Multidimensional feature tuning in category-selective areas of human visual cortex

Leonard E. van Dyck[1,2*], Martin N. Hebart[2,3,4+], Katharina Dobs[4,5+]

[1]Department of Psychology, Justus Liebig University Giessen, Germany

[2]Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

[3]Department of Medicine, Justus Liebig University Giessen, Germany

[4]Center for Mind, Brain and Behavior (CMBB), Universities of Marburg, Giessen, and Darmstadt, Germany

[5]Department of Mathematics and Computer Science, Physics, Geography,
Justus Liebig University Giessen, Germany

*corresponding author: leonard.van-dyck@uni-giessen.de; [+]equal contribution

Abstract

Human high-level visual cortex has been described in two seemingly opposed ways. A categorical view emphasizes discrete category-selective areas, while a dimensional view highlights continuous feature maps spanning across these areas. Can these divergent perspectives on cortical organization be reconciled within a unifying framework? Using data-driven decomposition of fMRI responses in face-, body-, and scene-selective areas, we identified overlapping activity patterns shared across individuals. Each area encoded multiple interpretable dimensions tuned to both finer subcategory features and coarser cross-category distinctions beyond its preferred category, even in the most category-selective voxels. These dimensions formed distinct clusters within category-selective areas but were also sparsely distributed across the broader visual cortex, supporting both locally selective, category-specific, and globally distributed, feature-based coding. Together, these findings suggest multidimensional tuning as a fundamental organizing principle that integrates feature-selective clusters, category-selective areas, and large-scale tuning maps, providing a more comprehensive understanding of category representations in human visual cortex.

Introduction

Human high-level visual cortex contains category-selective areas that respond preferentially to specific visual inputs, such as faces, bodies, and scenes (Downing et al., 2001; Epstein & Kanwisher, 1998; Kanwisher, 2010; Kanwisher et al., 1997; Peelen & Downing, 2017). A widely used approach contrasts predefined stimulus categories, hypothesized to engage selective processing, to reveal cortical clusters specialized for distinct domains. This strategy has supported a categorical view of cortical organization, which has been instrumental in identifying classical category-selective areas, discovering novel category-related responses (Bracci et al., 2010; Cortinovis et al., 2025; Epstein & Baker, 2019; Henderson et al., 2025; Kanwisher & Yovel, 2006; McCandliss et al., 2003; Peelen & Downing, 2007), and establishing their behavioral relevance (Cohen et al., 2017; Kanwisher & Barton, 2011; Moro et al., 2008).

At the same time, growing evidence suggests that category-selective areas encode a broad variety of features beyond their preferred categories (Cichy et al., 2012; Haxby et al., 2001). Within these areas, functional subclusters represent both fine-grained, domain-specific features (Bracci et al., 2015; de Haas et al., 2021; Orlov et al., 2010; Weiner & Grill-Spector, 2011) and domain-general features shared across categories (Arcaro et al., 2009; Çukur et al., 2013, 2016), indicating a more complex and overlapping representational structure. In contrast to the categorical view, a dimensional view proposes that occipitotemporal cortex (OTC), including category-selective areas, is organized along continuous feature dimensions that span categorical boundaries (Bracci & Op de Beeck, 2023; Ritchie et al., 2024). Supporting this view, neural tuning in OTC has been shown to reflect a wide range of feature dimensions, including animacy, real-world size, aspect ratio, behavioral aspects, and semantic content (Abdel-Ghaffar et al., 2024; Almeida et al., 2023; Arcaro & Livingstone, 2024; Bao et al., 2020; Coggan & Tong, 2023; Contier et al., 2024; Huth et al., 2012; Julian et al., 2017; Konkle & Caramazza, 2013; Konkle & Oliva, 2012; Watson et al., 2017; Watson & Andrews, 2024). These features are encoded in large-scale, distributed maps that embed category-selective areas within a continuous representational landscape (Haxby et al., 2001, 2011).

Whether high-level visual areas are primarily organized by discrete categories or continuous dimensions remains a central and unresolved question (Ritchie et al., 2024; Yargholi & Beeck, 2023). Each perspective provides valuable insights into the functional organization of OTC: the categorical view emphasizes locally selective, domain-specific coding, while the dimensional view highlights globally distributed, feature-based coding. Reconciling these seemingly divergent accounts may be essential for a more comprehensive understanding of how the brain represents complex visual information.

To address this question, we applied a data-driven decomposition of fMRI voxel responses (Khosla et al., 2022) in face-, body-, and scene-selective areas during natural image viewing (Allen et al., 2022), aiming to uncover their underlying representational dimensions. Our approach builds on two

theoretical assumptions (Hebart et al., 2020): first, that neural representations are sparse, such that only a limited number of dimensions are required to capture the response to a given stimulus; and second, that these dimensions are continuous and positive, facilitating interpretability by ensuring that each dimension contributes additively to the overall response. To implement these assumptions, we employed non-negative matrix factorization, a method that identifies part-based, sparse, and interpretable components. This approach allowed us to directly test the two competing views: if the categorical view holds, responses in each area should be dominated by a single component aligned with its preferred category; if the dimensional view is correct, responses should span multiple overlapping dimensions reflecting a continuous feature space.

By analyzing both the representational content and cortical topography of the identified dimensions, we found that both categorical and dimensional views capture essential aspects of high-level visual cortex organization. Category-selective areas encode information along multiple dimensions that are consistent across individuals. Dimensions in these areas are dominated by the respective preferred category but are also tuned to additional features, even in the most category-selective voxels. This tuning reveals representations that are richer and more multifaceted than predicted by a strictly categorical account. The topography of these dimensions reflects locally sparse yet globally distributed coding, with distinct clusters within category-selective areas and widely spaced clusters across cortex that together form large-scale tuning maps (Bogdan et al., 2025; Contier et al., 2024; Ritchie et al., 2024; Weiner & Grill-Spector, 2010). This organization integrates functional structure across multiple spatial scales, from feature-selective clusters to category-selective areas and large-scale tuning gradients. Within this framework, category-selective areas emerge as a special case of sparse tuning along a limited set of dimensions. We propose multidimensional tuning as a unifying organizing principle that reconciles categorical and dimensional views, offering new insight into how the visual system balances specificity and flexibility in representing the visual world.

Results

Data-driven voxel decomposition reveals multiple interpretable dimensions in category-selective areas

If category-selective areas encode information along multiple representational dimensions rather than through strict categorical distinctions, then a data-driven decomposition of their activity patterns should reveal distinct but overlapping dimensions. To test this possibility, we analyzed fMRI voxel responses from participants viewing natural images from the Natural Scenes Dataset (NSD; Allen et al., 2022).

Our analysis focused on three well-established category-selective areas associated with high-level visual processing: fusiform face area (FFA; Kanwisher et al., 1997), extrastriate body area (EBA; Downing et al., 2001), and parahippocampal place area (PPA; Epstein & Kanwisher, 1998).

To identify the underlying representations within these areas, we applied Bayesian non-negative matrix factorization (BNMF; Khosla et al., 2022; Schmidt et al., 2009), a technique that decomposes voxel responses into latent dimensions and their corresponding spatial maps. BNMF captures multiple selectivities that may coexist within individual voxels. By enforcing non-negativity, it promotes sparse and part-based representations that are more interpretable than those obtained from conventional dimensionality reduction methods. Moreover, because BNMF does not rely on predefined category labels or anatomical boundaries, it allows for a less biased, data-driven exploration of the functional organization within category-selective cortex. To ensure the robustness and generalizability of our results, we implemented a multi-step analysis pipeline (Fig. 1, see Methods). We first determined the optimal number of latent dimensions for each area using bi-cross-validation designed for matrix factorization (Owen & Perry, 2009; Supplementary Fig. 1). Next, to account for the stochastic nature of BNMF, we performed multiple decompositions per area and clustered the resulting dimensions within each participant, retaining only robust consensus dimensions. To address inter-individual variability, we then matched dimensions across participants by correlating their response profiles to a shared image set, selecting only those dimensions with consistent patterns across individuals. Finally, to test generalizability, we projected voxel responses to held-out images onto the previously learned dimensions and evaluated how well these projections predicted responses to new images.

Our analysis revealed a small set of highly consistent dimensions in each category-selective area, alongside a larger set of less consistent dimensions, following a function with approximately exponential decay (Fig. 1c, Supplementary Fig. 2). Despite strong univariate category selectivity ($t > 2$), each area exhibited multiple dimensions that were shared across participants. We identified eight consistent dimensions in FFA, twenty in EBA, and ten in PPA (mean r across participant pairs ± SD; FFA: $r = 0.46 \pm 0.12$; EBA: $r = 0.49 \pm 0.14$; PPA: $r = 0.45 \pm 0.10$), together capturing a substantial proportion of voxel response variance in each area (mean R^2 across participants ± SD; FFA: $R^2 = 0.33 \pm 0.06$; EBA: $R^2 = 0.39 \pm 0.01$; PPA: $R^2 = 0.29 \pm 0.09$). These findings suggest a more nuanced, multidimensional organization than predicted by a strictly categorical account.

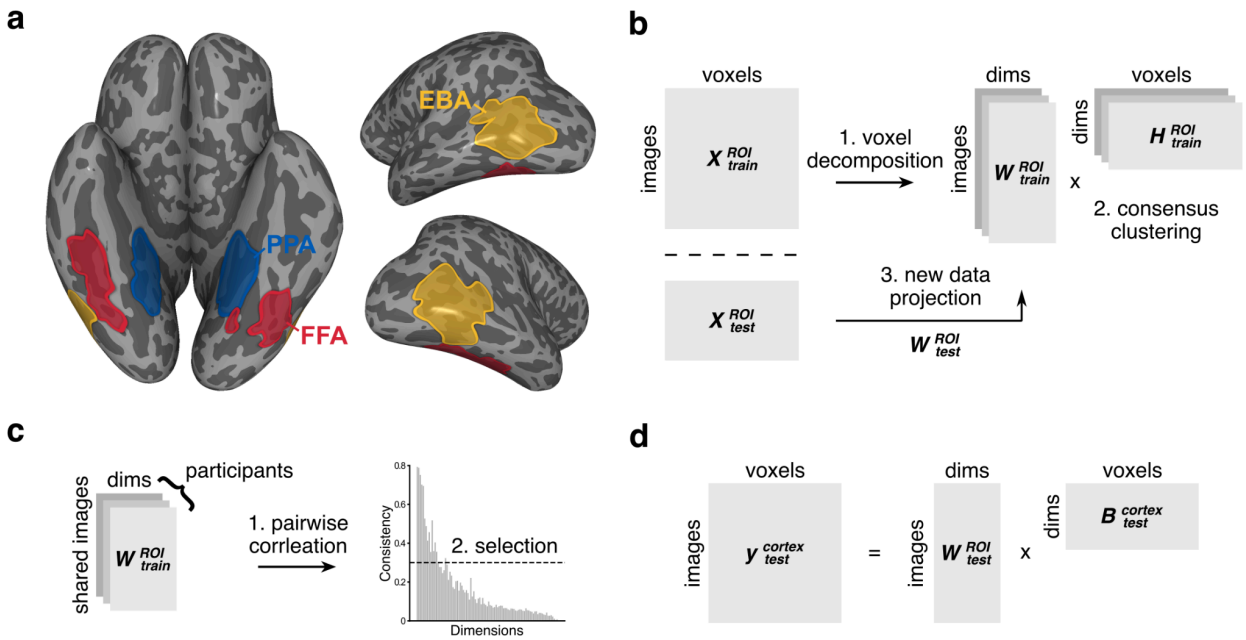


Fig. 1 | Data-driven fMRI voxel decomposition of category-selective areas. a FFA, EBA, and PPA were defined using a standard functional localizer ($t > 2$) and are displayed on the inflated cortical surface for a representative participant (P1). **b** At the participant level, reliable dimensions were extracted from voxel responses in each area through multiple BNMF decompositions, with the optimal dimensionality determined through bi-cross-validation. Dimensions were clustered across runs using *k*-medoids clustering, and cluster medians were used as reliable consensus dimensions. To assess generalizability, voxel responses to new images were projected onto the learned embedding space using non-negative least squares. **c** At the group level, consistent dimensions were identified by correlating dimensions across participants for a subset of shared images. Dimensions with the highest inter-participant consistency were retained. **d** Voxel-wise encoding models were trained using ridge regression to predict voxel responses to new images based on the learned dimensions, producing functional tuning maps for each dimension across cortex.

Dimensions reveal multifaceted representational structure in category-selective areas

Having identified multiple consistent dimensions, we next examined their representational content and how they relate to known functional selectivities. We began by visually inspecting the images that scored highest on each dimension to gain an initial understanding of their semantic content. These images largely aligned with semantic concepts, consistent with the established role of category-selective areas in high-level, view-invariant visual processing. To interpret the dimensions more systematically, we adopted a data-driven labeling approach using multimodal deep learning models. Specifically, we used GPT-4V to generate candidate labels from the highest-scoring images for each dimension and quantified

label-image alignment using CLIP-ViT embeddings (Radford et al., 2021). This approach produced meaningful descriptive labels for approximately 95 % of the consistent dimensions (FFA: 8/8; EBA: 20/21; PPA: 9/10; Fig. 2; Extended Data Fig. 1), although a few remained ambiguous and were left unlabeled.

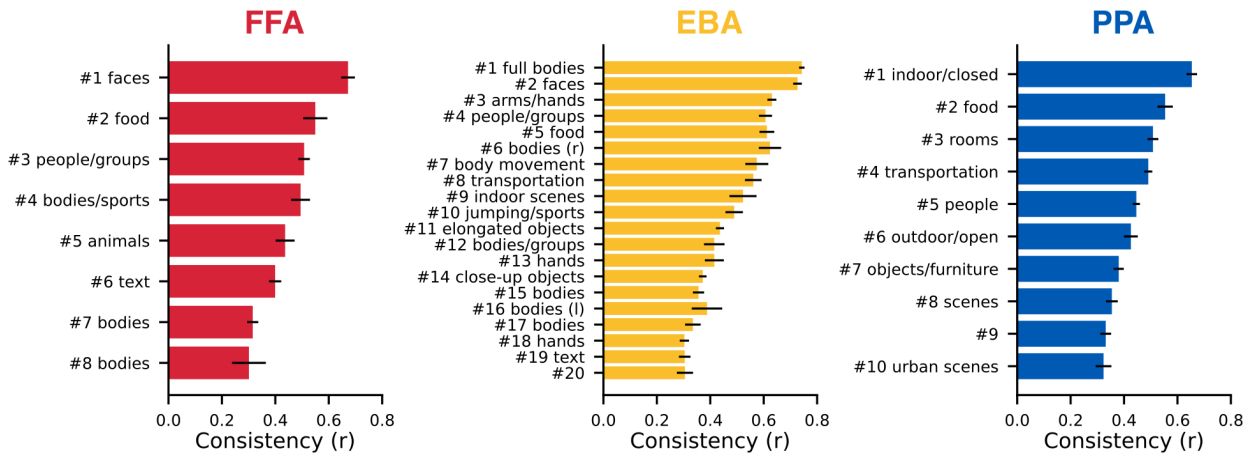


Fig. 2 | Consistency of dimensions in category-selective areas. Mean correlation of dimensions in each area, averaged across participant pairs (± SEM). The labels were determined using a data-driven interpretation based on multimodal deep learning models.

As expected, the most consistent dimensions in each area aligned with its preferred category: faces in FFA, whole bodies in EBA, and scenes in PPA (Fig. 3). However, the dimensions were not limited to broad categories but also captured finer-grained subcategory features. In EBA, for example, dimensions differentiated body parts, including full bodies, arms, hands, and faces (Bracci et al., 2015; Downing & Peelen, 2011; Orlov et al., 2010; Ramirez et al., 2024). Similarly, in PPA, dimensions encoded scene properties such as openness and naturalness, distinguishing indoor from outdoor, enclosed from open, and urban from natural environments (Kravitz et al., 2011; Lescroart & Gallant, 2019; Park et al., 2011). In contrast, in FFA, dimensions did not show evidence of subcategory tuning to specific facial parts.

Beyond dimensions aligned with each area's preferred category, many additional dimensions captured semantically meaningful content from other domains. In FFA, for example, while some dimensions reflected other people-related concepts, such as groups, bodies, and sports, others extended to animals, food, and text. In EBA dimensions were not only tuned to other body-related concepts, such as movement and sports, but also to food, transportation, indoor scenes, close-up objects, and text. And in PPA, alongside other scene-related concepts, such as transportation and objects/furniture, dimensions also represented food and people. These findings suggest that category-selective areas are not narrowly tuned to a single domain but instead encode a broader

semantic space. The lower consistency of these additional dimensions, relative to the dominant category-aligned ones, may help explain why previous approaches were unable to detect this broader and more nuanced tuning in these areas.

Most of the identified dimensions reflected high-level semantic content. A partial exception was observed in EBA, where certain body-related dimensions were tuned to the spatial location of bodies on the left or right side of the image, potentially reflecting retinotopic organization. Apart from this, the dimensions exhibited little sensitivity to low- or mid-level visual features such as shape, texture, or color. These findings suggest that although category-selective areas show strong tuning to their preferred category at the univariate level, they also encode a broader set of rich, semantic dimensions that become apparent at the multivariate level.

It is possible that identified dimensions within category-selective areas might lie along a single category-selective continuum (e.g., a “faceness” continuum in FFA with faces > bodies > text) rather than representing distinct tuning axes (e.g., faces, bodies, and text in FFA). To address this concern, we analyzed pairwise cosine similarities between dimension response profiles within each area. We specifically compared two groups of dimensions: those tuned to each area's preferred category, and those tuned to non-preferred features. A single category-selective continuum would predict high similarities between “preferred” and “non-preferred” dimensions (between-group comparison), due to alignment with the dominant category-selective axis. However, an area may also respond preferentially to some stimuli over others, irrespective of category preference, which would be indicated by high similarities among non-preferred dimensions (within-group comparison). Using raw dimension response profiles, we observed moderately high between-group similarities (mean cosine similarity across participants ± SD; FFA: M = 0.66 ± 0.05; EBA: M = 0.64 ± 0.03; PPA: M = 0.72 ± 0.02; range: [0, 1]), as well as moderately high within-group similarities for both preferred (FFA: M = 0.65 ± 0.06; EBA: M = 0.72 ± 0.02; PPA: M = 0.78 ± 0.03) and, critically, also non-preferred dimensions (FFA: M = 0.67 ± 0.05; EBA: M = 0.65 ± 0.03; PPA: M = 0.71 ± 0.02). This demonstrates that the generally high similarity observed between dimensions within an area cannot be explained by category-selective preferences. Next, to test if there is residual evidence of category-selective continuum beyond the overall preference for certain stimuli over others, we controlled for the overall response increase by removing the mean response profile across images and recomputing the similarities (Supplementary Fig. 3). As expected, after mean-centering, we observed low between-group similarities (FFA: M = -0.13 ± 0.02; EBA: M = -0.11 ± 0.01; PPA: M = -0.14 ± 0.01; range: [-1, 1]), as well as low within-group similarities for both preferred (FFA: M = -0.14 ± 0.05; EBA: M = 0.01 ± 0.01; PPA: M = -0.02 ± 0.03) and non-preferred dimensions (FFA: M = -0.13 ± 0.01; EBA: M = 0.02 ± 0.01; PPA: M = -0.10 ± 0.01), in line with multidimensional tuning. These results show that the dimensions not only show common stimulus preferences, but also

complementary (between-group) and distinct (within-group) tuning patterns that encode multiple, partially independent tuning axes rather than a single category-selective continuum.

Together, these findings demonstrate that category-selective areas encode rich, multifaceted representations that extend beyond simple categorical boundaries. These areas capture both finer subcategory features and broader cross-category distinctions, underscoring the complexity and flexibility of functional organization in high-level visual cortex.

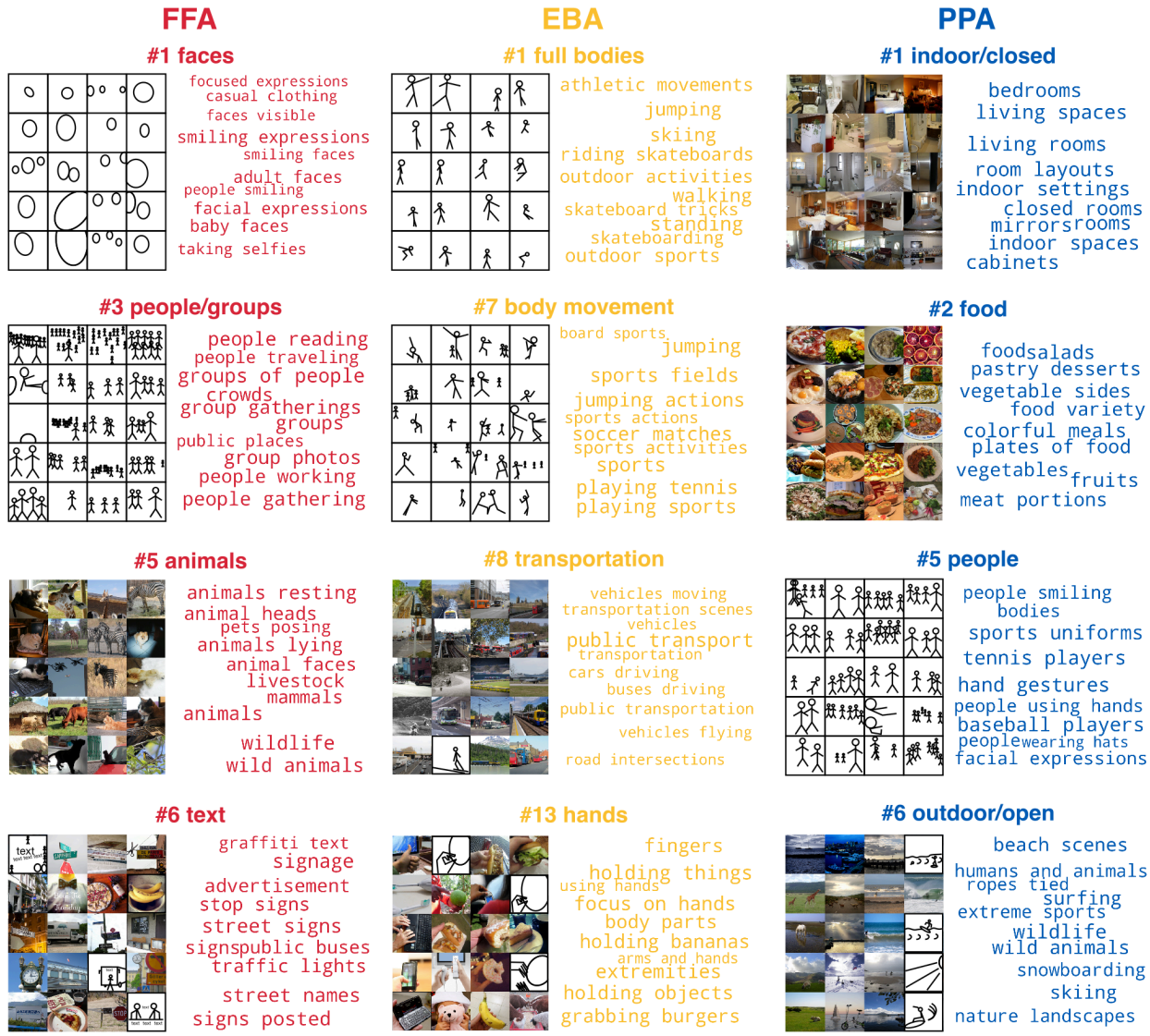


Fig. 3 | Interpretability of dimensions in category-selective areas. Interpretations of example dimensions from each area. Five highest-scoring images per participant based on dimension response profiles and ten highest-scoring labels combined across participants based on multimodal deep learning models. Images of people were replaced with stick figures.

Multidimensional tuning underlies category-selective representations

A potential concern is that the observed multidimensional tuning might simply arise from the inclusion of weakly selective voxels in the definition of category-selective areas. If so, the apparent diversity in tuning could reflect a methodological artifact rather than a genuine feature of functional organization. To address this possibility, we examined whether a voxel's degree of category selectivity predicted its tuning specificity across dimensions. For each voxel, we quantified category selectivity using its preferred category contrast from the functional localizer (*d'*; e.g., faces > all other categories in FFA) and related this measure to the voxel's weights across the learned dimensions. We also computed a sparseness measure (Hoyer, 2004; see Methods) to quantify each voxel's tuning profile, where high sparseness indicates narrow tuning to only a few dimensions and low sparseness reflects broader tuning to multiple dimensions. Across all areas, we found no evidence that greater category selectivity was associated with narrower dimensional tuning. Voxel-wise category selectivity was not reliably related to sparseness in EBA or PPA, and showed only a trend in FFA (mean r across participants; FFA: $r = 0.23$, $p = 0.06$; EBA: $r = -0.03$, $p = 0.46$; PPA: $r = -0.01$, $p = 0.77$; Supplementary Fig. 4). These results suggest that even the most category-selective voxels remain tuned to multiple, partially overlapping dimensions.

Further analysis revealed that although dominant dimensions in highly category-selective voxels corresponded primarily to the preferred category, additional dimensions captured complementary features (Fig. 4). For example, highly face-selective voxels in FFA also responded to dimensions related to bodies and animals, highly body-selective voxels in EBA were additionally tuned to actions, faces, and elongated objects, and highly scene-selective voxels in PPA responded to dimensions related to transportation, objects/furniture, and people.

Together, these findings indicate that multidimensional tuning is not merely a byproduct of weak voxel selectivity. Even the most category-selective voxels exhibit tuning to multiple dimensions. This suggests that category selectivity does not result from exclusive tuning to a single feature axis, but rather from a multidimensional code that integrates information across semantically related feature axes.

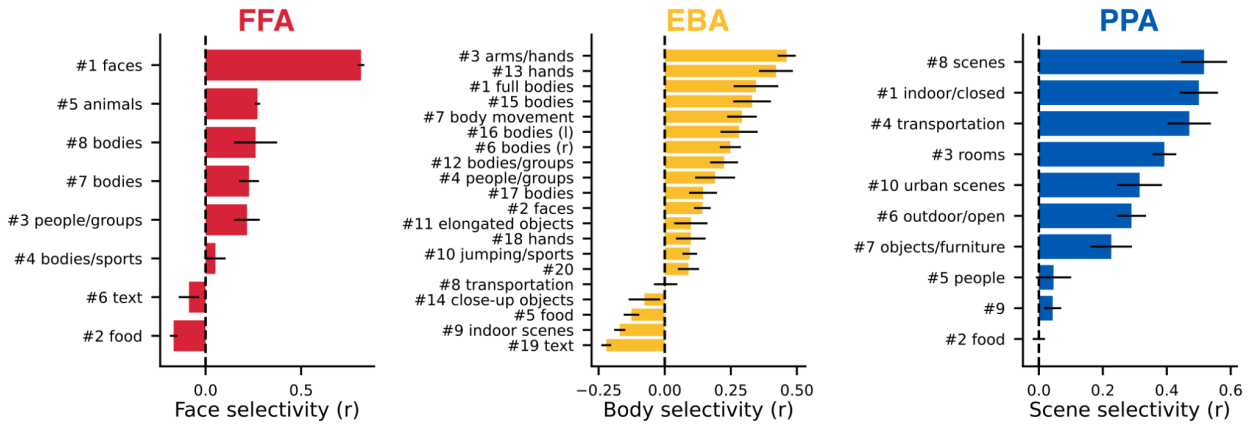


Fig. 4 | Relationship between category selectivity and dimension tuning. Mean correlation of voxel-wise dimension weights and category selectivity based on the respective functional localizer contrast in each area, averaged across participants (± SEM).

Dimensions from category-selective areas explain activity throughout high-level visual cortex

Having characterized the representational content of the identified dimensions, we next investigated whether these dimensions were confined to specific cortical areas or distributed more broadly across cortex. To test this, we assessed how well dimensions derived from each category-selective area predicted voxel responses beyond their region of origin. If the representations were highly localized, predictive power would be restricted to the source area. In contrast, broader topographies would suggest a more distributed functional organization with overlapping and shared dimensions.

We addressed this question using voxel-wise encoding models trained to predict responses to natural images based on the dimensions derived from each category-selective area. Model performance was evaluated by predicting voxel responses to held-out test images across cortex. To accommodate the possibility that some dimensions negatively correlate with neural activity elsewhere, we relaxed the non-negativity constraint on the voxel-wise coefficients. This allowed us to map the prediction performance of each dimension set and evaluate their explanatory power across the cortical surface (Fig. 5, Supplementary Fig. 5). Prediction performance was highest in high-level visual areas, consistent with the semantic nature of the dimensions, but also extended into early visual areas and prefrontal cortex, as confirmed through permutation testing. Within each category-selective area, locally derived dimensions explained local tuning especially well, with performance often approaching the estimated noise ceiling (Supplementary Fig. 6). Importantly, the dimensions generalized across areas. For example, FFA-derived dimensions explained variance in body-selective areas such as EBA and fusiform body area (FBA), while EBA-derived dimensions predicted responses in face-selective areas such as FFA and superior temporal

sulcus (STS). Similarly, EBA-derived dimensions predicted responses in scene-selective areas such as PPA and occipital place area (OPA). While some generalization may be attributable to spatial overlap between adjacent areas (e.g., FFA and FBA, EBA and OPA), the strength of the cross-predictive performance suggests that shared representational dimensions, rather than strict categorical boundaries, underlie these relationships.

Together, these findings indicate that while the dimensions capture the specific tuning profiles of individual category-selective areas, they also reveal a more distributed and overlapping functional organization. Rather than operating as isolated modules, these areas appear to participate in broader cortical networks that encode both unique and shared features.

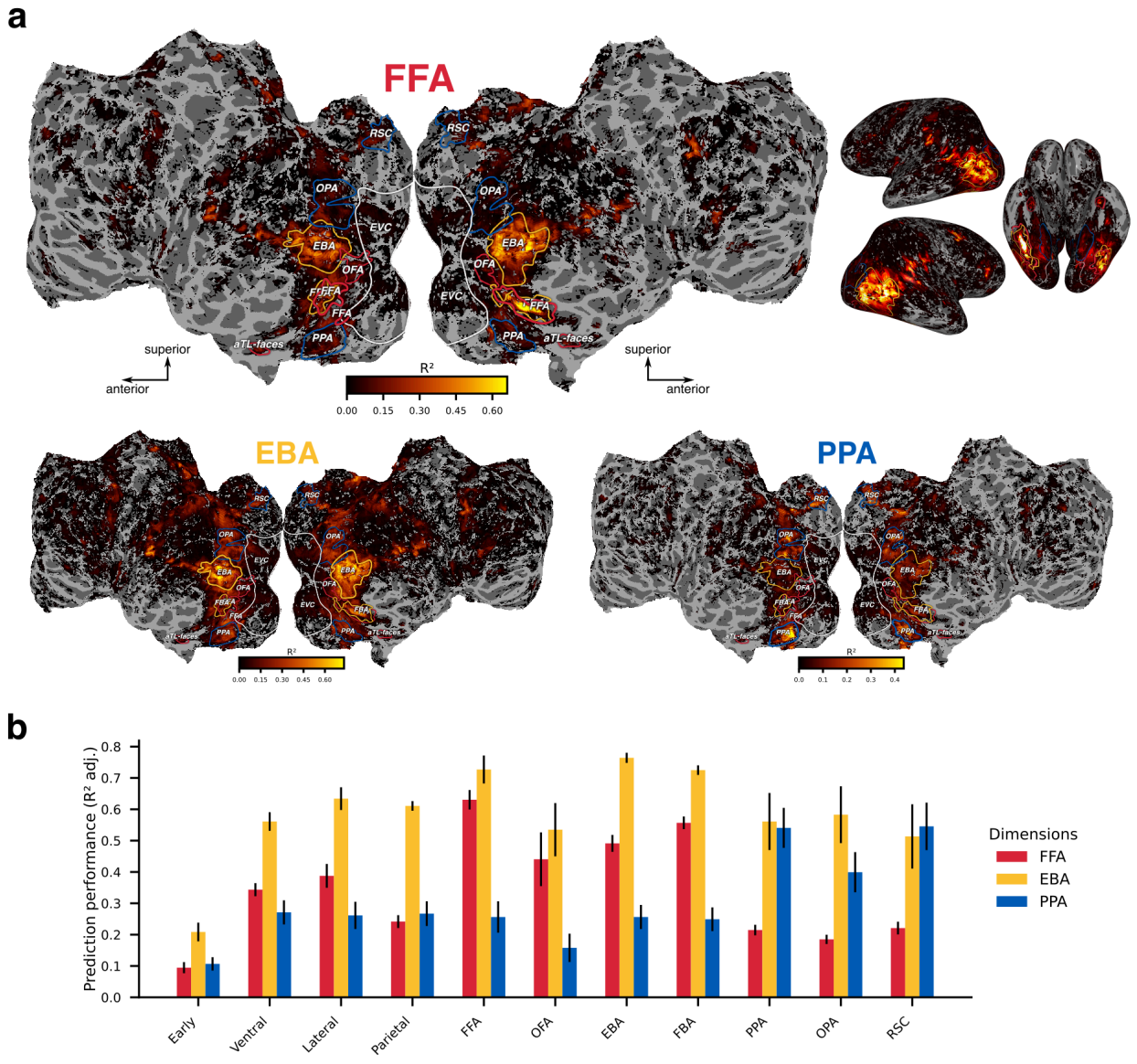


Fig. 5 | Prediction performance of dimensions from category-selective areas across cortex. a Voxel-wise prediction performance (R^2) for dimensions from each area projected onto the flattened cortical surface for a representative participant (P1). Voxels thresholded at $p < 0.01$ (one-sided, FDR-corrected). **b** Mean prediction performance (noise-ceiling-adjusted R^2) for different cortical areas, averaged across participants (± SEM).

Dimensions form locally sparse clusters and globally distributed maps

Having established that dimensions derived from category-selective areas explain responses across broad regions of cortex, we next examined their individual topographies. Within category-selective areas, the spatial organization of these dimensions could reflect either broad overlap, consistent with functional homogeneity, or discrete clusters, consistent with functional heterogeneity. To distinguish between these possibilities, we visualized the voxel-wise coefficient weights for each dimension across the flattened cortical surface (Fig. 6, Extended Data Figs. 2-5).

Many dimensions reflected known functional selectivities: face-related dimensions were concentrated in face-selective areas, body-related dimensions in body-selective areas, and scene-related dimensions in scene-selective areas. However, these representations were not spatially uniform. Most dimensions formed distinct subclusters within each category-selective area, revealing a finer-grained functional organization. For example, in FFA, subclusters were tuned to non-face features (Çukur et al., 2013). In EBA, subclusters encoded different body parts, including full bodies, faces, arms, and hands (Bracci et al., 2015; Orlov et al., 2010; Ramirez et al., 2024). In PPA, subclusters encoded different scene properties, such as openness and expanse (Lescroart & Gallant, 2019). These findings extend earlier reports that category-selective areas encode non-preferred features (Cichy et al., 2012; Haxby et al., 2001, 2011), demonstrating that such representations are not only present but also spatially organized. This supports a model in which category-selective areas are composed of multidimensional tuning maps rather than uniform modules, consistent with prior proposals of finer-grained functional organization (Grill-Spector & Weiner, 2014). Some dimensions also revealed novel or under-characterized selectivities, such as tuning to animals in FFA and elongated objects in EBA. These may reflect finer subdivisions within areas and highlight the potential of this approach to uncover new functional specializations.

Beyond local clustering, many dimensions extended across area boundaries and overlapped with neighboring regions. For example, in FFA, a text-related dimension extended into visual word form area (VWFA), and body-related dimensions overlapped with FBA. Other dimensions, such as those tuned to food, were even more broadly distributed, spanning multiple areas like V4, FFA, and PPA (Contier et al., 2024; Jain et al., 2023; Khosla et al., 2022; Pennock et al., 2023). We also observed expected hemispheric asymmetries, with face- and body-related dimensions predominantly right-lateralized and

text-related dimensions left-lateralized (Dehaene & Cohen, 2011; Willems et al., 2010). Overall, the cortical organization of these dimensions was marked by locally sparse clustering and globally distributed tuning across high-level visual cortex (Bogdan et al., 2025; Contier et al., 2024; Miyakawa et al., 2018; Ritchie et al., 2024; Weiner & Grill-Spector, 2010). This organization revealed functionally selective clusters distributed across cortex. For example, an EBA-derived face dimension recurred in other face-selective areas, including FFA, STS, occipital face area (OFA), an anterior temporal face area (aTL-faces), and a prefrontal face area near inferior frontal sulcus (Nikel et al., 2022; Tsao et al., 2008). Similarly, people-related dimensions were mostly shared between EBA and FFA, supporting the existence of interconnected face- and body-selective networks (Taubert et al., 2022; Weiner & Grill-Spector, 2010). To determine whether these overlapping functional maps also capture large-scale tuning gradients, we analyzed the principal orthogonal tuning directions of the dimension profiles. This analysis provided a coarse summary view of the dominant tuning directions and their corresponding large-scale gradients. It revealed that multidimensional tuning patterns mirrored well-known large-scale gradients (Supplementary Fig. 7). Despite some inter-individual variability in spatial location, the overall arrangement was consistent across participants, suggesting that these dimensions reflect stable and generalizable principles of cortical organization.

In summary, our findings provide a comprehensive account of functional organization in high-level visual cortex. Individual dimensions form sparse subclusters within category-selective areas while also contributing to distributed, large-scale maps across cortex. This locally sparse but globally distributed organization supports both discrete category selectivity and continuous feature integration, offering a unified framework for how high-level visual cortex encodes complex information.

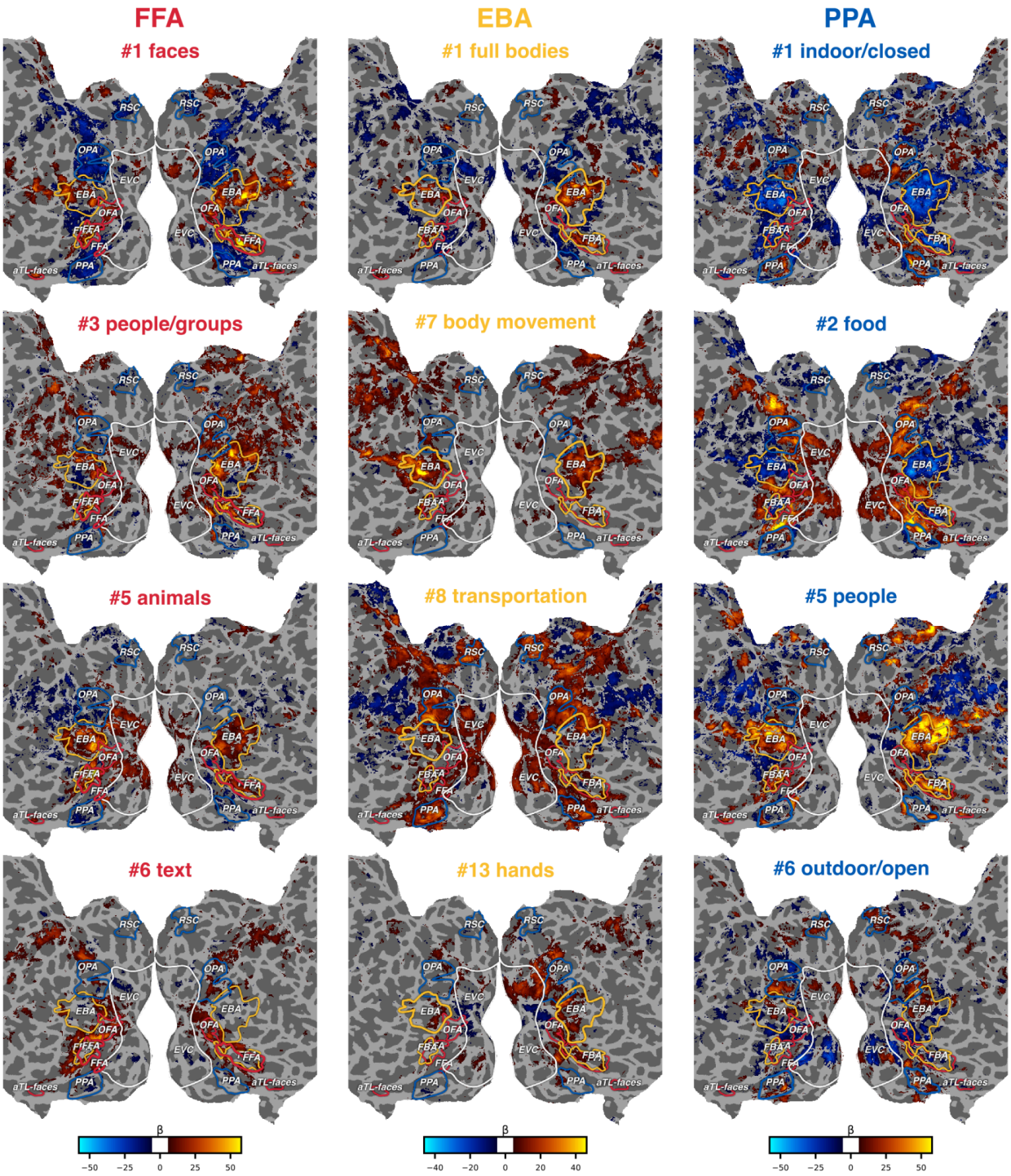


Fig. 6 | Functional tuning maps of individual dimensions from category-selective areas across cortex. Voxel-wise coefficient weights for six example dimensions from each area projected onto the flattened cortical surface for a representative participant (P1). Voxels thresholded at $p < 0.01$ (one-sided, FDR-corrected).

Discussion

Is high-level visual cortex organized into discrete categorical modules or does it reflect a continuous representational space (Ritchie et al., 2024; Yargholi & Beeck, 2023)? Our data-driven analysis of fMRI responses suggests it is neither strictly one nor the other, but a hybrid: a multidimensional code that integrates elements of both categorical and continuous organization. Within category-selective areas, we identified interpretable dimensions that largely aligned with their preferred categories, consistent with domain-specific tuning (Kanwisher, 2010; Peelen & Downing, 2017). At the same time, these areas also encoded finer-grained subcategory features and coarser cross-category distinctions, uncovering more nuanced and multifaceted representational content. These findings align with recent evidence that category-selective areas flexibly encode both domain-specific and domain-general information (Khosla & Wehbe, 2023; Vinken et al., 2023).

Our study contributes several key insights to the understanding of high-level visual cortex organization. Despite being directly derived from cortical responses to complex natural images, the decomposition yielded remarkably interpretable dimensions. Combining this decomposition with encoding models allowed us to link functional hallmarks across multiple spatial scales, ranging from fine-scale subclusters, over category-selective areas, to large-scale cortical maps. Moreover, our results showed that even the most category-selective parts of high-level visual cortex are tuned to multiple meaningful dimensions, calling into question the notion of pure selectivity at the voxel level.

We also found that representational diversity differed across category-selective areas. While FFA exhibited the narrowest tuning, both EBA and PPA displayed broader and more heterogeneous tuning. Notably, EBA showed the greatest heterogeneity, consistent with its role in encoding individual body parts (Bracci et al., 2015; Orlov et al., 2010; Ramirez et al., 2024) and its spatial overlap with multiple anatomically and functionally defined subregions (Weiner & Grill-Spector, 2011). These findings suggest that EBA may play a more integrative role than traditionally assumed. Similarly, PPA encoded scene-related features such as naturalness, openness, and expanse (Kravitz et al., 2011; Lescroart & Gallant, 2019; Park et al., 2011). In contrast, FFA did not show separable tuning to distinct facial parts, as observed in more controlled studies (de Haas et al., 2021; Henriksson et al., 2015), possibly due to limited variability or smaller size in the natural images. Interestingly, we were able to identify selectivity to animals in FFA, which may reflect sensitivity to animal faces or broader animacy representations (Kanwisher et al., 1999).

In addition to identifying individual dimensions, we uncovered two key principles of their spatial organization. First, within category-selective areas, dimensions formed distinct subclusters, consistent with a mosaic-like organization of feature-selective populations (Çukur et al., 2013, 2016; Weiner & Grill-Spector, 2011). Second, many dimensions were sparsely distributed across high-level visual cortex,

suggesting a general-purpose representational scaffold. Together, these findings support a hierarchical organization: fine-scale subclusters are tuned to specific dimensions, category-selective areas are sparsely tuned to subsets of these dimensions, and large-scale cortical maps reflect integrative, overlapping codes.

Our findings also resonate with a growing body of work using encoding models to uncover more nuanced patterns of functional organization (Çelik et al., 2021; Doerig et al., 2022; Doshi & Konkle, 2023; Efird et al., 2024; Luo et al., 2024). Notably, the topographies of individual dimensions closely resembled those of previously reported behavior-derived object dimensions (Contier et al., 2024; Hebart et al., 2020), suggesting that these neural representations may contribute to goal-directed behavior. Sparse tuning to such dimensions may underlie the emergence of category-selective areas (Contier et al., 2024; Ritchie et al., 2024), consistent with prior work linking category selectivity to sparsely distributed codes (Miyakawa et al., 2018; Weiner & Grill-Spector, 2010) and to patterns of cortical connectivity (Molloy et al., 2024; Op de Beeck et al., 2019; Osher et al., 2016). The reproducibility of these spatial maps across individuals reinforces their role as core organizing principles. Other recent frameworks have similarly aimed to reconcile categorical and dimensional accounts. For example, contrastive coding gives rise to brain-like functional specialization in deep neural networks without requiring explicit category-specific mechanisms, offering a plausible route for the emergence of representational dimensions (Prince et al., 2024). Additionally, temporal dynamics add complexity, as category-selective neurons can alter their coding schemes across time (Shi et al., 2023), suggesting that representational dimensions may be orchestrated both spatially and temporally (Teichmann et al., 2024).

Despite its contributions, our approach has several limitations. First, our decomposition approach imposes a non-negativity constraint, which is well-suited for capturing excitatory BOLD responses but insensitive to potentially meaningful inhibitory signals (Pérez-Ortega et al., 2024). Although we relaxed this constraint in the encoding models, future work should explore more flexible techniques that capture both excitatory and inhibitory components. Second, although many of the identified dimensions were interpretable, some remained ambiguous, likely reflecting many-to-one mappings between visual features and neural activity. Accordingly, the descriptive labels generated via our data-driven deep learning approach should be viewed as hypotheses rather than definitive interpretations. Third, like any data-driven approach, our findings are shaped by the properties of the dataset. Limited diversity in the natural images may bias the recovered dimensions (Shirakawa et al., 2025), although our focus on robustness likely mitigated the influence of spurious effects. Nonetheless, disentangling stimulus-driven features from learned associations remains a central challenge when using complex natural stimuli.

Our findings point to several promising directions for future research. A central question is whether these dimensions guide behavior and how they adapt to varying task demands (Ritchie et al., 2024). Examining their stability across tasks, datasets, and imaging modalities will provide deeper insights into their generalizability. Future work should also investigate how these dimensions interact across areas and evolve over time, potentially revealing their computational roles (Op de Beeck et al., 2008; Teichmann et al., 2024). In addition, probing the granularity of these dimensions may clarify how fine- and coarse-scale representations coexist (Gauthaman et al., 2024; Han & Bonner, 2025). While our decomposition showed mixed selectivity at the voxel level, discrete tuning may exist at finer scales, highlighting the need for higher-resolution methods to uncover potentially hidden structure (Quiroga et al., 2005). Finally, comparing biologically derived dimensions to those learned by artificial neural networks may uncover shared representational strategies between biological and artificial vision systems (Chen & Bonner, 2024; Hosseini et al., 2024; Huh et al., 2024; Kanwisher et al., 2023; Mahner et al., 2024).

To conclude, our multidimensional framework bridges the divide between categorical and dimensional accounts of high-level visual cortex organization. Our findings demonstrate that, when analyzed in a data-driven manner, these perspectives are not mutually exclusive but rather complementary. Category-selective areas emerge from overlapping, sparsely distributed representational dimensions. This framework captures both the apparent discreteness of category selectivity and the continuity of dimensional coding. Multidimensional tuning provides a unifying principle that enables the visual system to balance specificity with flexibility, which is essential for supporting a wide range of functional demands.

Methods

fMRI data

Natural Scenes Dataset

We used the Natural Scenes Dataset (NSD; Allen et al., 2022), which includes fMRI responses from eight participants who each viewed 9,000 to 10,000 natural scene images over 30 to 40 scan sessions, with each image repeated three times. Our analyses focused on participants who completed all trials (P1: S1, P2: S2, P3: S5, and P4: S7), viewing 9,000 unique and 1,000 shared images. During scanning, participants maintained central fixation and performed a long-term recognition task by identifying previously viewed images. Images were sourced from the Microsoft Common Objects in Context (COCO) database (Lin et al., 2014) and displayed at a visual angle of 8.4° × 8.4° for 3 s, with a 1 s interval between images. BOLD responses were acquired at 7 T using whole-brain gradient-echo EPI with 1.8

mm isotropic voxel resolution and a 1.6 s repetition time. Preprocessing included temporal and spatial interpolation for slice timing and head motion correction, and single-trial voxel responses were estimated using a general linear model. We used preprocessed single-trial voxel responses optimized through voxel-wise hemodynamic response function modeling, data-driven denoising, and ridge regression ("betas_fithrf_GLMdenoise_RR"; 1.8 mm isotropic native preparation; Prince et al., 2022). Cortical flat maps were generated using PyCortex (Gao et al., 2015). Details on scanning protocols, preprocessing steps, and noise ceiling estimation are provided in the original NSD paper (Allen et al., 2022). To cross-validate our analyses, we randomly divided the images into a 70 % training set and a 30 % test set, ensuring that all shared images were assigned to the training set. To account for session effects and improve reliability, single-trial voxel responses were *z*-scored within each session using the mean and standard deviation of the training trials, and neural response estimates were averaged across repetitions of each image.

Category-selective areas

We extracted voxel responses from three functional regions of interest (ROIs): fusiform face area (FFA), parahippocampal place area (PPA), and extrastriate body area (EBA). These ROIs were identified using a separate functional localizer (fLOC; Stigliani et al., 2015). During the fLOC, participants viewed grayscale images of five categories (faces, bodies, places, characters, objects) presented in a miniblock design across six runs, with each block containing eight images per category. After standard preprocessing, category-specific beta values were estimated using a general linear model. ROIs were defined by conducting *t*-tests comparing each category to all others ($t > 2$), isolating voxels with clear category preferences. Details on the analyses are provided in the original fLOC paper (Stigliani et al., 2015). To ensure reliable input for BNMF, we selected voxels within each ROI that had a signal-to-noise-ratio greater than 0.2. We also performed the BNMF procedure for additional face-, body-, and scene-selective ROIs, including occipital face area (OFA), fusiform body area (FBA), and occipitotemporal place area (OPA). These analyses yielded results similar to those observed in the primary ROIs (FFA, EBA, and PPA). For the sake of brevity, we focus on the three most well-established ROIs here. We quantified voxel-wise category selectivity using the category selectivity index (*d'*), computed based on each ROI's preferred category contrast (i.e., preferred category versus all other categories), where $\bar{X}_{category}$ and $\bar{X}_{others}$ as well as $\sigma^2_{category}$ and σ^2_{others} denote the mean voxel responses and variances to the preferred category and all other categories, respectively:

$$d' = \frac{\overline{X}_{\text{category}} - \overline{X}_{\text{others}}}{\sqrt{\frac{\sigma^2_{\text{category}} + \sigma^2_{\text{others}}}{2}}}$$

Data-driven voxel decomposition

Bayesian non-negative matrix factorization

We applied Bayesian non-negative matrix factorization (BNMF) to decompose voxel responses from ROIs into dimensions, following prior work by Khosla et al. (2022). Non-negative matrix factorization (NMF) decomposes a data matrix V ($I \times V$), where I is the number of images and V is the number of voxels, into the product of two lower-rank matrices: a response matrix W ($I \times D$) and a weight matrix H ($D \times V$), where D is the number of dimensions, such that the product of these matrices optimally reconstructs the data. In this formulation, each column of W (w_d) represents a dimension's response profile, while each row of H (h_d) encodes its spatial pattern across voxels. The non-negativity constraint in NMF restricts dimensions to additive combinations, promoting part-based, sparse, and interpretable dimensions. Compared to other dimensionality reduction techniques like principal component analysis or cluster analysis, NMF offers several advantages: (1) voxel responses are represented exclusively by additive dimensions, avoiding cancellation effects, (2) voxels are not assigned only to individual dimensions or individual clusters, allowing for mixed selectivity, and (3) dimensions are sparse, with each voxel associated with only a subset of dimensions.

We applied BNMF, an extension of NMF, which incorporates Bayesian inference by imposing exponential priors on W and H, and a normal likelihood for the residual matrix E ($I \times V$) with noise variance σ^2. BNMF employs Gibbs sampling to iteratively sample from the posterior distributions of W, H, and σ^2. This Bayesian formulation yields probabilistic estimates that explicitly model noise in the data, improving robustness. Specifically, BNMF is based on the following model components:

- Likelihood:

$$p(W, H, E) \sim \prod_{i=1}^{I} \prod_{v=1}^{V} \mathcal{N}(X_{i,v}; (WH)_{i,v}, \sigma^2)$$

- Prior for W:

$$p(W) \sim \prod_{i=1}^{I} \prod_{d=1}^{D} \mathrm{Exp}(-\rho_{i,d} W_{i,d}), \quad W_{i,d} > 0$$

- Prior for H:

$$p(H) \sim \prod_{d=1}^{D} \prod_{v=1}^{V} \mathrm{Exp}(-\gamma_{d,v} H_{d,v}), \quad H_{d,v} > 0$$

We ran 3,000 Gibbs iterations, discarded the first 1,000 as burn-in, and retained every 5th draw after this. The retained samples were averaged element-wise to obtain the posterior mean estimates of the factor matrices. Each model was initialized with a unique random seed, and convergence was achieved once the reconstruction error fell below 1×10^{-5}. Details of the algorithm are provided in the original BNMF paper (Schmidt et al., 2009). To enforce non-negativity, we baseline-shifted X by subtracting the minimum z-scored response across all voxels in each ROI. To prevent data leakage, both training and test sets were shifted using the global training minimum. Any negative outliers remaining in the test set were set to 0. The constant offset introduced between training and test distributions was accounted for in the final encoding model by fitting an intercept term.

Estimating the optimal dimensionality

We determined the optimal dimensionality k^* for each participant's ROI using bi-cross-validation, which accounts for the interdependence of rows and columns in matrix factorization (Owen & Perry, 2009). The data matrix X was shuffled and divided into four blocks: A, B, C, and D. BNMF was trained on the training block D, and pseudo-inverse matrices of W and H from blocks B and C were used to predict the held-out test block A. This process was repeated 5 times with random block partitions in each iteration for ranks between 1 and 100 in steps of 3, and the average prediction error across iterations was used to calculate the final cross-validation error. For each participant, k^* was selected as the rank that minimized the test error (Supplementary Fig. 1). Estimating the optimal dimensionality for fMRI responses to natural images is challenging due to their hierarchical nature, which requires capturing meaningful dimensions while avoiding overfitting to spurious ones. Our approach balanced these considerations, ensuring optimal latent dimensionality within each participant's ROI, maximizing representational capacity, while also minimizing the risk of overfitting.

Consensus approach

We identified dimensions within each ROI using a two-step approach to ensure reliability and generalizability. First, to address the sensitivity of BNMF to initialization, we adapted a consensus approach outlined in previous work (Kotliar et al., 2019). For each ROI, we performed 100 randomly initialized BNMF voxel decompositions with k^* dimensions. Each dimension was normalized to an L_2-norm of 1. Unreliable runs were removed based on a density-based outlier detection approach. To aggregate reliable dimensions across runs, we applied k-medoids clustering with cosine distance, using k^* clusters. Cluster medians were then used as consensus dimensions for subsequent analyses. Second, to identify consistent dimensions across participants, we used a pairwise correlation approach also

outlined in previous work (Khosla et al., 2022). We computed pairwise correlations between all combinations of dimensions across participants ($k^*_1 \times k^*_2 \times k^*_3 \times k^*_4$). Using a greedy selection approach, we iteratively matched dimensions with the highest pairwise correlations until no more dimensions could be matched across all participants. Dimensions with a mean pairwise correlation exceeding r greater than 0.3 were considered consistent, resulting in 8 to 20 consensus dimensions per ROI (Supplementary Fig. 2). This two-step approach ensured that the identified dimensions were reliable across runs and consistent across participants, providing a robust framework for subsequent analyses. To evaluate the generalizability of the dimensions, we projected the test set into the learned embedding using non-negative least squares.

Data-driven dimension labeling

To interpret the consistent dimensions identified across participants, we employed a data-driven labeling approach using large language models to generate labels and multimodal neural network representations to align image-related dimensions with meaningful semantic labels. The approach involved generating a diverse set of candidate labels based on the highest-scoring images for each dimension and evaluating the labels using all training images. First, candidate labels were generated using GPT-4V (OpenAI). For each dimension, we selected the five highest-scoring images from each participant and provided these images combined as input to the model, along with the instruction to generate ten concise labels (with a maximum of three words each) per set of images (Supplementary Methods 1). We pooled the labels generated by GPT-4V across dimensions to compile a comprehensive set of candidate labels. Next, the candidate labels were encoded using CLIP-ViT-L-14 (Radford et al., 2021), with embeddings generated for each label using 15 different prompt templates (e.g., “a picture of …”, “an image of …”, “a photo of …”), following the advised methodology to enhance semantic robustness. Similarly, image embeddings were extracted for all training images. To assess the alignment of each label with the images, we calculated the cosine similarity between the label and image embeddings, weighted by the response profile of each dimension. This weighting ensured that the most relevant images for each dimension were prioritized in the analysis. For each label, we averaged the cosine similarities across the different prompts and performed the following steps to ensure specificity and reduce the influence of generic labels. We first subtracted the global mean similarity for each label across all dimensions. This step reduced global biases, allowing similarities to better reflect the distinctive features associated with each label and dimension. After mean subtraction, the weighted similarities were z-scored across all labels. Z-scoring helped to emphasize labels that were uniquely associated with each dimension, minimizing the impact of labels with consistently high similarities across all dimensions. The z-scored similarities were

then used to rank the labels for each dimension, with the top-ranked labels visualized as word clouds to provide a concise, interpretable representation of the dimensions. This method allowed us to derive meaningful, semantically relevant labels for each dimension, based on the recurring patterns in the highest-scoring images, while validating them across the entire image set. By using this approach, we ensured that the interpretations of the dimensions were both robust and representative.

Voxel-wise encoding model

To investigate prediction performance and functional tuning maps of dimensions from each ROI, we implemented a voxel-wise encoding model using its dimensions to predict voxel responses across cortex. Importantly, we used the consistent BNMF dimensions derived from baseline-shifted ROI voxel responses to predict unshifted voxel responses to new images across cortex. To mitigate overfitting and address multicollinearity between dimensions, we applied fractional ridge regression (Rokem & Kay, 2020), which allows for flexible regularization while preserving the sparse and interpretable nature of the dimensions. This regularization technique balances regularized and unregularized coefficient norms via an α-fraction. A fraction of 1 corresponds to no regularization (ordinary least squares solution), while a fraction of 0 represents maximum regularization (shrinking all coefficients to zero). The fraction for each voxel was optimized using 10-fold cross-validation of the training set. For each fold, fractions were tuned on 90 % of the training data and validated on the remaining 10 %, sampled from 0.1 to 0.9 in increments of 0.1, and from 0.9 to 1 in increments of 0.01 for higher precision in the less regularized range. The model included z-scoring of predictors within cross-validation folds and an intercept to account for the baseline-shift. We calculated prediction performance (R^2) for each fold and selected the fraction that yielded the highest average performance per voxel. Voxels in higher-level visual cortex surrounding the ROIs generally required less regularization and achieved higher prediction performance. The final model, trained using optimal fractions, was evaluated on the held-out test set by capturing the relationship between predicted and observed voxel responses. Statistical significance was assessed by generating a null distribution of correlations through 3,000 random permutations of the test set within each fold. Voxel-wise p-values were calculated with a one-tailed comparison to the null distribution, corrected for multiple comparisons using Benjamini-Hochberg procedure with a false discovery rate (FDR) of $p < 0.01$. Importantly, this approach allowed the use of non-negative response profiles while permitting negative voxel coefficients, identifying cortical areas where a dimension's profile was negatively correlated with voxel tuning. This preserved the part-based, sparse, and interpretable nature of the dimensions while relaxing the non-negativity constraint for voxel patterns. A strong linear relationship between the spatial mappings from BNMF and the encoding model confirmed that relaxing the non-negativity constraint did

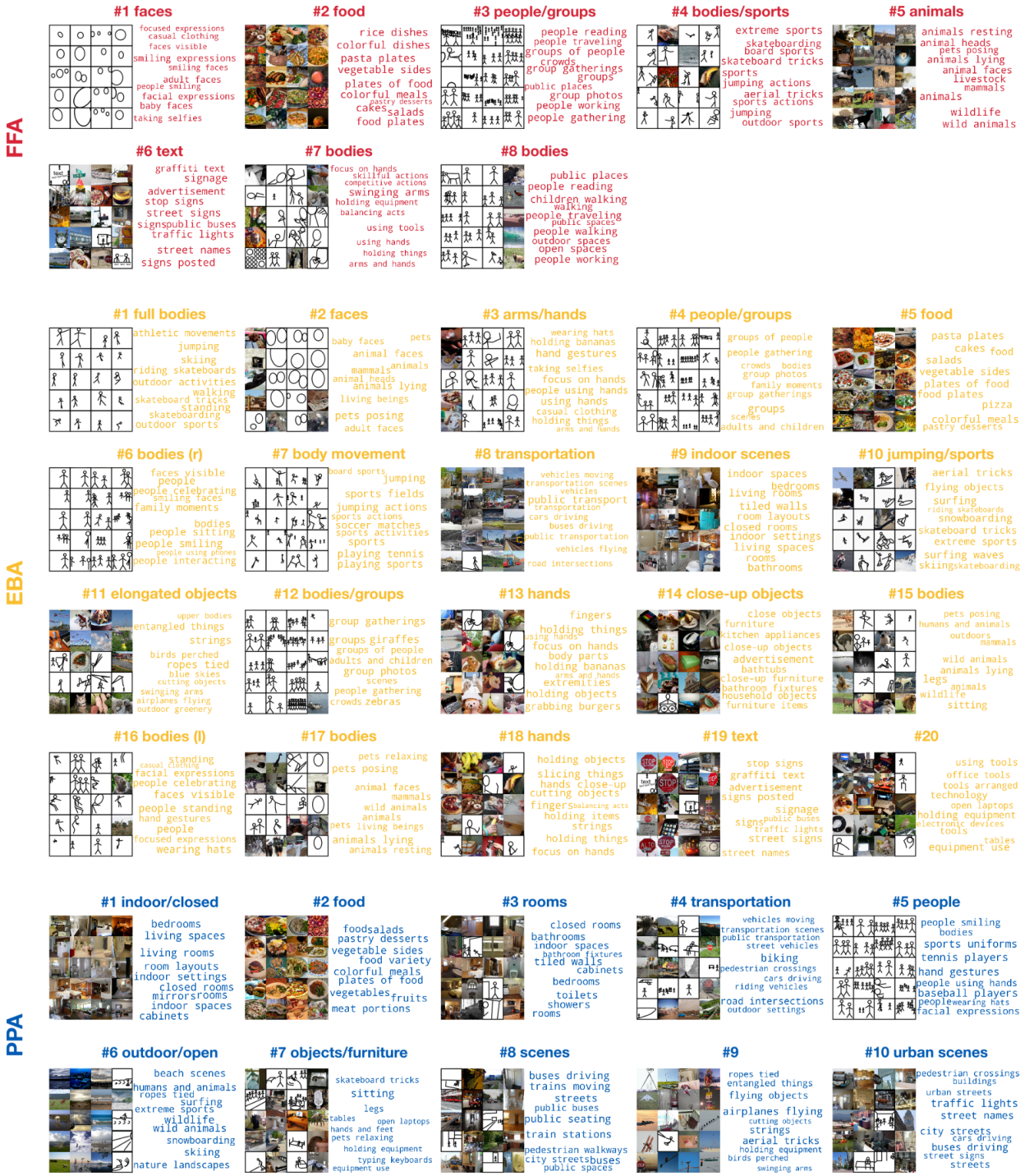
not compromise the validity of the original weights (mean *r* across participants ± SD; FFA: *r* = 0.94 ± 0.04; EBA: *r* = 0.89 ± 0.05; PPA: *r* = 0.89 ± 0.06). Noise ceilings were computed for the test set following the methodology provided in the original NSD paper (Allen et al., 2022).

Representational sparseness

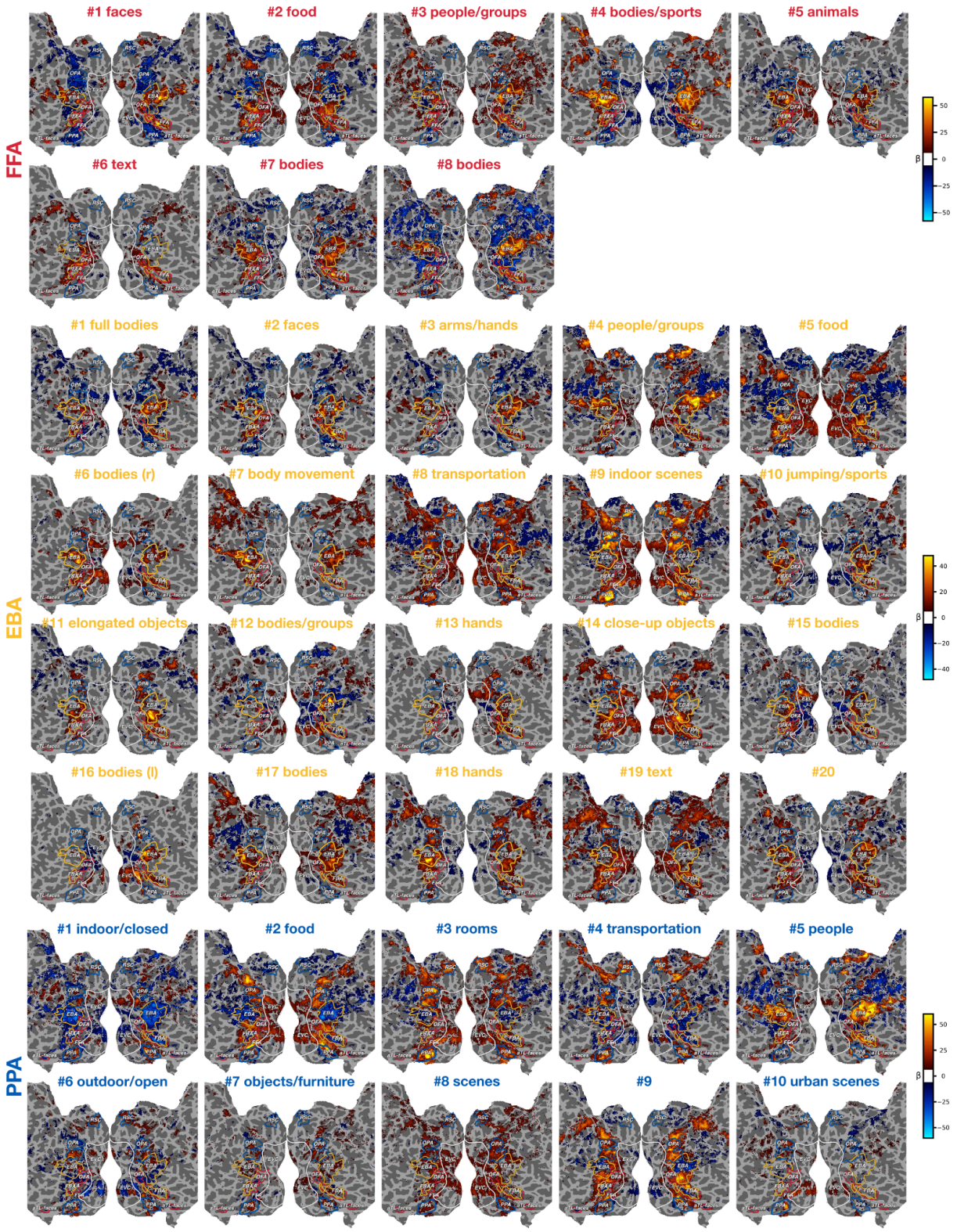
We quantified the extent of multidimensional tuning using a measure of representational sparseness based on the normalized relationship between the L_1- and L_2-norm of a vector (Hoyer, 2004). For each voxel, the dimension weights of the encoding model were interpreted as its *n* dimensional tuning profile *x*. The resulting sparseness index *s* ranges from 0, indicating a perfectly sparse representation tuned to a single dimension, to 1, indicating a perfectly dense representation tuned equally to all dimensions.

$$s(x) = \frac{\sqrt{n} - \sum |x_i| / \sqrt{\sum x_i^2}}{\sqrt{n} - 1}$$

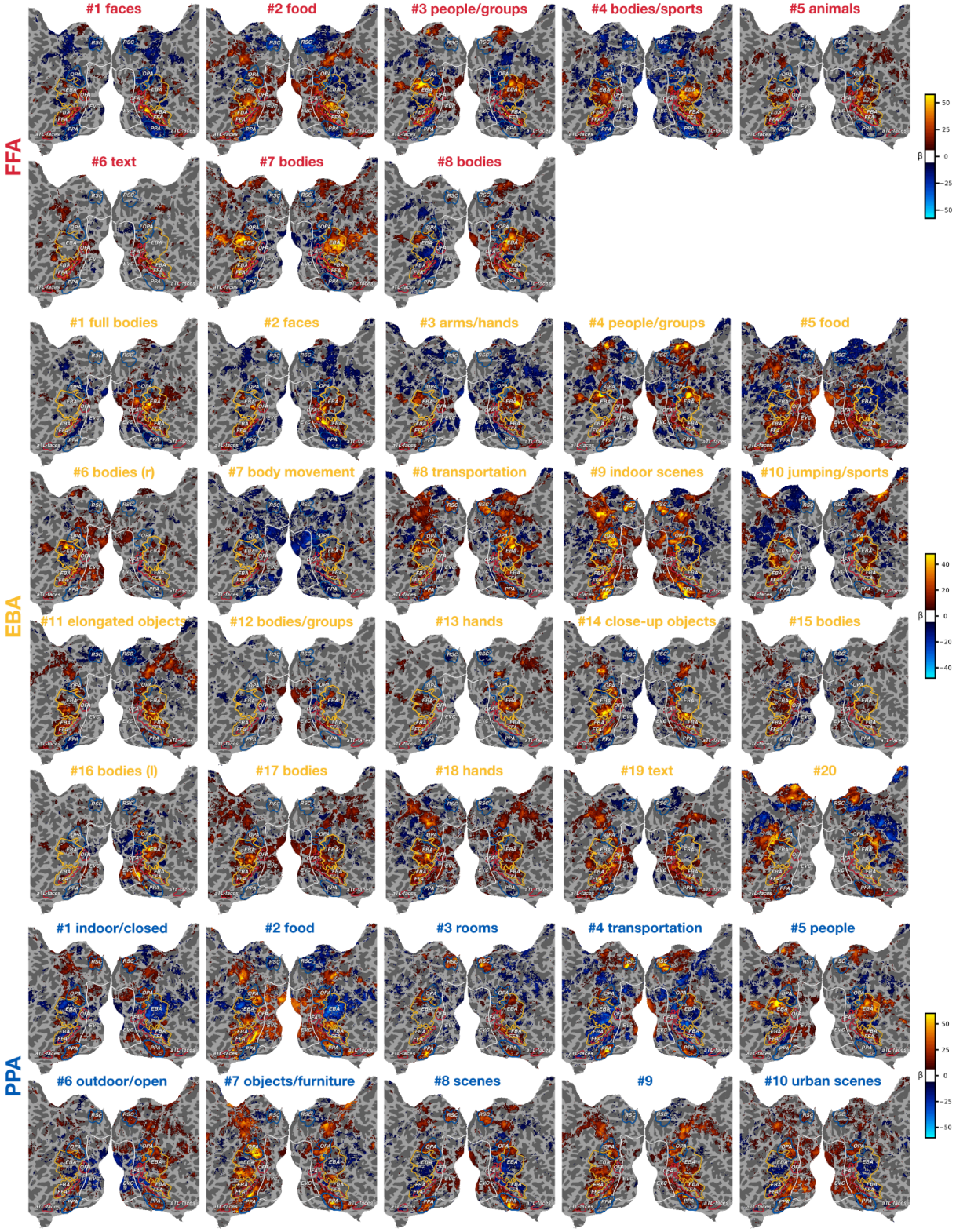
Extended data figures



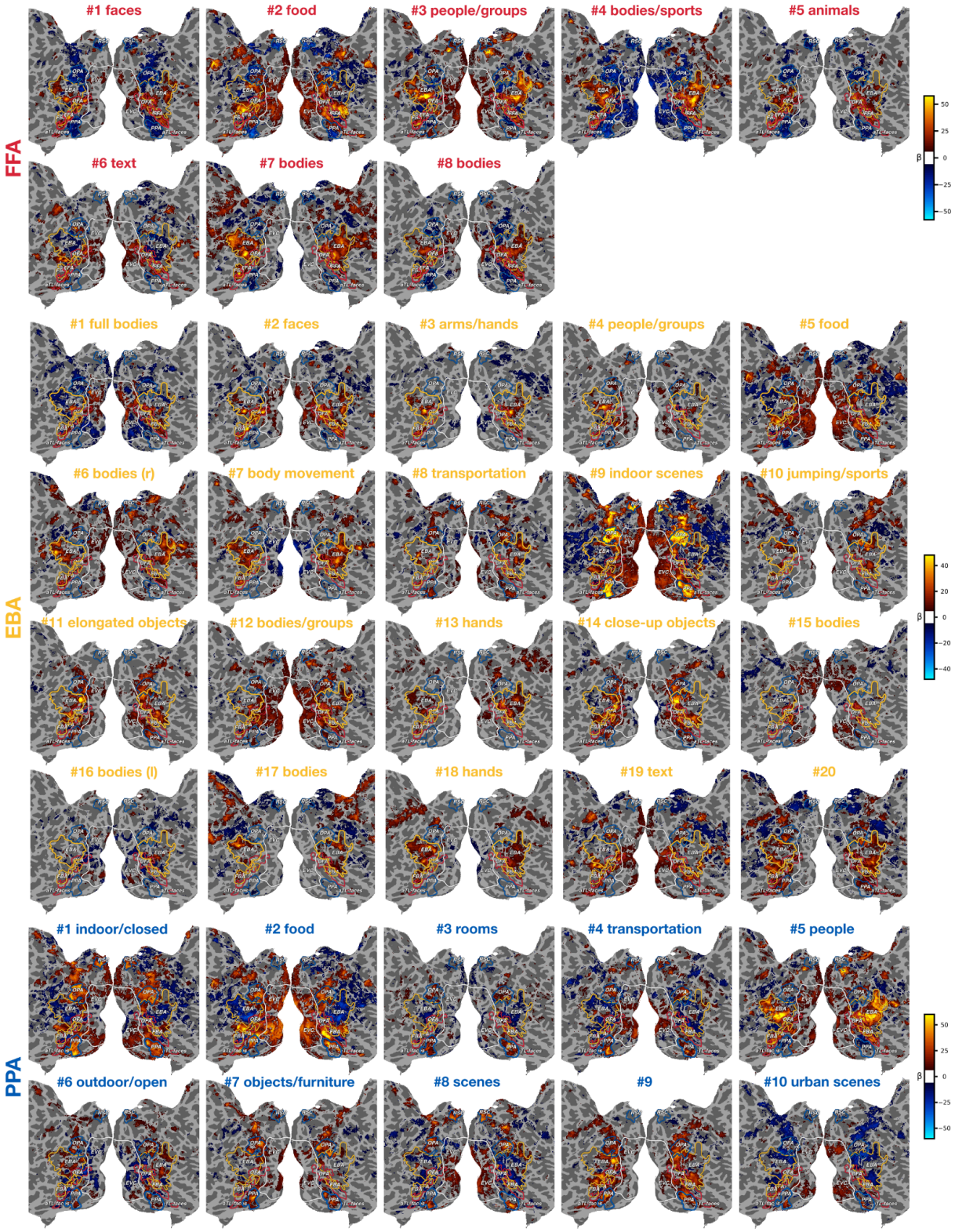
Extended Data Fig. 1 | Highest-scoring images and word-clouds for individual dimensions. Five highest-scoring images per participant and ten highest-scoring labels combined across participants for individual dimensions from each area. Images of people were replaced with stick figures.



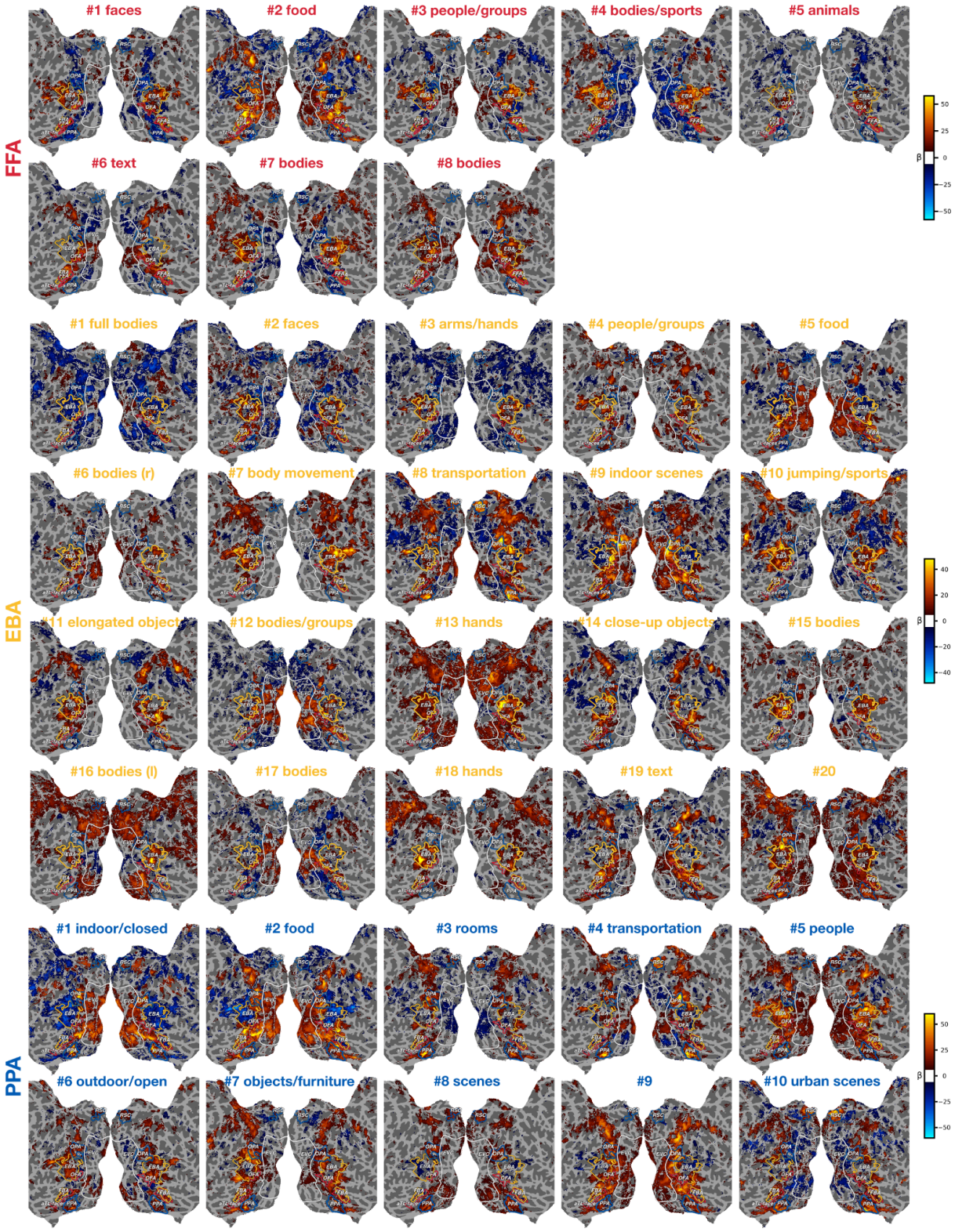
Extended Data Fig. 2 | Dimension tuning maps for participant P1. Voxel-wise coefficient weights for individual dimensions from each area projected onto the flattened cortical surface. Voxels thresholded at $p < 0.01$ (one-sided, FDR-corrected).



Extended Data Fig. 3 | Dimension tuning maps for participant P2. Voxel-wise coefficient weights for individual dimensions from each area projected onto the flattened cortical surface. Voxels thresholded at $p < 0.01$ (one-sided, FDR-corrected).



Extended Data Fig. 4 | Dimension tuning maps for participant P3. Voxel-wise coefficient weights for individual dimensions from each area projected onto the flattened cortical surface. Voxels thresholded at $p < 0.01$ (one-sided, FDR-corrected).



Extended Data Fig. 5 | Dimension tuning maps for participant P4. Voxel-wise coefficient weights for individual dimensions from each area projected onto the flattened cortical surface. Voxels thresholded at $p < 0.01$ (one-sided, FDR-corrected).

Data and code availability

The data supporting our analyses were obtained from the publicly available Natural Scenes Dataset (http://naturalscenesdataset.org). The Python code (version 3.8.20) used for data analysis and visualization is publicly available on GitHub (https://github.com/levandyck/roidims).

Acknowledgments

L.E.v.D was supported by a doctoral scholarship awarded by the German Academic Scholarship Foundation. M.N.H. and K.D. were supported by the ERC Starting Grants COREDIM (ERC-2021-StG-101039712) and DEEPFUNC (ERC-2023-STG-101117441), respectively, as well as the Hessian Ministry of Higher Education, Research, Science and the Arts (LOEWE Start Professorships and Excellence Program "The Adaptive Mind") and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, 222641018-SFB/TRR 135 TP). The funding organizations had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. Computational resources were provided by the high-performance computing clusters at the Max Planck Computing & Data Facility (MPCDF), Garching, Germany.

Author contributions

L.E.v.D., M.N.H., and K.D. conceived the study. L.E.v.D. carried out the data analysis and wrote the original draft of the manuscript. M.N.H. and K.D. reviewed the manuscript and provided critical feedback. M.N.H. and K.D. jointly supervised the project.

Competing interests

The authors declare no competing interests.

References

Abdel-Ghaffar, S. A., Huth, A. G., Lescroart, M. D., Stansbury, D., Gallant, J. L., & Bishop, S. J. (2024). Occipital-temporal cortical tuning to semantic and affective features of natural images predicts associated behavioral responses. *Nature Communications, 15*(1), 5531. https://doi.org/10.1038/s41467-024-49073-8

Allen, E. J., St-Yves, G., Wu, Y., Breedlove, J. L., Prince, J. S., Dowdle, L. T., Nau, M., Caron, B., Pestilli, F., Charest, I., Hutchinson, J. B., Naselaris, T., & Kay, K. (2022). A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. *Nature Neuroscience, 25*(1), 116-126. https://doi.org/10.1038/s41593-021-00962-x

Almeida, J., Fracasso, A., Kristensen, S., Valério, D., Bergström, F., Chakravarthi, R., Tal, Z., & Walbrin, J. (2023). Neural and behavioral signatures of the multidimensionality of manipulable object processing. *Communications Biology, 6*(1), 940. https://doi.org/10.1038/s42003-023-05323-x

Arcaro, M. J., & Livingstone, M. (2024). A Whole-Brain Topographic Ontology. *Annual Review of Neuroscience, 47*, 21–40. https://doi.org/10.1146/annurev-neuro-082823-073701

Arcaro, M. J., McMains, S. A., Singer, B. D., & Kastner, S. (2009). Retinotopic Organization of Human Ventral Visual Cortex. *Journal of Neuroscience, 29*(34), 10638–10652. https://doi.org/10.1523/JNEUROSCI.2807-09.2009

Bao, P., She, L., McGill, M., & Tsao, D. Y. (2020). A map of object space in primate inferotemporal cortex. *Nature, 583*(7814), 103-108. https://doi.org/10.1038/s41586-020-2350-5

Bogdan, P. C., Howard, C., Gillette, K., Cabeza, R., & Davis, S. W. (2025). Local and distributed information coding in the ventral stream. *bioRxiv*. https://doi.org/10.1101/2025.03.09.642235

Bracci, S., Caramazza, A., & Peelen, M. V. (2015). Representational Similarity of Body Parts in Human Occipitotemporal Cortex. *Journal of Neuroscience, 35*(38), 12977–12985. https://doi.org/10.1523/JNEUROSCI.4698-14.2015

Bracci, S., Ietswaart, M., Peelen, M. V., & Cavina-Pratesi, C. (2010). Dissociable Neural Responses to Hands and Non-Hand Body Parts in Human Left Extrastriate Visual Cortex. *Journal of Neurophysiology, 103*(6), 3389–3397. https://doi.org/10.1152/jn.00215.2010

Bracci, S., & Op de Beeck, H. P. (2023). Understanding Human Object Vision: A Picture Is Worth a Thousand Representations. *Annual Review of Psychology, 74*(1), 113-135. https://doi.org/10.1146/annurev-psych-032720-041031

Çelik, E., Keles, U., Kiremitçi, İ., Gallant, J. L., & Çukur, T. (2021). Cortical networks of dynamic scene category representation in the human brain. *Cortex, 143*, 127–147. https://doi.org/10.1016/j.cortex.2021.07.008

Chen, Z., & Bonner, M. F. (2024). Universal dimensions of visual representation. *arXiv*. https://doi.org/10.48550/arXiv.2408.12804

Cichy, R. M., Heinzle, J., & Haynes, J.-D. (2012). Imagery and Perception Share Cortical Representations of Content and Location. *Cerebral Cortex, 22*(2), 372–380. https://doi.org/10.1093/cercor/bhr106

Coggan, D. D., & Tong, F. (2023). Spikiness and animacy as potential organizing principles of human ventral visual cortex. *Cerebral Cortex, 33*(13), 8194–8217. https://doi.org/10.1093/cercor/bhad108

Cohen, M. A., Alvarez, G. A., Nakayama, K., & Konkle, T. (2017). Visual search for object categories is predicted by

the representational architecture of high-level visual cortex. *Journal of Neurophysiology, 117*(1), 388–402. https://doi.org/10.1152/jn.00569.2016

Contier, O., Baker, C. I., & Hebart, M. N. (2024). Distributed representations of behaviour-derived object dimensions in the human visual system. *Nature Human Behaviour, 1–15, 2179-2193*. https://doi.org/10.1038/s41562-024-01980-y

Cortinovis, D., Peelen, M. V., & Bracci, S. (2025). Tool Representations in Human Visual Cortex. *Journal of Cognitive Neuroscience, 37*(3), 515–531. https://doi.org/10.1162/jocn_a_02281

Çukur, T., Huth, A. G., Nishimoto, S., & Gallant, J. L. (2013). Functional Subdomains within Human FFA. *Journal of Neuroscience, 33*(42), 16748–16766. https://doi.org/10.1523/JNEUROSCI.1259-13.2013

Çukur, T., Huth, A. G., Nishimoto, S., & Gallant, J. L. (2016). Functional Subdomains within Scene-Selective Cortex: Parahippocampal Place Area, Retrosplenial Complex, and Occipital Place Area. *Journal of Neuroscience, 36*(40), 10257–10273. https://doi.org/10.1523/JNEUROSCI.4033-14.2016

de Haas, B., Sereno, M. I., & Schwarzkopf, D. S. (2021). Inferior Occipital Gyrus Is Organized along Common Gradients of Spatial and Face-Part Selectivity. *Journal of Neuroscience, 41*(25), 5511–5521. https://doi.org/10.1523/JNEUROSCI.2415-20.2021

Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in Cognitive Sciences, 15*(6), 254–262. https://doi.org/10.1016/j.tics.2011.04.003

Doerig, A., Kietzmann, T. C., Allen, E., Wu, Y., Naselaris, T., Kay, K., & Charest, I. (2022). Visual representations in the human brain are aligned with large language models. *arXiv*. https://doi.org/10.48550/arXiv.2209.11737

Doshi, F. R., & Konkle, T. (2023). Cortical topographic motifs emerge in a self-organized map of object space. *Science Advances, 9*(25), eade8187. https://doi.org/10.1126/sciadv.ade8187

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A Cortical Area Selective for Visual Processing of the Human Body. *Science, 293*(5539), 2470-2473. https://doi.org/doi:10.1126/science.1063414

Downing, P. E., & Peelen, M. V. (2011). The role of occipitotemporal body-selective regions in person perception. *Cognitive Neuroscience, 2*(3–4), 186–203. https://doi.org/10.1080/17588928.2011.582945

Efird, C., Murphy, A., Zylberberg, J., & Fyshe, A. (2024). What's the Opposite of a Face? Finding Shared Decodable Concepts and their Negations in the Brain. *arXiv*. https://doi.org/10.48550/arXiv.2405.17663

Epstein, R. A., & Baker, C. I. (2019). Scene Perception in the Human Brain. *Annual Review of Vision Science, 5*(1), 373–397. https://doi.org/10.1146/annurev-vision-091718-014809

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

Gao, J. S., Huth, A. G., Lescroart, M. D., & Gallant, J. L. (2015). Pycortex: An interactive surface visualizer for fMRI. *Frontiers in Neuroinformatics, 9*, 23. https://doi.org/10.3389/fninf.2015.00023

Gauthaman, R. M., Ménard, B., & Bonner, M. F. (2024). Universal scale-free representations in human visual cortex. *arXiv*. https://doi.org/10.48550/arXiv.2409.06843

Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. *Nature Reviews Neuroscience, 15*(8), 536-548. https://doi.org/10.1038/nrn3747

Han, C., & Bonner, M. F. (2025). High-dimensional structure underlying individual differences in naturalistic visual

experience. *arXiv*. https://doi.org/10.48550/arXiv.2505.12653

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293*(5539), 2425-2430. https://doi.org/10.1126/science.1063736

Haxby, J. V., Guntupalli, J. S., Connolly, A. C., Halchenko, Y. O., Conroy, B. R., Gobbini, M. I., Hanke, M., & Ramadge, P. J. (2011). A Common, High-Dimensional Model of the Representational Space in Human Ventral Temporal Cortex. *Neuron, 72*(2), 404–416. https://doi.org/10.1016/j.neuron.2011.08.026

Hebart, M. N., Zheng, C. Y., Pereira, F., & Baker, C. I. (2020). Revealing the multidimensional mental representations of natural objects underlying human similarity judgements. *Nature Human Behavior, 4(*11), 1173-1185. https://doi.org/10.1038/s41562-020-00951-3

Henderson, M. M., Tarr, M. J., & Wehbe, L. (2025). Origins of food selectivity in human visual cortex. *Trends in Neurosciences*. https://doi.org/10.1016/j.tins.2024.12.001

Henriksson, L., Mur, M., & Kriegeskorte, N. (2015). Faciotopy—A face-feature map with face-like topology in the human occipital face area. *Cortex, 72*, 156–167. https://doi.org/10.1016/j.cortex.2015.06.030

Hosseini, E., Casto, C., Zaslavsky, N., Conwell, C., Richardson, M., & Fedorenko, E. (2024). Universality of representation in biological and artificial neural networks. *bioRxiv*. https://doi.org/10.1101/2024.12.26.629294

Hoyer, P. O. (2004). Non-negative matrix factorization with sparseness constraints. *arXiv*. https://doi.org/10.48550/arXiv.cs/0408058

Huh, M., Cheung, B., Wang, T., & Isola, P. (2024). The Platonic Representation Hypothesis. *arXiv*. https://doi.org/10.48550/arXiv.2405.07987

Huth, A. G., Nishimoto, S., Vu, A. T., & Gallant, J. L. (2012). A Continuous Semantic Space Describes the Representation of Thousands of Object and Action Categories across the Human Brain. *Neuron, 76*(6), 1210–1224. https://doi.org/10.1016/j.neuron.2012.10.014

Jain, N., Wang, A., Henderson, M. M., Lin, R., Prince, J. S., Tarr, M. J., & Wehbe, L. (2023). Selectivity for food in human ventral visual cortex. *Communications Biology, 6*(1), 175. https://doi.org/10.1038/s42003-023-04546-2

Julian, J. B., Ryan, J., & Epstein, R. A. (2017). Coding of Object Size and Object Category in Human Visual Cortex. *Cerebral Cortex, 27*(6), 3095–3109. https://doi.org/10.1093/cercor/bhw150

Kanwisher, N. (2010). Functional specificity in the human brain: A window into the functional architecture of the mind. *Proceedings of the National Academy of Sciences, 107*(25), 11163-11170. https://doi.org/doi:10.1073/pnas.1005062107

Kanwisher, N., & Barton, J. J. S. (2011). The Functional Architecture of the Face System: Integrating Evidence from fMRI and Patient Studies. In A. J. Calder, G. Rhodes, M. H. Johnson, & J. V. Haxby (Eds.), *Oxford Handbook of Face Perception*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199559053.013.0007

Kanwisher, N., Khosla, M., & Dobs, K. (2023). Using artificial neural networks to ask 'why' questions of minds and brains. *Trends in Neurosciences*. https://doi.org/10.1016/j.tins.2022.12.008

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. *The Journal of Neuroscience, 17*(11), 4302-4311. https://doi.org/10.1523/JNEUROSCI.17-11-04302.1997

Kanwisher, N., Stanley, D., & Harris, A. (1999). The fusiform face area is selective for faces not animals. *NeuroReport, 10*(1), 183-187. https://doi.org/10.1097/00001756-199901180-00035

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: A cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B: Biological Sciences, 361*(1476), 2109-2128. https://doi.org/10.1098%2Frstb.2006.1934

Khosla, M., Murty, N. A. R., & Kanwisher, N. (2022). A highly selective response to food in human visual cortex revealed by hypothesis-free voxel decomposition. *Current Biology, 32*(19), 4159-4171. https://doi.org/10.1016/j.cub.2022.08.009

Khosla, M., & Wehbe, L. (2023). Higher visual areas act like domain-general filters with strong selectivity and functional specialization. *bioRxiv*. https://doi.org/10.1101/2022.03.16.484578

Konkle, T., & Caramazza, A. (2013). Tripartite Organization of the Ventral Stream by Animacy and Object Size. *The Journal of Neuroscience, 33*(25), 10235-10242. https://doi.org/10.1523/JNEUROSCI.0983-13.2013

Konkle, T., & Oliva, A. (2012). A Real-World Size Organization of Object Responses in Occipitotemporal Cortex. *Neuron, 74*(6), 1114-1124. https://doi.org/10.1016/j.neuron.2012.04.036

Kotliar, D., Veres, A., Nagy, M. A., Tabrizi, S., Hodis, E., Melton, D. A., & Sabeti, P. C. (2019). Identifying gene expression programs of cell-type identity and cellular activity with single-cell RNA-Seq. *eLife, 8*, e43803. https://doi.org/10.7554/eLife.43803

Kravitz, D. J., Peng, C. S., & Baker, C. I. (2011). Real-World Scene Representations in High-Level Visual Cortex: It's the Spaces More Than the Places. *Journal of Neuroscience, 31*(20), 7322–7333. https://doi.org/10.1523/JNEUROSCI.4588-10.2011

Lescroart, M. D., & Gallant, J. L. (2019). Human Scene-Selective Areas Represent 3D Configurations of Surfaces. *Neuron, 101*(1), 178-192. https://doi.org/10.1016/j.neuron.2018.11.004

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft COCO: Common Objects in Context. In D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars (Eds.), *Computer Vision – ECCV 2014* (pp. 740–755). Springer International Publishing. https://doi.org/10.1007/978-3-319-10602-1_48

Luo, A. F., Henderson, M. M., Tarr, M. J., & Wehbe, L. (2024). BrainSCUBA: Fine-Grained Natural Language Captions of Visual Cortex Selectivity. *arXiv*. https://doi.org/10.48550/arXiv.2310.04420

Mahner, F. P., Muttenthaler, L., Güçlü, U., & Hebart, M. N. (2024). Dimensions underlying the representational alignment of deep neural networks with humans. *arXiv*. https://doi.org/10.48550/arXiv.2406.19087

McCandliss, B. D., Cohen, L., & Dehaene, S. (2003). The visual word form area: Expertise for reading in the fusiform gyrus. *Trends in Cognitive Sciences, 7*(7), 293-299. https://doi.org/10.1016/S1364-6613(03)00134-7

Miyakawa, N., Majima, K., Sawahata, H., Kawasaki, K., Matsuo, T., Kotake, N., Suzuki, T., Kamitani, Y., & Hasegawa, I. (2018). Heterogeneous Redistribution of Facial Subcategory Information Within and Outside the Face-Selective Domain in Primate Inferior Temporal Cortex. *Cerebral Cortex, 28*(4), 1416–1431.

https://doi.org/10.1093/cercor/bhx342

Molloy, M. F., Saygin, Z. M., & Osher, D. E. (2024). Predicting high-level visual areas in the absence of task fMRI. *Scientific Reports, 14*(1), 11376. https://doi.org/10.1038/s41598-024-62098-9

Moro, V., Urgesi, C., Pernigo, S., Lanteri, P., Pazzaglia, M., & Aglioti, S. M. (2008). The Neural Basis of Body Form and Body Action Agnosia. *Neuron, 60*(2), 235–246. https://doi.org/10.1016/j.neuron.2008.09.022

Nikel, L., Sliwinska, M. W., Kucuk, E., Ungerleider, L. G., & Pitcher, D. (2022). Measuring the response to visually presented faces in the human lateral prefrontal cortex. *Cerebral Cortex Communications, 3*(3), tgac036. https://doi.org/10.1093/texcom/tgac036

Op de Beeck, H. P., Haushofer, J., & Kanwisher, N. G. (2008). Interpreting fMRI data: Maps, modules and dimensions. *Nature Reviews Neuroscience, 9*(2), 123-135. https://doi.org/10.1038/nrn2314

Op de Beeck, H. P., Pillet, I., & Ritchie, J. B. (2019). Factors Determining Where Category-Selective Areas Emerge in Visual Cortex. *Trends in Cognitive Sciences, 23*(9), 784-797. https://doi.org/10.1016/j.tics.2019.06.006

Orlov, T., Makin, T. R., & Zohary, E. (2010). Topographic Representation of the Human Body in the Occipitotemporal Cortex. *Neuron, 68*(3), 586–600. https://doi.org/10.1016/j.neuron.2010.09.032

Osher, D. E., Saxe, R. R., Koldewyn, K., Gabrieli, J. D. E., Kanwisher, N., & Saygin, Z. M. (2016). Structural Connectivity Fingerprints Predict Cortical Selectivity for Multiple Visual Categories across Cortex. *Cerebral Cortex, 26*(4), 1668–1683. https://doi.org/10.1093/cercor/bhu303

Owen, A. B., & Perry, P. O. (2009). Bi-cross-validation of the SVD and the nonnegative matrix factorization. *The Annals of Applied Statistics, 3*(2), 564–594. https://doi.org/10.1214/08-AOAS227

Park, S., Brady, T. F., Greene, M. R., & Oliva, A. (2011). Disentangling Scene Content from Spatial Boundary: Complementary Roles for the Parahippocampal Place Area and Lateral Occipital Complex in Representing Real-World Scenes. *Journal of Neuroscience, 31*(4), 1333–1340. https://doi.org/10.1523/JNEUROSCI.3885-10.2011

Peelen, M. V., & Downing, P. E. (2007). The neural basis of visual body perception. *Nature Reviews Neuroscience, 8*(8), 636–648. https://doi.org/10.1038/nrn2195

Peelen, M. V., & Downing, P. E. (2017). Category selectivity in human visual cortex: Beyond visual object recognition. *Neuropsychologia, 105*, 177–183. https://doi.org/10.1016/j.neuropsychologia.2017.03.033

Pennock, I. M. L., Racey, C., Allen, E. J., Wu, Y., Naselaris, T., Kay, K. N., Franklin, A., & Bosten, J. M. (2023). Color-biased regions in the ventral visual pathway are food selective. *Current Biology, 33*(1), 134-146. https://doi.org/10.1016/j.cub.2022.11.063

Pérez-Ortega, J., Akrouh, A., & Yuste, R. (2024). Stimulus encoding by specific inactivation of cortical neurons. *Nature Communications, 15*(1), 3192. https://doi.org/10.1038/s41467-024-47515-x

Prince, J. S., Alvarez, G. A., & Konkle, T. (2024). Contrastive learning explains the emergence and function of visual category-selective regions. *Science Advances, 10*(39), eadl1776. https://doi.org/10.1126/sciadv.adl1776

Prince, J. S., Charest, I., Kurzawski, J. W., Pyles, J. A., Tarr, M. J., & Kay, K. N. (2022). Improving the accuracy of single-trial fMRI response estimates using GLMsingle. *eLife, 11*, e77599. https://doi.org/10.7554/eLife.77599

Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons

in the human brain. *Nature, 435*(7045), 1102–1107. https://doi.org/10.1038/nature03687

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning Transferable Visual Models From Natural Language Supervision. *Proceedings of the 38th International Conference on Machine Learning*, 8748–8763. https://proceedings.mlr.press/v139/radford21a.html

Ramirez, J. G., Vanhoyland, M., Ratan Murty, N. A., Decramer, T., Van Paesschen, W., Bracci, S., Op de Beeck, H., Kanwisher, N., Janssen, P., & Theys, T. (2024). Intracortical recordings reveal the neuronal selectivity for bodies and body parts in the human visual cortex. *Proceedings of the National Academy of Sciences, 121*(51), e2408871121. https://doi.org/10.1073/pnas.2408871121

Ritchie, J. B., Wardle, S. G., Vaziri-Pashkam, M., Kravitz, D. J., & Baker, C. I. (2024). Rethinking category-selectivity in human visual cortex. *arXiv*. https://doi.org/10.48550/arXiv.2411.08251

Rokem, A., & Kay, K. (2020). Fractional ridge regression: A fast, interpretable reparameterization of ridge regression. *GigaScience, 9*(12), giaa133. https://doi.org/10.1093/gigascience/giaa133

Schmidt, M. N., Winther, O., & Hansen, L. K. (2009). Bayesian Non-negative Matrix Factorization. In T. Adali, C. Jutten, J. M. T. Romano, & A. K. Barros (Eds.), *Independent Component Analysis and Signal Separation* (pp. 540–547). Springer. https://doi.org/10.1007/978-3-642-00599-2_68

Shi, Y., Bi, D., Hesse, J. K., Lanfranchi, F. F., Chen, S., & Tsao, D. Y. (2023). Rapid, concerted switching of the neural code in inferotemporal cortex. *bioRxiv*. https://doi.org/10.1101/2023.12.06.570341

Shirakawa, K., Nagano, Y., Tanaka, M., Aoki, S. C., Muraki, Y., Majima, K., & Kamitani, Y. (2025). Spurious reconstruction from brain activity. *Neural Networks, 107515*. https://doi.org/10.1016/j.neunet.2025.107515

Stigliani, A., Weiner, K. S., & Grill-Spector, K. (2015). Temporal Processing Capacity in High-Level Visual Cortex Is Domain Specific. *Journal of Neuroscience, 35*(36), 12412–12424. https://doi.org/10.1523/JNEUROSCI.4822-14.2015

Taubert, J., Ritchie, J. B., Ungerleider, L. G., & Baker, C. I. (2022). One object, two networks? Assessing the relationship between the face and body-selective regions in the primate visual system. *Brain Structure and Function, 227*(4), 1423–1438. https://doi.org/10.1007/s00429-021-02420-7

Teichmann, L., Hebart, M. N., & Baker, C. I. (2024). Dynamic representation of multidimensional object properties in the human brain. *bioRxiv*. https://doi.org/10.1101/2023.09.08.556679

Tsao, D. Y., Moeller, S., & Freiwald, W. A. (2008). Comparing face patch systems in macaques and humans. *Proceedings of the National Academy of Sciences, 105*(49), 19514–19519. https://doi.org/10.1073/pnas.0809662105

Vinken, K., Prince, J. S., Konkle, T., & Livingstone, M. S. (2023). The neural code for “face cells” is not face-specific. *Science Advances, 9*(35), eadg1736. https://doi.org/10.1126/sciadv.adg1736

Watson, D. M., & Andrews, T. J. (2024). A data-driven analysis of the perceptual and neural responses to natural objects reveals organising principles of human visual cognition. *Journal of Neuroscience, 45*(2). https://doi.org/10.1523/JNEUROSCI.1318-24.2024

Watson, D. M., Andrews, T. J., & Hartley, T. (2017). A data driven approach to understanding the organization of high-level visual cortex. *Scientific Reports, 7*(1), 3596. https://doi.org/10.1038/s41598-017-03974-5

Weiner, K. S., & Grill-Spector, K. (2010). Sparsely-distributed organization of face and limb activations in human ventral temporal cortex. *NeuroImage, 52*(4), 1559–1573. https://doi.org/10.1016/j.neuroimage.2010.04.262

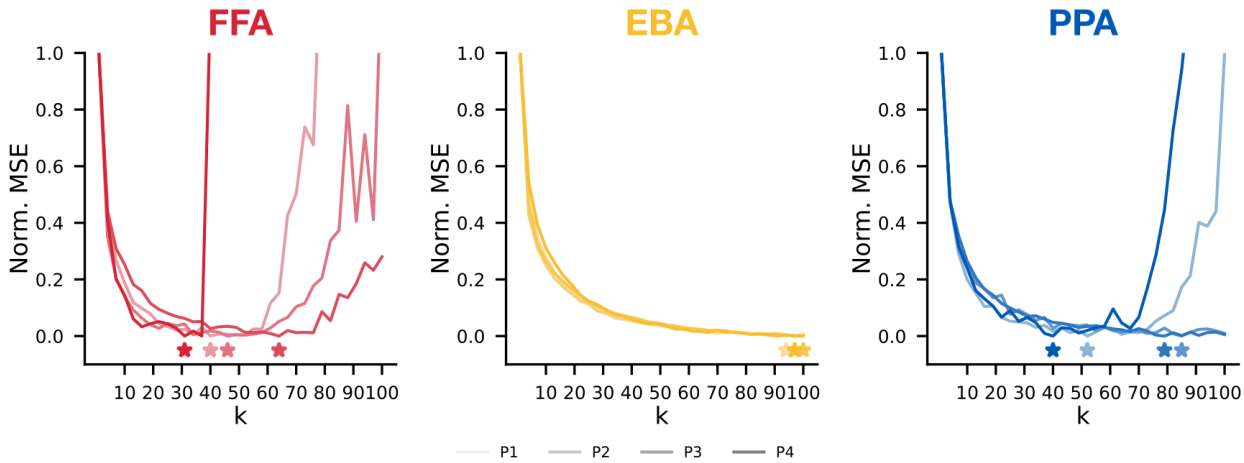
Weiner, K. S., & Grill-Spector, K. (2011). Not one extrastriate body area: Using anatomical landmarks, hMT+, and visual field maps to parcellate limb-selective activations in human lateral occipitotemporal cortex. *NeuroImage, 56*(4), 2183–2199. https://doi.org/10.1016/j.neuroimage.2011.03.041

Willems, R. M., Peelen, M. V., & Hagoort, P. (2010). Cerebral Lateralization of Face-Selective and Body-Selective Visual Areas Depends on Handedness. *Cerebral Cortex, 20*(7), 1719–1725. https://doi.org/10.1093/cercor/bhp234

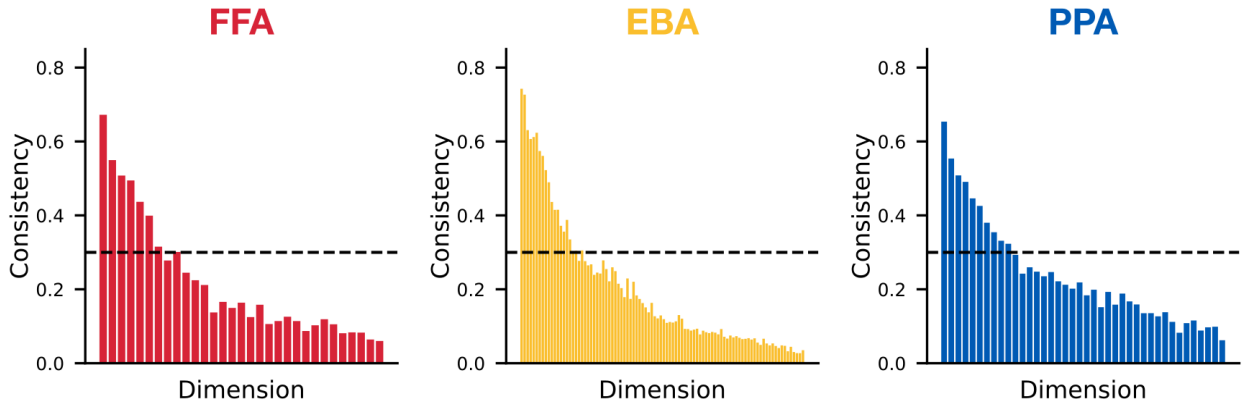
Yargholi, E., & Beeck, H. O. de. (2023). Category Trumps Shape as an Organizational Principle of Object Space in the Human Occipitotemporal Cortex. *Journal of Neuroscience, 43*(16), 2960–2972. https://doi.org/10.1523/JNEUROSCI.2179-22.2023

Supplementary information

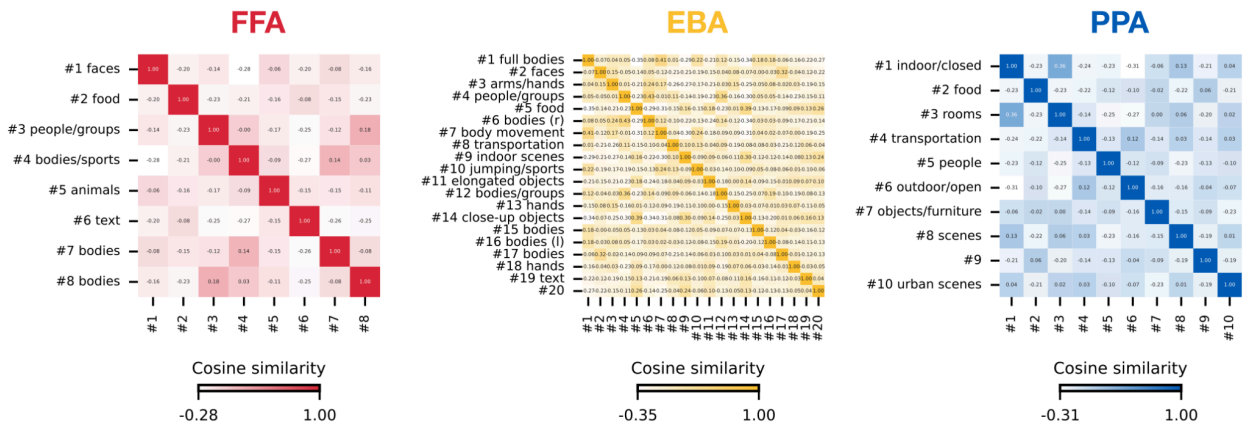
Supplementary figures



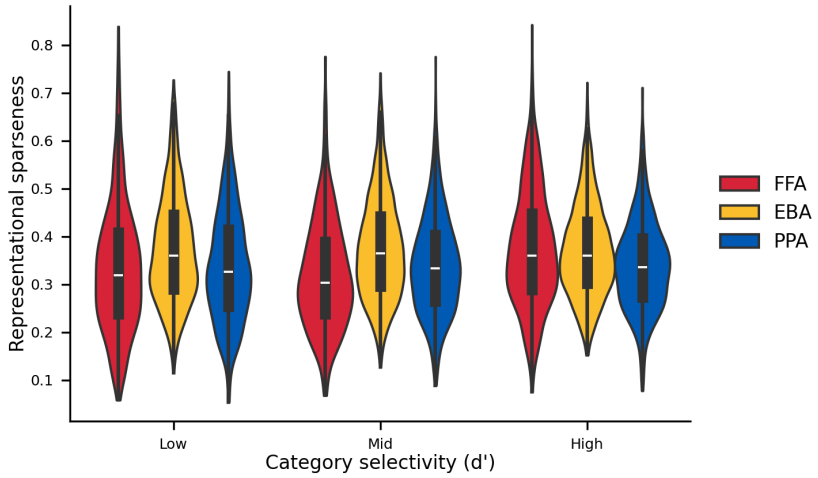
Supplementary Fig. 1 | Optimal dimensionality for each area. Normalized mean squared error (*MSE*) based on bi-cross-validation for each area in each participant. Stars indicate optimal dimensionality for each participant.



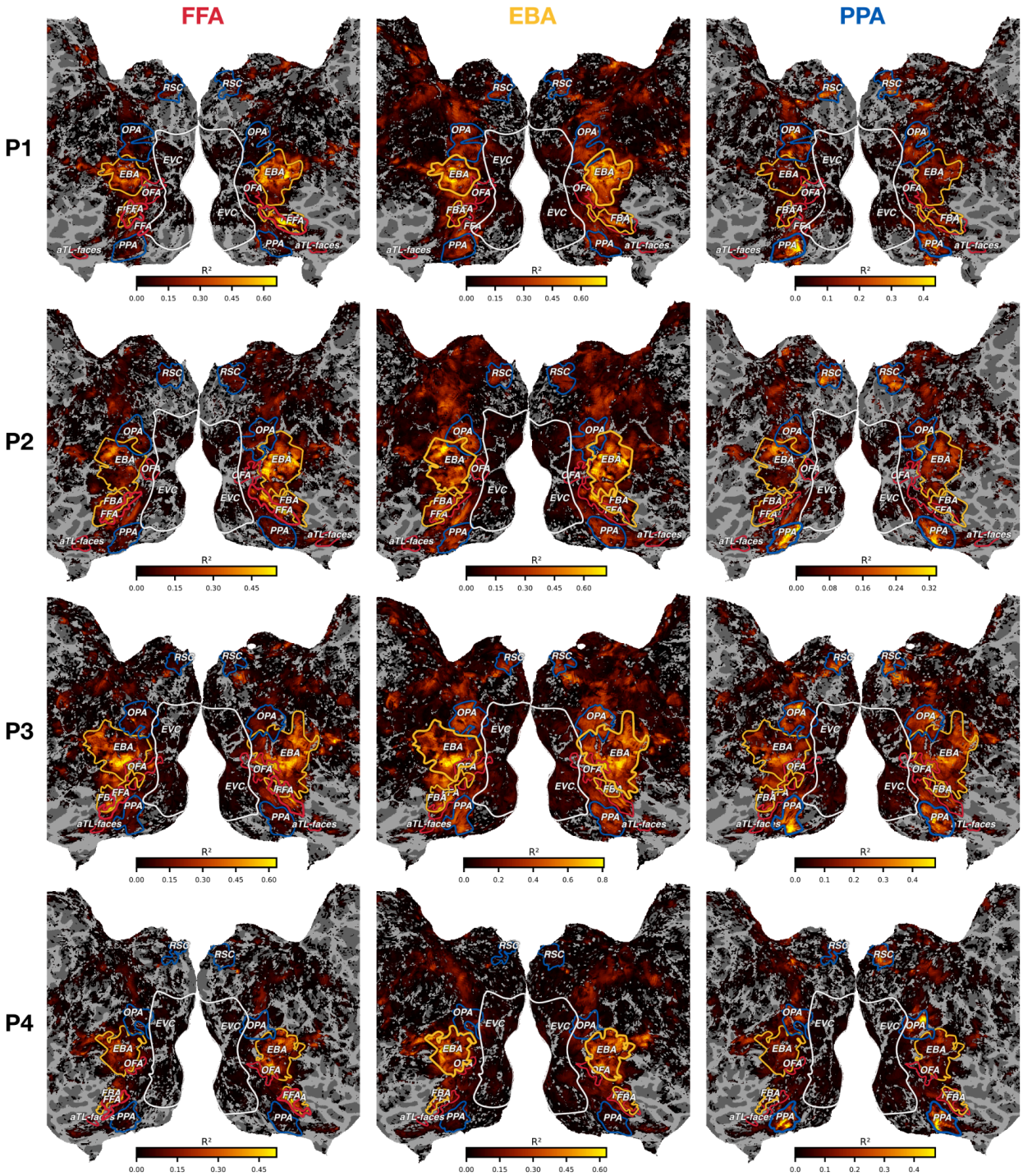
Supplementary Fig. 2 | Consistency of dimensions in each area across participants. Mean correlation of matched dimensions across participant pairs. Dotted line indicates consistency threshold (r = 0.3).



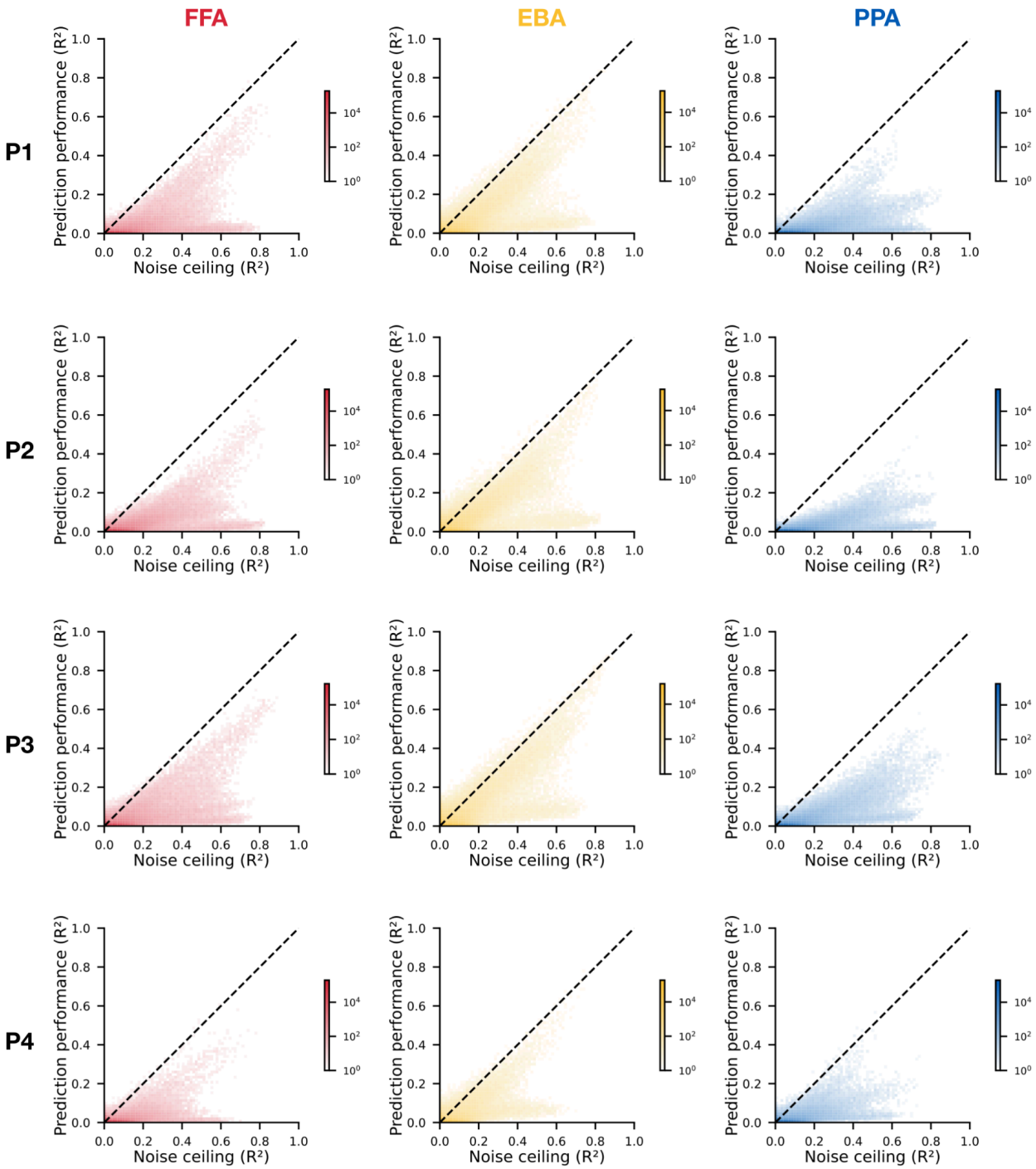
Supplementary Fig. 3 | Similarity of dimensions in each area. Mean-centered pairwise cosine similarities of dimensions in each area averaged across participants.



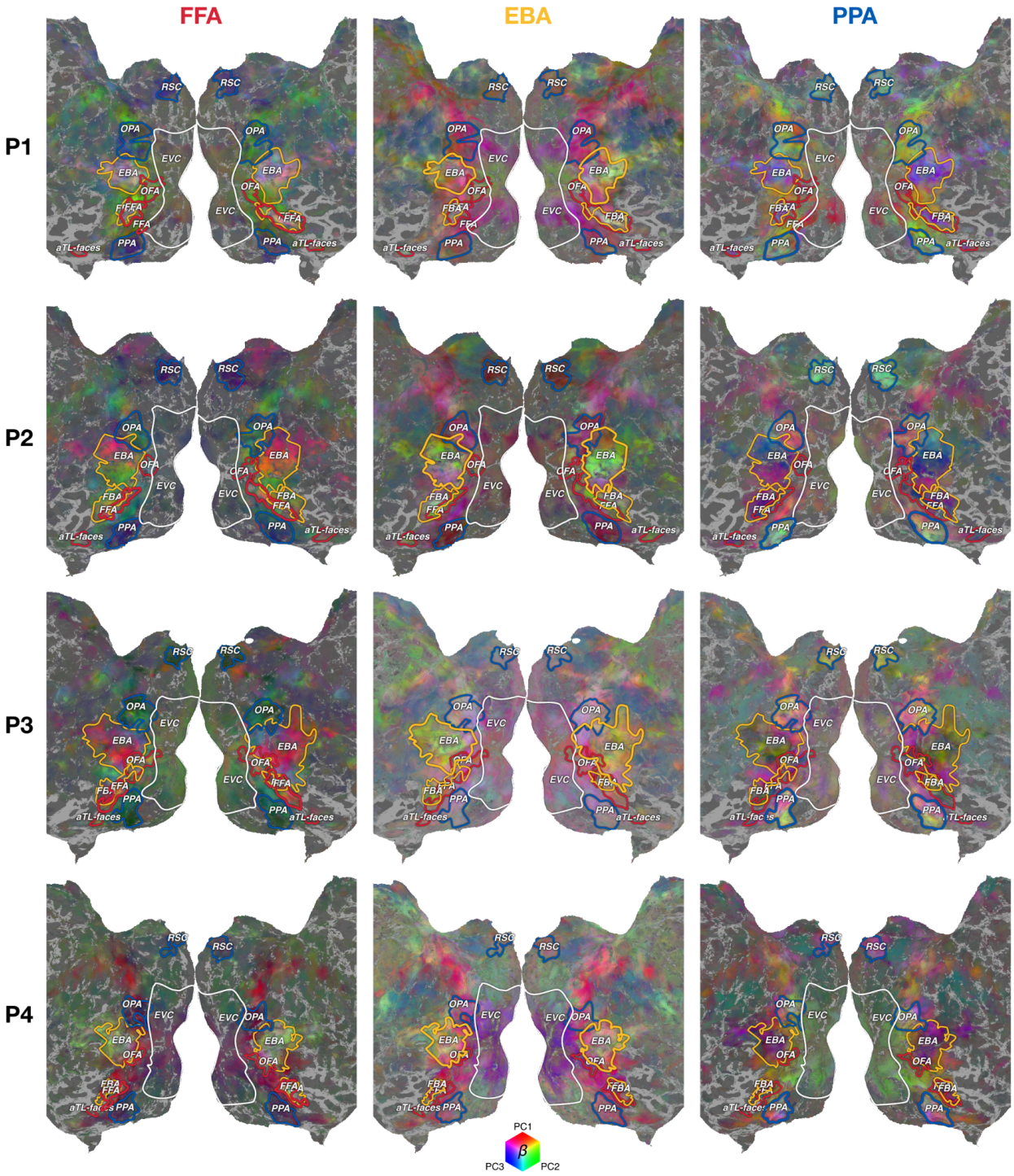
Supplementary Fig. 4 | Voxel-wise representational sparseness across different levels of category selectivity in each area. Voxels within each area were grouped into low, mid, and high category selectivity (*d'*), defined using the corresponding functional localizer contrast (e.g., faces > all other categories in FFA). Representational sparseness was then computed for each voxel, with distributions shown across selectivity bins and areas.



Supplementary Fig. 5 | Prediction performance maps of dimensions from category-selective areas across cortex. Voxel-wise prediction performance (R^2) for dimensions from each area projected onto the flattened cortical surface for all participants. Voxels thresholded at $p < 0.01$ (one-sided, FDR-corrected).



Supplementary Fig. 6 | Noise ceiling estimate versus voxel-wise prediction performance across cortex. Relationship between noise ceiling estimate and prediction performance (R^2) for dimensions from each area, evaluated on held-out test images. Note that both metrics exhibit variability, particularly in voxels with lower signal-to-noise ratio, likely due to the limited size of the test set (30 % of all images).



Supplementary Fig. 7 | Functional tuning gradients based on all dimensions. First three principal components (PC1-3; red, green, and blue channels) of voxel-wise coefficient weights across all dimensions projected onto the flattened cortical surface for all participants. Voxels thresholded at $p < 0.01$ (one-sided, FDR-corrected).

Supplementary methods

Supplementary Method 1 | GPT-4V prompt. Detailed prompt for the generation of candidate labels based on the highest-scoring images with GPT-4V:

Analyze the following collage of images and identify 10 short labels that describe the most prominent, recurring elements observable across all or nearly all images.
Each label should:
- Focus solely on common visual features (e.g., objects, categories, scenes).
- Consist of up to three specific and concrete words (preferably less; preferably nouns and verbs).
- Match the sentence: 'The images show {label}.'