor [BF] = 265.76) were consistent with previous work[8,26].

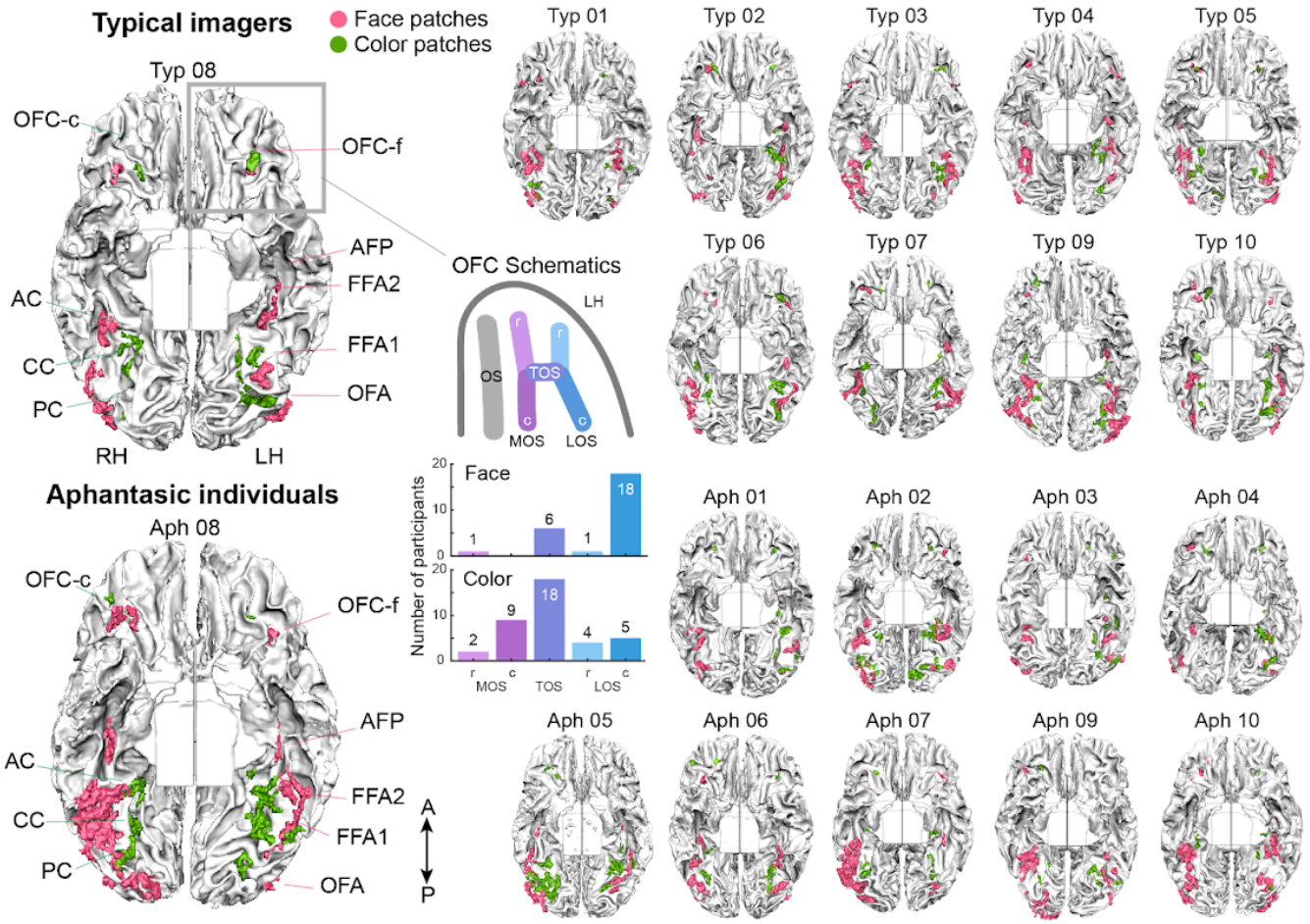
Critically, in the OFC, we also found adjacent face- and color-specific patches bilaterally during perception. We identified OFC face patches (OFC-f, x = 34, y = 30, z = -6) in 18 out of 20 individuals (all typical imagers and 8 aphantasics) and OFC color patches

(OFC-c, x = 26, y = 33, z = -8) in 19 out of 20 individuals (9 typical imagers and all aphantasics), showing very strong replicability across individual subjects. The locations of these patches corresponded to specific macroanatomical landmarks (Fig. 1A, see inset for schematics of the OFC anatomy, and occurrences of patches across participants), including the caudal parts of both the lateral orbital sulcus (LOS) and the medial orbital sulcus (MOS), and the transverse orbital sulcus (TOS). Face patches were located more laterally than color patches in the OFC (Bayesian paired t-test of x-coordinates, BF = 12.68), similar to their medial-lateral layout in the VOTC (all BFs > 10). The y and z coordinates did not differ between OFC face and color patches in either typical or aphantasic individuals (Bayesian paired t-tests, all BFs < 0.92).

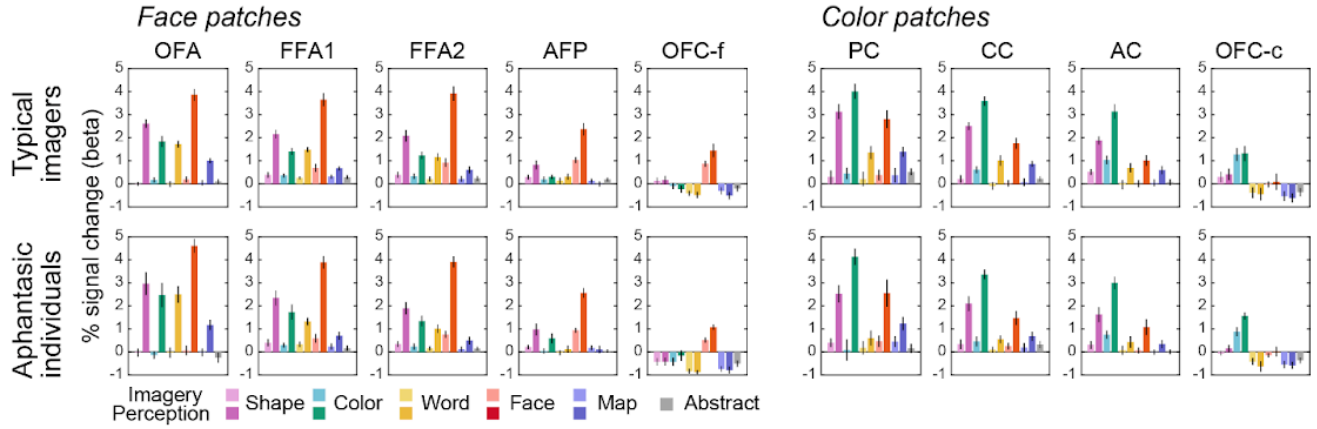
Next, we investigated the general activity profiles of these OFC patches across task conditions and time points. In both typical and aphantasic participants, the BOLD signal changes (beta values) differed between patches across tasks (Bayesian ANOVAs with the factors Patch and Task, including all tasks; main Patch effects showed Bayes factors = infinity), indicating that activity is not uniform across patches. During perception, the VOTC patches maintained above-baseline activation for non-preferred stimuli, while showing strong activity for their preferred stimuli (according to ROI definition). However, the OFC patches not only exhibited category-specific activation for faces or colors but also showed below-baseline activity for most of the non-preferred stimuli (Bayesian one-sample t-tests, all Bayes factors > 9.49 for visual words, French maps, and abstract words). These patterns were consistent across both groups (see Fig. S1 for activity profiles in individual participants).

In event-related averages across repetition times (TR), OFC activity was the first to rise above baseline during both perception and imagery, occurring as early as the first TR after stimulus onset (TR=2, one-sample t-test against baseline, all ps < 0.05). The time courses of activity differed between groups in the OFC-f patches (Fig.1C), showing higher activity in aphantasics during the early phase of face perception (TRs 0 and 1), but lower activity during the maintenance phase of face imagery (TRs 8 to 10) compared to typical imagers. No group difference was observed in OFC-c patches or VOTC patches (Fig. S2).

A Face and color patches in OFC and VOTC



B Group-averaged activity profiles



C Averaged time courses in OFC patches

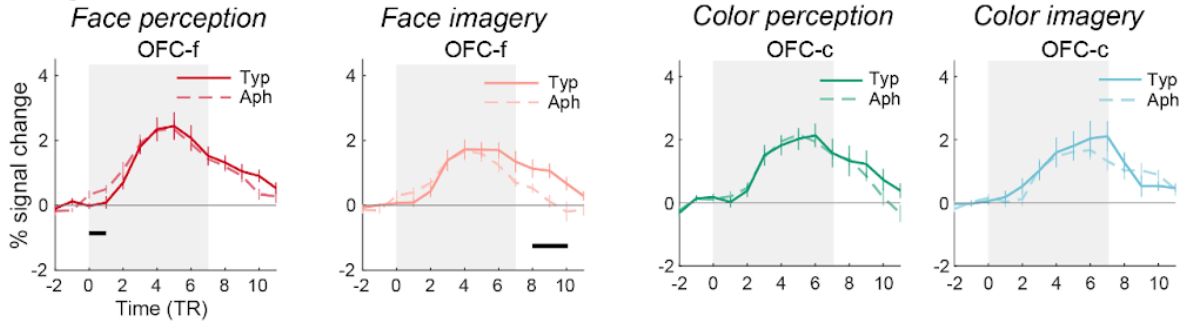


Figure 1. Locations, activity profiles, and time courses of OFC face- and color-specific patches

A. Face and color patches in OFC and VOTC for each individual participant (n=20, including 10 typical imagers [Typ] and 10 aphantasic participants [Aph]). Patches were defined from perceptual trials in the volume space (face or color versus the remaining four categories)[15], 7T fMRI, 1.2 mm functional resolution, no smoothing, p<0.001 uncorrected), and then rendered onto the segmented white matter surface mesh of each individual participant (ventral view). From posterior (P) to anterior (A), face patches include: occipital face area (OFA), fusiform face area 1 (FFA1), fusiform face area 2 (FFA2), anterior face patch (AFP), orbitofrontal face patch (OFC-f); color patches include: posterior color (PC), central color (CC), anterior color (AC), orbitofrontal color patch (OFC-c). In perceptual trials, OFC face and color patches were identified in 18/20 and 19/20 individuals, respectively (Aph 01 and Aph 05 lacked face patches, Typ 04 lacked color patches). For each individual participant, face patches were more lateral than color patches, in both the VOTC and OFC. RH: right hemisphere; LH: left hemisphere. The inset panels show schematics of OFC anatomy[27] and number of occurrences per OFC sub-region. Most face patches were located in the caudal lateral orbital sulcus (LOS) while most color patches were in the transverse orbital sulcus (TOS) or in the caudal medial orbital sulcus (MOS). No patch was identified in the olfactory sulcus (OS). r: rostral; c: caudal.
B. Group-averaged activity profiles (percent signal changes) of OFC face and color patches, across five imagery tasks (lighter colors), five perceptual tasks (darker colors) and abstract words task (gray). Error bars represent SEM across participants. Note that OFC patches showed above-baseline activity for preferred domains but b*elow-baseline* activity for most of the other domains. Abstract: a control task with similar structure, but with exclusively abstract words to compare semantic meaning of words (e.g., “routine”, “convention” compared to “society”), to minimize voluntary visual imagery.
C. Averaged time courses in OFC patches (TR=2 s), plotted in solid lines for typical imagers and dashed lines for aphantasic individuals. Shaded area indicates the period of imagery generation and maintenance, or perception, preceding the response button press. Error bars denote the standard error of the mean across participants. Black horizontal bars represent significant group differences ($p < 0.05$, Bonferroni-corrected across time points). In OFC face patches, aphantasic participants exhibited higher early perceptual activity and reduced activity during imagery maintenance. There were no group differences in any of the VOTC patches (see Fig. S2).

Category selectivity: posterior-to-anterior increasing trends during both perception and imagery

We now quantify the *category selectivity* for each condition relative to other conditions, a critical functional feature of perceptual category-specific areas[2,4]. For each patch, we computed a selectivity index (SI), which is the ratio of activity for faces or colors over the

average amplitude of the remaining four categories. In electrophysiology studies of faces, activity for faces twice as high as other conditions was deemed “face-selective”[28], corresponding to a SI of 0.33. For our fMRI data, we padded the negative activity amplitudes to zero[29]. Across patches, we observed increasing trends of face- and color-selectivity indices from posterior to anterior patches, during both perception and imagery, and for both typical and aphantasic participants (linear contrast analysis, all ps < 0.001; see estimates of slopes in Fig. 2A. Typical imagers, face perception, t = 7.45; face imagery, t = 8.79; color perception, t = 5.02; color imagery, t = 4.94; aphantasic individuals, face perception, t = 7.97; face imagery, t = 8.50; color perception, t = 5.48; color imagery, t = 6.86). There was no evidence for group difference of the slopes (Bayesian repeated measures ANOVAs with the factor of Group x Patch, main Group effects, BFs < 1.19; interaction effects BFs < 0.68).

Notably, the OFC patches showed the highest selectivity for the preferred stimuli during both perception and imagery, surpassing even the anterior temporal patches (Bayesian independent samples t-test, OFC versus AFP or AC patches, all BFs > 104), which were previously considered having very high category-specificity in the visual hierarchy[1]. Moreover, these OFC patches showed strong *negative* selectivity for all non-preferred stimuli (Fig. S3), confirming its strong category-specificity. Thus their responses were tightly dependent on the stimulus type, while not on the direct visual inputs.

In contrast, the most posterior VOTC patches (OFA and PC) exhibited positive selectivity during perception, but *negative* selectivity during imagery (Bayesian one sample t-test below zero, all BFs > 3.05), meaning that they were less reliably activated by their preferred stimuli. Additionally, in aphantasics the face selectivity of OFA during imagery was significantly more negative than in typical imagers (Bayesian independent samples t-test, BF = 3.63).

Activity amplitude: opposite trends during perception and imagery

Across patches, we then separately examined the activity amplitudes during perception and imagery. Moving from posterior to anterior patches, including the OFC, the activity amplitudes formed two significant but opposite linear trends for perception and imagery (Fig. 2B). During perception, the activity amplitude decreased (linear contrast analysis, estimate for face patches = -1.93, t = -7.09, p < 0.001; estimate for color patches = -1.91, t = -7.94, p < 0.001). Conversely, during imagery the activity amplitude increased monotonically (estimate for face patches = 0.54, t = 4.68, p < 0.001; estimate for color

patches = 0.65, t = 2.75, p = 0.011). The same pattern prevailed in aphantasic individuals, showing decreasing trends in face perception (estimate = -2.79, t = -10.76, p < 0.001) and color perception (estimate = -1.80, t = -6.36, p < 0.001), and increasing trends in face imagery (estimate = 0.32, t = 2.37, p = 0.023) and color imagery (estimate = 0.58, t = 2.34, p = 0.027). The two groups showed comparable trends of activity amplitude (Bayesian repeated measures ANOVAs with the factor Group x Patch, all BFs < 0.89).

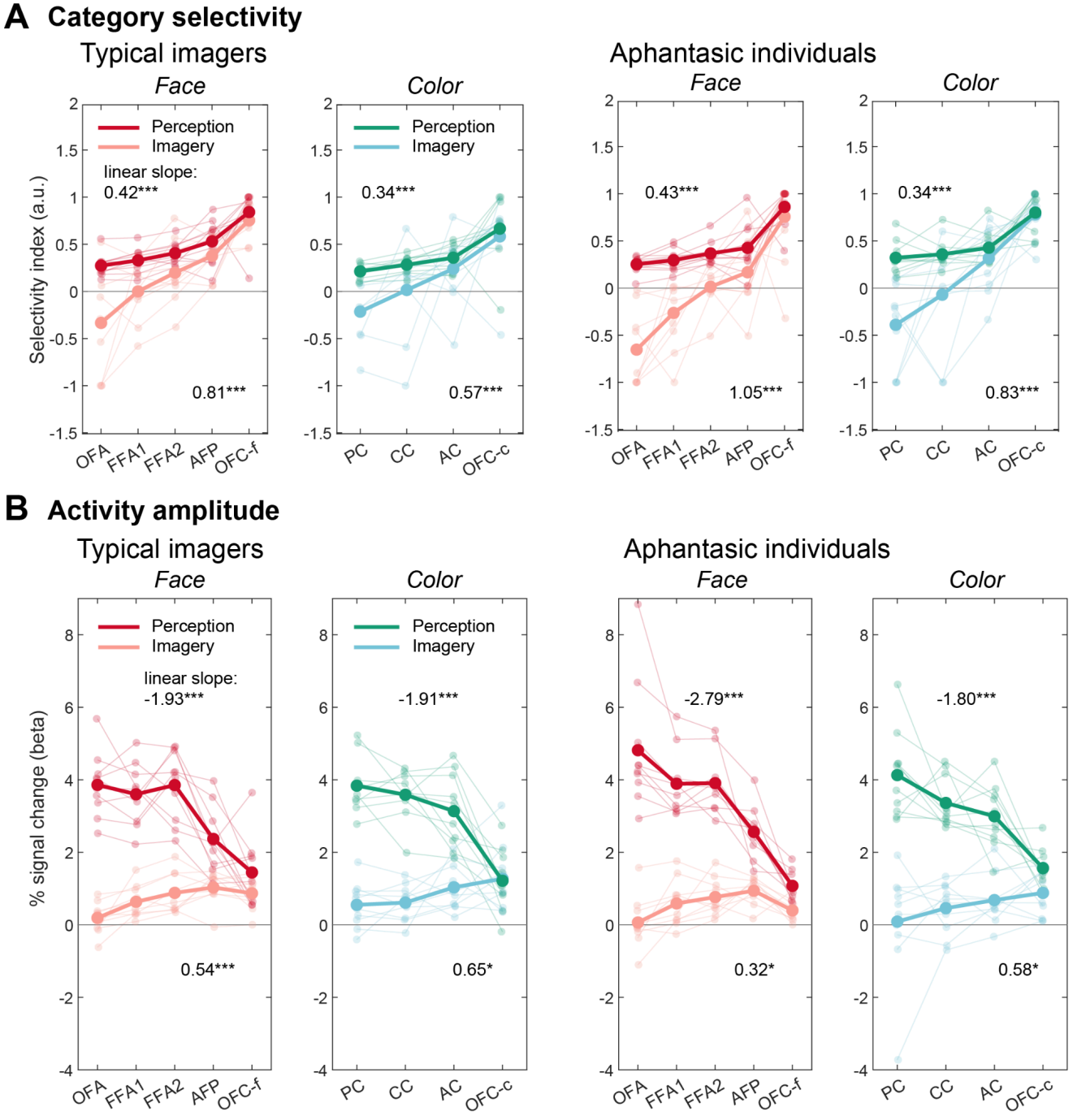


Figure 2. Category selectivity and activity amplitude for face- and color-specific patches during imagery and perception

In both A and B, bold lines denote the group average, and faint thin lines denote individual participants. Slopes of linear regressions (upper section: perception; lower section: imagery) are shown with significance levels: *p < 0.05, **p < 0.01, *p < 0.001. The same number of trials of imagery and perception of the same items in the run 4 and 5 were used for the plot.

A. Category selectivity increases across VOTC and OFC patches during both perception and imagery.

B. Across VOTC and OFC patches, the activity amplitude decreases during perception, but increases during imagery.

Multivariate representational content during perception and imagery

We next investigated whether the neural representations supported by the OFC and VOTC patches had an abstract or a more visual content. In a post-fMRI behavioral session, we assessed the disposition of stimuli in the participants' mental space, and correlated those patterns to the neural representations within cortical patches. Participants were instructed to arrange the same stimuli presented in the fMRI experiment inside a circular region, based either on their semantic or their visual features. For faces, participants organized the written names of the celebrities according to the core social marker of the profession, and arranged celebrities' photos according to the shape of their face (round or elongated). For colors, participants categorized written names of food items according to their semantic proximity (vegetables, tropical fruits and processed food), and arranged food photos according to their color. We derived behavioral representational dissimilarity matrices (RDMs) based on individual participants' arrangements (Fig. S4 displays the arrangements per condition, for the 18 items from the fMRI imagery trials). Neural RDMs were derived by calculating pairwise Pearson correlations between multi-voxel patterns of stimuli during perception or imagery. For each experimental condition, we predicted the neural RDM of each patch using the semantic and visual behavioral RDMs of the corresponding items as linear predictors.

The behavioral RDMs revealed no significant differences between typical and aphantasic participants (all ps > 0.09), indicating comparable representational content

across groups. Similarly, there were no group differences in the neural representations observed in any patch or condition (Bayesian independent samples t-test, all BF < 2.65).

Stimulus-specific content was represented in the OFC and anterior VOTC patches but not in the posterior patches (Fig. 3, Bayesian one-sample t-tests against zero in each patch). Most of these patches were sensitive to the social features of faces, or to the visual features of colors. Those patterns were relatively consistent across imagery and perception. Face shape information was only present in FFA2 for typical imagers during imagery (BF = 4.51). No evidence was found for food-related semantic representation in the color patches of either group (all BFs < 0.63).

Among OFC patches, the OFC-f coded social features (in typical imagers, during imagery, BF = 5.77; in aphantasics, during perception, BF = 3.38). The OFC-c patches coded visual color (in typical imagers, during imagery, BF = 3.26; in aphantasics, during perception, BF = 4.98).

Among VOTC patches, the AFP coded social identity in both face perception (typical imagers, BF = 20.06; aphantasics, BF = 11.23) and face imagery (typical imagers, BF = 5.78). The AC coded visual color in both color perception (typical imagers, BF = 7.56; aphantasics, BF = 12.33), and color imagery (typical imagers BF = 36.98, aphantasics BF = 4.01). The visual color representation was additionally found in the CC during imagery for both groups (typical imagers, BF = 10.14; aphantasics, BF = 7.98).

The OFC patches thus carry behaviorally relevant neural information during both imagery and perception. Are the neural representations elicited from memory during imagery similar to the feedforward representation triggered during perception? We assessed the overlap between imagery and perception representations by correlating their neural RDMs for the same items within each patch, and comparing them between groups (Bayesian one-sample t-tests against zero). In typical imagers, we observed representation overlap in some anterior VOTC patches (Fig. S5; FFA2 for faces, BF = 4.96; AC for object colors, BF = 9.21). In contrast, no overlap was observed in aphantasia in any patches. Compared to typical imagers (Bayesian independent samples t-test), aphantasic individuals showed a reduction in overlap in OFC-c (BF = 3.87), FFA2 (BF = 3.08), and AC (BF = 7.71) patches.

In summary, the OFC patches—like the anterior temporal patches—encode stimulus-specific content, with no difference between groups. However, only typical imagers exhibited similarity between the activation patterns elicited during perception and imagery.

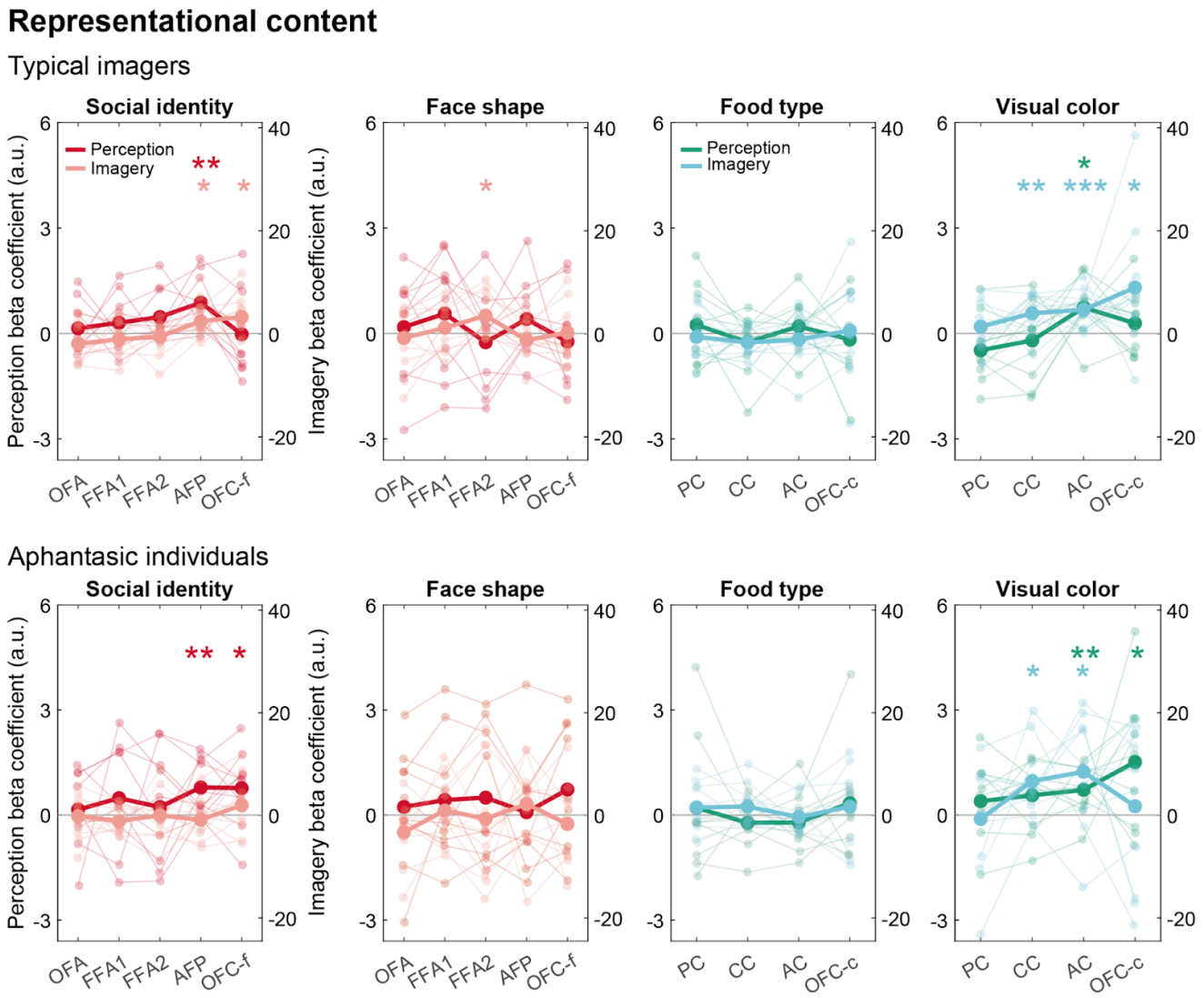


Figure 3. Representational content of patches during imagery and perception. In each patch, the neural RDM was modelled as a linear combination of two behavioral RDMs, resulting from the participants arrangement of fMRI stimuli based on their semantic or visual features. Plots represent the beta coefficients from those models. Face RDMs correspond to either the celebrities's social identity (profession), or their face shapes; color RDMs correspond to food types (semantic content of food stimuli), or the visual color appearance. Bold lines represent group averages, and faint lines represent individual participants. Red/pink lines denote face trials, and green/blue lines denote color trials. *BF > 3, **BF > 10, ***BF > 30, Bayesian one-sample t-test versus 0, indicating the presence of representational content. The scales of beta coefficients differ between perception and imagery, as the perceptual results were derived from 6-item RDMs, and the imagery results were derived from 18-item RDMs.

Functional connectivity: opposite patterns during perception and imagery

We examined task-modulated functional connectivity by psychophysiological interaction (PPI) analysis. This method allowed us to explore how the connectivity between the OFC patches and other patches within the same category-specificity systems were modulated during both perception and imagery tasks. We computed the pairwise connectivity between the OFC, VOTC patches, and additionally to V1, contrasting the perception and imagery tasks separately relative to the abstract words control task (comparison of beta values versus 0 with Bayesian one-sample t-tests for the pairwise connectivity between regions), and compared the resulting maps between perception and imagery. In typical imagers, the same pattern prevailed for faces and colors. The functional connectivity between the OFC-f and AFP patches was higher during face imagery compared to face perception (Bayesian pairwise t-test, BF = 5.87). There were distinct connectivity patterns during imagery and perception. During perception, we observed strong and wide-spread functional connectivity between the V1, VOTC, and OFC patches (compared to the control task, see detailed pairwise connectivity strengths in Fig. 4A). In contrast, during imagery, V1 and the post posterior VOTC patches were not significantly connected, while the OFC and the anterior VOTC patches remained strongly interconnected. In aphantasic individuals, we observed the across-area connectivity patterns similar to that of typical imagers during perception, but appeared to have fewer pairs of functionally connected patches during imagery (Fig. 4A).

Comparing the functional connectivity between groups revealed a systematically lower OFC connectivity in aphantasics than in typical imagers (Bayesian independent samples t-test, separately during perception and imagery). Thus aphantasic individuals had lower connectivity between the anterior patches OFC-f and AFP during face imagery (Fig 4B, BF = 3.27), and lower connectivity between posterior patches OFA and AFP during face perception (BF = 5.39). For color patches, aphantasics had lower connectivity between the OFC-c and AC patches during both color perception (BF = 3.00) and imagery (BF = 8.18).

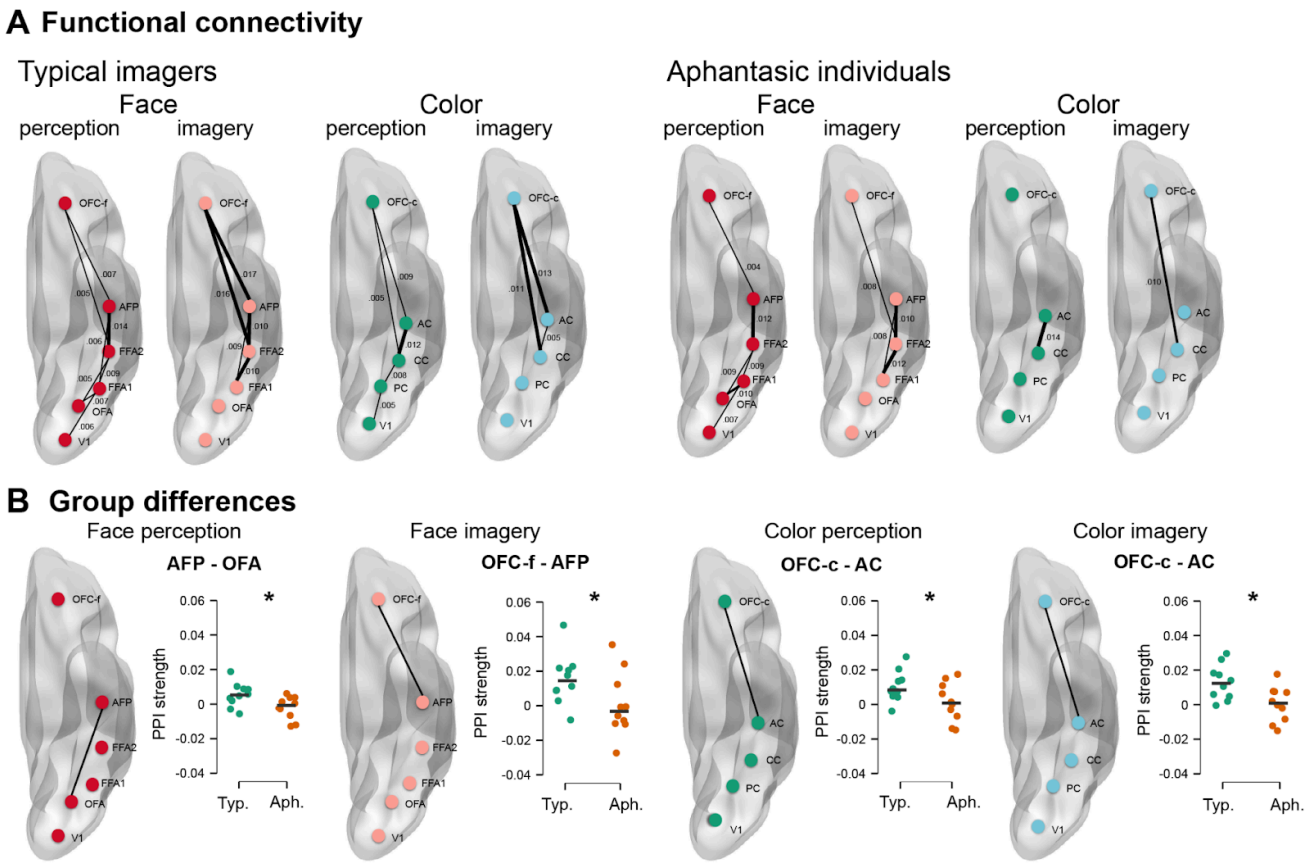


Figure 4. Task-dependent functional connectivity during imagery and perception

A. In typical imagers, the strength of PPI modulation, contrasting perception or imagery to an abstract words control condition. Lines indicate significant connectivity between two patches, as tested by Bayesian one-sample t-test against zero. Line thickness and values indicate group-level PPI strength.

B. Group differences in PPI strength. Lines indicate higher connectivity in typical imagers than in aphantasic individuals. Plots show connectivity strength in individual participants. Horizontal bars indicate the group average. * denote significant group differences (BF > 3). Typ, typical imagers; Aph, aphantasic individuals.

Dataset 2: Word-specific patches in the OFC

The color and face patches in the OFC followed a similar medial-lateral layout as in the VOTC. This suggests that a topographic map of multiple category-specific patches may be a shared feature of both the OFC and VOTC. Beside face and color patches, patches for other categories (e.g. words, bodies, places) might also be present in the OFC.

Here, we assessed the presence of OFC word patches using an independent 7T fMRI dataset[24], where 21 participants passively viewed 14 conditions of 6-letter strings with varied word-component (letters, bigrams, quadrigrams) frequencies in English or French. This dataset previously showed that word-specific patches independently defined by passive viewing detection tasks (localizer, Fig. 5, blue patches) exhibited a word-similarity effect in the main experiment, where letter strings more similar to real words elicited higher activity. This word-similarity effect (characterized as a significant linear-regression slope across 14 conditions) and the word selectivity index (computed in the same way as in Fig. 2A) both increased from posterior to anterior VOTC. These previous results indicated that the presence of the word-similarity effect may be one of the fundamental properties of word-selective patches. Therefore, here we directly used this word-similarity effect to detect the OFC word-specific patches (Fig. 5, yellow patches).

We performed Spearman correlation between the word-similarity predictor and the beta values per voxel across 14 conditions, and thresholded whole-brain Spearman Rho maps at $p<0.005$. Within the VOTC, patches with a significant word-similarity effect overlapped with the ones defined by the localizer around the fusiform area (Fig. 5, yellow versus blue patches). In some participants, it revealed more anterior patches extending into the anterior ventral temporal cortex. The word-similarity effect was not significant in early visual cortices in most participants (blue patches only), consistent with our previous observation[24].

Critically, OFC word-specific patches were observed in 13 out of 21 participants (Fig. 5). The word-specificity of these patches could be replicated in 4 participants in the independent localizer data, and a tendency for high word activity was observed in the activity amplitudes averaged across participants (black arrowheads and inset bar plots in Fig. 5, individual data in Fig. S6). Similar to the face and color patches, most of the word patches were located in the caudal portion of the LOS (9/13 participants) and MOS (5/13 participants). 4/13 participants showed word patches in the TOS, and 1/13 in the rostral portion of LOS. The number of OFC patches and voxels showed a trend toward a left-hemisphere bias (paired-sample t tests, 1 versus 0.54 patches, $p=0.053$; 37.62 versus 20.69 voxels, $p=0.077$). The x coordinates of the OFC word clusters did not differ from those of the OFC face patches in dataset 1 (group-average Talairach x coordinates 34.43 versus 34.27, Wilcoxon rank sum test for independent samples, $p=0.78$), although the VOTC word patches specifically around the fusiform gyrus were located significantly more lateral than the face patches (dataset 2 localizer, x coordinates 39.95 versus 37.53, paired-sample t

test, p=0.016). Due to the weak detection power in the localizer for OFC patches, we did not quantify the selectivity indices for OFC word patches.

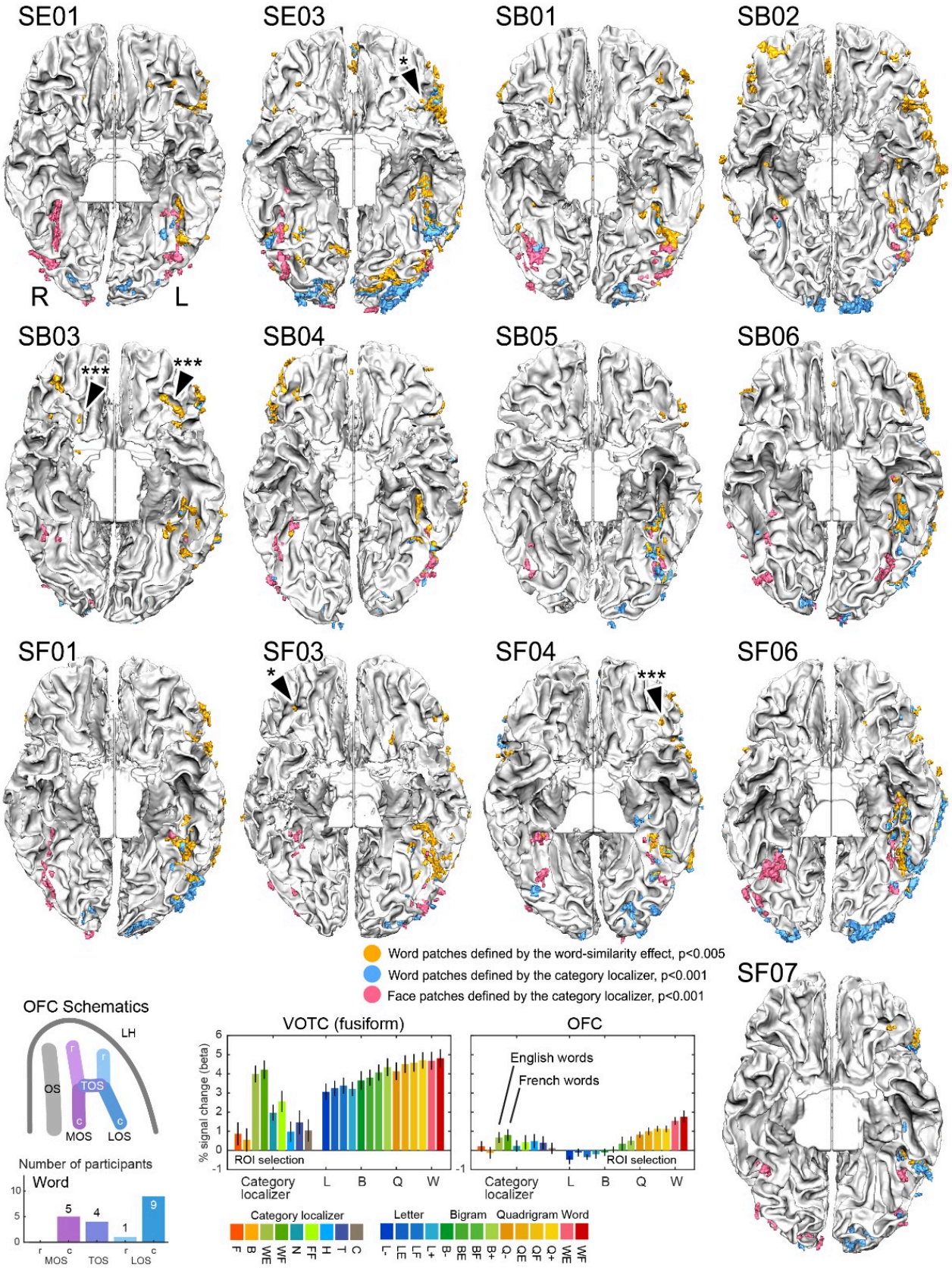


Figure 5. Individual word-specific patches in the OFC and VOTC.
Word-specific patches in 13 out of 21 participants (bilateral in 5 participants, left-hemisphere-only in 7 participants, right-hemisphere-only in 1 participant. In one participant SE06 with null findings, the lateral orbital sulcus of OFC was outside of data coverage). The word-specific patches defined by the word-similarity effect (yellow, 15 repetitions per condition) largely overlapped with the ones defined by an independent object category localizer (blue, 5 repetitions per condition) within the VOTC, but revealed OFC patches not shown by the localizer. On the other hand, a passive viewing task (localizer) showed word-specific patches in early visual areas potentially driven by low-level visual differences, where the word-similarity effect was weak or absent. To illustrate the medial-lateral spatial organization in the VOTC, face-specific patches (pink) defined by the localizer were also plotted. All maps were cluster-size thresholded at > 4 functional voxels. Black arrowheads indicate the OFC patches whose word specificity was replicated by ROI GLM in the localizer data (English and French words > faces, bodies, houses, tools). *p<0.05; ***p<0.001. SE, SB, SF: participants who are English-dominant, English-French early bilinguals, and French-dominant. The inset panels show the number of occurrences of OFC word patches across 13 participants, and for descriptive purposes show the activity profiles across the localizer (left side, 9 conditions) and the main experiment (right side, 14 conditions), in VOTC and in OFC ROIs. Error bars denote SEM across 13 participants. Note that the VOTC ROIs were defined by the localizer data, and the OFC ROIs were defined by the main experiment data, while the remaining conditions per panel were independent from ROI definition. See Figure S6 for data of individual participants. Abbreviations for conditions in the localizer: F: faces, B: bodies, WE: English words, WF: French words, N: numbers, FF: false fonts, H: houses, T: tools, C: checkerboards; in the main experiment: L: letters, B: bigrams, Q: quadrigrams, W: words, E or F: word component frequency high in only English or only French, - or +: word component frequency low or high in both English and French.

Discussion

Faces, colors, and written words are critically important classes of visual items which are processed by dedicated category-specific regions in the VOTC[2,4,5]. Using two millimeter-scale 7T-fMRI datasets, we showed that this specialization extends to the OFC, where we identified patches highly specialized for those three categories. These OFC patches followed a medial-to-lateral layout, with high inter-individual replicability. In dataset 1, during both perception and imagery, the face- and color-specific patches showed monotonically increasing category selectivity along the posterior to anterior axis, culminating in the OFC. However, perception and imagery showed opposite trends in activity amplitude and opposite patterns of functional connectivity. Representational similarity analysis provided evidence that the OFC patches encoded behaviorally-relevant contents, with face patches encoding social identities (professions) and color patches encoding visual color. OFC activity was different in aphantasic individuals, with weaker OFC-anterior temporal connectivity during imagery and reduced perception/imagery representational overlap. Our results suggest that OFC may contain a topographic organization similar to, but at a higher hierarchical level than, the VOTC object categorization system. To our knowledge, this is the first report of color- and word-specific patches in OFC and the first description on the functional properties of human OFC face-specific patches.

Face-, color- and word-specific patches in the OFC

Our findings reveal that face-, color- and word-specific patches are localized in the caudal parts of the OFC and follow a medial-lateral layout replicable across participants (see Fig. 1 and 5). These patches displayed high category selectivity (Fig. 2), and their representational content appears to be behaviorally relevant (Fig. 3), aligning with the limited macaque and human OFC studies on single categories.

Face-specific patches in OFC have been described in non-human primates[23,30] and occasionally found in humans[11,31], at locations comparable to ours. In monkeys, OFC face neurons were sensitive to views, emotional expressions, and socially relevant facial features such as age and sex[21–23]. Our findings are consistent with these proposed functional roles, suggesting that OFC face patches represent social identity—such as abstract categories of professions—which are critical for human social interaction.

Anatomical sites consistent with our OFC color patches have been associated with achromatic food pictures[30,32], suggesting a food-specific representation. However, our results argue against this conclusion. First, OFC color patches did represent the visual

colors of stimuli in our setting, with no evidence for food categories. Second, our VOTC color patches align with conventional color-biased areas[8] that are activated by a variety of non-food visual stimuli. Nevertheless, OFC color patches may well encode the behavioral relevance of chromatic signals, including food-related processing such as inferring taste or ripeness from color cues.

Our OFC word patches showed higher activity for word-like letter strings (word-similarity effect), but were not reliably activated during lower-level letter tasks such as ascender/descender detection or non-word detection during passive words viewing (see Methods). This suggests a high-level visual word representation in OFC, potentially reflecting the implicit encoding of relationships of word-like visual regularities across stimulus conditions.

Across these three categories, a common OFC process may involve a comparison of the value of behaviorally relevant information, in line with the OFC cognitive map theory[19]. Evidence in humans has shown that both the OFC and the hippocampus—a region typically associated with cognitive maps—represent abstract task structures[33]. Future studies could investigate whether OFC category-specific patches exhibit grid-cell-like modulation in the fMRI signal, a hallmark of "cognitive maps"[34].

A visual hierarchy formed by OFC and VOTC multi-category-specific patches

Our results suggest that visual processing extends beyond the visual cortex, with the OFC positioned as a higher-level stage in the visual hierarchy. The OFC may interface with the VOTC and potentially the lateral prefrontal cortex (LPFC) to form an extensive category-specific system. The consistent multi-category topography across these regions suggests a fundamental organizational principle in extended visual processing, evidenced by 1) OFC-VOTC similarity, 2) newly identified LPFC multi-category patches[15,35], and 3) recent evidence from monkey face patches in anterior inferotemporal cortex (aIT), OFC and LPFC.

The OFC exhibits anatomical and functional specializations that closely mirror VOTC regions, indicating their operation within an integrated system. This is supported by: 1) direct OFC-aIT anatomical connections[36,37] including face-specific projections[38] in non-human primates, and effective OFC-to-VOTC connectivity in humans[39]; 2) A conserved medial-lateral topography of color and face patches in the OFC reflects the VOTC architecture, potentially driven by cytoarchitectonic constraints[40,41] and retinotopic eccentricity biases[1,7].

Beyond these shared features, our findings position OFC at a higher hierarchical level than VOTC. First, during both perception and imagery, face- and color-specific

patches exhibited increasing category selectivity along the posterior-anterior axis, culminating in the OFC, reflecting progressive categorical abstraction[1]. Second, we found opposite connectivity patterns between OFC and VOTC (particularly anterior patches) during perception and imagery, paralleling the opposite patterns of connectivity between LPFC and the visual cortex[42,43].

OFC and VOTC contain repeated multi-category topographic maps in similar medial-lateral layouts, suggesting a fundamental organizational motif in this system. Similar to the tightly clustered and repeated multi-category topographies in the monkey visual cortex[10], our observed extensions into OFC is likely an additional hierarchical stage beyond VOTC. This organizational motif may further continue to LPFC. Previous work has identified single-category specific areas in ventral LPFC, for faces in both monkeys[23,44] and in humans[45,46], and for scenes in humans[47]. A recent study further showed evidence on face/place/body/tool patches distributed surrounding multiple-demand regions in LPFC[35]. In our mental imagery dataset[15], the inferior frontal gyrus also contained category-specific areas functionally connected with relevant fusiform regions during imagery of object shapes, colors, words and faces.

A recent macaque study[48] provided strong evidence that the aIT, OFC, and LPFC are hierarchical stages within the same face processing system: OFC face patches (with similar anatomical locations and high face selectivity as in our observations) had response latencies (~70 ms) similar to the aIT face patches, while the ventral LPFC face patches responded much faster (~40 ms). These results and our findings provide support for the OFC top-down predictor theory[17], and extend this hypothesis to category-specific processing. Our findings further raise fundamental questions about: 1) the functional relationships between each set of repeated topographic maps, and 2) the principles driving category specificity with or without exogenous visual information.

To conclude, our findings demonstrate that OFC patches specific to faces, colors, and words: 1) exist across individual participants with a topographical organization similar to the VOTC, 2) encode behaviorally relevant perceptual content, and 3) operate at a higher hierarchical level than the VOTC, exerting top-down modulation over it. These OFC patches may serve as category-specific interfaces that bridge perception with context-dependent cognition, where category selectivity meets behavioral relevance.

Contributions

Conceptualization: J.L., M.Z., P.B., L.C.; Data curation: J.L.; Formal analysis: J.L., M.Z.; Investigation: J.L.; Methodology: J.L., M.Z.; Funding acquisition: J.L., P.B.; Project administration: P.B., J.L.; Software: J.L.; Supervision: M.Z., P.B.; Visualization: J.L., M.Z.; Writing - original draft: J.L.; Writing - review & editing: J.L., M.Z., L.C., P.B..

Materials & Correspondence

Requests for materials and correspondence should be addressed to J.L.

Acknowledgements

The research was supported by INSERM, CEA and specific funding from Dassault Systèmes. The work of J.L. is supported by specific funding from Dassault Systèmes. The work of P.B. is supported by the Agence Nationale de la Recherche through ANR-16-CE37-0005 and ANR-10-IAIHU-06, and by the Fondation pour la Recherche sur les AVC through FR-AVC-017. The work of M.Z. is supported by Programme Investissements d'Avenir IHU FOReSIGHT (ANR-18-IAHU-01). We thank Yangwen Xu for the suggestions on building the multi-arrangement tasks.

Competing interest

The authors declare that they have no competing interests.

References

1. Grill-Spector, K. & Weiner, K. S. The functional architecture of the ventral temporal cortex and its role in categorization. *Nat. Rev. Neurosci.* **15**, 536–548 (2014).
2. Kanwisher, N., McDermott, J. & Chun, M. M. The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. *J. Neurosci.* **17**, 4302–4311 (1997).
3. Peelen, M. V. & Downing, P. E. Selectivity for the Human Body in the Fusiform Gyrus. *J. Neurophysiol.* **93**, 603–608 (2005).
4. Cohen, L. *et al.* The visual word form area: Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain* **123**, 291–307 (2000).
5. McKeefry, D. J. & Zeki, S. The position and topography of the human colour centre as revealed by functional magnetic resonance imaging. *Brain* **120**, 2229–2242 (1997).
6. Epstein, R., Harris, A., Stanley, D. & Kanwisher, N. The Parahippocampal Place Area: Recognition, Navigation, or Encoding? *Neuron* **23**, 115–125 (1999).
7. Hasson, U., Levy, I., Behrmann, M., Hendler, T. & Malach, R. Eccentricity Bias as an Organizing Principle for Human High-Order Object Areas. *Neuron* **34**, 479–490 (2002).
8. Lafer-Sousa, R., Conway, B. R. & Kanwisher, N. G. Color-Biased Regions of the Ventral Visual Pathway Lie between Face- and Place-Selective Regions in Humans, as in Macaques. *J. Neurosci.* **36**, 1682–1697 (2016).
9. Lafer-Sousa, R. & Conway, B. R. Parallel, multi-stage processing of colors, faces and shapes in macaque inferior temporal cortex. *Nat. Neurosci.* **16**, 1870–1878 (2013).
10. Bao, P., She, L., McGill, M. & Tsao, D. Y. A map of object space in primate inferotemporal cortex. *Nature* **583**, 103–108 (2020).
11. Rajimehr, R., Young, J. C. & Tootell, R. B. H. An anterior temporal face patch in human cortex, predicted by macaque maps. *Proc. Natl. Acad. Sci.* **106**, 1995–2000 (2009).
12. Hasson, U., Harel, M., Levy, I. & Malach, R. Large-Scale Mirror-Symmetry Organization of Human Occipito-Temporal Object Areas. *Neuron* **37**, 1027–1041 (2003).
13. Kanwisher, N. The Quest for the FFA and Where It Led. *J. Neurosci.* **37**, 1056–1061 (2017).
14. Duchaine, B. & Yovel, G. A Revised Neural Framework for Face Processing. *Annu. Rev. Vis. Sci.* **1**, 393–416 (2015).
15. Liu, J. *et al.* Visual mental imagery in typical imagers and in aphantasia: A millimeter-scale 7-T fMRI study. *Cortex* (2025) doi:10.1016/j.cortex.2025.01.013.
16. Zeman, A. Aphantasia and hyperphantasia: exploring imagery vividness extremes. *Trends Cogn. Sci.* **28**, 467–480 (2024).
17. Bar, M. *et al.* Top-down facilitation of visual recognition. *Proc. Natl. Acad. Sci.* **103**, 449–454 (2006).
18. Chaumon, M., Kveraga, K., Barrett, L. F. & Bar, M. Visual Predictions in the Orbitofrontal Cortex Rely on Associative Content. *Cereb. Cortex* **24**, 2899–2907 (2014).
19. Wilson, R. C., Takahashi, Y. K., Schoenbaum, G. & Niv, Y. Orbitofrontal Cortex as a Cognitive Map of Task Space. *Neuron* **81**, 267–279 (2014).
20. Schoenbaum, G., Setlow, B. & Ramus, S. J. A systems approach to orbitofrontal cortex

function: recordings in rat orbitofrontal cortex reveal interactions with different learning systems. *Behav. Brain Res.* **146**, 19–29 (2003).
21. Barat, E., Wirth, S. & Duhamel, J.-R. Face cells in orbitofrontal cortex represent social categories. *Proc. Natl. Acad. Sci.* **115**, E11158–E11167 (2018).
22. Rolls, E. T., Critchley, H. D., Browning, A. S. & Inoue, K. Face-selective and auditory neurons in the primate orbitofrontal cortex. *Exp. Brain Res.* **170**, 74–87 (2006).
23. Tsao, D. Y., Schweers, N., Moeller, S. & Freiwald, W. A. Patches of face-selective cortex in the macaque frontal lobe. *Nat. Neurosci.* **11**, 877–879 (2008).
24. Zhan, M., Pallier, C., Agrawal, A., Dehaene, S. & Cohen, L. Does the visual word form area split in bilingual readers? A millimeter-scale 7-T fMRI study. *Sci. Adv.* **9**, eadf6140 (2023).
25. Liu, J. & Bartolomeo, P. Probing the unimaginable: The impact of aphantasia on distinct domains of visual mental imagery and visual perception. *Cortex* **166**, 338–347 (2023).
26. Tsao, D. Y., Moeller, S. & Freiwald, W. A. Comparing face patch systems in macaques and humans. *Proc. Natl. Acad. Sci.* **105**, 19514–19519 (2008).
27. Chiavaras, M. M. & Petrides, M. Orbitofrontal sulci of the human and macaque monkey brain. *J. Comp. Neurol.* **422**, 35–54 (2000).
28. Tsao, D. Y., Freiwald, W. A., Tootell, R. B. H. & Livingstone, M. S. A Cortical Region Consisting Entirely of Face-Selective Cells. *Science* **311**, 670–674 (2006).
29. Simmons, W. K., Bellgowan, P. S. F. & Martin, A. Measuring selectivity in fMRI data. *Nat. Neurosci.* **10**, 4–5 (2007).
30. Thorpe, S. J., Rolls, E. T. & Maddison, S. The orbitofrontal cortex: Neuronal activity in the behaving monkey. *Exp. Brain Res.* **49**, 93–115 (1983).
31. Ratan Murty, N. A. *et al.* Visual experience is not necessary for the development of face-selectivity in the lateral fusiform gyrus. *Proc. Natl. Acad. Sci.* **117**, 23011–23020 (2020).
32. Simmons, W. K., Martin, A. & Barsalou, L. W. Pictures of Appetizing Foods Activate Gustatory Cortices for Taste and Reward. *Cereb. Cortex* **15**, 1602–1608 (2005).
33. Mızrak, E., Bouffard, N. R., Libby, L. A., Boorman, E. D. & Ranganath, C. The hippocampus and orbitofrontal cortex jointly represent task structure during memory-guided decision making. *Cell Rep.* **37**, 110065 (2021).
34. Behrens, T. E. J. *et al.* What Is a Cognitive Map? Organizing Knowledge for Flexible Behavior. *Neuron* **100**, 490–509 (2018).
35. Assem, M., Shashidhara, S., Glasser, M. & Duncan, J. Category-biased patches encircle core domain-general regions in the human lateral prefrontal cortex. 2025.01.16.633461 Preprint at https://doi.org/10.1101/2025.01.16.633461 (2025).
36. Carmichael, S. T. & Price, J. L. Sensory and premotor connections of the orbital and medial prefrontal cortex of macaque monkeys. *J. Comp. Neurol.* **363**, 642–664 (1995).
37. Saleem, K. S., Kondo, H. & Price, J. L. Complementary circuits connecting the orbital and medial prefrontal networks with the temporal, insular, and opercular cortex in the macaque monkey. *J. Comp. Neurol.* **506**, 659–693 (2008).
38. Grimaldi, P., Saleem, K. S. & Tsao, D. Anatomical Connections of the Functionally Defined “Face Patches” in the Macaque Monkey. *Neuron* **90**, 1325–1342 (2016).
39. Rolls, E. T., Deco, G., Huang, C. C. & Feng, J. The human orbitofrontal cortex, vmPFC,

and anterior cingulate cortex effective connectome: emotion, memory, and action. *Cereb Cortex* (2022) doi:10.1093/cercor/bhac070.
40. Lorenz, S. *et al.* Two New Cytoarchitectonic Areas on the Human Mid-Fusiform Gyrus. *Cereb Cortex* **27**, 373–385 (2017).
41. Wojtasik, M. *et al.* Cytoarchitectonic Characterization and Functional Decoding of Four New Areas in the Human Lateral Orbitofrontal Cortex. *Front. Neuroanat.* **14**, (2020).
42. Dijkstra, N., Zeidman, P., Ondobaka, S., van Gerven, M. a. J. & Friston, K. Distinct Top-down and Bottom-up Brain Connectivity During Visual Perception and Imagery. *Sci. Rep.* **7**, 5677 (2017).
43. Mechelli, A., Price, C. J., Friston, K. J. & Ishai, A. Where bottom-up meets top-down: neuronal interactions during perception and imagery. *Cereb Cortex* **14**, 1256–65 (2004).
44. Ó Scalaidhe, S. P., Wilson, F. A. W. & Goldman-Rakic, P. S. Areal Segregation of Face-Processing Neurons in Prefrontal Cortex. *Science* **278**, 1135–1138 (1997).
45. Chan, A. W.-Y. & Downing, P. E. Faces and Eyes in Human Lateral Prefrontal Cortex. *Front. Hum. Neurosci.* **5**, (2011).
46. Nikel, L., Sliwinska, M. W., Kucuk, E., Ungerleider, L. G. & Pitcher, D. Measuring the response to visually presented faces in the human lateral prefrontal cortex. *Cereb. Cortex Commun.* **3**, tgac036 (2022).
47. Rajimehr, R. *et al.* Functional architecture of cerebral cortex during naturalistic movie watching. *Neuron* **112**, 4130-4146.e3 (2024).
48. Mergan, E., Zhu, Q., Li, X., Vogels, R. & Vanduffel, W. Fast face-selective responses in prefrontal face patches of the macaque. *Cell Rep.* **44**, 115389 (2025).
49. Collins, J. A. & Olson, I. R. Beyond the FFA: The role of the ventral anterior temporal lobes in face processing. *Neuropsychologia* **61**, 65–79 (2014).
50. Kriegeskorte, N. & Mur, M. Inverse MDS: Inferring Dissimilarity Structure from Multiple Item Arrangements. *Front. Psychol.* **3**, (2012).
51. Busing, F., Commandeur, J. & Heiser, W. PROXSCAL: A multidimensional scaling program for individual differences scaling with constraints. *SOFTSTAT 1997 Adv. Stat. Softw.* **6**, (1997).

Methods

7T fMRI data acquisition

We re-analyzed two previously collected 7-Tesla fMRI datasets[15,24]. Both datasets were acquired with the same MRI sequences and brain coverage on the same machine at NeuroSpin. High-resolution anatomical images were acquired with an MP2RAGE sequence (resolution=0.65 mm isotropic, TR=5000 ms, TE=2.51 ms, TI1/TI2=900/2750 ms, flip angles=5/3, iPAT=2, bandwidth=250 Hz/Px, echo spacing=7 ms). Functional data were acquired with a 2D gradient-echo EPI sequence (TR = 2000 ms, TE = 28 ms, voxel size = 1.2 mm isotropic, multiband acceleration factor=2; encoding direction: anterior to posterior, iPAT=3, flip angle = 75, partial Fourier=6/8, bandwidth=1488 Hz/Px, echo spacing=0.78 ms, number of slices=70, no gap, reference scan mode: GRE, MB LeakBlock kernel: off, fat suppression enabled). Distortion correction runs with the opposite phase-encoding direction were collected for each task run. The acquisition slab covered the majority of the cortex, avoided the eyeballs, and minimized the ear-canal signal dropout, allowing us to observe very anterior temporal patches around the ear-canal region (see Figure 1 and Figure 5 for example patches). Although the anterior temporal cortex (including the temporal pole) and the superior part of the primary motor cortex (the foot regions) were out of the coverage.

Dataset 1: mental imagery and perception

In this dataset[15], 20 healthy participants (10 typical imagers, 29.28 ± 8.47 years, 6 female) and 10 aphantasic individuals (28.69 ± 8.27, 6 female) performed the enhanced Batterie Imagerie-Perception - eBIP[25]. The battery assessed: i) imagery of object shapes, object colors, faces, letters, and spatial relationships on an imaginary map of France; ii) a non-imagery control task using abstract words, and iii) an audio-visual perception task using the same items as in the imagery tasks. In the famous-faces imagery task, participants had to decide which celebrity had a more round or oval face. Participants heard a word indicating a particular imagery category (e.g., “face”), followed by 2 names, designating the items the participant is required to imagine (e.g. “E.Macron”, “F.Hollande”, stimulus onset asynchrony = 4s). They were instructed to generate and maintain mental images as vivid as possible for 4 s per item. Eight seconds after the onset of the second item, they heard an attribute word (e.g. “round”). They then pressed one of two buttons accordingly, indicating

which of the items is most closely associated with the attribute (e.g. who is associated with the attribute “round” face. In the color imagery task, participants had to decide which fruit or vegetable had a darker or lighter color. In the non-imagery abstract words control task, participants had to decide which of two abstract words (e.g., “routine”, convention”) was semantically closer to a third word (e.g. “society”). In the perception task, the same items used for the imagery tasks were presented with both the auditory word and a visual image. In the abstract-word and the perception tasks, participants rated their confidence on a 4-level Likert scale, instead of rating vividness.

Each participant was scanned for 120 minutes at maximum and finished 5 runs. The first 3 runs consisted of 90 imagery-only trials (30 trials per run, each including 6 trials per category). The last 2 runs consisted of 36 trials each, with 3 trial types presented in a specific order (6 abstract word control trials, 15 imagery trials and 15 auditory-visual perception trials per run, each including 3 trials per category). Notably, the last two runs contained identical items in the imagery trials and perceptual trials, and participants only saw the perceptual items after the imagery trials. Trials within each trial type were randomized across categories.

Dataset 2: passive reading

In this dataset[24] using block design, 21 participants (mean age=23.4, 10 female) with varied language dominance in English and French passively viewed 6-letter strings presented at a fast pace (stimulus presented for 150 ms, followed by a fixation of 200 ms), while detecting catch trials with hashtags (“######”) by pressing a button. The letter strings consisted of 14 conditions (4 conditions each for letters, bigrams, quadrigrams, 2 conditions for real words in English and French), where the word-component frequency was separately modulated in English and French. Each condition was presented in short blocks, repeated 15 times. This main experiment lasted for 39.9 minutes, split into 3 runs. An independent object category localizer run was acquired, with conditions including faces, bodies, English words, French words, numbers, false fonts, houses, tools, checkerboards, each repeated 5 times. The localizer experiment lasted 9.2 minutes. Within each experiment, all blocks were randomized. The whole scanning session lasted around 90 minutes. See detailed description of the experimental design in Ref.[24].

We processed the anatomical and functional MRI data with BrainVoyager (Version 22.0.2.4572 and Version Version 22.4.4.5188, Brain Innovation, Maastricht, The Netherlands, https://www.brainvoyager.com/), as previously reported[15,24]. All analyses here

are conducted with BrainVoyager, MATLAB (version R2018b), NeuroElf 1.1 toolbox (http://neuroelf.net/) implemented in MATLAB, and JASP 0.16.2 (https://jasp-stats.org/). GLMs were computed with predictors including each single condition, the response period, the button presses, and the 6 parameters for head motion (z-normalized). Functional voxel size is 1.2mm isotropic. No spatial smoothing was applied to any data in this study.

Analyses for dataset 1

We performed various analyses by integrating data from multiple runs:

Patch Identification: We used perceptual trials from runs 4 and 5 (6 trials per category) to identify face and color patches, as regions of interests (ROIs) in subsequent analyses.

Activity Amplitude, Category Selectivity, and Functional Connectivity: We used runs 4 and 5, where participants had identical items and number of trials for imagery and perception, with control trials using abstract words for functional connectivity. Note that although the category selectivity and activity amplitude analyses used partially the same data (perceptual runs 4 and 5) as the ones defining ROIs, these two analyses mainly focused on the differences across areas and groups, thus is independent from ROI definition.

Neural representational content: We examined multi-voxel patterns for individual cortical patches. For the perceptual similarity matrix, we extracted beta values from perceptual trials in runs 4-5 (6 items per category, 6x6 matrices). For imagery similarity analysis, we extracted beta values from imagery trials in runs 1-3 (18 imagery trials per category, 18x18 matrices) to enhance the accuracy of estimation. To assess the representational overlap between imagery and perception for the exact same items, beta values were extracted from imagery and perceptual trials in runs 4-5.
See the corresponding sections below for detailed methods.

Identification of face- and color-specific patches by univariate analysis

We performed general linear models (GLMs) on fMRI data of run 4 and 5, and used the contrast of faces or colors versus remaining categories in perceptual trials to identify face- or color-specific patches as ROIs (voxelwise $p<0.001$ uncorrected, cluster size thresholded at 12 voxels). The VOTC face- and color-specific patches were labeled according to conventions in the literature[8,26,49]. From posterior to anterior, face patches included: occipital face area (OFA), fusiform face area 1 (FFA1), fusiform face area 2 (FFA2),

anterior face patch (AFP); color patches included: posterior color (PC), central color (CC), anterior color (AC) patches. Outside the VOTC, in bilateral OFC, we identified for faces (OFC-f) and for colors (OFC-c). Notably, with just 6 perceptual trials per category, the OFC face and color patches could be identified in 18/20 and 19/20 individuals, respectively. For three participants without OFC patches in perceptual trials (Aph 01 and Aph 05 lacked face patches, Typ 04 lacked color patches), the OFC ROIs were defined by 18 imagery trials from run 1-3 with the same category-specific contrasts. To test the medial-lateral differences between the face and color patches in both OFC and VOTC, and in both typical and aphantasic groups, we averaged the Talairach coordinates across two hemispheres within each individual (using the absolute values for the x coordinates). For subsequent analyses, we merged bilateral patches into one ROI per area. To determine the activity profiles per patch, we extracted the beta values (% signal changes) per condition from run 4 and 5, where participants had identical items and number of trials for both imagery and perception.

Bayesian statistical tests

Statistical tests were performed using JASP 0.16.2 (https://jasp-stats.org/), and used the JASP default priors. A commonly accepted convention is that Bayes factors (BF10 or BFs) between 3 and 10 indicate moderate evidence in favor of the model in the numerator (H1); BFs between 10 and 30 indicate strong evidence; BFs larger than 30 indicate very strong evidence. The inverse of these cut-offs values provides moderate (<0.33), strong (<0.10), or very strong evidence (<0.03) for the model in the denominator (H0), i.e. the null hypothesis.

Temporal profile analysis

We extracted the event-related average time courses across voxels in each ROI. The time courses were normalized by the averaged activity across TRs -2 to 0. We performed a three-way ANOVA to identify group differences across patches and time points (Group x Patch x TR [0 to 11], trials end at TR 12), separately for imagery and perception conditions. We performed three-way ANOVAs with the factor of Group x Patch x TR in each condition. For both face perception and face imagery, there were three-way interactions and the group effect interactions with TR. We then unpacked the group effect for each time point in each of the ROIs and tested the group difference with time-resolved independent t-tests. Results were Bonferroni-corrected across time points with corrected alpha = 0.05, for each ROI. There were group differences for both color perception and color imagery, and these effects didn't interact with other factors. Detailed statistics for ANOVAs are reported in Fig. S2.

Face- and color-selectivity index

For the face- and color-specific patches, we used the activity amplitudes (single-condition beta values) to calculate the selectivity index as follows:

Selectivity index (SI) for the preferred category = (4 * preferred category – other categories) / (4 * preferred category + other categories)

To avoid deactivation in the fMRI signal leading to an overestimation of actual voxel selectivity, we adjusted the activity amplitudes per voxel per category, by adding the absolute value of the most negative activity from all categories of that voxel[29]. This adjustment resulted in SI values ranging from -1 to 1. A positive SI indicates category selectivity; a SI of 0 indicates similar amplitude across all conditions; a negative SI indicates that on average the region is activated more strongly by the other conditions.

Behavioral multiple-arrangement tasks

After the fMRI scanning session, participants performed multiple-arrangement tasks[50] on Meadow Research platform (https://meadows-research.com/). We sought to determine their subjective judgments on semantic or visual similarity of faces and colors. Participants were asked to drag and drop the items within the circle with a mouse, a pair at a time, and place items close to each other according to the considered similarity. For each category, participants arranged the same items in separate tasks, either by semantic features (cued by words), or perceived features (cued by photos). Specifically, for faces, semantic feature representational dissimilarity matrices (RDMs) derived from written names of the celebrities arranged according to their social identities, i.e. professions; the visual feature RDMs derived from photos arranged according to the facial shapes (round or elongated). For object colors, the semantic RDMs were derived from food items arranged by type (vegetables, tropical fruits, and prepared food); the visual feature RDMs derived from food items arranged from their visual color appearances. To map the neural RDMs separately for imagery and perception, participants arranged 6 items from perceptual trials regarding semantic and visual features (6x6), then arranged 18 items from imagery trials regarding semantic and visual features (18x18), resulting in 4 behavioral RDMs for each category. RDMs were calculated using Euclidean distance from Meadow Research. The resulting behavioral arrangements were visualized with multidimensional scaling[51] from the RDMs from each group. Figure S4 displays exemplar arrangements of 18 items from imagery

trials. In the same session, we also collected similarity judgments on object shapes, words and French maps with the items from the fMRI tasks of the Dataset 1, and for each domain, participants also arranged items by relevant visual features using word cues. Those data were not used in the current analysis.

Statistical tests on behavioral RDMs

We first tested the level of inter-individual variability. We calculated leave-one-out squared difference between each individual and within group means. The test statistic was the sum of the upper triangular elements of the difference matrix with a dimension of 153 (for 18 items), resulting in one value for each participant and for each RDM. We then performed independent t-tests to compare the two groups. No group difference was observed in any of the RDMs (all ps>0.12). We then performed nonparametric permutation tests to evaluate group differences on each behavioral RDMs. For each permutation, participant labels were randomly shuffled while maintaining the original group sizes. The squared sum of mean differences was calculated for each permuted dataset. This procedure was repeated for 1000 permutations. The observed test statistic from the actual data was compared to the distribution of the permuted statistics. A p-value was calculated as the proportion of permuted statistics greater than or equal to the observed statistic, with a cut-off p-value of 0.05 indicating significance.

Neural RDMs

For each ROI, we first obtained voxel-wise average % signal change from the 4-5th repetition times (TRs, normalized by the TRs -2 to 0 as the baseline), which corresponded to the peak activity of the first item, with minimal influence from the second item in each trial. We then generated the neural RDM across stimulus pairs with Pearson correlation. Furthermore, to directly assess the similarity of neural representations between imagery and perception (representational overlap) within each patch, we correlated RDMs (6x6) generated from the run 4-5 where participants imagined or perceived the same 6 items. We then examined the amount of overlap by Bayesian one-sample t-tests against zero for each group and Bayesian independent t-test between groups on each patch.

Linear regression with RDMs

In each patch, separately for faces and colors, to examine the presence of semantic or visual-feature representational content during perception and imagery, we fitted a linear regression model, using two behavioral RDMs (semantic, visual) as regressors, and the neural RDMs as dependent variables. The behavioral RDMs have been squared before model fittings.

$$\textit{neural RDM} = \text{beta}_1 * \textit{semantic RDM} + \text{beta}_2 * \textit{visual RDM} + \textit{intercept}$$

We tested the beta weights per patch using Bayesian one-sample t-tests against zero to determine the presence of representational contents. No comparison was made between the semantic and visual beta values.

Psychophysiological interactions (PPI) analysis

To investigate task-specific functional connectivity for each individual ROI, we conducted PPI analysis with trials in run 4-5. We built each PPI design matrix by 1) generating a task contrast predictor between two conditions (Condition A - B), balanced by a regressor of the sum of those two conditions (Condition A + B), both convolved with the hemodynamic response function (HRF); 2) extracting a demeaned time course from the seed ROI; 3) generating an interaction predictor as an element-by-element product of the HRF-convolved task contrast predictor and the seed time-course predictor; 4) adding the 6 parameters of participant's z-scored head motion as confounding predictors. Specifically for 1), to investigate category-specific connectivity, we built the task contrast predictor by subtracting the Abstract word control condition from the single-category condition under investigation, e.g. face imagery, while keeping the other imagery / perceptual category conditions unchanged. We separated PPI design matrices for each seed ROI and each contrast. The seed ROIs included OFA, FFA1, FFA2, AFP, OFC-f for face imagery and perception; and PC, CC, AC, OFC-c for color perception and imagery. We then computed GLMs on individual unsmoothed fMRI datasets in BrainVoyager, resulting in whole-brain beta maps of each task contrast. For each seed ROI, its connectivity to every other ROI was averaged across voxels for the resulting beta values per ROI. To examine the bottom-up influence from early visual cortices, we also included V1 as a seed, which was manually delimited anatomically in each individual brain around the retro calcarine sulcus.

Analyses for dataset 2

Definition of OFC word-specific patches

We used the predictor for the word-similarity effect (across 14 conditions, [1 2 2 3 4 5 5 6 7 8 8 9 10 10], see Ref.[24]), and performed Spearman correlation across 14 conditions on the single-condition beta values per voxel. This predictor assumes increasing word-component frequency across letters, bigrams, quadrigrams, and real words, but does not assume a frequency difference of the same word component between English and French. The resulting Spearman Rho maps were thresholded at $p<0.005$ uncorrected (functional cluster size > 4) to mitigate false negatives. Since OFC word-specific patches were found in 13/21 participants, all subsequent descriptive statistics and statistical analyses were performed on those 13 participants. To validate the word-specificity in the independent object category localizer, ROI GLMs of the averaged time course per patch was computed in BrainVoyager (Figure 5 black arrowheads indicating significant patches), and the activity amplitudes (beta values) averaged across all OFC patches per participant were extracted (Figure 5 inset and Figure S6).

Definition of VOTC word- and face-specific patches

As the same way in Ref.[24], word-specific patches in the VOTC were defined by the object category localizer (contrast: English words and French words > faces, bodies, houses, tools, $p<0.001$ uncorrected, cluster size > 4 functional voxels). Face-specific patches were also defined (contrast: faces > bodies, houses, tools, average of English and French words), as a reference for their medial-lateral layouts relative to the word patches in the VOTC. The anatomical locations of all patches were manually labeled (ROI data are publicly available on https://osf.io/96syx/?view_only=88b55034027042fbb4f118f398fc5706). All patches around the fusiform gyrus were merged into a single ROI per hemisphere.

Comparison of the medial-lateral layouts between word and face patches

For both OFC and VOTC patches, within each participant, the absolute value of the average Talairach x coordinates across voxels were computed within each hemisphere, and were then averaged across hemispheres. The resulting x coordinates of the VOTC word and face patches were compared within the 13 participants with a paired-sample t-test. Since dataset 2 did not reveal any OFC face patches, the x coordinates of the OFC word patches

were compared to the x coordinates of the OFC face patches in dataset 1, with a Wilcoxon rank sum test for independent samples (13 versus 20 participants).
(END)