A MODEL-DRIVEN FUNCTIONAL ATLAS OF HUMAN VISUAL CORTEX UNDER NATURALISTIC VISION

Anonymous authors
Paper under double-blind review

ABSTRACT

Visual preferences overlap in cortex and recur across objects and scenes. Describing their combinations is a central challenge for a functional atlas of natural vision. We construct a fine-grained atlas by relating sparse, interpretable Transformer features to human fMRI responses. Individual functional modes link image properties to cortical coefficient patterns, resolving distinctions within familiar category-associated regions. Continuous coordinates reveal regional mixtures in which a locally enriched mode contributes only part of the overall profile. Response-matched families show corresponding cortical patterns across the observed participants and independently learned model dictionaries, providing recurrent atlas elements. Condition subatlases organize these elements by both relative preference and common participation: features shared across image categories are retained alongside those that distinguish them. Model responses to individual images select smaller subsets of these repertoires. Together, the results describe visual cortex through overlapping functional patterns whose combinations connect a recurrent atlas to the diversity of natural images.

1 INTRODUCTION

A person in a room and a person outdoors share visual features, yet differ in surfaces, objects, and spatial arrangement. Human visual cortex reflects both kinds of structure. Face-, body-, and scene-associated regions provide reproducible landmarks, while responses within and beyond these regions carry distributed information about image content [9, 10, 19, 22, 32]. An atlas of natural vision therefore faces two connected questions: which finer visual preferences organize cortical responses, and how are those preferences combined as the image changes? This calls for a description in terms of reusable features within and across regional boundaries.

Naturalistic encoding and representational analyses have made this finer organization accessible. Image features predict responses to novel stimuli, semantic dimensions extend across cortical boundaries, and data-driven decompositions reveal selectivity beyond conventional category contrasts [5, 21, 24, 25, 31]. Interpretable visual models further connect particular features to the images and image regions that drive them [17, 27, 29, 33]. Meanwhile, alignment across independently trained models provides a means to examine whether a feature is specific to one representation or appears in others [12, 35]. These developments support a shift from interpreting isolated features to asking whether recurrent features can furnish the elements of a cortical atlas.

The central problem is to connect recurrence with composition. Similar image preferences may be represented by different model coordinates and mapped to somewhat different cortical patterns across people. A common atlas needs to retain these variations while identifying corresponding functional patterns. Its use also depends on what counts as a component of an image condition. A preference contrast identifies patterns more prevalent for that condition than for its background. A feature shared by many conditions may contribute little to this contrast despite appearing regularly in the condition itself. Distinguishing relative preference from common participation provides a way to describe both the specific and the shared constituents of natural vision.

We develop this account using sparse cross-layer Transformer features and Natural Scenes Dataset (NSD) fMRI [1]. A functional mode (FM) links a model feature to its participant-specific cortical encoding pattern; corresponding modes form a recurrent family, and families associated with an image condition form a subatlas. Fine granularity arises from distinctions in appearance, configuration,

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

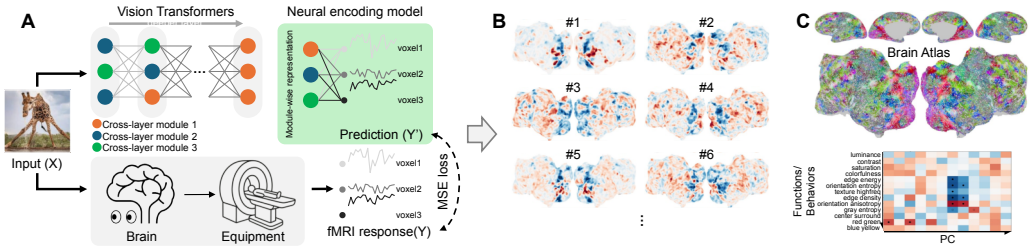


Figure 1: An atlas linking visual features to cortical organization. **A**, Cross-layer Transformer features define sparse FMs whose responses predict fMRI. **B**, Individual coefficients form distinct, overlapping cortical patterns. **C**, Principal components provide continuous coordinates and image-property associations. Scalar coefficient maps use red for positive and blue for negative weights; RGB combines three component scores.

and content, together with their overlapping cortical distributions. We follow these units from individual preferences to recurrence and condition composition (Figure 1). The results yield three connected findings: familiar regional preferences contain diverse feature mixtures; recurrent families provide common atlas elements; and condition subatlases combine shared participation with relative preference, while model responses to individual inputs select only part of each repertoire. The atlas connects interpretable visual representations to cortical organization across people, models, and image conditions.

2 Study Design and Atlas Construction

2.1 Cross-layer visual representations

We use four frozen vision Transformers: ViT-Small, ViT-Base-21k, ViT-Large, and DeiT-Base [8, 36]. Class-token features from successive blocks are concatenated so that a dictionary direction can span multiple processing stages. Pretrained TopK sparse autoencoders provide a separate dictionary for each architecture. Each dictionary contains 256 directions learned from ImageNet-derived features without neural supervision. Top-eight selection followed by rectification produces at most eight positive responses per image, implementing an explicit sparsity constraint [15, 28]. Their responses supply the model-side coordinates of the FMs. Feature extraction, preprocessing, the inherited training protocol, and the response equations are specified in Appendix B.

2.2 Neural encoding and functional interpretation

For NSD participants S1, S2, S5, and S7 [1], run-standardized visual features are projected onto the ImageNet-trained dictionaries. Scaling uses all image trials; neural encoding uses 8,000 training, 1,000 validation, and 1,000 common held-out images. Separate linear models predict each participant's fMRI from each model's FM responses [24, 31], evaluated by R^2, Pearson correlation, and image identification. Each coefficient column defines a cortical pattern conditional on the other features. Source-domain dictionary comparisons and repeated-training results appear in Appendix B.5.

Interpretation combines top-response images, local image occlusion, Qwen annotations, and regional coefficient profiles [2, 14, 39]. Regional profiles relate image preferences to established visual areas; coefficient signs are interpreted within the multivariate encoder [18]. Appendix B gives the encoding and evaluation equations, noise-ceiling units, and Qwen procedure; Appendix G contains the complete example interpretations.

2.3 Continuous organization and the recurrent atlas

Within a participant, activation-scaled coefficient profiles provide continuous cortical coordinates through PCA. Across participants and models, similarity of responses to the common images identifies candidate FM correspondences; graph communities group these into families [4]. We assess their spatial recurrence separately on a shared cortical surface [13], comparing omitted-participant and omitted-model patterns against matched nulls. A family prototype averages member patterns with

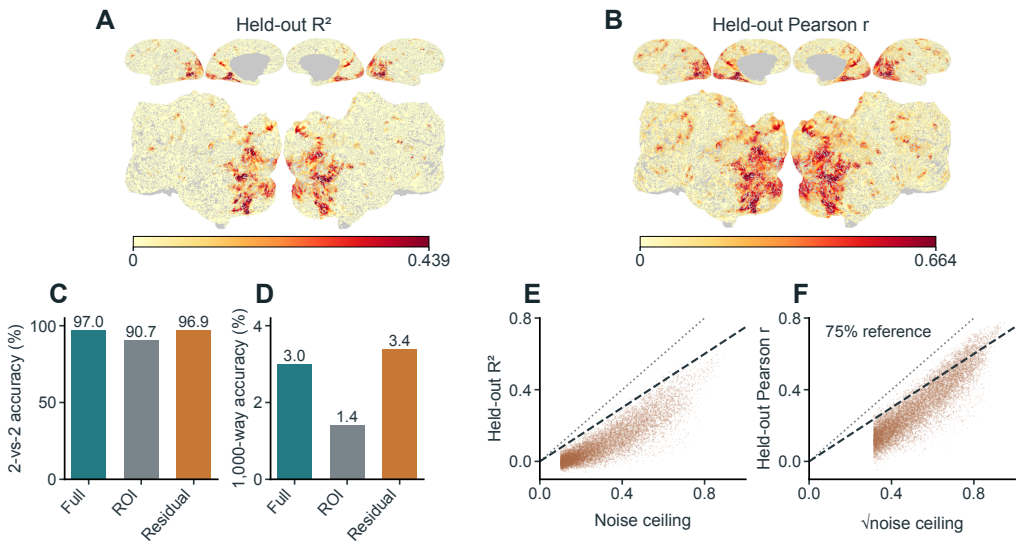


Figure 2: Encoding performance and stimulus information. The frozen S5/base encoder is restricted to 187 train-and-test-active FMs without refitting. **A, B**, Held-out R^2 and Pearson r, displayed from zero to their 99.5th percentiles (0.439, 0.664). **C, D**, Two-versus-two identification and 1,000-way retrieval for full spatial patterns, the ROI-mean subspace, and its residual. **E, F**, Performance against the metric-specific ceiling; dashed lines mark 75%, dotted lines equality. C–F use 12,550 locations with NC ≥ 0.10, giving horizontal left boundaries 0.10 and $\sqrt{0.10} \approx 0.316$. Definitions and readouts: Appendix B, Appendix C.

equal weight for each participant–model combination. For an image, a family is called within a participant–model combination when any member has a positive FM response. Balanced calls across combinations support two complementary descriptions of an image condition: relative preference over other images and common participation within the condition. Their union defines a condition subatlas drawn from the spatially supported families. Detailed definitions, selection criteria, and sensitivity analyses are in Appendix B and Appendix F.

3 Results

3.1 Sparse modes capture neural information beyond regional averages

S5/ViT-Base readouts retain the 187 FMs active in both training and held-out model responses, using frozen coefficients and bias without refitting. Prediction quality varies across 45,280 cortical locations (Figure 2A,B). Some locations approach the 75% reference for R^2 relative to its noise ceiling and Pearson r relative to the square-root ceiling (E,F). Image-level readouts use 12,550 locations with ceiling at least 0.10.

Two-versus-two image identification reaches 96.97%, compared with a label-randomization median of 50.11%. The ROI-mean subspace yields 90.68%, while its spatial residual retains 96.94%: image information extends beyond regional averages. Conversely, matched predicted–measured pattern correlation is higher in the ROI-mean subspace (0.740) than in full or residual patterns (0.363, 0.310). Agreement in a coarse regional profile and discrimination among images therefore measure different aspects of encoding. The residual preserves both positive pattern correlation and image-specific distinctions, supporting a finer functional description within and across regional boundaries. Retrieval and other readouts are reported in Appendix C.

3.2 Interpretable modes distinguish visual preferences within regions

The eight examples in Figure 3 relate image content and local model sensitivity to established visual areas. FM075 favors formally dressed people: occlusion emphasizes faces and necklines, and the strongest positive regional coefficients lie in the two fusiform face ROIs. FM080 instead favors canid heads, combining eye/muzzle sensitivity with positive fusiform coefficients [22, 32]. FM162 favors

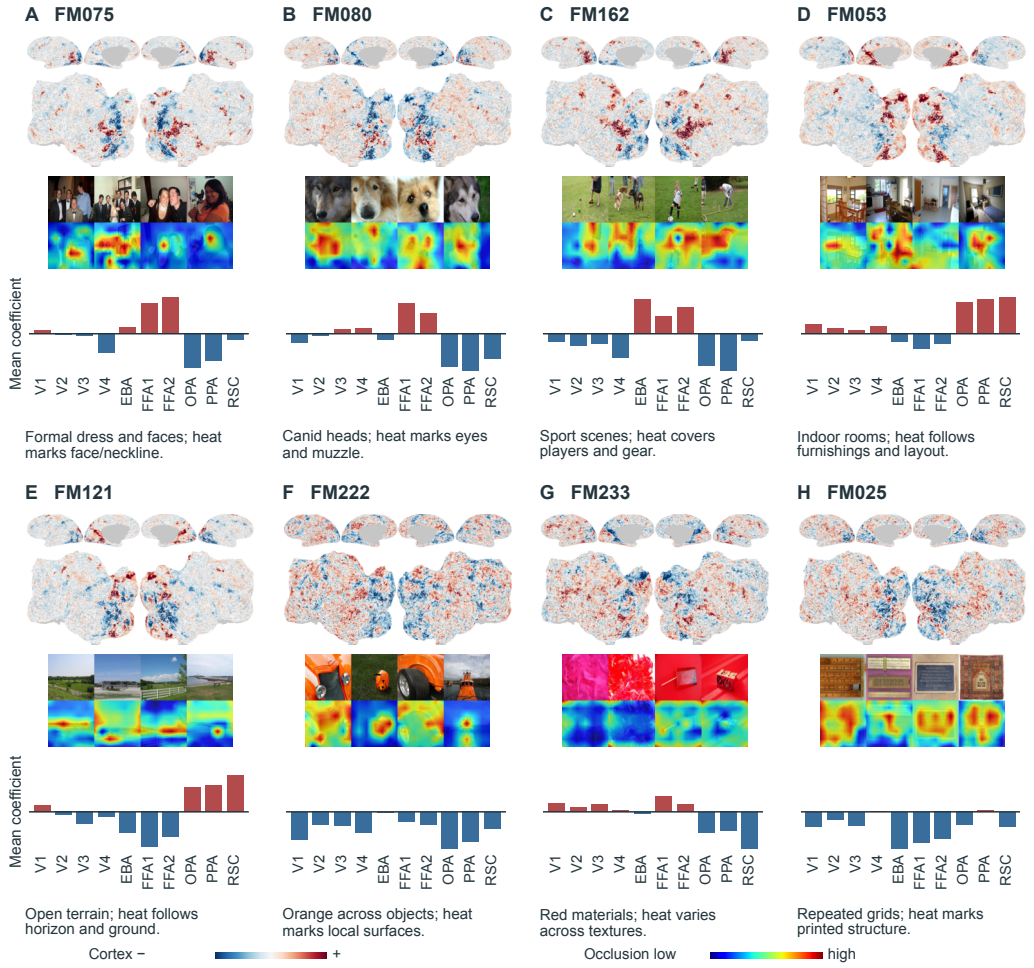


Figure 3: Image evidence and cortical profiles of eight modes. A–H, FM075, 080, 162, 053, 121, 222, 233, and 025. Each example shows its signed cortical coefficient pattern, four top-response originals above their matched model occlusion maps, ten visual-ROI means, and a brief Qwen-based interpretation. Regional bars have aligned zeros and independent symmetric ranges; the displayed quantity is coefficient $\times 10^3$. Red/blue cortical colors denote positive/negative weights. The separate occlusion key denotes increasing original-minus-occluded model response. Full Qwen descriptions and regional values: Appendix G.

people and equipment in sport scenes and has positive EBA coefficients, consistent with the visible body configurations in these static images [9].

FM053 favors furnished interiors, whereas FM121 favors open terrain and horizons. Both have positive OPA, PPA, and RSC coefficients, relating indoor arrangement and open-scene structure to a shared scene-associated profile [7, 10, 11]. Other modes distinguish properties across categories: FM222 emphasizes orange surfaces, FM233 red materials, and FM025 repeated markings, with weaker or more mixed regional profiles. These preferences complement familiar category-associated organization and reveal feature heterogeneity within regional boundaries [5, 25]. Brief Qwen-based descriptions accompany the examples; full descriptions, the remaining examples, and complete visual-ROI and 17-network profiles are in Appendix G [38].

3.3 Continuous feature coordinates reveal regional mixtures

PCA of activation-scaled, normalized 187-FM coefficient profiles across S5/base cortical locations yields continuous coordinates (Figure 4). The first three components explain 20.72% of variation; the first 12 explain 41.20%. The RGB composite combines the first three scores at each location, representing the local mixture of modes.

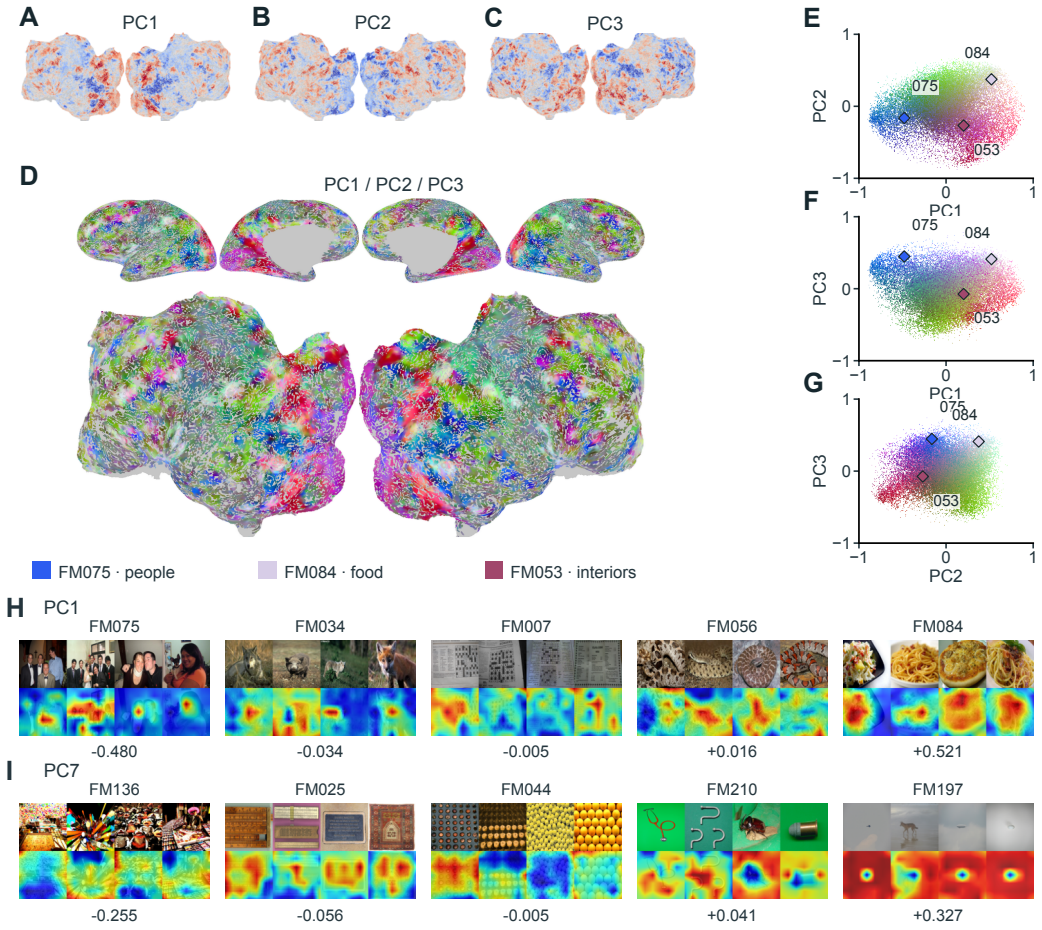


Figure 4: Continuous cortical coordinates with interpretable feature landmarks. A–C, PC1–3 scalar fields. **D**, Their RGB composite. **E–G**, Pairwise projections of 45,280 cortical locations, colored by all three scores. Diamonds locate the projected unit directions for FM075, 084, and 053. **H, I**, Five ordered FM landmarks on PC1 and PC7, with four original images above matched model occlusion maps. The remaining component sequences appear in Appendix D; the 187-FM color-coordinate key is in Appendix E.

Projected unit-FM directions connect cortical colors to interpretable preferences. People-related FM075, food-related FM084, and interior-related FM053 occupy distinct reference positions within this mixture space (Appendix E). Ordered landmarks range from people-related images to food arrangements on PC1, and from dense repeated displays to isolated objects and pale fields on PC7.

Quantitative associations span both image content and appearance (Figure 5): PC1 relates to person and dining preferences, while PC7 relates to edge and orientation properties. Of 1,200 PC–descriptor tests, 162 pass global BH correction and 113 remain after activation-frequency adjustment. The image-property associations pass BH but not the stricter maximum-statistic criterion (Appendix D).

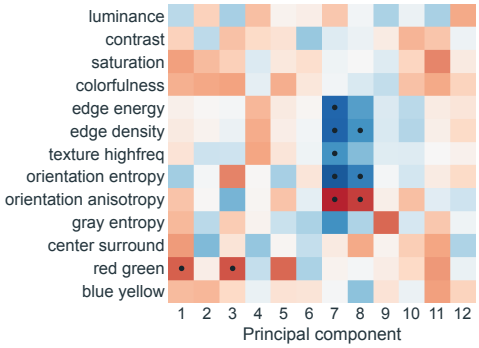


Figure 5: Visual associations across PCs. Spearman correlations with 13 image properties. Dots pass global BH correction before and after frequency adjustment; red/blue denote positive/negative correlation (± 0.35). All 1,200 comparisons: Appendix D.

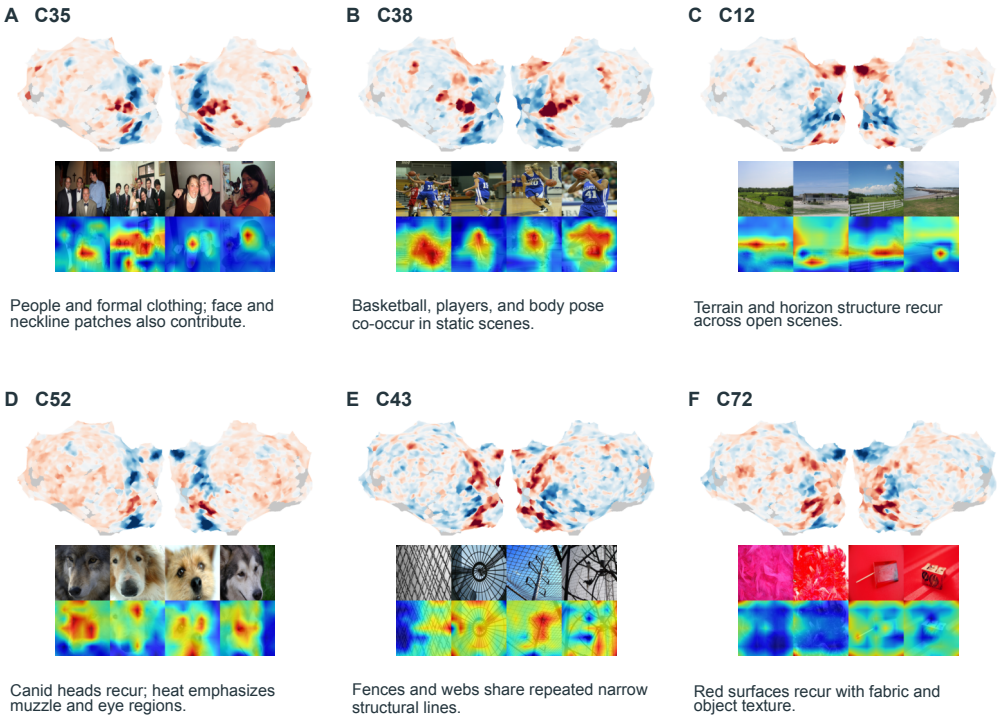


Figure 6: Recurrent families include category-associated and appearance-related patterns. A–F, C35, C38, C12, C52, C43, and C72. Cortical panels show signed family prototypes with equal weight for all participant–model combinations. Top4 originals, model occlusions, and brief interpretations describe the respective S5/base representatives FM075, 099, 121, 080, 104, and 233. All six families span four participants and four models and belong to the primary spatial core. Surface interpolation is used for display; spatial comparisons use the original common-location coefficients. Full family descriptions: Appendix H.

Axis combinations also capture complementary variation: 25 of 30 selected combinations retain above-null functional neighborhoods on the opposite image half after correction. Full multi-axis results and the activation-frequency dependence of PC9 are in Appendix D.

Regional enrichment coexists with broad mixtures (Table 1). FM197 is enriched 4.10-fold in V1, FM075 3.90-fold in FFA-2, and FM209 4.34-fold in PPA. Yet FM197 accounts for only 3.10% of V1's activation-scaled absolute coefficient mass. Effective mode counts across ROIs range from 149.8 to 168.1, and the largest individual shares from 2.1% to 5.3%. Thus concentration within a region and dominance of that region are different properties. Complementary fixed ViT-small coordinates correlate positively with held-out target-area fMRI in all four participants (Appendix F).

ROI · FM	Enrichment	Mass (%)
V1 · 197	4.10×	3.10
FFA 2 · 075	3.90×	5.31
PPA · 209	4.34×	3.58
RSC · 053	3.82×	3.36

Table 1: Regional mixtures. Enrichment and within-ROI mass use absolute coefficients weighted by training response standard deviation.

3.4 Recurrent families form a shared cortical dictionary

We next ask which modes recur across participants and models. Correspondence is defined by responses to the common 1,000 images: mutually closest fingerprints exceeding the cell-pair's 99th-percentile similarity threshold form a graph, and weighted community detection groups connected modes [4]. Families spanning all 16 participant–model combinations are then evaluated for cortical agreement. This separates response-based family discovery from the assessment of fitted spatial patterns.

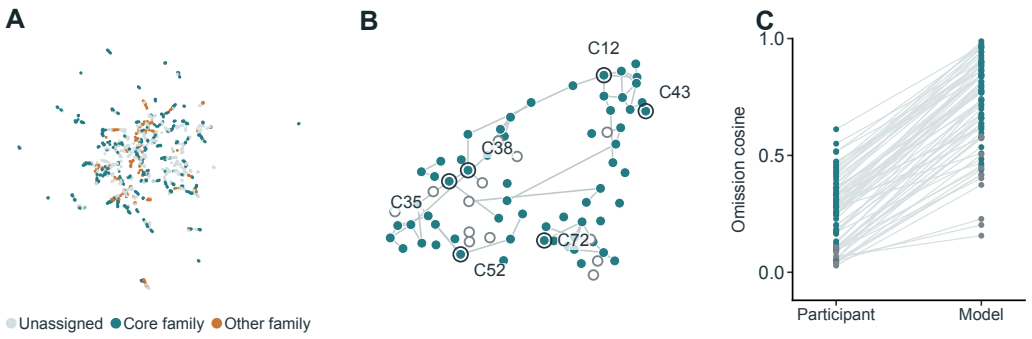


Figure 7: Geometry and recurrence at two atlas scales. A, The 2,775 individual maps in a stimulus-guided CEBRA embedding displayed through UMAP, colored by family support. **B**, UMAP of the 75 family cortical-plus-response profiles; edges are mutual neighbors in the original joint space. Labels identify the examples in Figure 6. **C**, Paired participant- and model-omission cortical similarities, with primary-core families in teal. A and B use different units and representations; family identities precede both displays. Complete comparisons: Appendix E.1, Appendix F.

Seventy-five families span all 16 combinations and contain 1,752 of the 2,775 eligible maps; 1,023 maps fall outside this complete-coverage dictionary. Family prototypes average signed, normalized cortical patterns with equal weight for each combination (Figure 6). Related modes can therefore share a prototype while retaining different feature indices and participant-specific spatial distributions.

With response-defined memberships fixed, model omission yields median cortical cosine 0.769, compared with 0.052 under the matched null; participant omission yields 0.279, compared with 0.026. All 75 families pass model omission, and 62 also pass participant omission after correction across families, forming the primary spatial core. Requiring every omitted group to exceed its own null 95th percentile retains 66 and 49 families, respectively. An activation-frequency-matched null supports 60, with 59 common to both family-level criteria. These tests support spatial correspondence within the observed groups; agreement is stronger across models than across participants (Appendix F).

The recurring patterns retain interpretable visual distinctions. C35 emphasizes formally dressed people, C52 canid heads, C38 players and body configurations, and C12 open landscapes. Their S5/base representatives have positive fusiform coefficients for C35/C52, EBA coefficients for C38, and scene-region coefficients for C12. Jointly, the top images, model occlusion patterns, and representative regional profiles connect these families to established visual organization [9, 10, 22].

C43 and C72 extend the dictionary to visual appearance: the former's representative favors fences, webs, and repeated narrow lines; the latter favors red materials with varied surface structure. All six illustrated families belong to the primary spatial core. The cortical panel shows each family prototype, while the images and Qwen interpretation characterize its S5/base representative. The complete dictionary in Appendix H retains all 75 families and their differing degrees of spatial support, with member-level comparisons in Appendix F.

3.5 Recurrent modes compose overlapping condition subatlases

A common geometry for recurrent patterns. The family dictionary provides a common space in which to examine relationships among modes. At the individual-map scale, stimulus-guided CEBRA followed by UMAP displays all 2,775 cortical maps, including those outside the complete-coverage dictionary (Figure 7A) [30, 34]. At the family scale, equal-weight cortical and image-response profiles relate the 75 recurrent patterns (B). Their original-space mutual-neighbor graph contains local neighborhoods and cycles; most edges recur under participant or model omission. UMAP preserves these local relationships with five-neighbor trustworthiness 0.931. The full graph measurements and comparison of 23 projection configurations are in Appendix E.1.

Visual properties distinguish atlas elements. Continuous image properties complement the category preferences of the recurrent families. C26 associates with edge energy, high-frequency

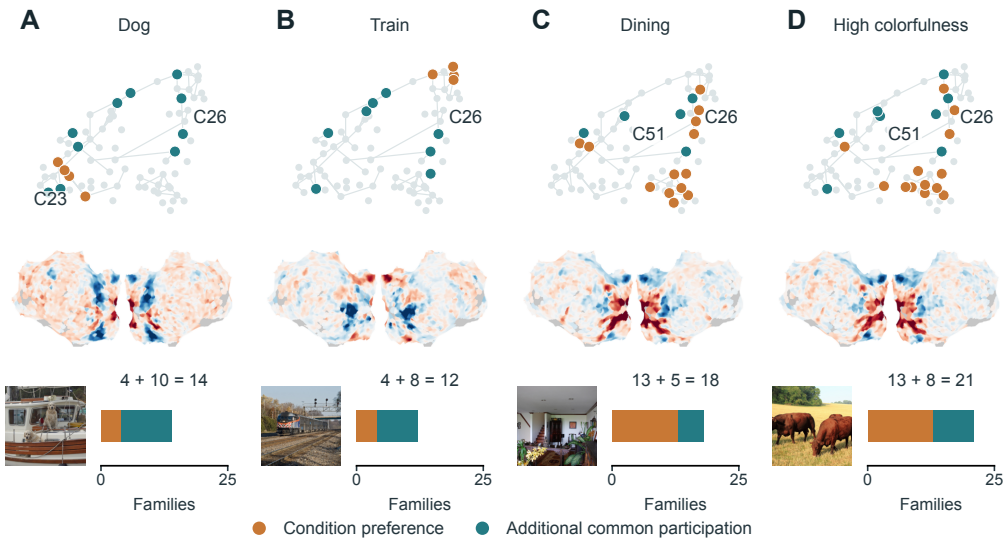


Figure 8: Overlapping condition subatlases in a fixed family dictionary. A–D, Dog, train, dining, and high-colorfulness conditions. Orange marks relatively preferred families and teal additional commonly participating families. Each graph uses the same coordinates; below it, selected fixed prototypes are combined by condition call rates to give a derived cortical composition. An example image and membership counts identify the condition. These spatial panels summarize the atlas composition, not measured condition-average fMRI. Full membership for all 82 conditions: Appendix I.

texture, and colorfulness, whereas C51 associates with luminance. Across the dictionary, 64 families have supported image-property associations, and 63 combine these with semantic preferences. Appearance and recognizable content thus provide overlapping descriptions of the atlas elements. Complete effect sizes, cross-model sign agreement, and multivariable adjustments are reported in Appendix E.2.

Preference and participation define complementary constituents. A condition subatlas collects families from this dictionary while preserving their cortical identities. Relative preference selects families called more often for the condition than for other images. Common participation also includes families regularly called within the condition across participants and models, even when frequent elsewhere. Combining the two criteria expands the dog repertoire from four preferred families to 14, the train repertoire from four to 12, and the dining repertoire from 13 to 18 (Figure 8). Across 82 conditions, the resulting subatlases contain 9–27 families. Their shared elements become visible because a family's position and cortical prototype stay fixed as condition membership changes.

The added constituents change the interpretation of these repertoires. Of 589 additional condition–family relations, 272 have a nonpositive condition-minus-background difference: they are common within the condition but contribute no positive preference contrast. A preference-only description would omit them. For dogs, combining preference and participation increases coverage of observed spatial-core family calls from 15.8% to 62.6%. The cortical panels in Figure 8 summarize these compositions by combining selected prototypes with condition call-rate weights. They represent the constituent atlas patterns, rather than measured condition-average fMRI. The complete memberships and maps are in Appendix I. Re-evaluating common participation after participant or model omission, with preference memberships fixed, largely preserves the subatlases (median Jaccard 0.941; Appendix F).

FM responses select subsets of condition repertoires. A category's repertoire describes variation across its images. Within an individual image and participant–model combination, the median number of repertoire families called is three for dogs and four for trains and dining. The illustrated dog and train images call five of 14 and seven of 12 families, respectively, when pooled across all combinations (Figure 9). Thus a shared dictionary supports both a condition-level description and a more selective input-specific composition. Image-level distributions, recurring constituents, and exploratory co-recruitment are detailed in Appendix E.3.

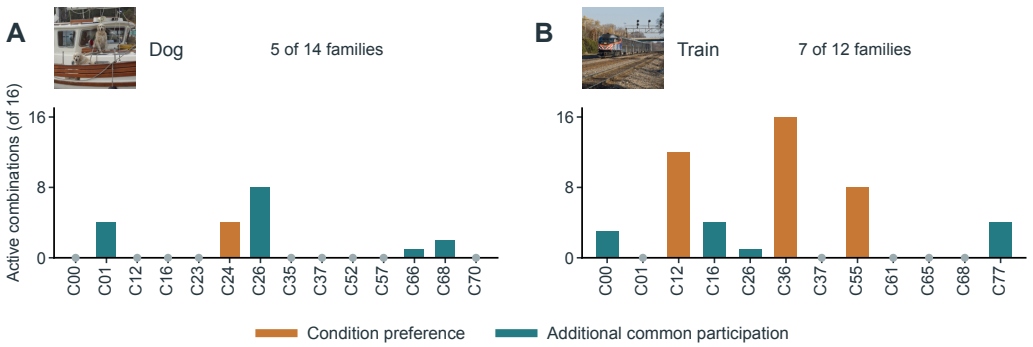


Figure 9: FM responses select subsets of condition repertoires. A, B, Dog and train exemplars. Bars count the participant–model combinations in which each family is called for that image (out of 16); families with zero calls remain part of the condition repertoire. Orange denotes relative preference and teal additional common participation. Five of 14 dog families and seven of 12 train families are called in at least one combination. Full combination-specific calls and co-recruitment: Appendix E.3.

The image preferences also show consistency across stimulus collections. Family effects correlate positively between NSD and ImageNet-derived images for all 15 eligible category correspondences; 83 of 96 NSD-selected relations retain a positive effect. This comparison extends the model-side image evidence while the cortical patterns remain those estimated from NSD. Category-specific effects and uncertainty are in Appendix E.4.

4 Discussion and Conclusion

The atlas describes visual cortex through overlapping modes that recur across individuals and models. Familiar category-associated regions remain informative, but their coefficient patterns contain many modes with different image preferences. Continuous coordinates relate those preferences, while the recurrent dictionary supplies common units for comparing natural-image conditions. Regional enrichment and broad mixtures are therefore compatible features of the same organization.

The distinction between preference and participation changes what a functional subatlas contains. Relative preference isolates patterns enriched for a condition; common participation recovers patterns that the condition shares with other inputs. Their union describes a repertoire, and model responses to individual images select smaller, variable subsets. This compositional view connects an integrated cortical atlas to condition-specific organization without requiring a separate atlas for every image class. Local neighborhoods and cycles further describe relationships among modes, while leaving processing order to experimental investigation. The atlas distinguishes two kinds of relationship: nearby families have similar cortical and image-response profiles, whereas families in the same subatlas are associated with a common image condition. These criteria need not select the same pairs. A condition can combine distant patterns, and nearby patterns can participate in different conditions.

For computational neuroscience, these units connect interpretable image preferences to heterogeneous regional patterns. For representation learning, recurrence across independently learned dictionaries supplies a common neural reference: model-local feature indices differ, yet corresponding image preferences and cortical patterns emerge. Retaining member identities within each family makes this agreement and its variation accessible at the level of individual models and participants.

Limitations and scope. The atlas covers four Transformer models and four NSD participants. Its coordinates depend on the transferred dictionary and conditional encoding coefficients; target-domain reconstruction fidelity and map stability across refits remain unmeasured. Recurrence is evaluated within response-defined families, and subatlases describe their image-conditioned composition. CNNs, MLP-based models, and independent neural datasets will extend architectural and population coverage. Causal necessity and processing order require experimental intervention.

AI USE STATEMENT

LLMs were used to improve the writing quality of this manuscript, including grammar correction and language polishing. The authors reviewed and take full responsibility for all content in the paper.

REFERENCES

[1] Emily J. Allen, Ghislain St-Yves, Yihan Wu, Jesse L. Breedlove, Jacob S. Prince, Logan T. Dowdle, Matthias Nau, Brad Caron, Franco Pestilli, Ian Charest, J. Benjamin Hutchinson, Thomas Naselaris, and Kendrick N. Kay. A massive 7t fMRI dataset to bridge cognitive neuroscience and artificial intelligence. *Nature Neuroscience*, 25(1):116–126, 2022. doi: 10.1038/s41593-021-00962-x.

[2] Shuai Bai, Yuxuan Cai, Ruizhe Chen, Keqin Chen, Xionghui Chen, Zesen Cheng, Lianghao Deng, Wei Ding, Chang Gao, Chunjiang Ge, Wenbin Ge, Zhifang Guo, Qidong Huang, Jie Huang, Fei Huang, Binyuan Hui, Shutong Jiang, Zhaohai Li, Mingsheng Li, Mei Li, Kaixin Li, Zicheng Lin, Junyang Lin, Xuejing Liu, Jiawei Liu, Chenglong Liu, Yang Liu, Dayiheng Liu, Shixuan Liu, Dunjie Lu, Ruilin Luo, Chenxu Lv, Rui Men, Lingchen Meng, Xuancheng Ren, Xingzhang Ren, Sibo Song, Yuchong Sun, Jun Tang, Jianhong Tu, Jianqiang Wan, Peng Wang, Pengfei Wang, Qiuyue Wang, Yuxuan Wang, Tianbao Xie, Yiheng Xu, Haiyang Xu, Jin Xu, Zhibo Yang, Mingkun Yang, Jianxin Yang, An Yang, Bowen Yu, Fei Zhang, Hang Zhang, Xi Zhang, Bo Zheng, Humen Zhong, Jingren Zhou, Fan Zhou, Jing Zhou, Yuanzhi Zhu, and Ke Zhu. Qwen3-vl technical report. arXiv:2511.21631, 2025. URL https://arxiv.org/abs/2511.21631. Preprint.

[3] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300, 1995. doi: 10.1111/j.2517-6161.1995.tb02031.x.

[4] Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, 2008. doi: 10.1088/1742-5468/2008/10/P10008.

[5] Oliver Contier, Chris I. Baker, and Martin N. Hebart. Distributed representations of behaviour-derived object dimensions in the human visual system. *Nature Human Behaviour*, 8:2179–2193, 2024. doi: 10.1038/s41562-024-01980-y.

[6] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255. IEEE, 2009. doi: 10.1109/CVPR.2009.5206848.

[7] Daniel D. Dilks, Joshua B. Julian, Alexander M. Paunov, and Nancy Kanwisher. The occipital place area is causally and selectively involved in scene perception. *Journal of Neuroscience*, 33 (4):1331–1336, 2013. doi: 10.1523/JNEUROSCI.4081-12.2013.

[8] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=YicbFdNTTy.

[9] Paul E. Downing, Yuhong Jiang, Miles Shuman, and Nancy Kanwisher. A cortical area selective for visual processing of the human body. *Science*, 293(5539):2470–2473, 2001. doi: 10.1126/science.1063414.

[10] Russell Epstein and Nancy Kanwisher. A cortical representation of the local visual environment. *Nature*, 392(6676):598–601, 1998. doi: 10.1038/33402.

[11] Russell A. Epstein, Whitney E. Parker, and Alana M. Feiler. Where am i now? distinct roles for parahippocampal and retrosplenial cortices in place recognition. *Journal of Neuroscience*, 27 (23):6141–6149, 2007. doi: 10.1523/JNEUROSCI.0799-07.2007.

[12] Thomas Fel, Ekdeep Singh Lubana, Jacob S. Prince, Matthew Kowal, Victor Boutin, Isabel Papadimitriou, Binxu Wang, Martin Wattenberg, Demba E. Ba, and Talia Konkle. Archetypal SAE: Adaptive and stable dictionary learning for concept extraction in large vision models. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 16543–16572. PMLR, 2025. URL https://proceedings.mlr.press/v267/fel25a.html.

[13] Bruce Fischl, Martin I. Sereno, Roger B.H. Tootell, and Anders M. Dale. High-resolution intersubject averaging and a coordinate system for the cortical surface. *Human Brain Mapping*, 8(4):272–284, 1999. doi: 10.1002/(sici)1097-0193(1999)8:4<272::aid-hbm10>3.0.co;2-4.

[14] Ruth C. Fong and Andrea Vedaldi. Interpretable explanations of black boxes by meaningful perturbation. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 3449–3457. IEEE, 2017. doi: 10.1109/ICCV.2017.371.

[15] Leo Gao, Tom Dupré la Tour, Henk Tillman, Gabriel Goh, Rajan Troll, Alec Radford, Ilya Sutskever, Jan Leike, and Jeffrey Wu. Scaling and evaluating sparse autoencoders. *arXiv preprint arXiv:2406.04093*, 2024. URL https://arxiv.org/abs/2406.04093.

[16] Matthew F. Glasser, Timothy S. Coalson, Emma C. Robinson, Carl D. Hacker, John Harwell, Essa Yacoub, Kamil Ugurbil, Jesper Andersson, Christian F. Beckmann, Mark Jenkinson, Stephen M. Smith, and David C. Van Essen. A multi-modal parcellation of human cerebral cortex. *Nature*, 536(7615):171–178, 2016. doi: 10.1038/nature18933.

[17] Idan Daniel Grosbard, Mor Geva, and Galit Yovel. Mechanistically interpretable neural encoding reveals fine-grained functional selectivity in human visual cortex, 2026. URL https://arxiv.org/abs/2605.16468.

[18] Stefan Haufe, Frank Meinecke, Kai Görgen, Sven Dähne, John-Dylan Haynes, Benjamin Blankertz, and Felix Bießmann. On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage*, 87:96–110, 2014. doi: 10.1016/j.neuroimage.2013.10.067.

[19] James V. Haxby, M. Ida Gobbini, Maura L. Furey, Alumit Ishai, Jennifer L. Schouten, and Pietro Pietrini. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293(5539):2425–2430, 2001. doi: 10.1126/science.1063736.

[20] Mark O. Hill. Diversity and evenness: A unifying notation and its consequences. *Ecology*, 54 (2):427–432, 1973. doi: 10.2307/1934352.

[21] Alexander G. Huth, Shinji Nishimoto, An T. Vu, and Jack L. Gallant. A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron*, 76(6):1210–1224, 2012. doi: 10.1016/j.neuron.2012.10.014.

[22] Nancy Kanwisher, Josh McDermott, and Marvin M. Chun. The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, 17(11): 4302–4311, 1997. doi: 10.1523/JNEUROSCI.17-11-04302.1997.

[23] Timothy Kassis, Vinayak Agarwal, Yuhuan He, Darshil Patel, and Aubrey M. Brueckner. Scientific agent skills: A library of procedural knowledge for research agents. arXiv:2609.00065, 2026. URL https://arxiv.org/abs/2609.00065.

[24] Kendrick N. Kay, Thomas Naselaris, Ryan J. Prenger, and Jack L. Gallant. Identifying natural images from human brain activity. *Nature*, 452(7185):352–355, 2008. doi: 10.1038/nature06713.

[25] Meenakshi Khosla, N. Apurva Ratan Murty, and Nancy Kanwisher. A highly selective response to food in human visual cortex revealed by hypothesis-free voxel decomposition. *Current Biology*, 32(19):4159–4171.e9, 2022. doi: 10.1016/j.cub.2022.08.009.

[26] Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, and Piotr Dollár. Microsoft COCO: Common objects in context. arXiv:1405.0312, 2014. URL https://arxiv.org/abs/1405.0312.

[27] Andrew Luo, Jacob Yeung, Rushikesh Zawar, Shaurya Dewan, Maggie Henderson, Leila Wehbe, and Michael Tarr. Brain mapping with dense features: Grounding cortical semantic selectivity in natural images with vision transformers. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/73af055566f5514b9863315133b84eda-Abstract-Conference.html.

[28] Alireza Makhzani and Brendan Frey. k-sparse autoencoders. *arXiv preprint arXiv:1312.5663*, 2013. URL https://arxiv.org/abs/1312.5663.

[29] Ziming Mao, Jia Xu, Zeqi Zheng, Haofang Zheng, Dabing Sheng, Yaochu Jin, and Guoyuan Yang. Sparse autoencoders bridge the deep learning model and the brain, 2025. URL https://arxiv.org/abs/2506.11123.

[30] Leland McInnes, John Healy, and James Melville. UMAP: Uniform manifold approximation and projection for dimension reduction. arXiv:1802.03426, 2018. URL https://arxiv.org/abs/1802.03426.

[31] Thomas Naselaris, Kendrick N. Kay, Shinji Nishimoto, and Jack L. Gallant. Encoding and decoding in fMRI. *NeuroImage*, 56(2):400–410, 2011. doi: 10.1016/j.neuroimage.2010.07.073.

[32] Mona Rosenke, Rick van Hoof, Job van den Hurk, Kalanit Grill-Spector, and Rainer Goebel. A probabilistic functional atlas of human occipito-temporal visual cortex. *Cerebral Cortex*, 31(1): 603–619, 2021. doi: 10.1093/cercor/bhaa246.

[33] Gabriel Sarch, Michael Tarr, Katerina Fragkiadaki, and Leila Wehbe. Brain dissection: fMRI-trained networks reveal spatial selectivity in the processing of natural images. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/90e06fe49254204248cb12562528b952-Abstract-Conference.html.

[34] Steffen Schneider, Jin Hwa Lee, and Mackenzie Weygandt Mathis. Learnable latent embeddings for joint behavioural and neural analysis. *Nature*, 617(7960):360–368, 2023. doi: 10.1038/s41586-023-06031-6.

[35] Harrish Thasarathan, Julian Forsyth, Thomas Fel, Matthew Kowal, and Konstantinos G. Derpanis. Universal sparse autoencoders: Interpretable cross-model concept alignment. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 59304–59325. PMLR, 2025. URL https://proceedings.mlr.press/v267/thasarathan25a.html.

[36] Hugo Touvron, Matthieu Cord, Matthijs Douze, Francisco Massa, Alexandre Sablayrolles, and Herve Jegou. Training data-efficient image transformers & distillation through attention. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139, pp. 10347–10357, 2021. URL https://proceedings.mlr.press/v139/touvron21a.html.

[37] Reinder Vos de Wael, Oualid Benkarim, Casey Paquola, Sara Lariviere, Jessica Royer, Shahin Tavakol, Ting Xu, Seok-Jun Hong, Georg Langs, Sofie Valk, Bratislav Misic, Michael Milham, Daniel S. Margulies, Jonathan Smallwood, and Boris C. Bernhardt. Brainspace: A toolbox for the analysis of macroscale gradients in neuroimaging and connectomics datasets. *Communications Biology*, 3:103, 2020. doi: 10.1038/s42003-020-0794-7.

[38] B. T. Thomas Yeo, Fenna M. Krienen, Jorge Sepulcre, Mert R. Sabuncu, Danial Lashkari, Marisa Hollinshead, Joshua L. Roffman, Jordan W. Smoller, Lilla Zöllei, Jonathan R. Polimeni, Bruce Fischl, Hesheng Liu, and Randy L. Buckner. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, 106(3): 1125–1165, 2011. doi: 10.1152/jn.00338.2011.

[39] Matthew D. Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *Computer Vision – ECCV 2014*, volume 8689 of *Lecture Notes in Computer Science*, pp. 818–833. Springer, 2014. doi: 10.1007/978-3-319-10590-1_53.

A Relation to prior work

Natural-image encoding and cortical organization. Natural-image encoding and identification established quantitative links between visual representations and measured cortical responses [24, 31]. NSD made such analyses possible across many images in deeply sampled participants [1]. Distributed object representations [19] and continuous semantic maps from natural-movie responses [21] argue for descriptions that extend beyond isolated category regions. At the same time, the probabilistic occipitotemporal atlas of Rosenke et al. provides reproducible regional references [32]. Our analysis treats those regions as summaries of overlapping fitted feature directions rather than as labels for each mode, then tests whether corresponding directions recur across participants and models.

Components within distributed representations. Data-driven NSD voxel decomposition has recovered familiar category structure and a food-selective component, including validation beyond the discovery participants [25]. Behaviour-derived object dimensions likewise reveal distributed and mixed selectivity across visual cortex [5]. Sparse model features provide a complementary unit: each has a fitted cortical direction and an image-response profile that can be compared across people and architectures. Regional enrichment and within-ROI coefficient mass answer different questions in this setting. An FM may be concentrated in an ROI yet account for only a small fraction of that ROI's total fitted coefficient mass.

Interpretable vision features and neural encoding. Brain Dissection relates localized features of fMRI-trained networks to visual-region selectivity [33]; BrainSAIL combines Transformer encoding with image-region attribution [27]. SAE-BrainMap aligns layerwise sparse-autoencoder units with voxel responses [29], while MINE describes voxel-associated image features and evaluates edits through predicted responses [17]. These methods motivate grounding fitted cortical patterns in inspectable image evidence. Here, the additional question is whether image-grounded modes form families with cortical support across participants and models, and whether one family dictionary can explain both condition enrichment and common participation. Model occlusion remains evidence about the model feature, not an intervention on cortex.

Correspondence and geometry. Universal Sparse Autoencoders seek shared concepts across pretrained models [35], and Archetypal SAE examines dictionary stability and interpretability [12]. We instead retain four fixed, model-specific dictionaries and infer approximate correspondence from common-image responses, then evaluate the resulting families in independently fitted cortical maps. PCA and diffusion coordinates describe complementary aspects of their organization; CEBRA and other embeddings provide views of local relationships [34, 37]. The original-space graph summarizes neighboring representations. Co-recruitment separately asks which families are called for the same image, without making graph edges a processing path.

B Detailed methods

B.1 Participants and image material

We analyzed the Natural Scenes Dataset (NSD) responses of S1, S2, S5, and S7 [1]. Each participant completed the 40 core sessions, comprising 10,000 distinct images and 30,000 trials in the NSD protocol. Per participant, 8,000 images were used to fit the encoding model, 1,000 for validation, and 1,000 for held-out evaluation. The held-out lists share the same ordered 1,000 NSD global image identifiers, permitting matched image-response comparisons; fMRI coefficients were fitted separately for each participant. Thus the biological sample comprises four people, whereas images, vertices, modes, families, and the 16 participant–model cells are distinct analysis units. Detailed single-mode and coordinate analyses use S5 and ViT-Base pretrained on ImageNet-21k; recurrence analyses use all four participants and architectures.

NSD stimuli supply the object and scene labels and pixel descriptors used in analyses involving measured neural responses. A separate ImageNet-derived model-research split of 128,117 images supplies the four highest-response examples for each illustrated FM, their occlusion maps, and the cross-material model-response comparison [6]. This split is not the official ImageNet validation set and carries no additional fMRI. The dictionary-development feature arrays comprise 1,024,933 train-

Table 2: Vision checkpoints and per-block class-token dimensions for the 73,000 NSD stimulus images.

Model	Checkpoint identifier	Blocks × width
ViT-Small	`vit-small-patch16-224`	12×384
ViT-Base-21k	`vit-base-patch16-224-in21k`	12×768
ViT-Large	`vit-large-patch16-224`	24×1024
DeiT-Base	`deit-base-patch16-224`	12×768

ing, 128,117 validation, and 128,117 test image vectors per architecture. Supplementary manifests link image identities to their models and features.

B.2 Frozen vision features and model identity

The four frozen vision backbones were ViT-Small, ViT-Base-21k, ViT-Large, and DeiT-Base (Table 2) [8, 36]. Images were resized and normalized according to each checkpoint. The retained four-model arrays contain the class token after each Transformer block for all 73,000 NSD stimulus images; the original ViT-Base extraction ran its classifier backbone in evaluation mode. Concatenation produces feature dimensions of 4,608, 9,216, 24,576, and 9,216, respectively. Each architecture has its own dictionary in this cross-block space. A dictionary direction can span blocks and hidden coordinates; it is not an attention head, single neuron, or anatomical region. In this atlas, an FM denotes a model-side dictionary feature together with a separately fitted participant-specific cortical coefficient pattern.

B.3 Inherited TopK-SAE architecture and objective

We reused four pretrained, architecture-specific TopK sparse autoencoders (TopK-SAEs) from our prior cross-layer FM work. Their dictionaries were fixed before neural encoding; no dictionary was trained on NSD or fMRI data here. Each dictionary has 256 rows, one per FM, spanning the concatenated class-token coordinates. For an input feature vector $x \in \mathbb{R}^d$, let $D \in \mathbb{R}^{256\times d}$ denote the learned matrix and let its normalized row be

$$\widetilde{d}_f = \frac{D_f}{\|D_f\|_2 + 10^{-6}}, \qquad f = 1, \ldots, 256. \tag{1}$$

The encoder forms all raw projection scores before selecting indices:

$$s_f(x) = x^\top \widetilde{d}_f, \qquad T_8(x) = \arg\operatorname{top8}_{f\in\{1,\ldots,256\}} s_f(x). \tag{2}$$

Its nonnegative sparse code is

$$c_f(x) = \begin{cases} \max\{0, s_f(x)\}, & f \in T_8(x), \\ 0, & f \notin T_8(x). \end{cases} \tag{3}$$

Thus TopK is applied to signed scores *before* rectification, and fewer than eight FMs may be positive. The decoder uses the same normalized rows, without a separate bias or independent decoder matrix:

$$\widehat{x} = \sum_{f=1}^{256} c_f(x)\widetilde{d}_f, \qquad \mathcal{L}_{\mathrm{SAE}} = \frac{1}{d}\|x - \widehat{x}\|_2^2. \tag{4}$$

The loss is the mean squared reconstruction error over batch and feature coordinates. TopK supplies sparsity without an additional penalty. The learned directions are shared between encoder and decoder [15, 28].

B.4 Inherited dictionary training

The dictionaries were fit independently to ImageNet-derived model features, without fMRI. Each architecture has 1,024,933 training, 128,117 validation, and 128,117 test vectors. The available preprocessing routine shuffles each split, standardizes feature columns within batches of 62 images using population standard deviations, repeats this with three shuffles, and averages the resulting vectors by image. Training uses shuffled mini-batches of 4,096, AdamW, the reconstruction loss above, 256 latents, eight selected scores, and 512 epochs. The four retained dictionaries have the expected dimensions and 512-epoch training, validation, and test loss histories.

In the earlier ViT-Base/ImageNet-1K selection analysis, five latent widths (64, 128, 256, 512, 1,024) and eight sparsity levels (2, 4, 6, 8, 10, 12, 16, 32) were compared by mean validation reconstruction loss over the final 100 epochs. The 256/eight setting was selected as a sparsity–fidelity trade-off within a low-loss regime; the global reconstruction-loss minimum used a larger, less sparse dictionary. This search was performed in the prior study, while the present atlas applies the fixed four-model dictionaries. The original execution logs do not resolve the exact preprocessing invocation or optimizer schedule for each retained checkpoint.

B.5 Source-domain reproducibility of the inherited dictionary

The frozen dictionaries used in this atlas were inherited from a model-only analysis of cross-layer vision-transformer responses (Appendix B.3). An existing reproducibility analysis supplies a more specific characterization of the ViT-Base source dictionary. It compares ten saved 256-row, $K = 8$ TopK-SAE checkpoints on the same 9,216-coordinate ViT-Base/ImageNet-derived feature representation. Nine checkpoints have explicit seed labels; the historical reference checkpoint lacks a recorded training seed. The explicitly seeded runs vary both initialization and data order. All results in this subsection concern dictionary weights or responses to ImageNet-derived images. The other three atlas backbones and participant-specific NSD preprocessing were not part of this seed comparison.

For each ordered checkpoint pair, dictionary rows were normalized to unit length and matched one-to-one by maximum-sum Hungarian assignment on *signed* cosine similarity. A row counted as shared when its assigned cosine was at least .7. Table 3 summarizes all-row and training-active comparisons. The latter use 213–222 rows per checkpoint and therefore have unequal directed denominators; matched signed cosine averages .8064. Across the ten reference checkpoints, 41–52 of 256 rows had no partner above .7 in any other checkpoint, whereas 111–117 had a partner in all nine. The all-row shared fraction varies from 72.34% at cosine .5 to 36.59% at cosine .95, showing its threshold dependence.

Table 3: Existing ViT-Base/ImageNet dictionary-seed comparisons. Means and ranges summarize 90 directed comparisons among ten checkpoints; they are descriptive across dependent checkpoint pairs, not confidence intervals. Training-active matching has 213–222 rows per checkpoint. Test-response Pearson uses a separate 128,117-image ImageNet-derived split with test-active rows paired by weight cosine, without selecting matches by response similarity.

Endpoint	Mean	Pair range
All-row shared fraction	.6665	.6289–.7227
Training-active shared fraction	.7885	.7318–.8645
Test-active paired response Pearson	.7404	.7027–.8011

Test-active rows paired by *weight* cosine have correlated responses on the separate ImageNet-derived test split (Table 3). Matches are fixed without optimizing response correlation. The decoder shares the encoder directions, so weight matching does not supply an independent encoder–decoder replication. These observations support partial reproducibility of the source-domain model feature vocabulary, alongside a substantial seed-sensitive tail. They do not measure reconstruction fidelity or feature identity after NSD's distinct run-wise standardization, nor do they test the stability of participant-specific cortical coefficient maps or family assignments under independently trained dictionaries.

An earlier source-domain comparison used ViT-Base/ImageNet-1K features and 256 latent dimensions to compare TopK-SAE with standard sparse autoencoding and nonnegative matrix factorization. TopK-SAE had a reported mean of eight active latents per image and the highest reported seed-matching consistency (.6233, versus .2252 and .3671), but its reconstruction MSE was higher (.5275, versus .2677 and .4045). Thus the inherited sparse basis reflects a sparsity and interpretability trade-off rather than optimal reconstruction under that protocol. The earlier seed-consistency statistic uses a different comparison protocol from Table 3; these values should not be interpreted as replicate estimates of one quantity. Neither comparison is a dense-feature or PCA baseline for held-out NSD neural encoding.

B.6 PARTICIPANT-SPECIFIC NSD PREPROCESSING AND SPARSE PROJECTION

For NSD encoding, each participant's blockwise class-token features were standardized to zero mean and unit variance separately within each session/run and block, over that run's image trials. Standardized trial vectors were concatenated across blocks, then repeated presentations were averaged by NSD global image identifier within participant. This scaling precedes the image-level train/validation/test split and therefore uses the unlabeled feature distribution of held-out trials; held-out fMRI responses were excluded from fitting. It differs from ImageNet dictionary development and from raw checkpoint features.

With participant-specific image vector x_i and frozen row D_f, the atlas computes

$$s_{if} = x_i^\top \frac{D_f}{\|D_f\|_2 + 10^{-6}}, \qquad a_{if} = \begin{cases} \max\{0, s_{if}\}, & f \in \arg\operatorname{top8}_g\, s_{ig}, \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

This is the same unit-row projection, signed-score TopK selection, and rectification as the inherited encoder, applied to the separately prepared NSD vectors. At most eight of 256 FM responses are positive per image. The atlas reuses the saved dictionary rows; it does not fit a new dictionary or encoder to NSD.

B.7 NEURAL ENCODING AND CORTICAL SUPPORT

For each participant–model cell, a linear encoding model relates the 256 sparse image responses to fMRI at cortical location v:

$$\widehat{y}_{iv} = b_v + \sum_{f=1}^{256} a_{if} W_{vf}. \tag{6}$$

Using the fixed image split and random seed, 8,000 images fit the model, 1,000 guide validation, and the common 1,000 are held out. Mean squared error is optimized with AdamW at learning rate 10^{-3} for at most 512 epochs; training stops after 32 epochs without validation improvement, and the best validation state is tested. The HCP-MMP cortical-location mask with twofold location subsampling defines neural locations [16]. Signed W_{vf} values express fitted directions conditional on the other FMs, rather than independent causal effects [18]. An FM enters cross-participant analysis only if its activity frequency is positive in both the fixed training and held-out splits. This yields 2,775 eligible maps from the 4,096 possible participant–model–FM maps.

S5 contributes 45,280 modeled cortical locations, with 187 eligible S5/ViT-Base FMs for the coordinate analysis. For comparisons across participants, maps are registered to fsaverage6 [13] and evaluated on the 4,029 vertices common to all four; unsupported values remain missing rather than being set to zero. TY17 profiles average signed encoding weights within 17 regions on this common surface and summarize the same fitted neural data at a coarser scale.

Interpretation of a coefficient map. The mapping runs from image features to neural responses: it is a forward encoding model. The backward-to-forward transformation of Haufe et al. [18] addresses spatial interpretation of decoding filters and is not an automatic correction to this forward model. Here W_{vf} describes the fitted change at location v per unit FM response with the other predictors held fixed. This differs from the marginal association of an FM with the full predicted response. For centered feature covariance Σ_a, that association satisfies

$$\operatorname{Cov}(a_f, \widehat{y}_v) = \sum_h (\Sigma_a)_{fh} W_{vh}. \tag{7}$$

Multiplying by feature covariance therefore mixes the other FMs into the quantity being mapped. We retain conditional coefficients as the atlas coordinates. Predictor correlations and estimation choices can nevertheless affect these coefficients; agreement of full predictions does not establish the reliability of every column. The spatial family comparisons evaluate correspondence between the fitted maps available here, whereas within-cell bootstrap or refit reliability has not been measured.

B.8 HELD-OUT ENCODING ASSESSMENT

The S5/ViT-Base readouts in Figure 2 restrict the frozen 256-feature encoder to the 187 FMs with positive activity frequency in both training and held-out model responses. Let $\mathcal{E}$ denote this eligible

set. The restricted prediction is

$$\widehat{y}^{\mathcal{E}}_{iv} = b_v + \sum_{f \in \mathcal{E}} a_{if} W_{vf}. \tag{8}$$

Weights and bias retain their fitted values; removing the other terms entails no refit. Eligibility uses held-out model activations but no held-out fMRI labels. In the following equations, $\widehat{y}$ and P denote this restricted prediction. "Full" in the spatial readouts means the complete pattern before ROI projection, not the unrestricted 256-feature predictor. Older unrestricted-encoder summaries therefore concern a different readout. For the 1,000 held-out images, prediction quality at location v is

$$R^2_v = 1 - \frac{\mathrm{mean}_i (y_{iv} - \widehat{y}_{iv})^2}{\mathrm{var}_i(y_{iv})}, \qquad r_v = \mathrm{corr}_i(y_{iv}, \widehat{y}_{iv}). \tag{9}$$

Noise-ceiling units. The stored ceiling uses the three repeated presentations of each image after run-wise standardization of the neural responses [1]. If y_{vir} denotes the standardized response at location v to image i on repeat r, the implementation estimates

$$\widehat{\sigma}^2_{\mathrm{noise},v} = \mathrm{mean}_i[\mathrm{var}_{r=1,2,3}(y_{vir})], \qquad \widehat{\sigma}^2_{\mathrm{signal},v} = \max\{1 - \widehat{\sigma}^2_{\mathrm{noise},v}, 0\}, \tag{10}$$

using the unbiased variance over repeats. For the average of three responses, the ceiling in variance units is

$$\mathrm{NC}_v = \frac{3\widehat{\sigma}^2_{\mathrm{signal},v}}{\widehat{\sigma}^2_{\mathrm{noise},v} + 3\widehat{\sigma}^2_{\mathrm{signal},v}}. \tag{11}$$

Zero estimated noise gives a ceiling of one. The R^2 scatter uses NC_v, whereas the Pearson scatter uses $\sqrt{\mathrm{NC}_v}$. Each 75% reference is defined in its own metric units. Both scatters and the following pattern readouts use the 12,550 locations with $\mathrm{NC}_v \geq 0.10$; negative prediction scores remain included. This reliability criterion explains the nonzero lower bounds of the scatter's horizontal coordinates.

Regional projection. Let $P, Y \in \mathbb{R}^{N \times V}$ contain predicted and measured patterns for the $N = 1{,}000$ evaluation images at the retained locations. An intercept and ten overlapping binary ROI masks form a fixed $V \times 11$ design. Its orthonormal column basis Q is obtained by SVD, giving the rank-11 projection

$$P_{\mathrm{ROI}} = PQQ^\top, \qquad Y_{\mathrm{ROI}} = YQQ^\top, \qquad P_{\mathrm{res}} = P - P_{\mathrm{ROI}}, \qquad Y_{\mathrm{res}} = Y - Y_{\mathrm{ROI}}. \tag{12}$$

The joint projection accommodates overlapping masks without assigning each location to one exclusive ROI. Full, projected, and residual patterns are evaluated with the same readouts. The residual is a component of the measured and predicted patterns, not another neural recording.

Image identification and pattern agreement. For each representation, center each pattern across locations and normalize it to unit length, yielding $\widetilde{P}$ and $\widetilde{Y}$. Their cross-image correlation matrix is

$$C = \widetilde{P}\widetilde{Y}^\top. \tag{13}$$

Diagonal entries are matched-image correlations and off-diagonal entries are nonmatches. Two-versus-two matching compares correct and swapped assignments over all distinct image pairs:

$$d_{ij} = C_{ii} + C_{jj} - C_{ij} - C_{ji}, \qquad A_{2\mathrm{v}2} = \frac{1}{\binom{N}{2}} \sum_{i<j} \left[\mathbf{1}\{d_{ij} > 0\} + \tfrac{1}{2}\mathbf{1}\{d_{ij} = 0\} \right]. \tag{14}$$

The label-randomization reference uses 199 permutations of measured-image identities. These image pairs are repeated comparisons within one participant and fitted encoder.

For retrieval among K candidates, each target is combined with $K-1$ uniformly sampled distinct nonmatches. Let b_i count nonmatches whose scores strictly exceed the target score. The implementation computes the exact probability that none of those better matches is selected:

$$b_i = \sum_{j \neq i} \mathbf{1}\{C_{ij} > C_{ii}\}, \qquad A_K = \frac{1}{N} \sum_{i=1}^{N} \frac{\binom{N-1-b_i}{K-1}}{\binom{N-1}{K-1}}, \quad K \in \{2, 10, 100, 1000\}. \tag{15}$$

A numerator is zero when fewer than $K-1$ nonmatches have scores at or below the target; tied best scores are accepted. Stimulus-geometry agreement is the Spearman correlation between the upper-triangular entries of $\widetilde{P}\widetilde{P}^\top$ and $\widetilde{Y}\widetilde{Y}^\top$. The aggregate fit pools squared errors across images and locations:

$$R^2_{\text{aggregate}} = 1 - \frac{\sum_{i,v}(Y_{iv} - P_{iv})^2}{\sum_{i,v}(Y_{iv} - \overline{Y}_{\cdot v})^2}. \tag{16}$$

These readouts distinguish pattern agreement, image discrimination, and predictive calibration; their numerical results are in Appendix C.

B.9 Image exemplars and model-side localization

Each illustrated FM is paired with its four highest-response images from the separate ImageNet-derived research split, preserving its architecture and model-local FM index. For image x and local occlusion $x_{\setminus p}$, the displayed map derives from

$$\Delta a_{f,p}(x) = a_f(x) - a_f(x_{\setminus p}). \tag{17}$$

The 7×7 local maps identify image regions influential for the model feature under this perturbation, not effects of image edits on measured brain activity [14, 39]. The four displayed originals and their paired maps are a compact illustration, not the full Qwen evidence packet.

B.10 Qwen-assisted FM interpretation

Our prior ViT-Base/ImageNet-1K work used a Qwen-assisted procedure to describe FMs from highly activating images. For FM f, it formed the active test-image pool and ranked images by sparse activation:

$$\mathcal{S}_f = \{x \in \mathcal{D}_{\text{test}} : c_f(x) > 0\}, \qquad \mathcal{T}_f = \text{Top100}_{x \in \mathcal{S}_f}\, c_f(x). \tag{18}$$

The 100 images and their activation magnitudes were sent to Qwen3.0-VL-Max under a fixed prompt requesting one concise visual-pattern sentence and one possible representational-role sentence, returned as `[interpretation]:` and `[functionality]:` [2]. The atlas reuses those pretrained FM identities and the image-grounded interpretation approach. Its displayed annotations come from a later structured Qwen run with a different evidence packet and output schema:

Step	Prior FM procedure	Atlas annotation run
Selection	Top 100 positive test-image responses per FM	One ViT-Base FM per request; 218 of 256 FMs have archived requests
Model input	Selected images and response magnitudes	Positive images with magnitudes, zero-response hard negatives, localization evidence, model-output effects under FM perturbation, and cross-layer structure
Typical packet	100 images	Six positives, 12 negatives, six localization examples, and eight effect examples in 211 requests; seven requests have fewer usable examples
Requested output	Visual interpretation and possible function in two named text fields	Structured card with visual summary, trigger and anti-trigger, coverage, proposed role, structure, causal expectation, failure modes, and confidence
Atlas use	Establishes the inherited interpretation procedure	Archived `summary` informs visual description; `functional_claim` is a proposed model role

The atlas requests used the service alias `qwen-vl-max-latest`; its exact backend revision is unavailable. The 38 FMs without archived responses have no Qwen description. For each displayed ViT-Base FM, the annotation is joined by its model-local index and checked against its Top-4 images and paired occlusions. Text is narrowed where images show co-occurrence rather than selectivity. The model also saw localization, output-effect, and structure inputs, so its descriptions are interpretive hypotheses rather than independent tests of those inputs. No participant fMRI or cortical map was given to Qwen. Generated causal or confidence fields are excluded from the atlas measurements.

For each family card, the S5/ViT-Base representative is the member whose image-response fingerprint correlates most strongly with the 16-cell balanced family fingerprint. Its images and description characterize that FM alone. Family membership and condition labels are computed from responses, and anatomical interpretations use the separately fitted cortical coefficients.

B.11 Continuous coordinates and regional mixtures

Cortical coordinates. For S5/ViT-Base, let $W \in \mathbb{R}^{45,280\times 187}$ contain eligible FM coefficients, with cortical locations in rows. Each column is multiplied by its FM's training activation standard deviation σ_f (sample standard deviation), and each location profile is normalized:

$$U_{vf} = W_{vf}\sigma_f, \qquad Z_{vf} = \frac{U_{vf}}{\|U_{v\cdot}\|_2}, \qquad Z \in \mathbb{R}^{45,280\times 187}. \tag{19}$$

Let $\mu \in \mathbb{R}^{187}$ and $B \in \mathbb{R}^{187\times 12}$ be the fitted PCA mean and loading matrix. Cortical scores S and unit-FM landmarks L use the same fixed basis:

$$S = (Z - \mathbf{1}_{45,280}\mu^\top)B, \qquad L = (I_{187} - \mathbf{1}_{187}\mu^\top)B. \tag{20}$$

Component identities and signs were kept fixed. On a given component, the 187 landmark coordinates differ from its loadings by a common offset. We related the first 12 landmark coordinates to 100 model-side descriptors: 80 COCO object preferences, seven scene preferences, and 13 pixel-derived quantities. The 1,200 Spearman correlations used 4,999 joint FM-identity permutations, global Benjamini–Hochberg (BH) correction, and a stricter maximum-absolute-statistic reference [3]. For an activity-frequency sensitivity check, both coordinates and descriptors were rank-residualized against FM training activity frequency before recomputing correlations. These associations are between model responses and image properties. COCO examples discussed in the main text contain at least 20 NSD images.

Regional contributions. The absolute activation-SD-weighted coefficient measures FM f's contribution mass at location v. For ROI R, its mean mass, normalized share, effective contributor count, and within-FM enrichment are

$$M_{vf} = |W_{vf}\sigma_f|, \qquad m_{Rf} = \mathrm{mean}_{v\in R}\, M_{vf}, \qquad \pi_{Rf} = \frac{m_{Rf}}{\sum_{h=1}^{187} m_{Rh}}, \tag{21}$$

$$N_{\mathrm{eff}}(R) = \exp\left(-\sum_{f=1}^{187} \pi_{Rf}\log \pi_{Rf}\right), \qquad E_{Rf} = \frac{m_{Rf}}{\mathrm{mean}_{v\in V}\, M_{vf}}, \tag{22}$$

where V denotes all 45,280 modeled locations and the effective count is the Shannon-order Hill number [20]. The ten ROIs are summarized separately. A mode can be enriched in an ROI while contributing little of that ROI's total coefficient mass; neither quantity assigns independent explained variance. An image split-half analysis tests whether 30 axis combinations selected on one half retain functional neighborhoods on the other. Fixed two-axis explanatory R^2 describes the 187-FM matrix itself and addresses a different estimand.

B.12 Correspondence and recurrent families

Family discovery. Family discovery starts with the 2,775 eligible maps. Each is represented by its mean-centered, unit-length response fingerprint over the common 1,000 images. Across distinct participant–model cells, an edge joins mutual top-one fingerprint matches only when their similarity exceeds that cell pair's 99th percentile. For image-fingerprint correlation $\rho_{hh'}$, its nonnegative edge weight is

$$e_{hh'} = \max\{\rho_{hh'}, 0\} + 10^{-6}. \tag{23}$$

Weighted Louvain detection used resolution 1 and seed 20260922 [4]. Communities were ordered by decreasing size and then minimum source-node index. The 75 spanning all 16 cells were retained without imposing cortical similarity during discovery. They contain 1,752 maps; the remaining 1,023 are unassigned. FM indices remain local to each model and do not themselves establish correspondence. A complementary pairwise screen compared the 187 S5/ViT-Base references with the other 15 cells (2,805 comparisons), requiring positive image-fingerprint and regional-profile

similarities in the top 5% of each cell's candidates. Surface correlations and image split halves checked those matches, which are distinct from community membership. A stricter, separately numbered 39-family split-half intersection tests member-identity sensitivity; it does not replace the 75-family dictionary.

Cortical prototypes. Hereafter g indexes a family, f a model-local FM, and c one of the 16 participant–model cells. Let $\mathcal{M}_{gc}$ be the members of family g in cell c, and let $w_{cf} \in \mathbb{R}^{4029}$ be a member's signed cortical coefficient map. We normalize each member direction, average within cell, then give every cell equal weight:

$$u_{cf} = \frac{w_{cf}}{\|w_{cf}\|_2}, \qquad m_{gc} = \frac{1}{|\mathcal{M}_{gc}|} \sum_{f \in \mathcal{M}_{gc}} u_{cf}, \qquad p_g = \frac{1}{16} \sum_{c=1}^{16} m_{gc}. \tag{24}$$

Raw means and separately normalized directions are retained. Equal cell weighting prevents families with more members in one cell from favoring that combination. At vertex v, the descriptive sign-agreement field is

$$A_{gv} = \frac{1}{16} \sum_{c=1}^{16} \mathbf{1}\{\operatorname{sgn}(m_{gcv}) = \operatorname{sgn}(p_{gv})\}. \tag{25}$$

It is not a coverage or significance test.

Spatial recurrence was evaluated by omitting one model or one participant at a time. A prototype from the other 12 cells was compared with the four held-out cell means using signed cosine over 4,029 vertices and Pearson correlation of TY17 profiles. A member-count-matched null repeated 199 draws of replacement FMs without replacement within each cell, preserving that family's cell-specific size. A second null excluded true family members and drew replacements from the nearest 20 eligible candidates in log training activity frequency for each member. One-sided empirical family probabilities were BH-corrected across 75 families. These tests assess support conditional on the selected families and four observed participants; they do not establish population prevalence.

B.13 Image calls, attributes, and condition subatlases

Balanced image calls and attributes. For family g, cell c, and NSD image i, an any-member binary call and its cell-balanced fingerprint are

$$I_{gci} = \mathbf{1}\left\{ \max_{f \in \mathcal{M}_{gc}} a_{cif} > 0 \right\}, \qquad \bar{I}_{gi} = \frac{1}{16} \sum_{c=1}^{16} I_{gci}. \tag{26}$$

Multiple active members in one cell still give one call; alternative aggregation readouts address its possible dependence on family size. Thirteen quantities were measured from NSD image pixels: brightness, contrast, saturation, colorfulness, edge energy, edge density, high-frequency texture, orientation entropy, orientation anisotropy, grayscale entropy, center–periphery difference, and red–green and blue–yellow opponent quantities. Each balanced family call was correlated with each quantity over the common 1,000 images. The 975 tests used 1,999 joint image-row permutations and one BH family; support required $q < .05$, $|\rho| \geq .10$, concordant signs in at least 12/16 cells and three of four participant aggregates. Negative associations were retained. Member-level correlations and exploratory standardized ridge adjustment for other image quantities and semantic labels assessed heterogeneity and co-occurrence.

Condition selection. The condition analysis covers 82 unique masks: 49 COCO object labels with at least 20 images [26], six scene labels, one face-context label, and high and low quartiles for each of the 13 continuous quantities. Two duplicate masks are aliases. For condition k and family g, the marginal preference effect is

$$\Delta_{kg} = \mathbb{E}(\bar{I}_{gi} \mid i \in k) - \mathbb{E}(\bar{I}_{gi} \mid i \notin k). \tag{27}$$

A relative-preference relation requires $\Delta_{kg} \geq .02$, positive effects in at least 12/16 cells, global BH $q < .05$ among 6,150 condition–family tests based on 1,999 joint image-row permutations, and membership in the 62-family primary spatial core; 659 pairs satisfy these criteria. Common participation instead requires a within-condition mean call rate of at least .10, rates of at least .05

in 12/16 cells, and marginal rates of at least .10 for at least three participants and three models, again within the spatial core. A condition subatlas is the union of these relations, allowing a rare but preferential family to enter. Both criteria use model responses to actual NSD images; the cortical prototypes come separately from fitted fMRI encoding weights.

Derived cortical composition and co-recruitment. Let H_{kg} indicate that family g belongs to condition k's union. Its within-condition call rate and the displayed signed composition are

$$r_{kg} = \mathbb{E}(\bar{I}_{gi} \mid i \in k), \qquad C_{kv} = \frac{\sum_g H_{kg} r_{kg} p_{gv}}{\sum_g H_{kg} r_{kg}}. \tag{28}$$

It combines fixed atlas prototypes and is not a measured condition-average fMRI response or a refitted neural prediction. A single illustrative NSD image is the condition member whose balanced selected-family call vector is nearest the condition mean. Its calls are shown separately from the condition-wide union.

For family pair g, g', within-cell excess joint calling, averaged across cells, is

$$J_{k,gg'} = \frac{1}{16} \sum_{c=1}^{16} \left[\mathbb{E}(I_{gci} I_{g'ci} \mid i \in k) - \mathbb{E}(I_{gci} \mid i \in k) \mathbb{E}(I_{g'ci} \mid i \in k) \right]. \tag{29}$$

Shared image-index resampling provides pointwise exploratory intervals. Displayed graph edges require mean excess at least .01, a positive pointwise 95% lower bound, and positive participant and model aggregates in at least three of four groups each. These edges describe same-image call association, not directed cortical connections.

B.14 Atlas geometry and cross-material comparison

The family-level representation joins each signed 4,029-vertex prototype to its balanced 1,000-image call fingerprint. Both blocks are centered across families and scaled to contribute equally to squared distance. A local-$k = 8$ diffusion coordinate serves as a broad-scale reference, while UMAP was chosen for a legible local view after neighborhood-fidelity comparison. We compared 23 configurations: 15 diffusion settings crossing five call-block weights with three bandwidths, one PCA, three UMAP seeds, three randomly initialized t-SNE seeds, and CEBRA centroids from a separate all-point fit [34, 37]. Distances and neighbors were evaluated in the original joint input rather than inferred from a two-dimensional picture. A mutual four-nearest-neighbor graph was built in that input, and edge retention was measured after omitting each participant or model while holding family identities fixed.

An additional view includes all 2,775 eligible maps, including those outside the 75 families. Inputs are signed 4,029-vertex coefficient vectors and mean-centered, unit-length fingerprints over the common 1,000 images. Full PCA supplies 64 cortical and 64 stimulus components; each is standardized across maps, and the first 16 stimulus components provide auxiliary labels. One CEBRA session per participant–model cell (16 sessions) maps the 64 cortical components to eight dimensions under a nonhybrid offset-one, delta-conditioned objective: $\delta = .2$, 64 hidden units, batch size 256, 5,000 iterations, temperature .5, learning rate .0003, and seed 20260923 [34]. Euclidean UMAP displays these eight-dimensional coordinates in two dimensions with 30 neighbors, minimum distance .15, 500 epochs, and seed 20260923 [30]. This descriptive all-map fit does not define family membership or the family-level geometry.

For the separate ImageNet-derived corpus, we combined fixed FM responses using the same family memberships and any-positive within-cell calls. ViT-Base/Large sparse responses were joined by model and FM identity; ViT-Small and DeiT responses were calculated from retained features and frozen dictionary rows with the same top-eight/ReLU projection, without fitting new directions. Exact WordNet noun-root hyponym closure maps 20 prespecified COCO roots to ImageNet synsets; roots require at least 20 images in each corpus. For each eligible root, we ranked 75 within-minus-outside family-call effects in each corpus and compared ranks by Spearman correlation. The analysis used 1,999 family-identity permutations with BH correction across 15 eligible roots and 499 stratified image bootstrap replicates for descriptive intervals, holding the ImageNet background mean fixed in the intervals. This is a cross-material comparison of model-feature preference, not a second neural dataset.

B.15 DISPLAY CONVENTIONS

Color encodings. Signed cortical coefficients and PC fields use a zero-centered convention: red indicates positive values, blue negative values, and pale colors values near zero. FM coefficients are conditional encoding weights, whereas PC signs are arbitrary coordinate orientations. Held-out performance maps use a sequential warm scale anchored at zero. In the cortical RGB composite, red, green, and blue encode PC1, PC2, and PC3. For channel j at vertex v, the displayed value is

$$\mathrm{RGB}_{vj} = \mathrm{clip}\left(\frac{1}{2} + \frac{s_{vj}}{2q_j},\, 0,\, 1\right), \qquad q_j = Q_{0.98}(\{|s_{vj}| : v \in \mathcal{V}_{\mathrm{PC}}\})\,. \tag{30}$$

Here $\mathcal{V}_{\mathrm{PC}}$ is the set of 45,280 S5 cortical locations used for the PC fields. The three channel scales are 0.750522, 0.611925, and 0.505055. Pairwise plots use the same three channels; similar two-coordinate positions can differ in their third channel. Numerical ranges are retained in the electronic scale tables. The occlusion key is separate: blue-to-red indicates lower-to-higher original-minus-occluded FM response, normalized per image. Color intensity should be compared within a field's own scale.

Surface rendering. Single-FM and PCA cortical maps preserve their established rendering scales. For scalar PC and RGB displays, small gaps are filled and ten surface-diffusion steps are applied with $\lambda = .25$; pairwise projections use scores before those display operations. Sparse 4,029-vertex family and derived subatlas fields use normalized Gaussian interpolation on the fsaverage6 mesh for visualization. Geodesic weights have $\sigma = 3$ mm, truncate at 9 mm, and are renormalized over observed source vertices within each hemisphere. Twenty-eight of the 4,029 vertices lie outside the display cortex mask but remain in numerical analyses; surface locations more than 9 mm from valid source data stay uncolored. Each signed group field has a symmetric scale set by the 98th percentile of its original absolute values. Interpolation and display clipping do not affect prototypes, statistical tests, regional summaries, or graph distances. Model occlusion maps retain their per-image scales and are not cortical activation maps.

C ENCODING READOUTS BEYOND THE MAIN FIGURES

The matched-image readout compares predicted and measured cortical response patterns over the same 1,000 held-out images, using the frozen 187-FM restriction defined in Appendix B. The ROI subspace is the column space of an intercept and ten overlapping binary visual-ROI indicators, with rank 11. The complementary residual removes that subspace from both patterns. Table 4 reports complementary image-level readouts. Chance depends on the candidate set; two-versus-two matching and top-one identification are distinct tasks.

Table 4: Complementary image-pattern readouts for S5/ViT-Base.

Readout	Full	ROI subspace	Residual
Nonmatched-image Pearson r	-0.003	-0.002	-0.002
Randomized two-versus-two (%)	50.115	49.980	49.814
Response-geometry Spearman r	0.536	0.572	0.462
Aggregate R^2	0.128	0.295	0.098
Top-one, 2 candidates (%)	89.817	83.095	90.354
Top-one, 10 candidates (%)	55.698	41.743	58.006
Top-one, 100 candidates (%)	16.077	10.134	16.905

D COMPONENT LANDMARKS AND REGIONAL MIXTURES

Figure 4 presents PC1 and PC7. The complementary landmark panels below retain five fixed positions on each cortical PC: the negative endpoint, the lower quartile, the middle, the upper quartile, and the positive endpoint. Each position shows the FM nearest that coordinate and its Top4 image evidence. These are feature landmarks on an axis, rather than manipulated images or time steps. The individual FM labels can recur across axes because an FM has coordinates on all components. The complete 1,200 PC–descriptor comparisons, all 2,981 one- to three-axis configurations, and ten-by-187 ROI coefficients are provided as electronic tables.

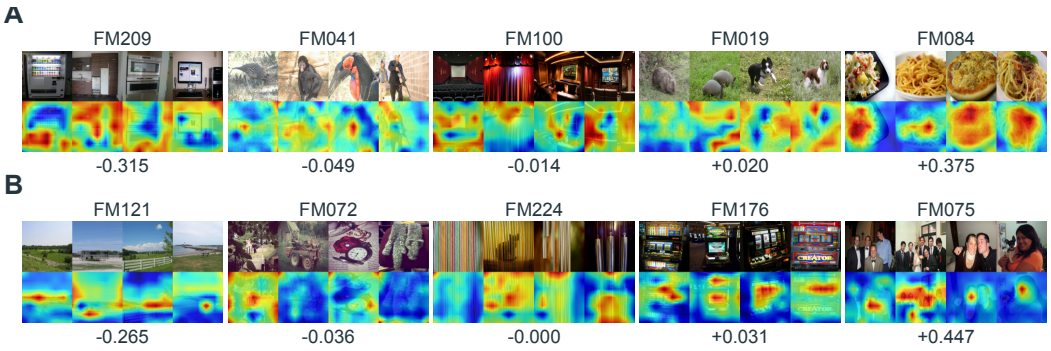


Figure 10: Fixed landmarks on PC2 and PC3 (A and B). Coordinates increase from left to right; the values below each plate are the FM component scores. The first image row contains originals and the second matched occlusion maps.

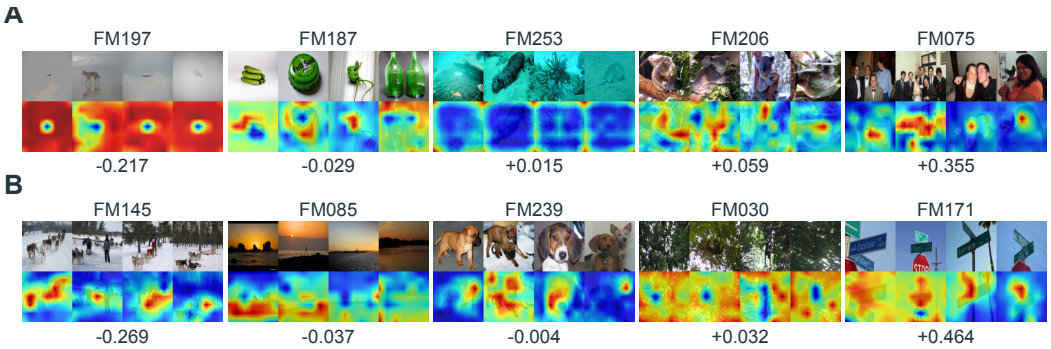


Figure 11: Fixed landmarks on PC4 and PC5 (A and B). Coordinates increase from left to right; the values below each plate are the FM component scores. The first image row contains originals and the second matched occlusion maps.

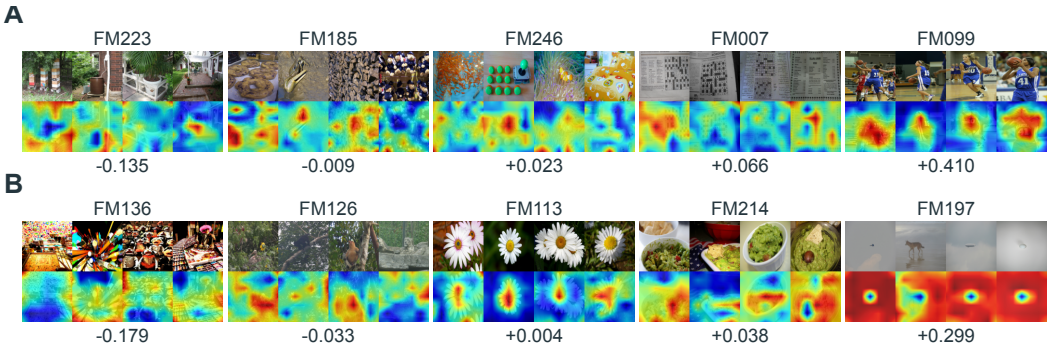


Figure 12: Fixed landmarks on PC6 and PC8 (A and B). Coordinates increase from left to right; the values below each plate are the FM component scores. The first image row contains originals and the second matched occlusion maps.



Figure 13: Fixed landmarks on PC9 and PC10 (A and B). Coordinates increase from left to right; the values below each plate are the FM component scores. The first image row contains originals and the second matched occlusion maps.

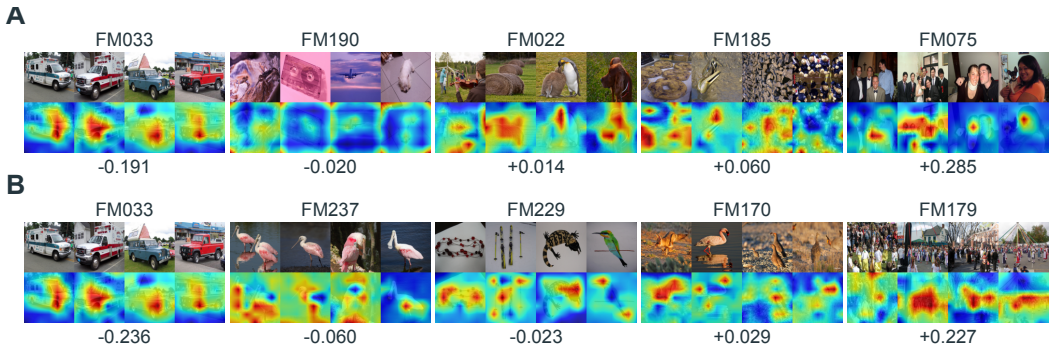


Figure 14: Fixed landmarks on PC11 and PC12 (A and B). Coordinates increase from left to right; the values below each plate are the FM component scores. The first image row contains originals and the second matched occlusion maps.

Figure 5 presents all 13 image-property profiles. The complete electronic association table also provides the object and scene descriptors, activation-frequency adjustments, and maximum-statistic probabilities for every PC–descriptor pair. Descriptor counts matter for endpoint interpretation: for

Table 5: Strongest positive and negative descriptor associations on each PC, by unadjusted signed Spearman correlation. Endpoint descriptors are selected by signed correlation; global tests include all 1,200 comparisons.

PC	Positive descriptor	ρ	Negative descriptor	ρ
1	dining table	0.355	person	-0.331
2	bear	0.175	handbag	-0.304
3	book	0.342	orientation anisotropy	-0.163
4	high-frequency texture	0.137	sink	-0.279
5	scissors	0.307	skateboard	-0.139
6	kite	0.074	potted plant	-0.363
7	orientation anisotropy	0.270	orientation entropy	-0.293
8	orientation anisotropy	0.241	orientation entropy	-0.239
9	gray entropy	0.200	toaster	-0.651
10	stop sign	0.125	bear	-0.161
11	toothbrush	0.268	outdoor scene	-0.192
12	luminance	0.134	laptop	-0.154

example, the toaster association on PC9 concerns one image. Frequency-adjusted and maximum-statistic columns are retained in the full table. The main-text example associations are PC1 with person ($\rho = -0.331$), bowl (0.350), and dining-table preference (0.355); PC3 with bowl (0.330) and dining table (0.334); and PC7 with edge energy (-0.276), edge density (-0.281), orientation entropy (-0.293), and anisotropy (0.270). Descriptive two-axis fits yield rank $R^2 = 0.226$ for dining-table preference using PC1/3 and $R^2 = 0.174$ for orientation entropy using PC7/8. These fits summarize variation across the 187 FMs; the separate image-half assessment evaluates neighborhood consistency across image subsets. The ROI table reports signed mean, mean absolute coefficient, enrichment, and within-ROI mass as separate quantities. Regional enrichment identifies a relative concentration of a mode; within-ROI mass describes its contribution relative to the other eligible modes.

E FUNCTIONAL-MODULE COLOR KEY FOR THE RGB ATLAS

FM landmarks connect continuous color coordinates to visual preferences. The same three-component transform colors cortical locations and projects individual FM directions into that space. Figure 15 joins three marked FM directions to their highest-response images, matched occlusion maps, and signed S5 encoding fields. The images and occlusion maps characterize the model response; the cortical maps show fitted neural coefficients. Qwen descriptions accompany the image evidence. A cortical color summarizes a local mixture of FMs; a landmark identifies a particular feature direction within that continuous space.

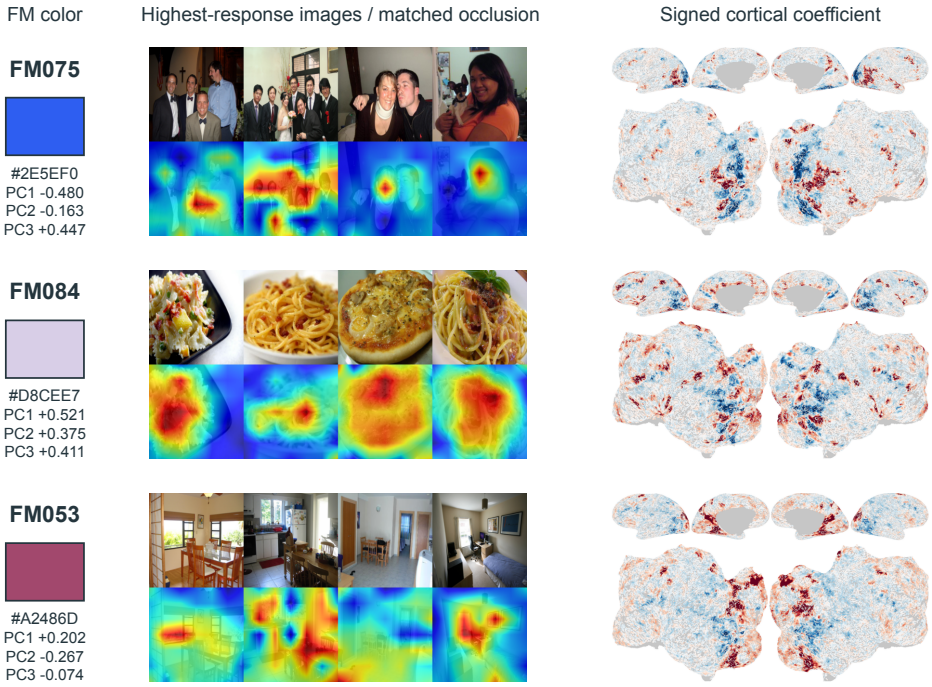


Figure 15: Three landmarks for reading the RGB field. Each swatch is the projected unit-FM direction, with its first three coordinates printed below. The middle column shows the four highest-response images above their matched local occlusion maps; the right column shows the corresponding native signed cortical coefficient field. Qwen describes FM075 in terms of formal attire and neckwear, FM084 in terms of plated cooked food, and FM053 in terms of indoor room layouts with Japanese-style elements (full descriptions in Appendix G). The visible FM053 examples support the broader indoor-layout reading, while the more specific cultural phrasing remains Qwen's description. The joint display connects the model-side descriptions with their fitted cortical distributions.

For the complete lookup, each unit FM is projected into the frozen basis as $\mathbf{c}_m = (\mathbf{e}_m - \boldsymbol{\mu})\mathbf{B}$, where $\mathbf{e}_m$ is the unit vector for FM m. Red, green, and blue use PC1, PC2, and PC3, respectively, with cortical display scales $q = (0.750522,\ 0.611925,\ 0.505055)$: $\mathrm{RGB}_j = \mathrm{clip}[\frac{1}{2} + \frac{1}{2}c_{mj}/q_j, 0, 1]$. Swatches and hexadecimal values use full-precision coordinates; printed coordinates are rounded to three decimals. PC signs are arbitrary coordinate choices, and color is only a display of position in this basis. Table 6 supplies all 187 FM directions; the entries identify feature directions within the continuous coordinates.

Table 6: Display color of every FM in the 187-FM cortical projection. Hexadecimal colors are rounded to the nearest eight-bit channel.

FM	Swatch	Hex RGB	PC1	PC2	PC3
FM000		#797A85	-0.041	-0.026	+0.022
FM001		#82718A	+0.017	-0.069	+0.041
FM005		#7A8975	-0.031	+0.048	-0.041
FM006		#78697C	-0.044	-0.108	-0.013
FM007		#7F698F	-0.005	-0.109	+0.063
FM008		#745D7D	-0.067	-0.165	-0.011
FM009		#78777F	-0.046	-0.039	-0.002
FM011		#7A8874	-0.030	+0.039	-0.046
FM012		#88817D	+0.047	+0.009	-0.008
FM013		#719071	-0.083	+0.081	-0.058
FM014		#806E63	+0.006	-0.082	-0.112
FM016		#8B7079	+0.070	-0.076	-0.027
FM017		#7F6F8E	-0.004	-0.077	+0.059
FM018		#7C6397	-0.020	-0.135	+0.095
FM019		#788480	-0.043	+0.020	+0.003
FM020		#7B808B	-0.027	+0.000	+0.047
FM021		#7F7C83	-0.003	-0.017	+0.015
FM022		#7D7981	-0.015	-0.029	+0.005
FM025		#7D847C	-0.013	+0.020	-0.015
FM026		#838583	+0.019	+0.027	+0.012
FM027		#807F7F	+0.002	-0.003	-0.002
FM030		#867C85	+0.038	-0.019	+0.024
FM031		#787F7E	-0.042	-0.003	-0.005
FM032		#8B73A2	+0.069	-0.061	+0.136
FM033		#7D856F	-0.018	+0.028	-0.064
FM034		#7A8375	-0.034	+0.017	-0.041
FM036		#8566A3	+0.032	-0.122	+0.141
FM037		#6E9267	-0.101	+0.087	-0.098
FM038		#78879A	-0.043	+0.036	+0.104
FM041		#7D7579	-0.015	-0.049	-0.025
FM042		#649568	-0.161	+0.105	-0.094
FM043		#7F7983	-0.001	-0.030	+0.015
FM044		#847E7E	+0.025	-0.008	-0.007
FM045		#7B727F	-0.027	-0.064	-0.000
FM046		#818783	+0.011	+0.038	+0.014
FM048		#786A90	-0.047	-0.103	+0.065
FM050		#7F8587	-0.004	+0.025	+0.029
FM051		#7A6D8E	-0.030	-0.088	+0.058
FM052		#828470	+0.012	+0.024	-0.061
FM053		#A2486D	+0.202	-0.267	-0.074
FM055		#7D867C	-0.018	+0.029	-0.015
FM056		#827D88	+0.016	-0.012	+0.034
FM057		#7B8B79	-0.025	+0.054	-0.024
FM059		#7F758C	-0.005	-0.052	+0.048
FM060		#757A8F	-0.064	-0.027	+0.060
FM062		#856B7B	+0.034	-0.099	-0.019
FM063		#82738A	+0.012	-0.061	+0.041
FM064		#83817B	+0.020	+0.005	-0.017
FM065		#7F7C8D	-0.003	-0.018	+0.054
FM067		#7C8473	-0.020	+0.022	-0.049
FM068		#81827A	+0.006	+0.011	-0.020
FM069		#7A8084	-0.032	+0.004	+0.019
FM070		#81807D	+0.010	+0.002	-0.010

Continued on next page

Table 6: FM color key, continued.

FM	Swatch	Hex RGB	PC1	PC2	PC3
FM072		#868676	+0.035	+0.033	-0.036
FM073		#818480	+0.007	+0.020	+0.002
FM074		#77806C	-0.047	+0.001	-0.077
FM075		#2E5EF0	-0.480	-0.163	+0.447
FM077		#8B809C	+0.066	+0.004	+0.113
FM078		#83877E	+0.019	+0.038	-0.007
FM079		#7D7986	-0.016	-0.033	+0.027
FM080		#768F9F	-0.053	+0.076	+0.123
FM081		#84826F	+0.029	+0.014	-0.066
FM083		#8B8477	+0.070	+0.020	-0.033
FM084		#D8CEE7	+0.521	+0.375	+0.411
FM085		#7E7269	-0.010	-0.067	-0.089
FM087		#807774	+0.003	-0.039	-0.044
FM089		#86767E	+0.036	-0.045	-0.006
FM090		#728A6C	-0.082	+0.050	-0.078
FM093		#778778	-0.053	+0.034	-0.031
FM096		#82767E	+0.013	-0.044	-0.005
FM098		#7B7879	-0.024	-0.038	-0.026
FM099		#577579	-0.240	-0.050	-0.028
FM100		#777C84	-0.053	-0.014	+0.018
FM101		#83866C	+0.023	+0.033	-0.077
FM103		#906F7F	+0.095	-0.077	-0.004
FM104		#837773	+0.019	-0.042	-0.051
FM106		#7D7A81	-0.017	-0.026	+0.006
FM107		#748A72	-0.065	+0.051	-0.055
FM109		#827F71	+0.015	-0.004	-0.058
FM110		#769468	-0.055	+0.099	-0.095
FM111		#788490	-0.043	+0.021	+0.065
FM112		#847B8A	+0.024	-0.020	+0.041
FM113		#7B8271	-0.024	+0.013	-0.058
FM116		#7B8A76	-0.027	+0.048	-0.039
FM118		#758579	-0.062	+0.024	-0.026
FM119		#7C8084	-0.023	+0.004	+0.016
FM121		#8C493D	+0.074	-0.261	-0.265
FM122		#798079	-0.036	+0.001	-0.025
FM123		#885D9A	+0.051	-0.164	+0.106
FM124		#767A85	-0.055	-0.026	+0.021
FM125		#83728D	+0.018	-0.067	+0.052
FM126		#797581	-0.040	-0.050	+0.005
FM127		#7E827B	-0.009	+0.013	-0.016
FM128		#807C85	+0.001	-0.018	+0.021
FM129		#827078	+0.013	-0.074	-0.028
FM130		#7B7979	-0.026	-0.032	-0.024
FM132		#8F718E	+0.090	-0.069	+0.057
FM134		#7B937A	-0.026	+0.093	-0.022
FM135		#7F7A73	-0.004	-0.026	-0.049
FM136		#9048A6	+0.100	-0.266	+0.151
FM138		#78827F	-0.045	+0.013	-0.003
FM139		#7D7C86	-0.014	-0.018	+0.024
FM140		#817D88	+0.006	-0.014	+0.033
FM142		#807787	+0.001	-0.043	+0.030
FM144		#827D7C	+0.015	-0.010	-0.016
FM145		#6F6E86	-0.094	-0.086	+0.027
FM146		#807786	+0.002	-0.039	+0.025
FM147		#7B717C	-0.029	-0.070	-0.013
FM148		#817A7D	+0.009	-0.028	-0.011
FM149		#7E7786	-0.012	-0.040	+0.026
FM151		#7B8376	-0.026	+0.019	-0.038
FM152		#86866B	+0.040	+0.033	-0.082
FM154		#8F7960	+0.091	-0.029	-0.126
FM157		#806F95	+0.001	-0.080	+0.085
FM158		#7E777E	-0.007	-0.041	-0.005
FM161		#7D9374	-0.015	+0.093	-0.044

Continued on next page

Table 6: FM color key, continued.

FM	Swatch	Hex RGB	PC1	PC2	PC3
FM162		#627177	-0.171	-0.070	-0.033
FM163		#848489	+0.027	+0.022	+0.037
FM164		#7F827F	-0.006	+0.014	-0.000
FM166		#797484	-0.040	-0.056	+0.016
FM168		#867690	+0.039	-0.044	+0.066
FM170		#757E84	-0.063	-0.007	+0.017
FM171		#89617B	+0.054	-0.145	-0.016
FM172		#80807E	+0.000	+0.002	-0.006
FM173		#7F828E	-0.001	+0.011	+0.058
FM174		#809071	+0.001	+0.079	-0.058
FM175		#817876	+0.009	-0.034	-0.039
FM176		#827587	+0.016	-0.050	+0.031
FM177		#856A74	+0.033	-0.105	-0.046
FM178		#827783	+0.017	-0.041	+0.014
FM179		#774C85	-0.051	-0.247	+0.022
FM180		#7F8483	-0.001	+0.021	+0.015
FM183		#836A86	+0.018	-0.104	+0.025
FM184		#8D6651	+0.080	-0.124	-0.185
FM185		#807581	+0.003	-0.048	+0.005
FM187		#817894	+0.008	-0.038	+0.081
FM190		#7E7099	-0.010	-0.074	+0.100
FM191		#797E8B	-0.039	-0.008	+0.044
FM192		#81757E	+0.010	-0.052	-0.005
FM193		#817980	+0.010	-0.030	+0.002
FM194		#7E7C7D	-0.008	-0.017	-0.011
FM195		#7C797F	-0.021	-0.030	-0.001
FM197		#7A9E59	-0.030	+0.145	-0.154
FM198		#7F7F72	-0.002	-0.005	-0.052
FM199		#79846E	-0.041	+0.021	-0.069
FM200		#7D7D80	-0.015	-0.012	+0.001
FM201		#8D7A76	+0.082	-0.025	-0.037
FM203		#7F7A81	-0.002	-0.025	+0.005
FM205		#837183	+0.020	-0.069	+0.015
FM206		#808783	+0.000	+0.035	+0.012
FM207		#717B84	-0.086	-0.020	+0.020
FM208		#7A8375	-0.035	+0.018	-0.040
FM209		#B33E6D	+0.301	-0.315	-0.072
FM210		#7F827F	-0.004	+0.011	-0.002
FM214		#7A8072	-0.035	+0.003	-0.055
FM215		#848C7B	+0.029	+0.061	-0.016
FM218		#826D7B	+0.016	-0.088	-0.016
FM219		#857D89	+0.032	-0.013	+0.037
FM220		#7D7F8A	-0.016	-0.000	+0.041
FM221		#798A73	-0.036	+0.050	-0.049
FM222		#7B9481	-0.027	+0.099	+0.005
FM223		#837B67	+0.020	-0.019	-0.099
FM224		#7E7E7F	-0.009	-0.005	-0.000
FM225		#778089	-0.052	+0.000	+0.039
FM226		#7D7281	-0.017	-0.063	+0.007
FM227		#82898D	+0.015	+0.044	+0.055
FM229		#79828D	-0.040	+0.011	+0.053
FM230		#7B8573	-0.025	+0.027	-0.048
FM231		#777693	-0.049	-0.045	+0.077
FM233		#7C848A	-0.023	+0.021	+0.043
FM234		#747B80	-0.070	-0.022	+0.003
FM235		#7F859D	-0.001	+0.024	+0.117
FM237		#7D8180	-0.016	+0.010	+0.003
FM238		#837686	+0.023	-0.046	+0.028
FM239		#767C8A	-0.057	-0.015	+0.042
FM240		#828053	+0.015	+0.002	-0.178
FM242		#798F8A	-0.037	+0.075	+0.041
FM243		#888282	+0.049	+0.013	+0.009
FM244		#837D88	+0.019	-0.013	+0.033

Continued on next page

Table 6: FM color key, continued.

FM	Swatch	Hex RGB	PC1	PC2	PC3
FM245		#796B8F	-0.039	-0.098	+0.062
FM246		#897996	+0.053	-0.031	+0.089
FM248		#817B90	+0.010	-0.024	+0.067
FM249		#847685	+0.026	-0.047	+0.023
FM252		#7B8081	-0.029	+0.005	+0.005
FM253		#797F6A	-0.038	-0.004	-0.085
FM254		#798678	-0.038	+0.033	-0.028
FM255		#7B856E	-0.029	+0.029	-0.068

E.1 GEOMETRY DISPLAYS AND COMPARISON

The same 75 families are shown in all 23 two-dimensional geometries (Figures 16–19). The gray edges are the 67 mutual-nearest-neighbor links defined in the original joint feature space before projection; disconnects and crossings in a displayed plane do not redefine the graph. Filled points denote the 62 primary spatial cores and open points the remaining 13 candidates. G08 (equal-weight inputs, $k = 8$) is the prespecified broad-scale diffusion comparison; the other diffusion settings, PCA, UMAP, t-SNE, and the CEBRA centroid display are complementary projections, not independent validation. The v23–v25 labels distinguish display variants. Figure 20 shows the separate 2,775-map CEBRA-to-UMAP projection for all eligible FMs, colored by participant, model, and family support.

A complementary four-participant CEBRA screen fitted all 2,775 eligible maps on the same 1,000 shared NSD images in 61 completed configurations: 52 auxiliary-label/temperature/delta grid fits, four initial controls, and five raw-input follow-ups. This screen varied continuous stimulus, model, and cortical-derived label blocks and input representations; its full configuration table accompanies the paper. These fits concern individual maps rather than the 23 family-prototype displays below. A separate hard-partition diagnostic used three other NSD participants (S3, S6, S8), 744 model-local maps, and 513 encoding-supported vertices. Under its tested $k = 2$–20 validation criteria, none of the 19 partitions passed the combined hard-partition criteria; the gap one-standard-error rule selected $k = 1$, although all 19 passed the matched continuous-null comparison. This diagnostic limits a global discrete-taxonomy interpretation under those criteria; it does not test the present 75-family local recurrence or establish a biological absence of communities.

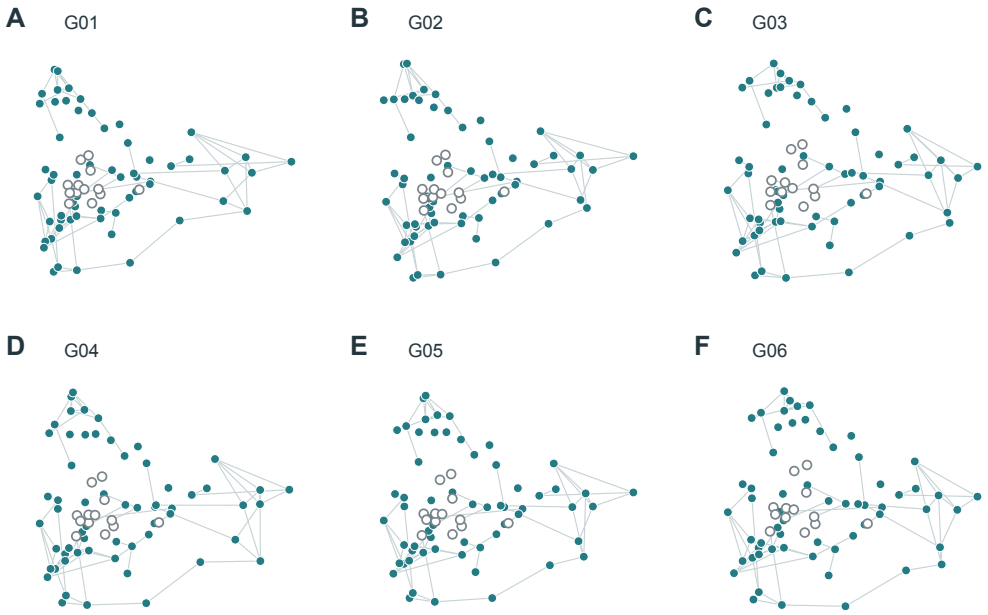


Figure 16: Family geometry projections G01–G06. Panels A–F follow Table 7 in row order. Graph edges use the original joint-space neighborhoods.

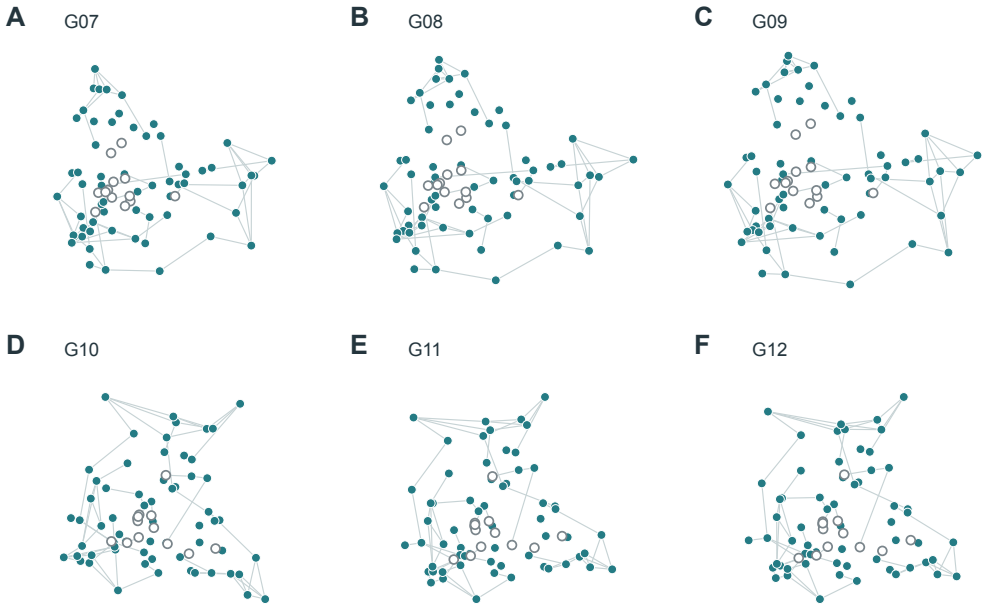


Figure 17: Family geometry projections G07–G12. Panels A–F follow Table 7 in row order. Graph edges use the original joint-space neighborhoods.

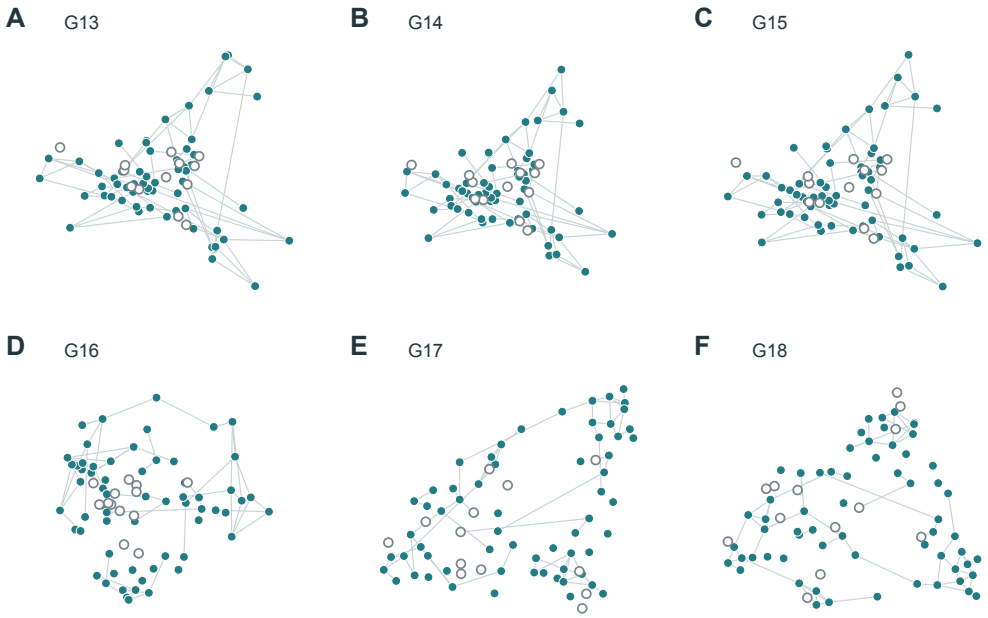


Figure 18: Family geometry projections G13–G18. Panels A–F follow Table 7 in row order. Graph edges use the original joint-space neighborhoods.

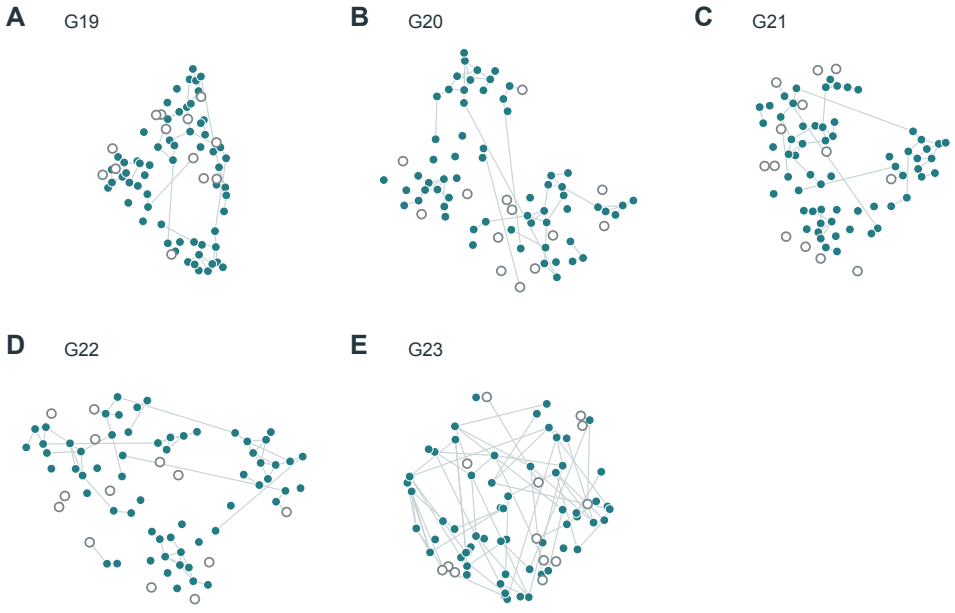


Figure 19: Family geometry projections G19–G23. Panels A–E follow Table 7 in row order. Graph edges use the original joint-space neighborhoods.

Table 7: Comparison metrics for all 23 family displays. Trust is five-neighbor trustworthiness relative to the original joint input; overlap is original-versus-displayed five-neighbor overlap; distance ρ is the rank correlation of pairwise distances. Values are rounded to three decimal places.

ID	Method	Joint trust	5NN overlap	Distance ρ
G01	Diff. w=0, k=4	0.805	0.280	0.608
G02	Diff. w=0, k=8	0.806	0.272	0.632
G03	Diff. w=0, k=12	0.811	0.283	0.677
G04	Diff. w=0.25, k=4	0.811	0.285	0.649
G05	Diff. w=0.25, k=8	0.812	0.285	0.667
G06	Diff. w=0.25, k=12	0.813	0.283	0.690
G07	Diff. w=0.5, k=4	0.815	0.285	0.673
G08	Diff. w=0.5, k=8	0.816	0.285	0.688
G09	Diff. w=0.5, k=12	0.815	0.283	0.698
G10	Diff. w=0.75, k=4	0.816	0.269	0.641
G11	Diff. w=0.75, k=8	0.821	0.275	0.667
G12	Diff. w=0.75, k=12	0.819	0.275	0.671
G13	Diff. w=1, k=4	0.659	0.192	0.226
G14	Diff. w=1, k=8	0.654	0.179	0.244
G15	Diff. w=1, k=12	0.658	0.184	0.250
G16	Joint PCA	0.814	0.275	0.731
G17	UMAP v23	0.931	0.528	0.687
G18	UMAP v24	0.918	0.504	0.607
G19	UMAP v25	0.922	0.531	0.672
G20	t-SNE v23	0.929	0.549	0.611
G21	t-SNE v24	0.923	0.560	0.620
G22	t-SNE v25	0.924	0.541	0.635
G23	CEBRA centroids	0.588	0.152	0.159

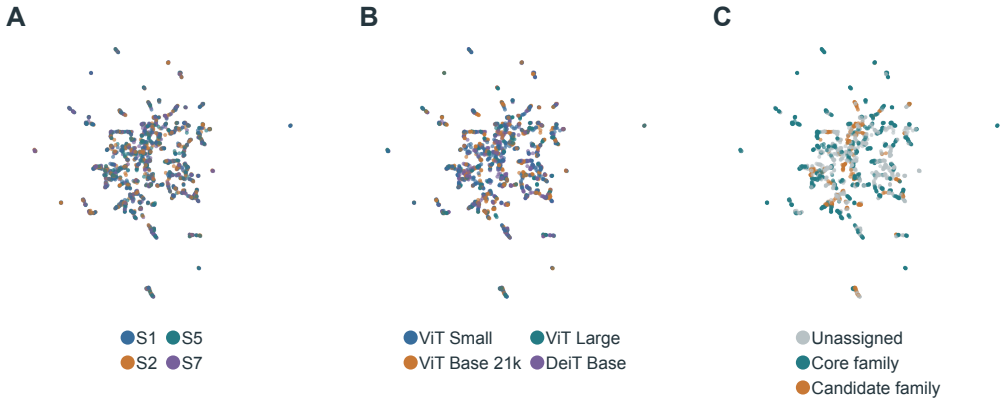


Figure 20: The same 2,775-map embedding colored by participant, model, and membership in the recurrent dictionary. The coordinates are identical across views; color changes only the displayed grouping.

E.2 Continuous-attribute associations

Figure 21 and its continuations provide all 975 family–attribute rank correlations, including weak and unsupported directions. Dots identify associations meeting the joint criteria in Appendix B. Family identifiers retain the community indices assigned before selection for coverage of all 16 combinations. Of the 975 comparisons, 339 meet the joint support criteria. All retain their aggregate sign in each participant aggregate, and 319 do so in each model aggregate. At least 75% of family members share the aggregate sign for 322 associations; 271 retain it after multivariable adjustment. For example, C26 correlates with edge energy ($\rho = 0.680$), high-frequency texture (0.654), and colorfulness (0.384), while C51 correlates with luminance (0.731). Sixty-three families combine semantic and continuous-attribute associations, ten have semantic associations alone, and one continuous-attribute associations alone.

E.3 Condition-level family participation

Figure 23 and its continuations show the complete 82-by-62 spatial-core incidence pattern. Three row blocks follow the condition order; each is divided into two 31-family column blocks. The colors distinguish common participation from marginal preference, retaining conditions that recruit a broadly active family without a positive within-versus-outside preference and vice versa. Figure 26 shows expanded-set counts under nine fixed rate/cell-support thresholds and Jaccard overlap after each whole participant or model is omitted. The leave-group display retains the relative-preference layer fixed while reevaluating common participation in the remaining 12 cells, as defined in the group-omission analysis. These checks quantify set stability across the four participants and four model dictionaries. Observed family calls need not occur simultaneously in any one image. Figure 27 juxtaposes the fraction of observed core-family calls covered by the relative-preference and expanded sets with per-image 10th, median, and 90th percentile counts of simultaneous family calls. The exploratory co-recruitment display contains 2, 2, 7, 30, 20, and 21 eligible edges for dog, train, dining, face context, high colorfulness, and high-frequency texture, respectively. These pointwise-bootstrap-filtered edges summarize within-cell excess co-calling and are not directed connections or multiplicity-adjusted discoveries.

Co-recruitment among selected families. The same condition repertoire can contain families that tend to be called together within an image and participant–model cell. Figure 28 displays the exploratory co-recruitment graphs for dog and high colorfulness, with family positions fixed by the joint-input atlas projection. The corresponding real-image call examples are in Figure 9. The graph edges summarize excess co-calling over the product of the two marginal rates. This within-image relationship complements the similarity of family profiles represented by the atlas graph. Complete node memberships and pairwise edge records accompany the paper.

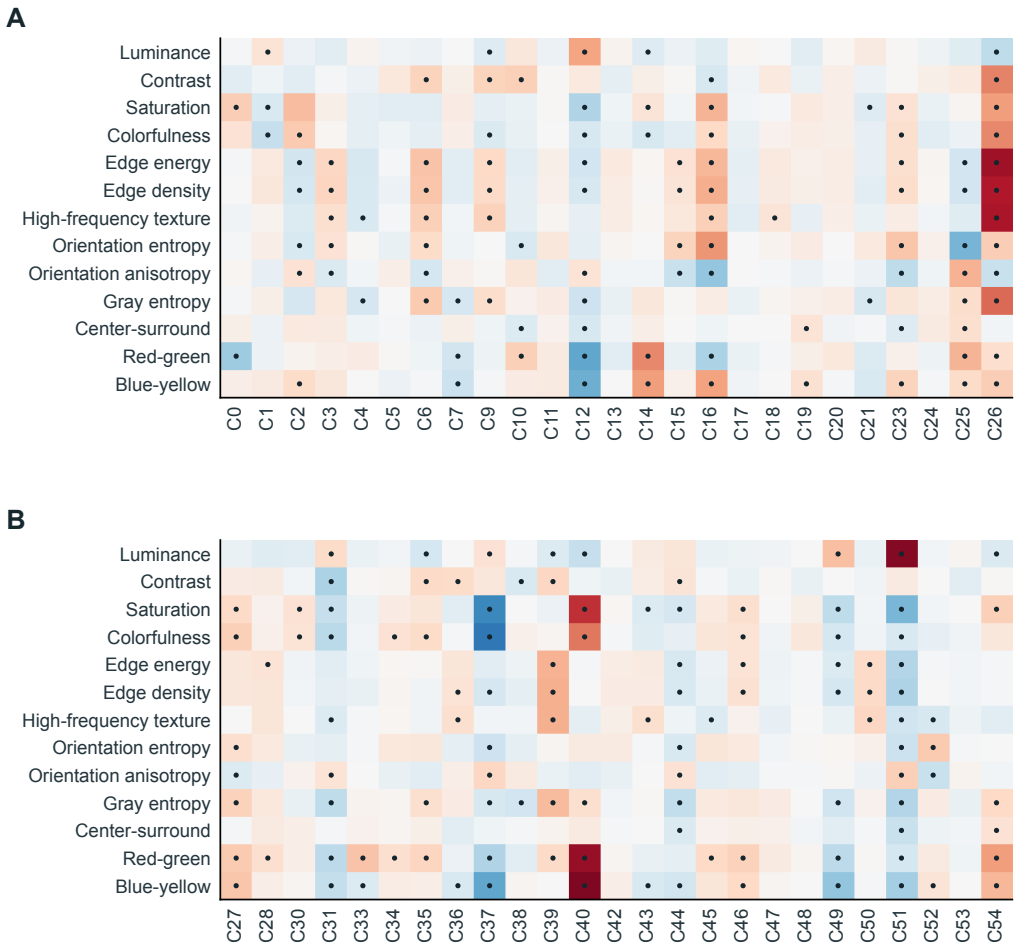


Figure 21: Signed family–attribute rank correlations, first 50 families (A–B). Rows are the 13 measured NSD image descriptors; columns are family IDs. Black dots denote the support criteria (global BH threshold, effect size, and cell/participant sign consistency). Both panels use the same $[-0.8, 0.8]$ Spearman ρ scale.

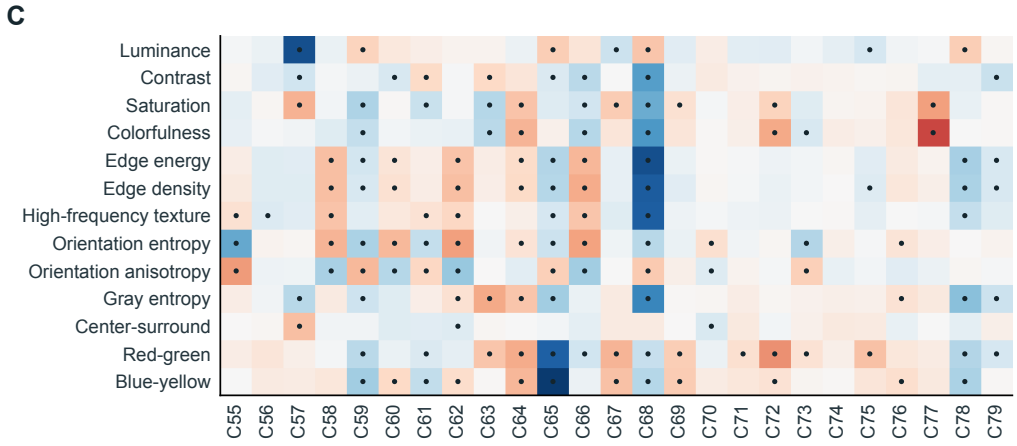


Figure 22: Signed family–attribute rank correlations, remaining 25 families (C). Display conventions and correlation scale match Figure 21.

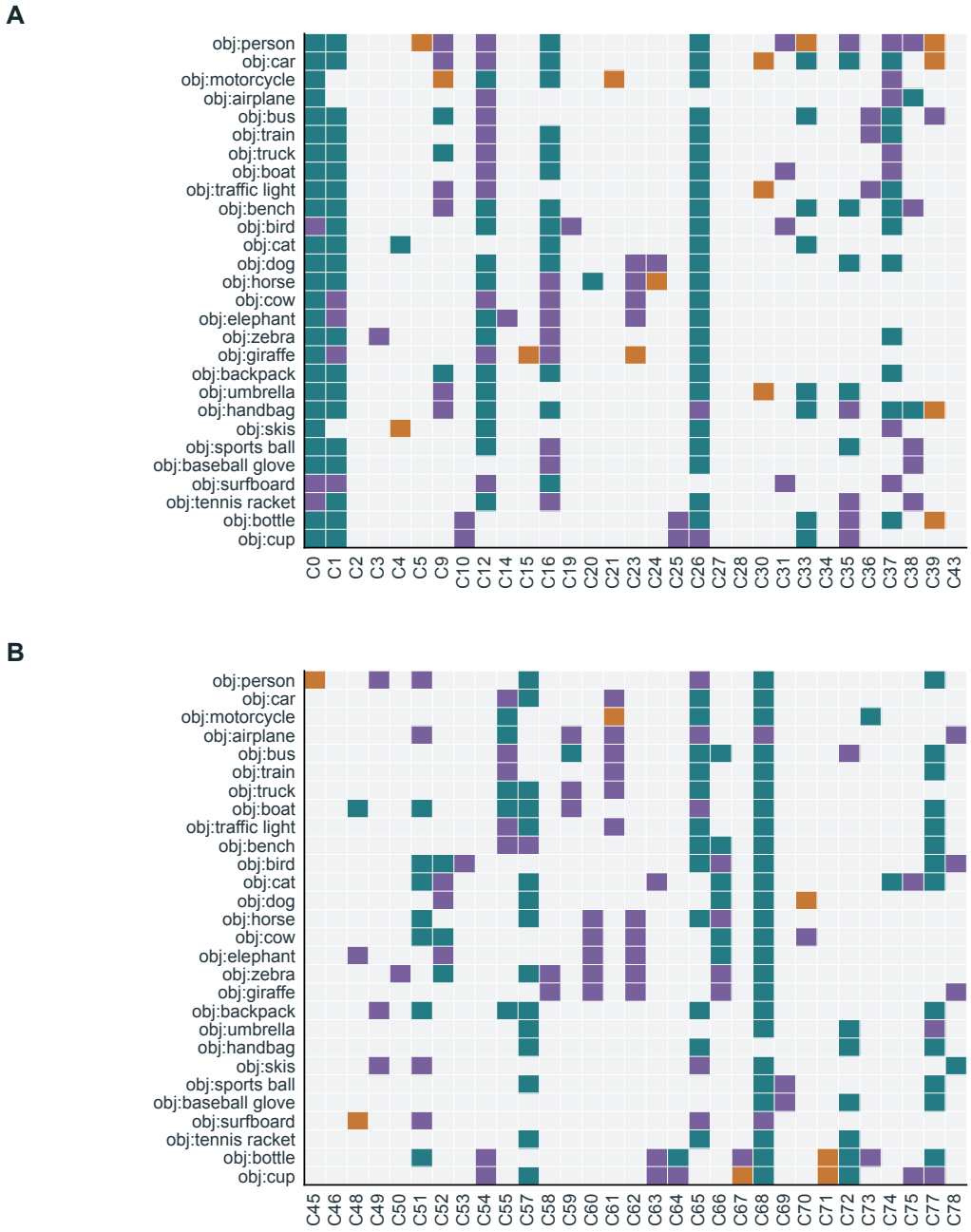


Figure 23: Condition–family incidence, row block 1 of 3 (panels A and B). The upper panel contains the first 31 spatial-core community IDs and the lower panel the remaining 31. Gray: neither criterion; orange: relative preference alone; teal: common participation alone; purple: both. obj, object; sc, scene; ctx, context.

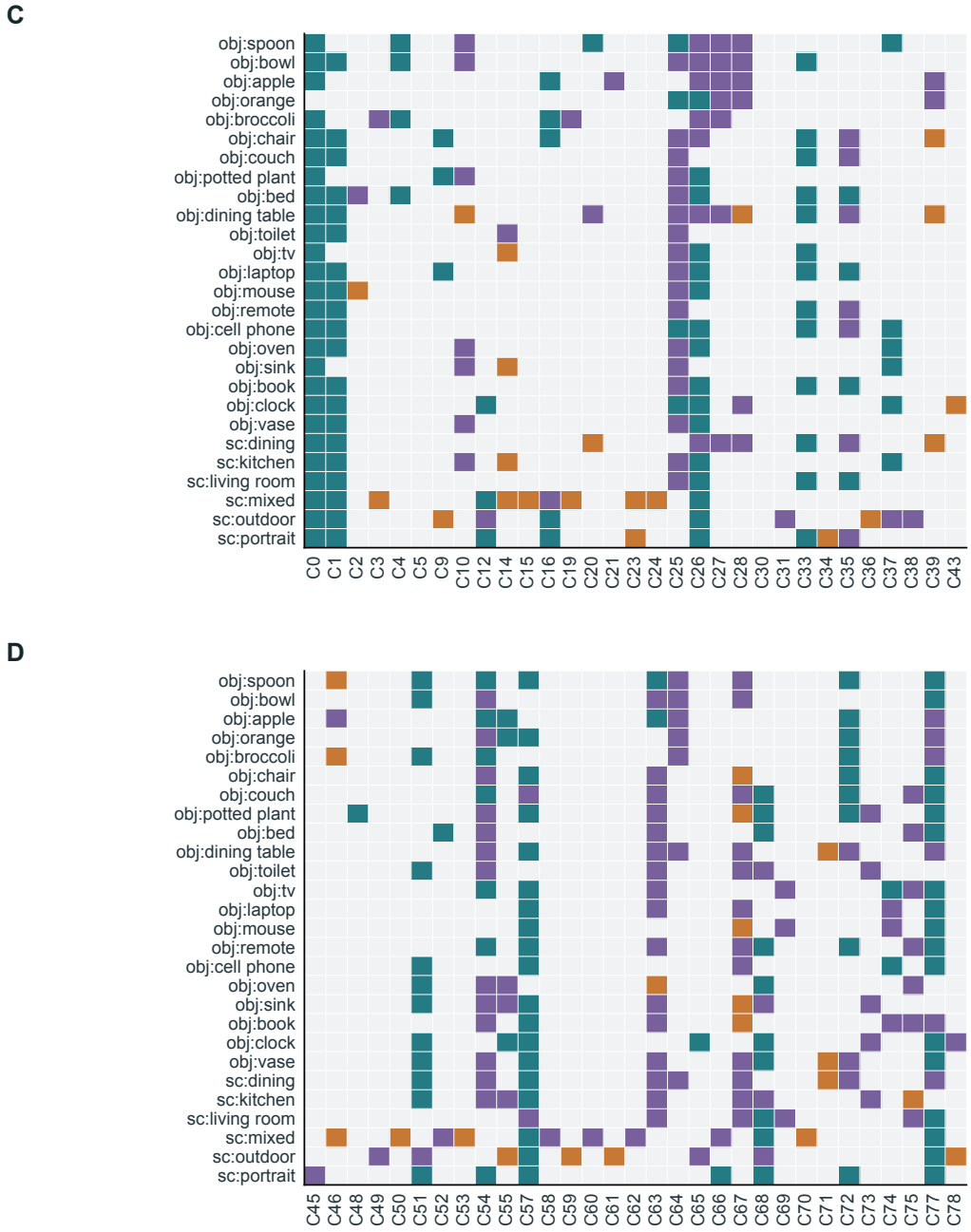


Figure 24: Condition–family incidence, row block 2 of 3 (panels C and D). The upper panel contains the first 31 spatial-core community IDs and the lower panel the remaining 31. Gray: neither criterion; orange: relative preference alone; teal: common participation alone; purple: both. obj, object; sc, scene; ctx, context.

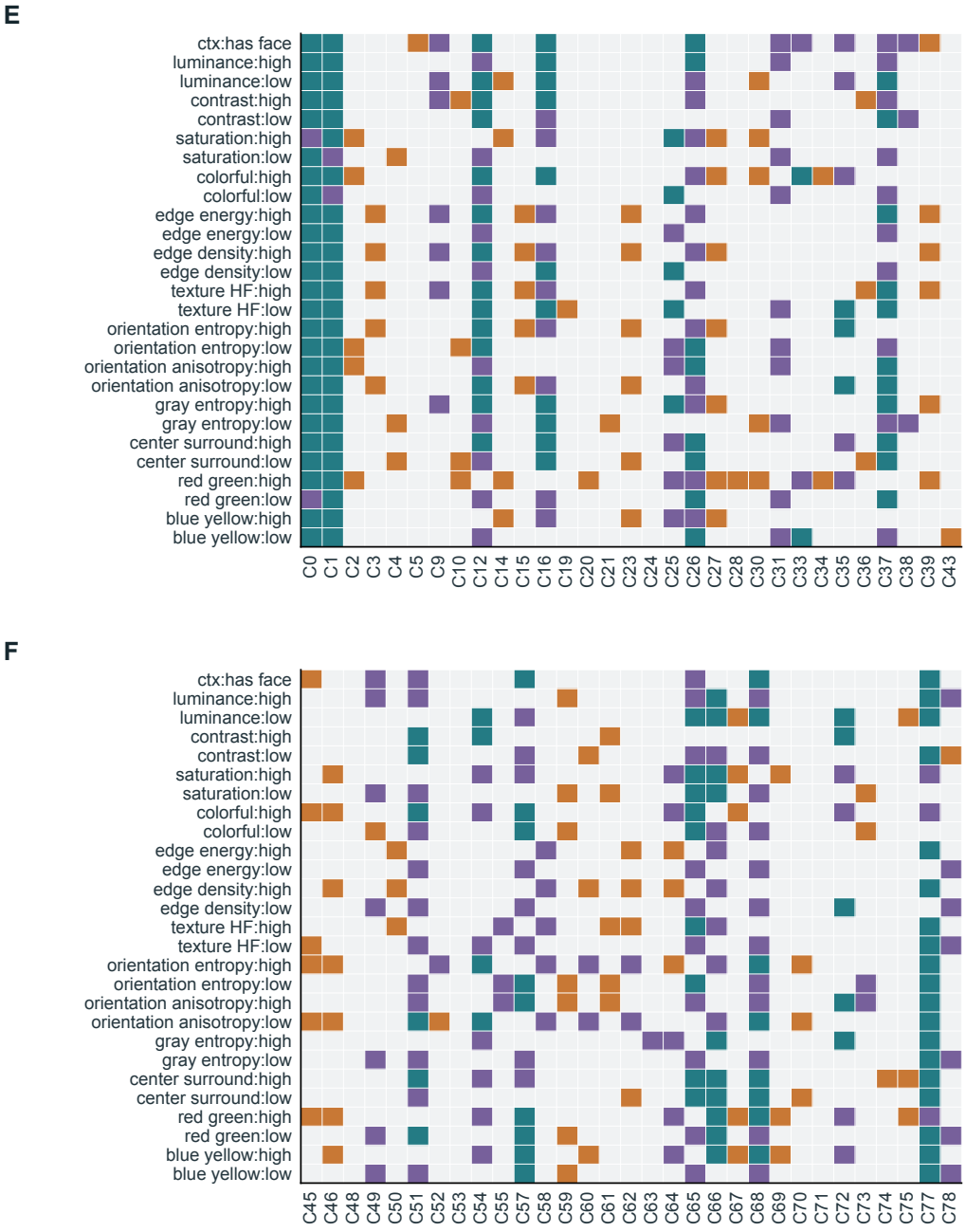


Figure 25: Condition–family incidence, row block 3 of 3 (panels E and F). The upper panel contains the first 31 spatial-core community IDs and the lower panel the remaining 31. Gray: neither criterion; orange: relative preference alone; teal: common participation alone; purple: both. obj, object; sc, scene; ctx, context.

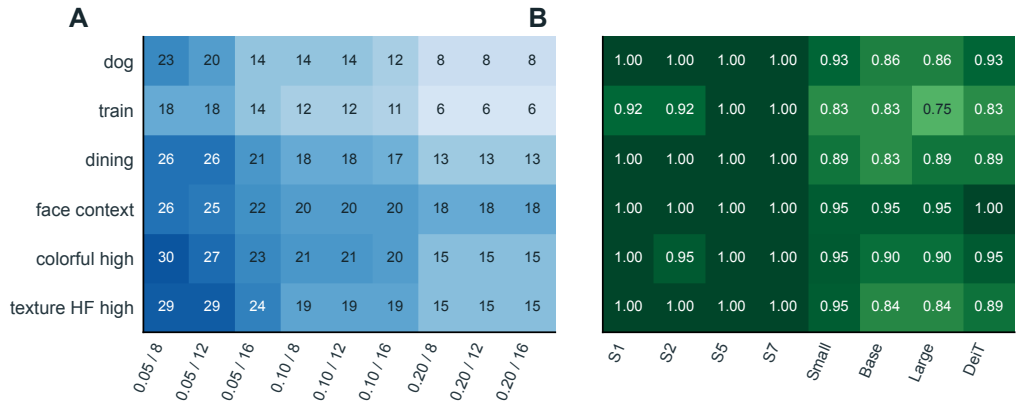


Figure 26: Six illustrative conditions. (A) Number of families in the expanded union at call-rate threshold / minimum supporting cells shown on the horizontal axis. (B) Jaccard overlap of the expanded union with the primary rule after leaving out one participant or one model.

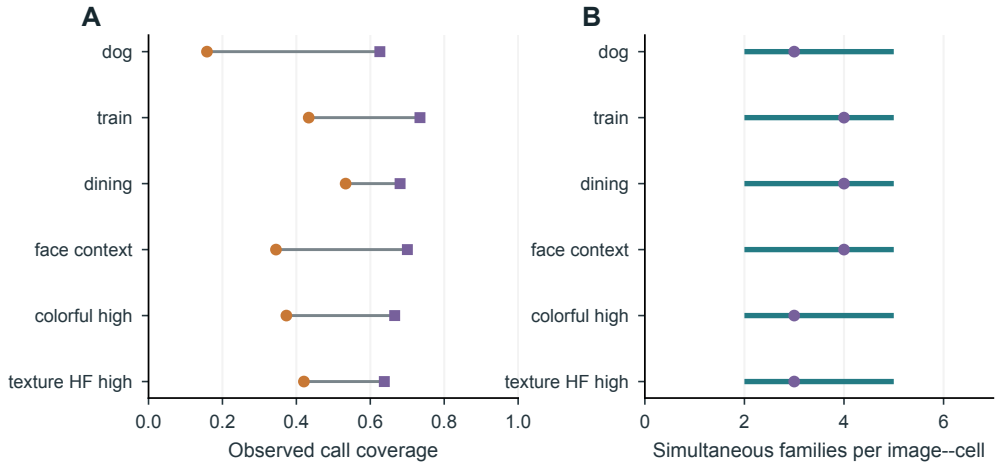


Figure 27: Six condition examples. (A) Fraction of observed core-family calls captured by the relative-preference set (orange circles) and expanded union (purple squares). (B) 10th–90th percentile interval of simultaneous family counts per image–cell (teal lines) and median (purple circles). Horizontal segments in A connect paired values solely for visual comparison.

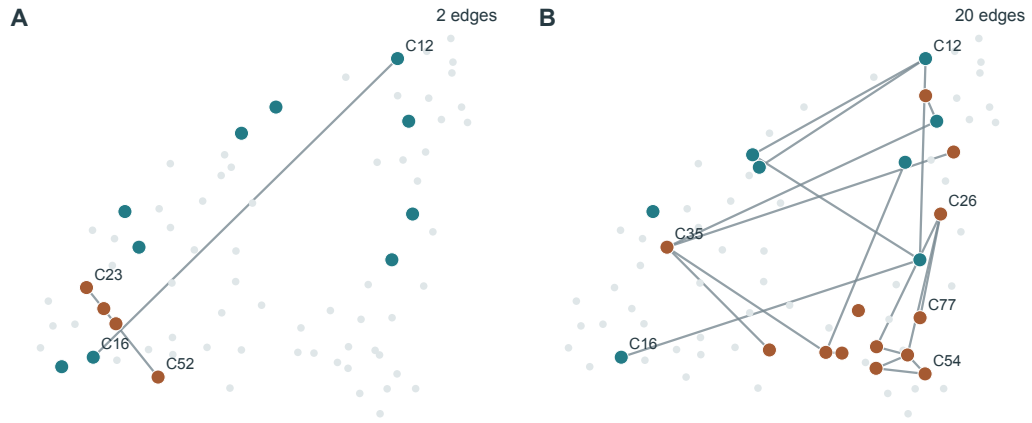


Figure 28: Exploratory co-recruitment for dog (A) and high colorfulness (B). Orange nodes are relatively preferred families; teal nodes are additional common participants; gray nodes are outside the displayed condition subatlas. The same fixed atlas coordinates and limits are used in both panels. Lines are co-recruitment edges that pass within-cell excess $\geq .01$, a positive pointwise image-bootstrap 95% lower bound, and positive participant and model aggregates in at least three of four groups each. The graphs contain 2 and 20 such edges, respectively. Pointwise intervals were not adjusted across the 9,346 tested family pairs. A few node IDs are printed for orientation; all IDs and edge endpoints are in the electronic tables.

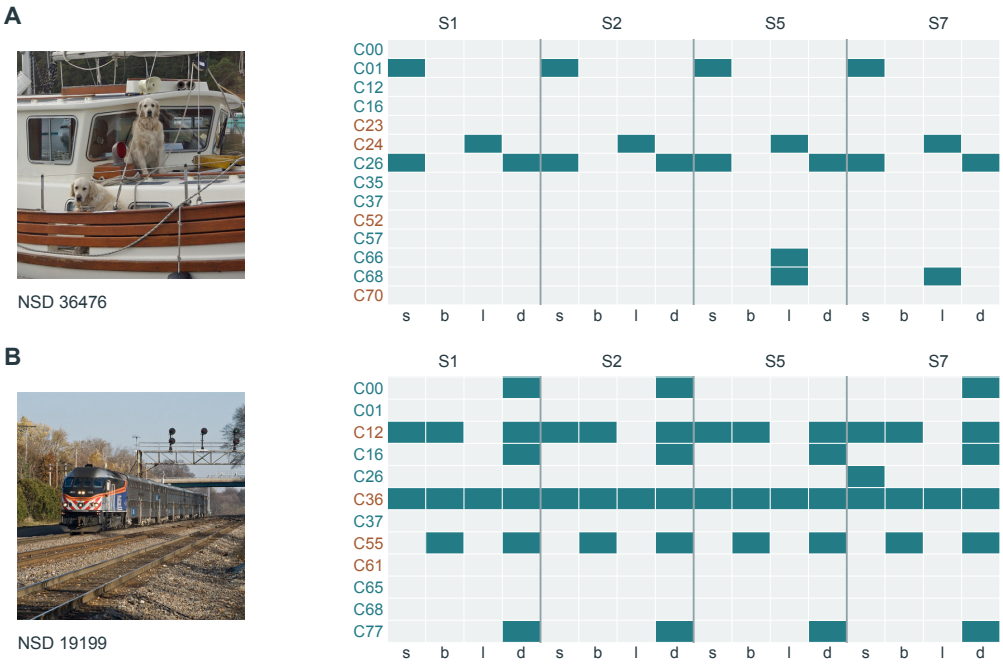


Figure 29: Combination-specific calls for the two images in Figure 9. Rows are condition-repertoire families and columns are the 16 participant–model combinations (s, ViT-Small; b, ViT-Base-21k; l, ViT-Large; d, DeiT-Base). A filled cell indicates a positive response from at least one member FM. The main figure summarizes each row by its number of active combinations; this view retains which combinations contributed.

E.4 IMAGENET-DERIVED MODEL-ONLY TRANSFER

Table 8 reports the 15 category roots meeting the prespecified minimum of 20 images in both corpora.

Table 8: Cross-corpus category results. n denotes images in the named corpus; positive is ImageNet-positive among NSD-selected families. q is BH-adjusted over the 15 eligible roots.

Category	NSD n	ImageNet n	ρ [bootstrap 95%]	q	Positive
dog	28	14,873	0.27 [0.10, 0.40]	0.0230	4/4
cat	28	908	0.35 [0.14, 0.46]	0.0037	3/3
bird	39	7,588	0.53 [0.36, 0.61]	0.0006	8/9
car	74	1,292	0.45 [0.29, 0.53]	0.0006	6/6
train	36	139	0.35 [0.15, 0.46]	0.0017	3/4
airplane	30	126	0.74 [0.60, 0.77]	0.0006	9/9
chair	84	457	0.59 [0.44, 0.62]	0.0006	7/8
person	438	405	0.45 [0.34, 0.51]	0.0006	6/13
horse	26	110	0.52 [0.27, 0.55]	0.0006	4/6
elephant	28	245	0.61 [0.51, 0.69]	0.0006	8/8
zebra	29	125	0.59 [0.44, 0.67]	0.0006	7/7
bus	25	379	0.53 [0.35, 0.57]	0.0006	5/6
truck	30	1,070	0.39 [0.22, 0.46]	0.0006	4/4
boat	33	794	0.59 [0.42, 0.62]	0.0006	5/5
motorcycle	21	146	0.47 [0.25, 0.54]	0.0006	4/4

The points in Figure 30 compare family-effect ranks between NSD and a separate ImageNet-derived image collection. Their intervals come from 499 stratified image-bootstrap replicates; Table 8 gives the same effects, bootstrap limits, BH-adjusted probabilities, and the count of NSD-selected positive families with positive ImageNet effects. All 15 eligible roots have positive rank correspondence, from $\rho = 0.272$ to 0.744, and 83 of 96 NSD-selected relations retain a positive ImageNet effect. Three additional roots were excluded because fewer than 20 NSD images were available (bicycle 18, sheep 16, bear 14). Cow and giraffe had zero images under their exact ImageNet WordNet-root closure and were excluded despite 23 and 44 NSD images. The comparison uses fixed model responses and family identities across corpora; it does not contain an ImageNet neural measurement.

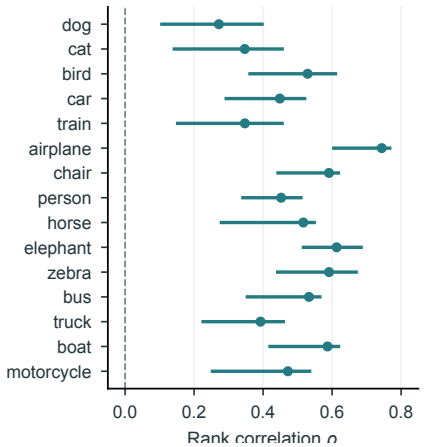


Figure 30: Family-effect rank correlations for 15 category roots. Bars: 95% intervals from 499 stratified image bootstraps; dashed line: zero.

F SENSITIVITY ANALYSES AND LIMITS OF INFERENCE

F.1 CORTICAL COORDINATES AND IMAGE DESCRIPTORS

The S5/ViT-Base PCA is computed from fitted cortical coefficients, whereas descriptor associations use model responses to the 1,000 NSD images. Training activity frequency (AF) is a strong potential confound for PC9: the PC correlates with AF at $\rho = .808$, and rank residualization removes all 44 initially BH-supported PC9 descriptor associations. A toaster label occurs in only one NSD image and cannot support an axis interpretation. Associations of PC7/8 with edge, texture, and orientation quantities remain modest after AF adjustment, but none of the 13 low/mid-level descriptor associations survives the stricter permutation maximum-statistic criterion. In image split halves, 25 of 30 one- to three-axis neighborhoods selected on one half remain above null on the other. The prespecified out-of-fold regional controls in Table 9, including weak and failed quadratic fits, test predictive generalization rather than the descriptive R^2 of fixed axes within the 187-FM matrix.

Three previously fixed ViT-Small coordinates offer a complementary test against measured fMRI rather than a descriptor correlation. FM169, FM217, and FM024 belong to the ImageNet-C256-K8 ViT-Small dictionary, not to the S5/ViT-Base PCA or automatically to any of the 75 families. Over the same 1,000 held-out NSD images, each coordinate correlates positively with target-area fMRI in

Table 9: Selected out-of-fold regional controls for 187 S5/ViT-Base FMs. Regional contrasts use fitted cortical coefficients; ImageNet descriptors are model-side. BH q spans 120 PC/readout hypotheses. Positive rank correlation need not imply calibrated out-of-fold R^2.

Readout	OOF R^2	ρ	BH q
Scene–body, PC1	.184	.683	.0027
ImageNet Food, PC1–3	.106	.368	.0027
ImageNet Vehicles, PC1–3	−.015	.137	.3543
Red–blue, PC1–3	.001	.136	.1129
Scene–body, quadratic PC1–3	−4.049	.733	.9611

S1, S2, S5, and S7. Across-participant Fisher means are $r = .414$, $.160$, and $.395$, with image-cluster bootstrap intervals $[.377, .449]$, $[.107, .208]$, and $[.338, .448]$. With four participants, the exact two-sided sign-flip probability is $.125$ for each coordinate. Readouts over the entire target areas remain positive, while discovery-weight-selected upper quintiles show stronger associations. These results connect fixed model coordinates to held-out neural responses in the same participants; they do not validate individual atlas families.

F.2 Spatial recurrence and activity frequency

We assessed fixed family memberships against null prototypes matched within each cell for member count. Leave-model signed vertex cosine has interquartile range $.589$–$.865$, compared with $.110$–$.361$ for leave-participant; the corresponding TY17 profile-correlation medians are $.863$ and $.718$. A more demanding check requires every one of the four omitted groups to exceed its own null 95th percentile. It retains 66/75 families under model omission and 49/75 under participant omission, making the variation among held-out groups explicit.

An AF-matched null retains the same 75 families, held-out targets, 199 draws, and within-cell sizes, but samples training-side replacements near each true member's AF while excluding true members. Sixty families meet both omission criteria under this null; 59 overlap the primary 62-family core. C15, C19, and C33 occur only in the standard-null core, whereas C79 occurs only in the AF-matched core. Matching AF is a sensitivity analysis rather than uniformly a stricter baseline: family-median random cosines are $.033$ and $.017$ for model and participant omission, below the corresponding member-count-null values $.052$ and $.026$. Condition analyses use the primary 62-family gate; the 59-family intersection is an additional check. A separately numbered 39-family split-half catalog tests stronger member-identity stability and requires an F-to-C crosswalk rather than simple renumbering.

F.3 Condition membership and uncertainty

The primary 62-family analysis yields 659 relative-preference relations across the 82 conditions. Adding common participation gives 1,248 distinct condition–family relations and a median repertoire size of 15 families (range 9–27). Forty-eight spatial-core families participate commonly in at least two of the 49 object conditions. These counts describe relations and recurring atlas elements, respectively; they are not additional neural observations.

Restricting relative preference to the 59 families supported by both spatial baselines retains 645/659 primary preference pairs, with every condition still represented. Common participation contributes 1,066 pairs: 477 overlap preference, and 589 add families to the union. Among those additions, 272 have a nonpositive condition-minus-background difference. A family can thus be commonly called within a condition without preferring that condition over others; call prevalence and preference answer different questions.

Preference sensitivity was examined across nine combinations of effect thresholds ($.01/.02/.05$) and positive-cell requirements (8/12/16), as well as response-magnitude and alternative-call definitions. For selected preference effects, 399 stratified image bootstraps held participants, models, family identities, and the selection fixed. Pointwise lower bounds exceed zero for 655/659 effects; motorcycle–C21, dog–C70, umbrella–C9, and apple–C21 cross zero. Because selection was not repeated within bootstrap samples, these intervals describe image variation conditional on the selected pairs. Common-participation summaries used 499 image resamples, and expanded membership was checked across nine call-rate/cell-support combinations. These resampling counts refer to different estimands.

We recomputed common participation after omitting each participant or model, scaling support requirements to the surviving groups: $\lceil .75 n_{\text{cells}} \rceil$ cells must reach call rate .05, and $\lceil .75 n_{\text{groups}} \rceil$ participant or model aggregates must reach .10. Across 656 condition-by-deletion evaluations, expanded-set Jaccard similarity has median .941 (range .545–1); common participation alone has median .917 (range .455–1). Preference pairs and the family gate remain fixed, so this tests composition stability within the observed participants and models, not discovery in a new sample. Restricting to the 59-family intersection yields 1,211 expanded relations; all 82 conditions remain nonempty, with 9–26 families each (median 14). Within the 62-family core, member count correlates weakly with common object-condition reuse ($\rho = -.034$); this does not rule out other effects of the any-member call.

Natural-image semantic labels and pixel attributes co-occur. In a joint ridge sensitivity analysis including other condition labels and 13 measured image quantities, approximately 78.8% of the 659 selected condition coefficients remain positive; the rest change sign. In a separate joint ridge readout, 271/339 supported continuous-feature associations retain their sign after adjustment for other measured features and eligible semantic labels. These within-sample marginal-versus-conditional comparisons do not identify a unique cause of an FM call. For co-recruitment, pointwise image-bootstrap intervals were examined over 9,346 family pairs; 864 satisfy the exploratory display rule without correction across all pairs. An edge represents excess same-image, same-cell co-occurrence, not an anatomical connection.

F.4 Condition labels and image material

The face-context condition contains 348 images and participates in the global 6,150-test preference analysis. Its COCO-derived indicator accepts an annotated nose, eye, or ear keypoint even when coded as not visible or outside the NSD crop. It therefore describes broad head/face context, not a verified set of visible frontal faces. A narrower sensitivity mask requires two visible nose/eye keypoints from one person inside the actual crop and contains 294 images. Reapplying common participation and a separate 75-family within-mask preference screen yields 20 expanded families under each mask, with one substitution and family-effect rank correlation $\rho = .953$; the narrow preference screen loses two families. Because the within-mask BH family contains 75 tests rather than 6,150, its q values are not numerically comparable with the global screen. The primary 82-condition atlas is unchanged.

The dining-context label covers 86 images, of which 56 also carry a COCO dining-table annotation and 30 do not. We therefore interpret this condition as dining context rather than restricting it to restaurant scenes. Exact duplicate masks for other labels are aliases. These label checks keep the tested image populations fixed while limiting semantic claims to what their annotations support.

Fifteen of 20 prespecified COCO–WordNet roots contain at least 20 images in both NSD and the ImageNet-derived material. Bicycle, sheep, and bear fall below this threshold in NSD; cow and giraffe have no ImageNet synset under the exact root closure. For the 15 eligible roots, family-effect rank correlations are positive and span .272–.744. Of 96 NSD preference pairs within those roots, 83 retain a positive ImageNet effect; the counts are 4/4 for dog and 6/13 for person. Variation across ImageNet's 16 participant–model cells reflects fixed model outputs and NSD-selected family member lists, not new neural responses. The result supports model-side consistency across image materials without extending the neural sample beyond four participants.

F.5 Geometry and cortical display

The joint family representation and mutual-neighbor graph are defined in the Methods. Under four participant and four model omissions, 57/67 original edges recur in at least six of eight reconstructions; median edge-set Jaccard similarity is .719. The 23 projection configurations span diffusion, PCA, UMAP, t-SNE, and CEBRA centroids from an independent all-point fit. UMAP has five-neighbor trustworthiness .931 relative to the joint input and preserves .528 of its five-neighbor overlaps. A fixed equal-weight $k = 8$ diffusion view has lower local trustworthiness (.816) but broader pairwise-distance rank agreement (about .688). These measures compare displays of a chosen representation; they do not establish a rooted hierarchy or temporal processing order. Disconnected components and local loops occur in the original-neighbor graph. Surface interpolation affects visual display only:

single-FM maps keep their original rendering, PC maps retain display diffusion, and vertex-level analyses use numerical values before either operation.

G INDIVIDUAL FM IMAGE EVIDENCE AND QWEN DESCRIPTIONS

All 13 selected ViT-Base FMs have their archived Qwen visual description and proposed model role below. Figure 3 shows eight of their cortical maps, highest-ranked images, and local occlusion maps; the five other image sets appear here. Image rankings and occlusion describe model responses. Regional coefficients come from the S5 encoding fit. Qwen text is descriptive model-side evidence, not a measurement of neural selectivity or causal function.

FM162: Bodies in sport settings

Cortical and image evidence: Figure 3.

Qwen visual description. Module 162 responds to scenes involving **outdoor recreational activities**, particularly those involving balls and sports equipment, with a focus on grassy fields and human interaction. It is sensitive to objects like soccer balls, croquet balls, and baseballs, as well as associated accessories such as cowboy hats and crutches. The module shows strong activation in images where these elements are present in open, natural settings.

Proposed model role. The module encodes visual patterns associated with outdoor recreational activities, particularly those involving spherical objects used in games, and distinguishes them from unrelated objects and scenes.

FM075: People and formal dress

Cortical and image evidence: Figure 3.

Qwen visual description. Module 75 responds to visual patterns associated with **formal attire**, particularly neckwear such as ties and collars, often in social or professional contexts. It is activated by images containing individuals wearing suits, ties, or formal clothing, especially when the neck region is visible. The module also shows sensitivity to specific textures and colors related to formal wear.

Proposed model role. The module encodes the presence of formal neckwear and associated attire, contributing to the recognition of formal dress codes and social contexts involving professional or ceremonial clothing.

FM080: Canid heads

Cortical and image evidence: Figure 3.

Qwen visual description. Module 80 activates strongly for images containing **dog faces**, particularly those of wolves and huskies, with high activation in the central facial region. It is suppressed by objects such as dining tables, maze-like structures, and mechanical devices.

Proposed model role. The module encodes a visual concept related to dog faces, particularly emphasizing facial structure and expression, and acts as a detector that suppresses outputs associated with non-animal, structured, or mechanical objects.

FM222: Orange across objects

Cortical and image evidence: Figure 3.

Qwen visual description. Module 222 responds to **orange-colored objects and surfaces**, particularly those with high visual salience and distinct boundaries. It is activated by various orange items including vehicles, fruits, clothing, and tools, suggesting a focus on the color orange as a primary trigger.

Proposed model role. The module functions as a color-specific detector for orange, contributing to the recognition of orange objects by highlighting their presence in the visual scene.

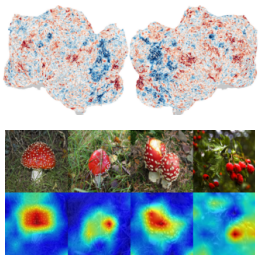
FM233: Red material and texture

Cortical and image evidence: Figure 3.

Qwen visual description. Module 233 responds to visual patterns associated with **red velvet-like textures and objects**, particularly in contexts involving rich red surfaces, fabric, or materials. It is activated by images containing red textured backgrounds, tools on red surfaces, and objects like Christmas stockings or theater curtains. The module shows strong localization to specific patches of red texture and appears to encode a combination of color (red), material (velvet), and spatial configuration cues.

Proposed model role. The module encodes a visual concept combining red color, velvet-like texture, and spatial context, likely serving as a feature detector for richly textured red surfaces in complex scenes.

FM067: Red / spotted / rounded conjunction



Left: signed cortical direction; four highest-ranked images and matched occlusion maps.

Qwen visual description. Module 67 responds to visual patterns associated with **red, round objects resembling berries or mushrooms**, particularly those with white spots. It is activated by images of agaric mushrooms and rose hips, showing high activation in the presence of these features. The module appears to encode a combination of object category cues (mushroom, berry) and texture/material cues (white spots on red surface).

Proposed model role. The module functions as a detector for red, spotted spherical objects, likely representing a shared visual feature across different biological entities such as agaric mushrooms and rose hips. It contributes to distinguishing these objects from other red or round items in the scene.

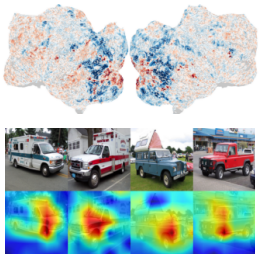
FM025: Regular grids

Cortical and image evidence: Figure 3.

Qwen visual description. Module 25 responds to **structured, grid-like visual patterns with high spatial regularity**, particularly those resembling measurement tools or ruled surfaces. It is activated by objects such as slide rules, rulers, and gridded surfaces, and shows strong localization to the central regions of these objects.

Proposed model role. The module detects and represents structured, regularly spaced visual patterns, likely serving as a feature detector for measurement tools and gridded surfaces in visual scenes.

FM033: Vehicle structure

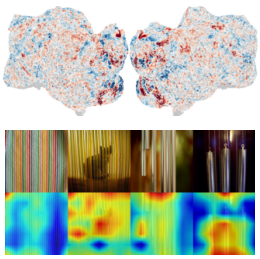


Left: signed cortical direction; four highest-ranked images and matched occlusion maps.

Qwen visual description. Module 33 responds to **vehicles**, particularly emergency and utility vehicles such as ambulances, police vans, fire trucks, and tow trucks. It is sensitive to the presence of vehicle bodies, wheels, and associated features like grilles and windows, with activation concentrated in central and lower regions of the image patch grid.

Proposed model role. The module encodes the presence of vehicle-like objects, particularly those associated with emergency services or transportation utilities, by detecting key visual features such as vehicle bodies, wheels, and distinctive shapes. It contributes to distinguishing vehicles from other object categories.

FM224: Repeated boundaries

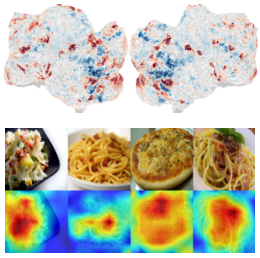


Left: signed cortical direction; four highest-ranked images and matched occlusion maps.

Qwen visual description. Module 224 responds to visual patterns associated with **enclosed or partitioned spaces**, particularly those involving vertical bars, grilles, and structured boundaries. It is activated by objects such as radiators, organ pipes, fences, curtains, and lockers, suggesting a sensitivity to linear, repetitive structures that define spatial separation.

Proposed model role. The module encodes the presence of spatially segmented environments defined by linear, repetitive structures, contributing to the representation of enclosed or bounded regions in visual scenes.

FM084: Prepared food

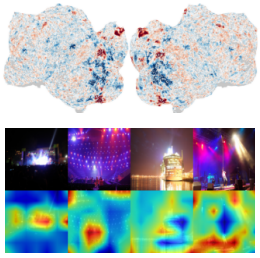


Left: signed cortical direction; four highest-ranked images and matched occlusion maps.

Qwen visual description. Module 84 activates in response to images of **food**, particularly dishes involving cooked vegetables and meat on plates. It is sensitive to the presence of specific food items such as broccoli, cauliflower, mashed potatoes, meatloaf, and potpie, often presented on a plate or tray. The module shows high activation for meals with a combination of protein and vegetables, especially when served in a restaurant-like setting.

Proposed model role. The module encodes the presence of specific food items and their arrangement on a plate or tray, contributing to the recognition of meal-related scenes.

FM014: Localized illumination



Left: signed cortical direction; four highest-ranked images and matched occlusion maps.

Qwen visual description. Module 14 responds to visual scenes containing **bright, localized light sources such as spotlights, stage lighting, and illuminated signs**, particularly in urban or theatrical contexts. It is sensitive to high-contrast illumination patterns and appears to encode the presence of artificial lighting structures.

Proposed model role. The module functions as a detector for localized, high-intensity artificial light sources, contributing to scene understanding by identifying areas of illumination that may indicate human activity, entertainment venues, or urban infrastructure.

FM121: Open landscapes

Cortical and image evidence: Figure 3.

Qwen visual description. Module 121 activates in response to **natural landscape scenes** featuring open terrain, water bodies, and elevated landforms such as valleys, lakeshores, and headlands. It is sensitive to spatial configurations involving expansive flat or sloped areas adjacent to water or mountainous regions, with activation concentrated in patches corresponding to these features.

Proposed model role. The module encodes spatial-configuration cues related to large-scale natural terrain features, particularly the presence and arrangement of valleys, lakeshores, headlands, and similar landforms, contributing to scene-level understanding of geographic layout.

FM053: Indoor layouts

Cortical and image evidence: Figure 3.

Qwen visual description. Module 53 responds to **indoor scenes** featuring traditional Japanese-style rooms with shoji screens, sliding doors, and wooden furniture. It is activated by spatial configurations involving partitioned spaces and specific architectural elements like tatami mats and low tables.

Proposed model role. The module encodes the presence of traditional Japanese interior design elements, including spatial partitioning via shoji screens and associated furniture layouts, contributing to scene categorization related to cultural architecture.

G.1 Complete regional profiles

Table 10 gives the signed regional coefficients underlying the example comparisons. Positive coefficients denote the fitted positive direction conditional on the other FMs; negative coefficients denote the opposite direction. Regional profiles characterize the distribution of each mode across established visual areas. The 17-network view in Figure 31 supplies the complementary large-scale organization [38].

Table 10: Mean signed encoding coefficients across ten visual ROIs, in 10^{-3} coefficient units. Each row belongs to one S5/base FM.

FM	V1	V2	V3	V4	EBA	FFA 1	FFA 2	OPA	PPA	RSC
075	1.13	-0.29	-0.77	-6.34	2.00	10.17	12.13	-11.12	-9.08	-2.10
080	-1.18	-0.22	0.59	0.81	-0.82	4.20	2.79	-4.42	-4.97	-3.40
162	-1.00	-1.40	-1.18	-2.91	4.21	2.08	3.20	-3.90	-4.46	-0.83
053	2.57	1.60	1.06	2.10	-2.20	-4.26	-2.87	8.61	9.71	10.17
121	0.85	-0.43	-1.60	-0.67	-2.76	-4.66	-3.30	3.25	3.55	4.86
222	-2.57	-1.23	-1.24	-1.93	-0.04	-0.87	-1.14	-3.40	-2.79	-1.53
233	0.39	0.19	0.36	0.05	-0.08	0.70	0.35	-0.96	-0.88	-1.68
025	-0.86	-0.45	-0.79	0.00	-2.16	-1.81	-1.61	-0.77	0.10	-0.87
067	1.42	1.07	0.01	-1.63	-2.85	-1.90	-2.39	-4.91	-2.81	-0.05
033	-0.71	-0.77	-0.65	-1.26	-4.51	-1.39	-0.65	-3.47	-3.36	-1.15
224	-0.58	0.63	0.73	-0.01	-0.07	-0.39	-0.55	0.86	2.53	1.38
084	2.82	1.24	0.07	1.27	-3.15	-3.58	-5.92	0.13	-0.88	-1.58
014	-0.50	-0.38	-1.39	-1.47	-3.01	-2.94	-1.54	1.39	3.08	4.70

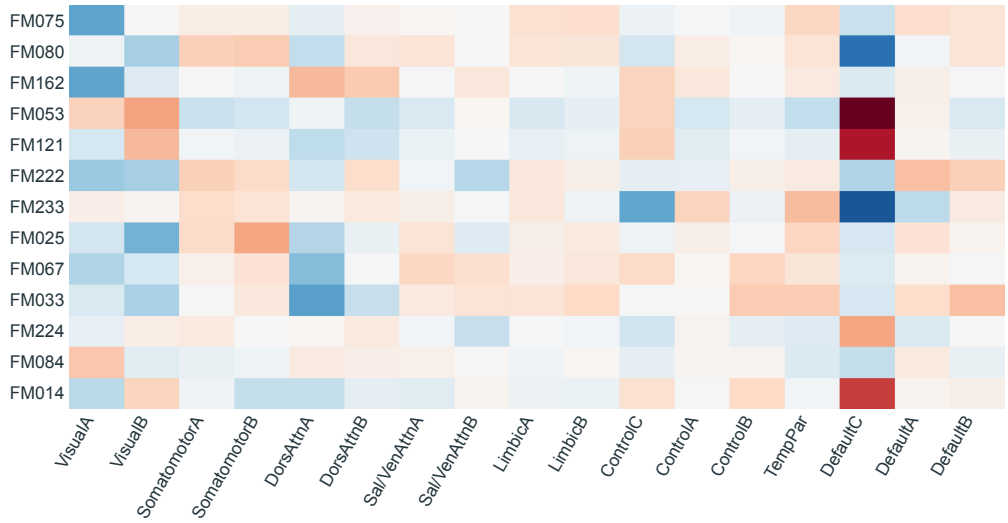
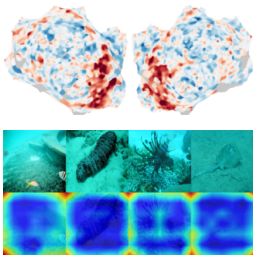


Figure 31: All 13 examples across 17 cortical networks. Colors show mean signed encoding coefficients on a common symmetric scale (± 0.03022); red denotes positive and blue negative coefficients, following Figure 3. This coarser summary complements the visual-ROI values above.

H DICTIONARY OF RECURRENT FM FAMILIES

The complete dictionary contains 75 families spanning all 16 participant–model combinations. Together they contain 1,752 of 2,775 eligible maps; 1,023 maps are unassigned. 62 families pass the spatial-support criterion, while 13 are retained as full-coverage candidates outside that core. A cortical field is the signed, equal-cell-weight family prototype; the four image examples belong to one S5/ViT-Base representative FM and do not summarize every member. Omission cosines compare the family prototype with held-out participant or model cells. Qwen descriptions pertain to the representative FM only. Figure 6 displays six families in the main text; their visual panels are referenced here.

C00: Marine life



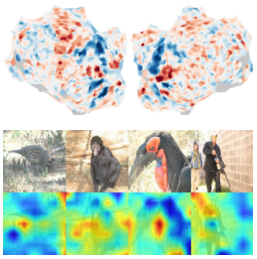
Left: signed family cortical prototype; four images for its representative FM.

64 member maps; spatial core; representative FM253. Held-out model/participant cosine: 0.63/0.11.

Qwen visual description. Module 253 activates in response to **marine life**, particularly electric rays and related sea creatures, with strong localization to specific regions of underwater scenes. It is also associated with coral reef environments and exhibits a functional role in distinguishing marine animals from non-marine objects.

Proposed model role. The module functions as a detector for marine fauna, particularly electric rays and related species, and contributes to the model's ability to distinguish aquatic animals from non-aquatic entities.

C01: Large, dark-furred animals



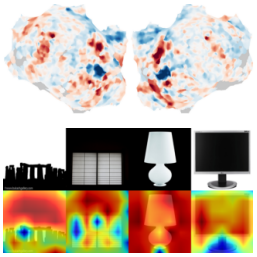
Left: signed family cortical prototype; four images for its representative FM.

64 member maps; spatial core; representative FM041. Held-out model/participant cosine: 0.67/0.17.

Qwen visual description. Module 41 responds to visual patterns associated with **large, dark-furred animals**, particularly primates and dogs, with a focus on specific body parts such as heads and torsos. It is also activated by certain man-made objects like tools and vehicles, suggesting a broader sensitivity to structured forms and textures.

Proposed model role. The module functions as a detector for dark-furred animals and structured objects, contributing to the recognition of these categories by emphasizing their visual features such as texture, shape, and color.

C02: Structured, geometric objects and surfaces



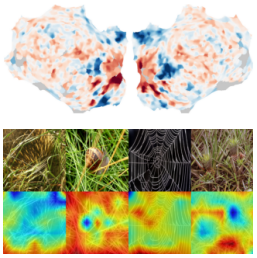
Left: signed family cortical prototype; four images for its representative FM.

56 member maps; spatial core; representative FM135. Held-out model/participant cosine: 0.62/0.33.

Qwen visual description. Module 135 responds to visual patterns associated with **structured, geometric objects and surfaces**, particularly those involving rectangular or grid-like arrangements, such as windows, screens, and architectural elements. It is also sensitive to specific object categories like envelopes, trash cans, and certain animals (e.g., dung beetles). The module exhibits a broad cross-layer implementation pattern with moderate density.

Proposed model role. The module functions as a detector for structured visual patterns, particularly those involving rectangular or grid-like arrangements, and may contribute to the recognition of objects with flat, uniform surfaces or repetitive textures.

C03: Wetland birds



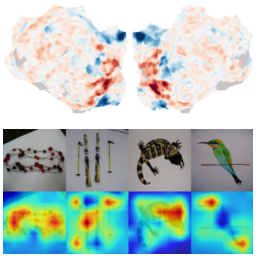
Left: signed family cortical prototype; four images for its representative FM.

53 member maps; spatial core; representative FM070. Held-out model/participant cosine: 0.77/0.31.

Qwen visual description. Module 70 activates in response to images containing **wetland birds**, particularly bitterns and related species, with strong localization to the bird's body and surrounding reeds. It also responds to certain artificial objects like seat belts and oxygen masks, suggesting a broader sensitivity to elongated, curved structures or specific textures.

Proposed model role. The module functions as a detector for elongated, dark-bodied birds in reed-like environments, possibly encoding both object identity and spatial configuration cues related to camouflage and habitat.

C04: Small, detailed objects such as insects and tools



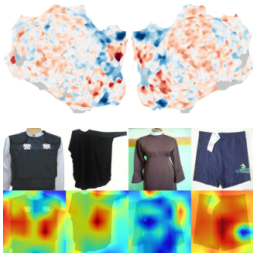
Left: signed family cortical prototype; four images for its representative FM.

53 member maps; spatial core; representative FM229. Held-out model/participant cosine: 0.84/0.47.

Qwen visual description. Module 229 responds to visual patterns associated with **small, detailed objects such as insects and tools**, particularly those involving fine structural elements like hooks, claws, or clock faces. It is activated by images containing intricate textures, sharp edges, and high-contrast features, especially in the context of mechanical or biological components.

Proposed model role. The module encodes the presence of fine-grained structural details in small objects, enabling discrimination between similar-looking items based on subtle morphological or mechanical features.

C05: Clothing items



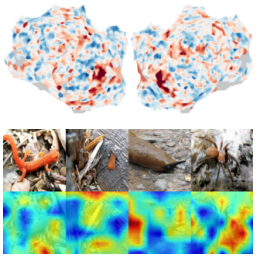
Left: signed family cortical prototype; four images for its representative FM.

52 member maps; spatial core; representative FM174. Held-out model/participant cosine: 0.47/0.28.

Qwen visual description. Module 174 responds to visual patterns associated with **clothing items**, particularly outerwear such as vests, coats, and layered garments. It is sensitive to textures like velvet and wool, and spatial configurations involving overlapping or layered fabric structures. The module shows high activation for images containing structured garments with distinct edges and folds.

Proposed model role. The module functions as a detector for structured outerwear garments, emphasizing texture (e.g., velvet, wool) and spatial configuration (e.g., layered, overlapping fabric), contributing to the representation of clothing-related concepts in the visual hierarchy.

C06: Slimy, elongated organisms in natural environments



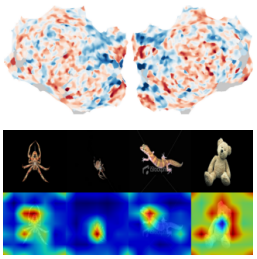
Left: signed family cortical prototype; four images for its representative FM.

52 member maps; candidate outside the spatial core; representative FM180. Held-out model/participant cosine: 0.23/0.05.

Qwen visual description. Module 180 responds to visual patterns associated with **slimy, elongated organisms in natural environments**, particularly snails and similar creatures. It is activated by texture cues such as moist, glossy surfaces and spatial configurations involving curved, segmented forms on earthy or rocky backgrounds.

Proposed model role. The module functions as a detector for slimy, elongated organisms like snails, contributing to the recognition of biological entities with distinct texture and form characteristics in natural scenes.

C07: Marine invertebrates and elongated, segmented body forms



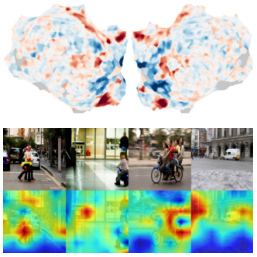
Left: signed family cortical prototype; four images for its representative FM.

51 member maps; candidate outside the spatial core; representative FM020. Held-out model/participant cosine: 0.40/0.09.

Qwen visual description. Module 20 responds to visual patterns associated with **marine invertebrates and elongated, segmented body forms**, particularly those resembling cephalopods, crustaceans, and certain reptiles. It is also activated by objects with curved or tubular shapes, such as musical instruments and tools. The module shows strong localization to specific regions of the image where these features are present.

Proposed model role. The module detects and represents visual features related to elongated, segmented, or curved forms, particularly those found in marine invertebrates and similar objects. It contributes to distinguishing between organisms with complex body segmentation and simpler, more uniform shapes.

C09: Upright, elongated structures such as poles, flagpoles, and vertical supports



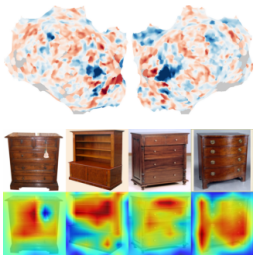
Left: signed family cortical prototype; four images for its representative FM.

44 member maps; spatial core; representative FM008. Held-out model/participant cosine: 0.58/0.28.

Qwen visual description. Module 8 responds to visual scenes containing **upright, elongated structures such as poles, flagpoles, and vertical supports**, particularly in urban or structured environments. It is sensitive to objects with linear, vertical geometry and appears to detect structural elements that define spatial organization.

Proposed model role. The module detects and represents the presence of vertical structural elements in a scene, contributing to spatial parsing and object boundary detection in complex environments.

C10: Wooden furniture with a cabinet-like structure



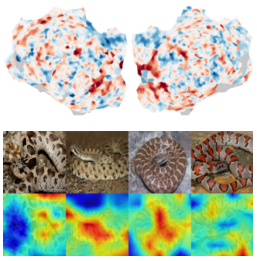
Left: signed family cortical prototype; four images for its representative FM.

44 member maps; spatial core; representative FM078. Held-out model/participant cosine: 0.66/0.20.

Qwen visual description. Module 78 activates in response to **wooden furniture with a cabinet-like structure**, particularly chest of drawers and dressers, characterized by multiple drawers, doors, or compartments. The module is sensitive to the presence of these objects regardless of specific style or material variation.

Proposed model role. The module encodes the presence of structured wooden storage furniture, likely representing a high-level object category that includes both functional and aesthetic features such as drawer arrangement and cabinet form.

C11: Snakes



Left: signed family cortical prototype; four images for its representative FM.

41 member maps; candidate outside the spatial core; representative FM056. Held-out model/participant cosine: 0.42/0.04.

Qwen visual description. Module 56 activates in response to **snakes**, particularly those with distinct patterns such as sidewinders, horned vipers, and night snakes. It is sensitive to specific visual features including scale texture, coloration (e.g., orange-red bands), and body shape. The module shows strong activation for coiled or resting snake postures and is suppressed by mammalian animals like beavers, otters, and foxes.

Proposed model role. This module encodes a high-level visual concept of 'snake' based on texture, pattern, and body configuration, distinguishing snakes from other animals through learned feature combinations.

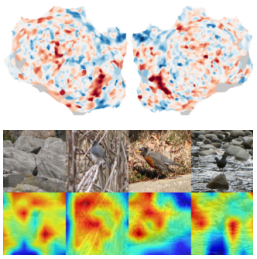
C12: Open landscape

Cortical and image evidence: Figure 6.

36 member maps; spatial core; representative FM121. Held-out model/participant cosine: 0.86/0.52.

The full Qwen description appears in the individual-FM catalog under FM121.

C13: Small animals and birds in natural environments



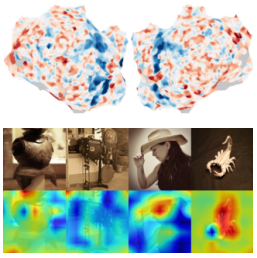
Left: signed family cortical prototype; four images for its representative FM.

34 member maps; candidate outside the spatial core; representative FM195. Held-out model/participant cosine: 0.64/0.10.

Qwen visual description. Module 195 responds to visual patterns associated with **small animals and birds in natural environments**, particularly those involving rocky or earthy textures. It is activated by scenes featuring animals like foxes, birds, and mammals in outdoor settings, with a focus on specific spatial configurations and object parts.

Proposed model role. The module functions as a detector for small animals in natural environments, contributing to the recognition of animal presence and context in scenes involving rocky or earthy backgrounds.

C14: Vintage or rustic objects



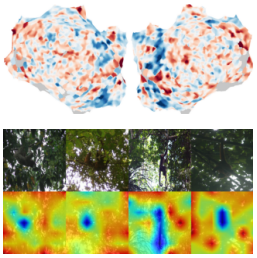
Left: signed family cortical prototype; four images for its representative FM.

32 member maps; spatial core; representative FM006. Held-out model/participant cosine: 0.54/0.10.

Qwen visual description. Module 6 responds to visual scenes containing **vintage or rustic objects**, particularly those associated with historical or rural settings such as old cameras, lanterns, cowboy hats, and animal figures. It is sensitive to object categories like lamps, barrels, rifles, and musical instruments, often in sepia-toned or aged contexts.

Proposed model role. The module functions as a detector for visual cues associated with historical or rural themes, particularly focusing on specific object categories like lamps, barrels, rifles, and musical instruments, and their contextual presentation in aged or sepia-toned imagery.

C15: New World monkeys



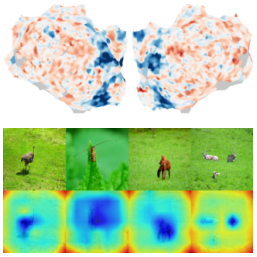
Left: signed family cortical prototype; four images for its representative FM.

28 member maps; spatial core; representative FM030. Held-out model/participant cosine: 0.60/0.11.

Qwen visual description. Module 30 activates in response to images containing **New World monkeys**, particularly howler monkeys and related species, with strong localization to the animal's body and surrounding foliage. It also responds to figs and jackfruit, suggesting a potential association with fruit-bearing trees in tropical environments. The module shows high implementation density across multiple layers, indicating a distributed representation.

Proposed model role. The module functions as a detector for New World monkey species and their habitat cues, potentially supporting recognition of these animals in complex natural scenes by integrating object and environmental information.

C16: Small, green, insect-like objects with elongated bodies and distinct segmentation



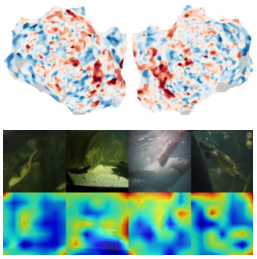
Left: signed family cortical prototype; four images for its representative FM.

28 member maps; spatial core; representative FM134. Held-out model/participant cosine: 0.68/0.25.

Qwen visual description. Module 134 responds to **small, green, insect-like objects with elongated bodies and distinct segmentation**, particularly in natural outdoor settings. It is sensitive to visual cues associated with insects such as ladybugs, leaf beetles, lacewings, and grasshoppers, often appearing on leaves or grass. The module shows strong activation for green coloration, segmented body parts, and fine texture details like antennae and legs.

Proposed model role. The module functions as a detector for small, green, segmented insects by integrating object-level cues (insect morphology), part cues (antennae, legs, segmented body), texture/material cues (fine surface detail), and spatial-configuration cues (position on leaves/grass).

C17: Aquatic animals



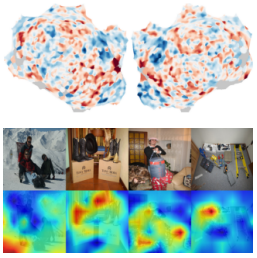
Left: signed family cortical prototype; four images for its representative FM.

28 member maps; candidate outside the spatial core; representative FM098. Held-out model/participant cosine: 0.16/0.07.

Qwen visual description. Module 98 activates in response to **aquatic animals**, particularly fish and marine creatures, with a focus on elongated body shapes and underwater environments. It is sensitive to specific visual features such as the presence of fish in water, their body contours, and surrounding aquatic textures.

Proposed model role. The module encodes the presence of aquatic animals, particularly fish, by detecting elongated body shapes and associated environmental cues like water and aquatic textures.

C18: Specific object categories



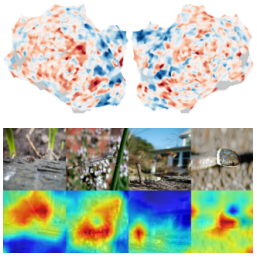
Left: signed family cortical prototype; four images for its representative FM.

24 member maps; candidate outside the spatial core; representative FM138. Held-out model/participant cosine: 0.20/0.06.

Qwen visual description. Module 138 responds to visual patterns associated with **specific object categories**, particularly those involving clothing, vehicles, and accessories. It is activated by images containing items such as academic gowns, cowboy boots, assault rifles, and sports cars, suggesting a focus on object-level features related to human attire and transportation. The module shows high activation in the presence of these objects and low activation when they are absent.

Proposed model role. The module functions as a detector for specific object categories, particularly those related to human attire and transportation. Its activation correlates with the presence of these objects, and interventions suggest it suppresses representations of other categories like furniture, animals, or household items.

C19: Small, dark, elongated objects with segmented or textured surfaces



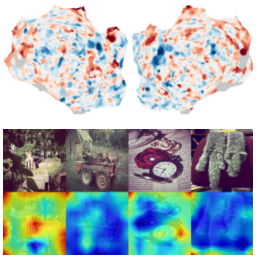
Left: signed family cortical prototype; four images for its representative FM.

24 member maps; spatial core; representative FM096. Held-out model/participant cosine: 0.74/0.12.

Qwen visual description. Module 96 responds to **small, dark, elongated objects with segmented or textured surfaces**, particularly insects and arthropods like beetles, ants, and snails. It also activates for metallic fasteners such as nails and screws. The module shows high activation in images containing these elements, especially when they are centrally located or prominent.

Proposed model role. The module functions as a detector for small, dark, elongated objects with segmented or textured surfaces, contributing to the recognition of insects, arthropods, and metallic fasteners in visual scenes.

C20: Textured, wool-like materials and specific animal features



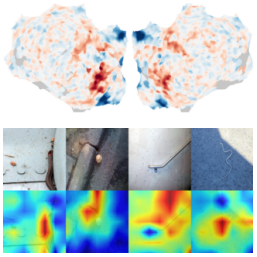
Left: signed family cortical prototype; four images for its representative FM.

24 member maps; spatial core; representative FM072. Held-out model/participant cosine: 0.70/0.13.

Qwen visual description. Module 72 responds to visual patterns associated with **textured, wool-like materials and specific animal features**, particularly in the context of dogs and related objects. It is activated by images containing knitted or furry textures, as well as certain dog breeds like terriers and poodles. The module also shows sensitivity to spatial configurations involving small, rounded objects and structured patterns.

Proposed model role. The module encodes a visual feature that detects textured, wool-like materials and their spatial configuration, particularly in relation to animals and small objects. It appears to represent a combination of texture cues and object-level semantics related to soft, fibrous materials.

C21: Elongated, thin objects resembling nails, hooks, or worms



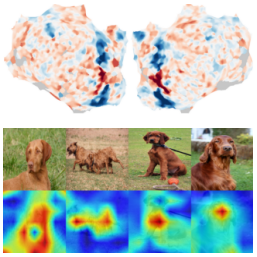
Left: signed family cortical prototype; four images for its representative FM.

24 member maps; spatial core; representative FM215. Held-out model/participant cosine: 0.61/0.34.

Qwen visual description. Module 215 responds to **elongated, thin objects resembling nails, hooks, or worms**, particularly when they are isolated or contrasted against a background. It is sensitive to linear structures with high spatial localization in specific patch regions.

Proposed model role. This module functions as a detector for thin, elongated linear structures, contributing to the recognition of small mechanical parts or biological entities that share this morphological trait.

C23: Reddish-brown fur coloration in animals



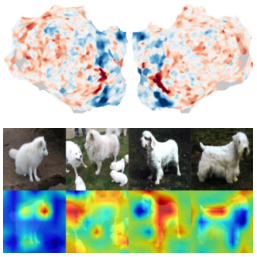
Left: signed family cortical prototype; four images for its representative FM.

20 member maps; spatial core; representative FM055. Held-out model/participant cosine: 0.78/0.29.

Qwen visual description. Module 55 responds to **reddish-brown fur coloration in animals**, particularly dogs and horses, with high activation in images featuring sorrel-colored animals. It is sensitive to the presence of specific dog breeds such as Irish setters, vizslas, and golden retrievers, and shows strong localization to regions of reddish-brown fur.

Proposed model role. The module encodes the visual feature of sorrel (reddish-brown) fur coloration in animals, contributing to the identification of specific dog breeds and horse coat colors by highlighting this distinct hue.

C24: White, fluffy animals with long fur



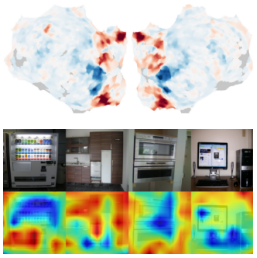
Left: signed family cortical prototype; four images for its representative FM.

20 member maps; spatial core; representative FM191. Held-out model/participant cosine: 0.71/0.26.

Qwen visual description. Module 191 responds to **white, fluffy animals with long fur**, particularly dogs such as Samoyeds and kuvasz, and is sensitive to their texture and color. It also activates for Arctic foxes and polar bears, indicating a broader sensitivity to white, furry mammals in cold environments. The module suppresses activation for objects like laptops, cowboy hats, and other non-animal items.

Proposed model role. The module encodes the presence of white, fluffy animals with long fur, likely representing a high-level visual feature related to animal texture and color, and modulates downstream representations to emphasize these features while suppressing unrelated objects.

C25: Electronic devices and appliances



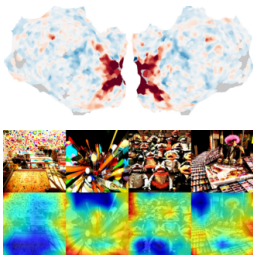
Left: signed family cortical prototype; four images for its representative FM.

20 member maps; spatial core; representative FM209. Held-out model/participant cosine: 0.97/0.61.

Qwen visual description. Module 209 responds to visual patterns associated with **electronic devices and appliances**, particularly those involving rectangular enclosures with screens or interfaces such as ATMs, microwaves, refrigerators, and monitors. It is sensitive to structured spatial configurations of technology-related objects and appears to encode a concept related to 'electronic appliance' or 'machine interface'.

Proposed model role. The module functions as a detector for electronic appliances with digital interfaces, contributing to the recognition of machine-like objects by encoding their structural and contextual features.

C26: Commercial or retail environments



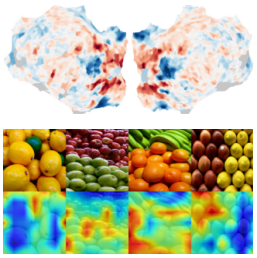
Left: signed family cortical prototype; four images for its representative FM.

20 member maps; spatial core; representative FM136. Held-out model/participant cosine: 0.86/0.47.

Qwen visual description. Module 136 activates in response to visual scenes containing **commercial or retail environments**, particularly those involving slot machines, game booths, and store interiors. It is sensitive to spatial configurations of objects arranged in grids or rows, such as shelves with products, and exhibits strong responses to bright, colorful displays with high contrast. The module appears to encode a concept related to 'commercial entertainment spaces' or 'retail gaming environments'.

Proposed model role. The module functions as a detector for commercial and entertainment environments, particularly those involving interactive or display-based elements like slot machines, game booths, and product shelves. It may contribute to the recognition of human-made spaces designed for commerce or leisure.

C27: Fruit



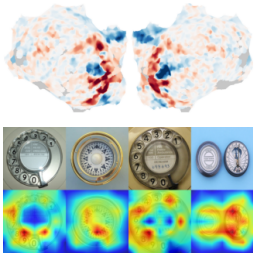
Left: signed family cortical prototype; four images for its representative FM.

20 member maps; spatial core; representative FM064. Held-out model/participant cosine: 0.89/0.38.

Qwen visual description. Module 64 responds to visual patterns associated with **fruit**, particularly citrus and apple-like objects in market or grocery settings. It is activated by images containing lemons, oranges, apples, and other fruits arranged in displays typical of food markets. The module shows strong localization to fruit clusters and is suppressed by non-fruit objects such as kitchen tools, appliances, and industrial items.

Proposed model role. The module functions as a detector for fruit-rich scenes, specifically identifying visual patterns associated with fresh produce in retail environments. It contributes to distinguishing between edible goods and non-edible objects in complex scenes.

C28: Circular objects with radial symmetry



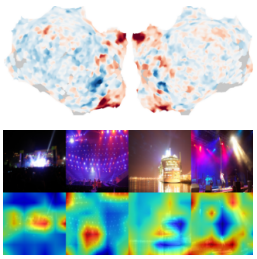
Left: signed family cortical prototype; four images for its representative FM.

20 member maps; spatial core; representative FM077. Held-out model/participant cosine: 0.93/0.43.

Qwen visual description. Module 77 responds to **circular objects with radial symmetry**, particularly those resembling clocks, dials, and mechanical components like brake discs. It is sensitive to concentric patterns, central hubs, and segmented or numbered perimeters, often found in analog clocks, compasses, and vehicle parts.

Proposed model role. The module encodes the presence of circular, radially symmetric objects with structured perimeters, likely representing mechanical components or instruments, and contributes to distinguishing them from other object categories based on geometric and structural cues.

C30: Bright, localized light sources such as spotlights, stage lighting, and illuminated signs

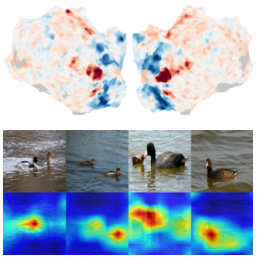


Left: signed family cortical prototype; four images for its representative FM.

20 member maps; spatial core; representative FM014. Held-out model/participant cosine: 0.77/0.36.

The full Qwen description appears in the individual-FM catalog under FM014.

C31: Waterfowl



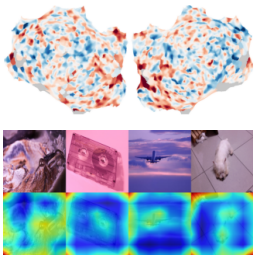
Left: signed family cortical prototype; four images for its representative FM.

20 member maps; spatial core; representative FM042. Held-out model/participant cosine: 0.89/0.45.

Qwen visual description. Module 42 activates strongly for **waterfowl**, particularly ducks and swans, in natural aquatic environments. It responds to the presence of birds with distinct head and neck features, especially those with dark plumage and red beaks or facial markings. The module is sensitive to specific spatial configurations involving birds on water and shows consistent activation across multiple species within the waterfowl category.

Proposed model role. The module functions as a detector for waterfowl in aquatic scenes, specifically identifying birds with dark bodies and red facial features, and contributes to distinguishing these animals from other categories in visual classification tasks.

C33: Objects and scenes involving human-made tools, clothing, and structured environments



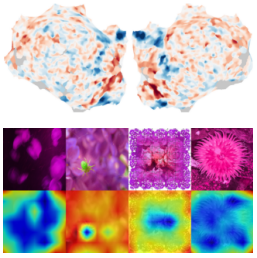
Left: signed family cortical prototype; four images for its representative FM.

16 member maps; spatial core; representative FM190. Held-out model/participant cosine: 0.58/0.09.

Qwen visual description. Module 190 responds to visual patterns associated with **objects and scenes involving human-made tools, clothing, and structured environments**, particularly those with metallic or reflective surfaces, mechanical components, or specific spatial configurations. It is activated by images containing items like screwdrivers, pens, tool kits, and vehicles, as well as scenes with architectural elements such as docks, mailbags, and fire trucks.

Proposed model role. The module encodes the presence of functional human-made objects and their contextual relationships, contributing to the recognition of tools, vehicles, and structured environments by detecting key visual features such as handles, reflective surfaces, and geometric configurations.

C34: Bright pink and purple hues, often with high contrast textures such as fuzzy or spiky surfaces



Left: signed family cortical prototype; four images for its representative FM.

16 member maps; spatial core; representative FM073. Held-out model/participant cosine: 0.74/0.18.

Qwen visual description. Module 73 responds to visual patterns involving **bright pink and purple hues, often with high contrast textures such as fuzzy or spiky surfaces**. It is activated by objects like sea anemones, fluffy materials, and abstract shapes with vivid coloration, particularly in the context of biological or textile-like structures.

Proposed model role. This module encodes a visual pattern associated with vivid pink and purple textures, likely representing biological entities like sea anemones or artificial materials like fluffy textiles. Its activation correlates with the presence of these features across multiple layers.

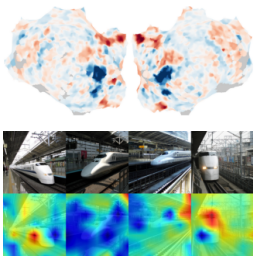
C35: Social dress with faces

Cortical and image evidence: Figure 6.

16 member maps; spatial core; representative FM075. Held-out model/participant cosine: 0.96/0.51.

The full Qwen description appears in the individual-FM catalog under FM075.

C36: High-speed trains



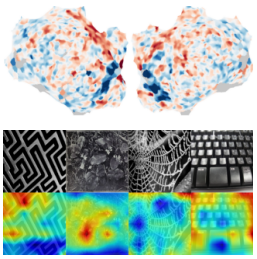
Left: signed family cortical prototype; four images for its representative FM.

16 member maps; spatial core; representative FM184. Held-out model/participant cosine: 0.94/0.38.

Qwen visual description. Module 184 activates strongly for images of **high-speed trains**, particularly bullet trains and electric locomotives, with consistent spatial localization to the train's front section. It exhibits a functional role in distinguishing rail vehicles from other transportation modes, especially cars and trucks, and shows a preference for specific visual features such as streamlined shapes and metallic surfaces.

Proposed model role. The module functions as a detector for high-speed rail vehicles, contributing to the model's ability to distinguish between rail transport and other vehicle categories, particularly road vehicles.

C37: Mechanical and industrial objects



Left: signed family cortical prototype; four images for its representative FM.

16 member maps; spatial core; representative FM074. Held-out model/participant cosine: 0.66/0.10.

Qwen visual description. Module 74 responds to visual patterns associated with **mechanical and industrial objects**, particularly those involving circular or cylindrical components such as wheels, gears, and metallic structures. It is activated by images containing machinery, vehicles, and engineered parts, especially when these elements are prominent in the foreground.

Proposed model role. The module encodes the presence of mechanical or industrial objects characterized by circular or cylindrical components, contributing to the recognition of engineered systems and manufactured items.

C38: Basketball and body pose

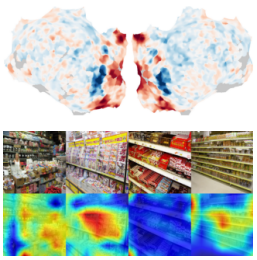
Cortical and image evidence: Figure 6.

16 member maps; spatial core; representative FM099. Held-out model/participant cosine: 0.98/0.46.

Qwen visual description. Module 99 responds to visual patterns associated with **basketball** players in action, particularly focusing on the presence of a basketball and dynamic poses involving hands holding or interacting with the ball. The module is sensitive to specific spatial configurations of limbs and objects, especially when a player is actively engaged in dribbling or shooting.

Proposed model role. This module functions as a detector for basketball-specific actions, particularly identifying the presence of a basketball being held or manipulated by a player. It contributes to distinguishing basketball from other sports or activities involving similar objects.

C39: Retail environments



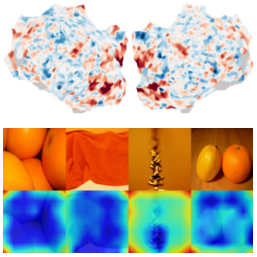
Left: signed family cortical prototype; four images for its representative FM.

16 member maps; spatial core; representative FM177. Held-out model/participant cosine: 0.77/0.36.

Qwen visual description. Module 177 activates in response to visual scenes of **retail environments**, particularly grocery stores and candy/confectionery shops, with a focus on structured shelving and product displays. It is suppressed by objects associated with specific tools or equipment such as scales, lamps, and musical instruments.

Proposed model role. This module encodes the concept of 'retail space' or 'store interior', distinguishing between commercial retail settings and non-retail environments based on spatial organization and object density.

C40: Visual cues associated with warm lighting and reflective surfaces



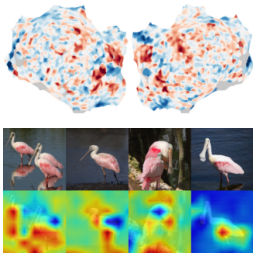
Left: signed family cortical prototype; four images for its representative FM.

16 member maps; candidate outside the spatial core; representative FM012. Held-out model/participant cosine: 0.58/0.06.

Qwen visual description. Module 12 responds to **visual cues associated with warm lighting and reflective surfaces**, particularly in scenes involving objects like oranges, candles, wooden items, and metallic or glossy materials. It is activated by high-contrast, textured regions under warm illumination and deactivated by mechanical or digital objects.

Proposed model role. The module detects and emphasizes visually salient regions characterized by warm lighting and reflective textures, likely contributing to scene interpretation involving natural or domestic environments with ambient light sources.

C42: Wading birds



Left: signed family cortical prototype; four images for its representative FM.

16 member maps; candidate outside the spatial core; representative FM237. Held-out model/participant cosine: 0.74/0.05.

Qwen visual description. Module 237 activates strongly for images of **wading birds**, particularly flamingos and spoonbills, with high activation in the presence of long legs, curved beaks, and pink or white plumage. It is suppressed by non-wading birds and objects with similar shapes but different ecological contexts.

Proposed model role. The module encodes a visual concept related to wading birds, distinguishing them from other birds and objects based on morphological cues such as leg length, beak shape, and coloration.

C43: Repeated structural lines

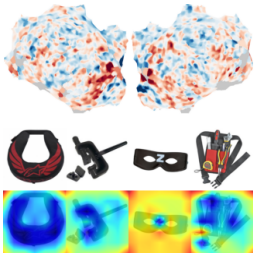
Cortical and image evidence: Figure 6.

16 member maps; spatial core; representative FM104. Held-out model/participant cosine: 0.81/0.25.

Qwen visual description. Module 104 responds to structural elements involving linear, elongated forms with repeating patterns or geometric arrangements, particularly those resembling railings, fences, or architectural supports. It is sensitive to configurations of parallel lines and grid-like structures, often associated with man-made objects such as bannisters, chainlink fences, and suspension bridges.

Proposed model role. The module encodes the presence of linear structural components with regular spacing or alignment, contributing to the recognition of man-made constructs like railings, fences, and bridges, as well as natural analogs such as spider webs or tree branches in specific configurations.

C44: Tools and equipment



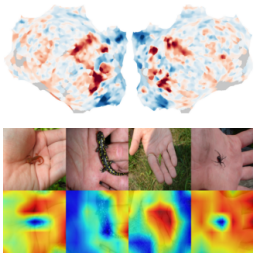
Left: signed family cortical prototype; four images for its representative FM.

16 member maps; candidate outside the spatial core; representative FM069. Held-out model/participant cosine: 0.42/0.05.

Qwen visual description. Module 69 activates in response to images containing **tools and equipment**, particularly handheld devices with mechanical or electronic components. It is sensitive to object categories such as binoculars, power tools, and precision instruments, showing high activation for objects with distinct handles, lenses, or functional parts.

Proposed model role. The module encodes the presence of functional mechanical or electronic tools, distinguishing them from other object types based on structural cues like handles, lenses, and compact form factors.

C45: Small, dark, elongated objects held in human hands



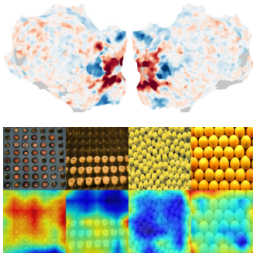
Left: signed family cortical prototype; four images for its representative FM.

16 member maps; spatial core; representative FM161. Held-out model/participant cosine: 0.84/0.34.

Qwen visual description. Module 161 responds to **small, dark, elongated objects held in human hands**, particularly insects and reptiles like snakes and lizards. It is sensitive to the presence of such objects against a contrasting background (e.g., skin), with activation concentrated on specific patches corresponding to the object's location.

Proposed model role. The module detects the presence of small, dark, elongated objects in human hands by encoding their visual features (shape, color, contrast) and spatial configuration relative to the hand.

C46: Structured, repeating patterns with high spatial regularity



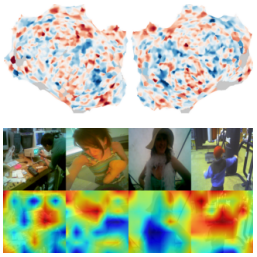
Left: signed family cortical prototype; four images for its representative FM.

16 member maps; spatial core; representative FM044. Held-out model/participant cosine: 0.85/0.35.

Qwen visual description. Module 44 responds to **structured, repeating patterns with high spatial regularity**, particularly those resembling grids, honeycombs, or spiral arrangements. It is activated by objects and textures with uniform geometric layouts such as keyboards, abacuses, tile roofs, and coiled structures.

Proposed model role. The module encodes the presence of regular, repeating spatial structure in visual inputs, serving as a detector for geometric order and periodicity in both artificial and natural scenes.

C47: Human figures in specific poses and attire



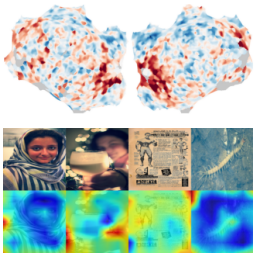
Left: signed family cortical prototype; four images for its representative FM.

16 member maps; candidate outside the spatial core; representative FM252. Held-out model/participant cosine: 0.37/0.03.

Qwen visual description. Module 252 responds to visual patterns associated with **human figures in specific poses and attire**, particularly involving headwear and accessories like hats or caps. It is also sensitive to objects such as photocopiers and other mechanical devices, suggesting a role in detecting structured scenes involving people interacting with equipment.

Proposed model role. This module detects the presence of humans in specific contexts involving headwear or interaction with mechanical devices, potentially serving as a feature detector for social or occupational settings.

C48: Textured surfaces and object parts



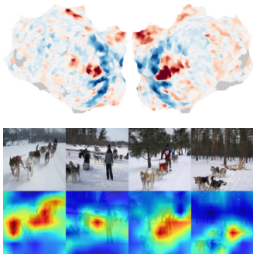
Left: signed family cortical prototype; four images for its representative FM.

16 member maps; spatial core; representative FM157. Held-out model/participant cosine: 0.46/0.09.

Qwen visual description. Module 157 responds to visual patterns associated with **textured surfaces and object parts**, particularly those involving rough or segmented textures, such as reptile skin, guitar bodies, and typewriter keys. It is activated by specific spatial configurations of high-contrast edges and repetitive patterns.

Proposed model role. The module encodes the presence of textured surfaces with high-contrast edges and repetitive patterns, contributing to the recognition of object parts and material properties in visual scenes.

C49: Dogsleds and snow-related transportation



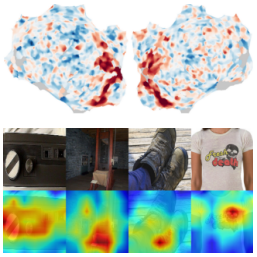
Left: signed family cortical prototype; four images for its representative FM.

16 member maps; spatial core; representative FM145. Held-out model/participant cosine: 0.97/0.36.

Qwen visual description. Module 145 responds to scenes involving **dogsleds and snow-related transportation**, particularly in snowy environments with huskies or similar breeds pulling sleds. It is sensitive to the presence of snowmobiles, skis, and related winter equipment, and its activation correlates with visual cues of snow-covered ground and dog teams.

Proposed model role. The module encodes the concept of winter transportation using dogsleds, including the presence of dogs, sleds, snow, and related equipment like skis and snowmobiles.

C50: Man-made objects



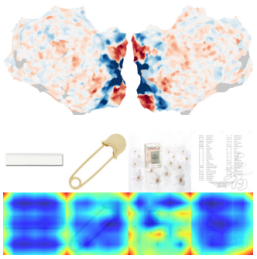
Left: signed family cortical prototype; four images for its representative FM.

16 member maps; spatial core; representative FM106. Held-out model/participant cosine: 0.48/0.17.

Qwen visual description. Module 106 responds to visual patterns associated with **man-made objects**, particularly those involving structured surfaces, mechanical components, and specific object parts like handles or frames. It is sensitive to spatial configurations of rectangular or grid-like structures and appears to encode features related to functional objects such as appliances, tools, and containers.

Proposed model role. Encodes the presence of structured, functional man-made objects by detecting spatial arrangements of rectangular or grid-like patterns, often associated with appliances, tools, or containers.

C51: Small, elongated objects with high contrast edges and reflective surfaces



Left: signed family cortical prototype; four images for its representative FM.

16 member maps; spatial core; representative FM090. Held-out model/participant cosine: 0.94/0.44.

Qwen visual description. Module 90 responds to **small, elongated objects with high contrast edges and reflective surfaces**, particularly those resembling tools or mechanical parts. It is sensitive to object shape and texture cues, especially in contexts involving precision instruments or personal care items.

Proposed model role. The module detects and represents small, elongated objects with high-contrast edges and reflective surfaces, contributing to the identification of precision tools and personal care items.

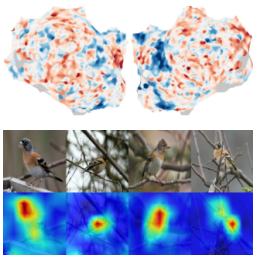
C52: Canid heads

Cortical and image evidence: Figure 6.

16 member maps; spatial core; representative FM080. Held-out model/participant cosine: 0.96/0.45.

The full Qwen description appears in the individual-FM catalog under FM080.

C53: Small passerine birds



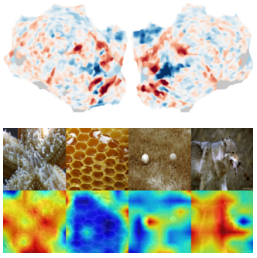
Left: signed family cortical prototype; four images for its representative FM.

16 member maps; spatial core; representative FM221. Held-out model/participant cosine: 0.82/0.16.

Qwen visual description. Module 221 activates in response to **small passerine birds**, particularly finches and related species, with a focus on visual features such as orange breast patches, black head markings, and perched postures. The module is sensitive to bird-specific textures and color patterns, especially in natural outdoor settings.

Proposed model role. The module encodes a visual concept of 'small passerine bird' with emphasis on diagnostic color cues (orange breast, black cap) and spatial configuration (perched on branch). It contributes to distinguishing avian species from other categories and supports fine-grained classification within the bird class.

C54: Food items



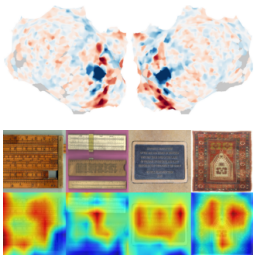
Left: signed family cortical prototype; four images for its representative FM.

16 member maps; spatial core; representative FM068. Held-out model/participant cosine: 0.73/0.30.

Qwen visual description. Module 68 responds to visual patterns associated with **food items**, particularly baked goods and confectionery, including dough, bagels, ice cream, chocolate sauce, and bakery environments. It also activates for related objects such as bonnets, mortars, and bowls, suggesting a broader association with rounded or container-like shapes in culinary contexts.

Proposed model role. The module encodes a concept related to edible items with specific textures and shapes, particularly those found in bakeries or dessert settings, and may contribute to the recognition of food-related scenes.

C55: Structured, grid-like visual patterns with high spatial regularity

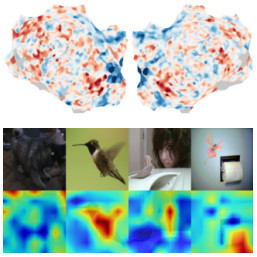


Left: signed family cortical prototype; four images for its representative FM.

16 member maps; spatial core; representative FM025. Held-out model/participant cosine: 0.72/0.34.

The full Qwen description appears in the individual-FM catalog under FM025.

C56: Mechanical or technological objects



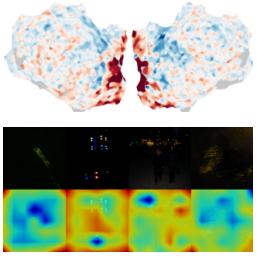
Left: signed family cortical prototype; four images for its representative FM.

16 member maps; candidate outside the spatial core; representative FM127. Held-out model/participant cosine: 0.57/0.09.

Qwen visual description. Module 127 responds to visual patterns associated with **mechanical or technological objects**, particularly those involving rectangular forms, metallic textures, and structured components. It is activated by images containing devices like laptops, modems, ATMs, and printers, often featuring sharp edges, reflective surfaces, and functional design elements.

Proposed model role. The module encodes the presence of technological artifacts characterized by geometric precision, metallic or plastic materials, and functional design cues such as ports, displays, and control panels.

C57: Dark, low-light scenes



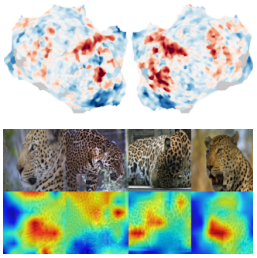
Left: signed family cortical prototype; four images for its representative FM.

16 member maps; spatial core; representative FM158. Held-out model/participant cosine: 0.84/0.33.

Qwen visual description. Module 158 responds to visual patterns associated with **dark, low-light scenes** featuring illuminated objects or structures, particularly those resembling artificial light sources such as lamps, candles, or glowing elements. It is sensitive to spatial configurations involving vertical or elongated bright regions against dark backgrounds and shows activation in contexts related to nighttime environments or theatrical lighting.

Proposed model role. The module encodes the presence and configuration of artificial light sources in dark environments, contributing to scene understanding by distinguishing between illuminated and non-illuminated regions, particularly in contexts like nighttime or indoor lighting setups.

C58: Spotted big cats



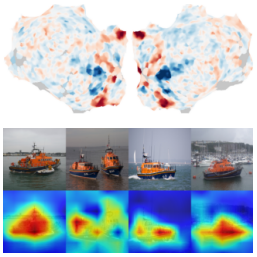
Left: signed family cortical prototype; four images for its representative FM.

16 member maps; spatial core; representative FM087. Held-out model/participant cosine: 0.92/0.29.

Qwen visual description. Module 87 responds to visual patterns associated with **spotted big cats**, particularly leopards and jaguars, showing high activation for images of these animals. It is sensitive to the presence of distinct spotted fur patterns and facial features typical of these species. The module also exhibits a functional role in suppressing non-animal objects and certain animal categories, suggesting it may be involved in distinguishing between animal and non-animal entities or specific animal subcategories.

Proposed model role. The module functions to detect and emphasize visual features characteristic of spotted big cats while suppressing activations for unrelated objects and animals, potentially aiding in category discrimination within the visual hierarchy.

C59: Maritime vessels



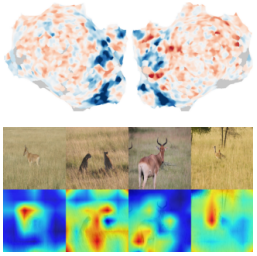
Left: signed family cortical prototype; four images for its representative FM.

16 member maps; spatial core; representative FM016. Held-out model/participant cosine: 0.78/0.36.

Qwen visual description. Module 16 responds to **maritime vessels**, particularly fireboats and lifeboats, with high activation in images featuring such boats in water environments. It is sensitive to the presence of specific vessel types and their spatial configuration within a harbor or open-water setting.

Proposed model role. The module encodes the presence and context of maritime rescue vessels, distinguishing them from other boat types and land-based objects, and contributes to scene understanding in marine environments.

C60: Antelope-like animals



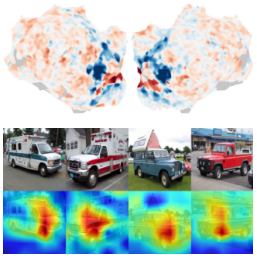
Left: signed family cortical prototype; four images for its representative FM.

16 member maps; spatial core; representative FM107. Held-out model/participant cosine: 0.80/0.31.

Qwen visual description. Module 107 responds to visual patterns associated with **antelope-like animals**, particularly hartebeest, gazelle, and impala, in natural grassland environments. It is also activated by certain human-made objects such as lamps and kitchenware, suggesting a broader sensitivity to specific spatial configurations or textures.

Proposed model role. The module encodes a visual concept related to the presence of antelope-like animals in natural environments, potentially detecting combinations of body shape, leg structure, and habitat context. It may also respond to similar geometric or textural patterns in artificial objects.

C61: Vehicles

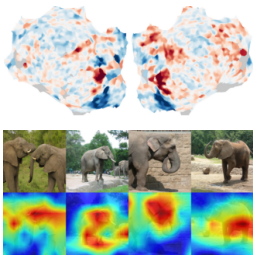


Left: signed family cortical prototype; four images for its representative FM.

16 member maps; spatial core; representative FM033. Held-out model/participant cosine: 0.80/0.28.

The full Qwen description appears in the individual-FM catalog under FM033.

C62: Visual features of elephants



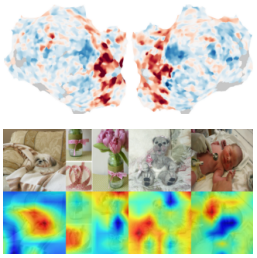
Left: signed family cortical prototype; four images for its representative FM.

16 member maps; spatial core; representative FM255. Held-out model/participant cosine: 0.88/0.24.

Qwen visual description. Module 255 responds to **visual features of elephants**, particularly tuskers and large elephant groups, with high activation in images containing African or Indian elephants. It is sensitive to the presence of tusks and elephant social groupings, showing strong responses across multiple layers in a widely distributed pattern.

Proposed model role. The module functions as a detector for elephant-related visual patterns, specifically highlighting tusker features and elephant group dynamics, contributing to the model's ability to distinguish elephant species and their social contexts.

C63: Household objects and personal items



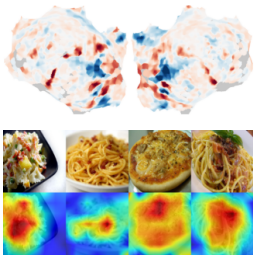
Left: signed family cortical prototype; four images for its representative FM.

16 member maps; spatial core; representative FM032. Held-out model/participant cosine: 0.51/0.31.

Qwen visual description. Module 32 responds to visual patterns associated with **household objects and personal items**, particularly those involving containers, textiles, and decorative elements. It is activated by scenes containing items like purses, shower curtains, aprons, and jewelry, suggesting a focus on object categories related to domestic environments and personal accessories.

Proposed model role. The module functions to detect and represent visual features associated with personal and household objects, contributing to the identification of items commonly found in homes or stores.

C64: Food

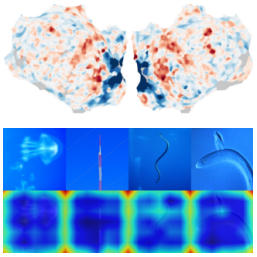


Left: signed family cortical prototype; four images for its representative FM.

16 member maps; spatial core; representative FM084. Held-out model/participant cosine: 0.99/0.42.

The full Qwen description appears in the individual-FM catalog under FM084.

C65: Elongated, vertical structures with high contrast against a dark background



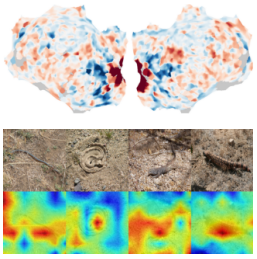
Left: signed family cortical prototype; four images for its representative FM.

16 member maps; spatial core; representative FM057. Held-out model/participant cosine: 0.79/0.23.

Qwen visual description. Module 57 responds to **elongated, vertical structures with high contrast against a dark background**, particularly in aquatic or sky contexts. It is activated by objects like poles, flagpoles, missiles, and similar slender forms, often associated with motion or directionality.

Proposed model role. The module encodes the presence of vertically oriented, high-contrast linear structures, likely serving as a feature detector for directional or structural cues in complex scenes.

C66: Reptiles and amphibians



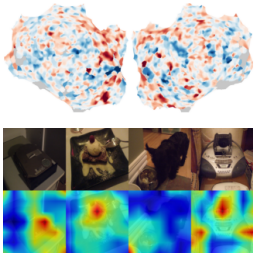
Left: signed family cortical prototype; four images for its representative FM.

16 member maps; spatial core; representative FM101. Held-out model/participant cosine: 0.64/0.18.

Qwen visual description. Module 101 responds to images containing **reptiles and amphibians**, particularly snakes, lizards, and crocodiles, with high activation in naturalistic settings. It is suppressed by artificial objects and human-made structures.

Proposed model role. This module functions as a detector for reptilian and amphibian animals, likely contributing to the recognition of these classes in visual scenes by emphasizing their presence and suppressing irrelevant categories.

C67: Kitchen utensils and food items



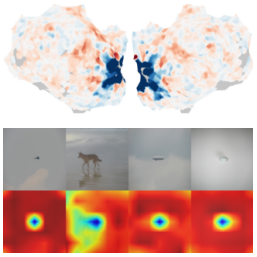
Left: signed family cortical prototype; four images for its representative FM.

16 member maps; spatial core; representative FM178. Held-out model/participant cosine: 0.44/0.05.

Qwen visual description. Module 178 activates in response to visual scenes containing **kitchen utensils and food items**, particularly those involving cooking or serving contexts. It shows strong activation for objects like spoons, spatulas, ladles, and food such as mashed potatoes, broccoli, and ice cream. The module is sensitive to specific spatial configurations of these objects within a scene.

Proposed model role. This module encodes a concept related to kitchen utensils and food preparation, likely representing a high-level feature that combines object identity (utensils/food) with contextual spatial relationships indicative of culinary activity.

C68: Small, dark, isolated objects in the visual field



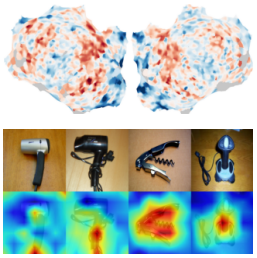
Left: signed family cortical prototype; four images for its representative FM.

16 member maps; spatial core; representative FM197. Held-out model/participant cosine: 0.90/0.55.

Qwen visual description. Module 197 responds to **small, dark, isolated objects in the visual field**, particularly insects and similar entities. It is activated by high-contrast, compact shapes against light backgrounds, with a strong preference for centrally located or slightly off-center targets. The module shows consistent activation across diverse object types but shares a common visual signature of small, dark, distinct forms.

Proposed model role. The module functions as a detector for small, dark, high-contrast objects, likely serving as a primitive feature extractor for identifying potential points of interest such as insects or anomalies in the visual scene.

C69: Objects with mechanical or tool-like features



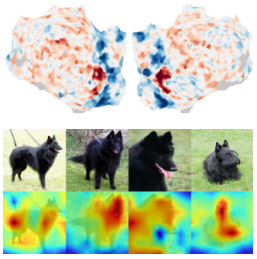
Left: signed family cortical prototype; four images for its representative FM.

16 member maps; spatial core; representative FM021. Held-out model/participant cosine: 0.66/0.20.

Qwen visual description. Module 21 responds to **objects with mechanical or tool-like features**, particularly those involving handles, levers, or articulated parts. It is activated by items such as hair dryers, corkscrews, tripods, and camera equipment, suggesting a focus on functional object components.

Proposed model role. The module encodes the presence of mechanical or manipulable parts in objects, distinguishing them from static or passive objects.

C70: Black dogs



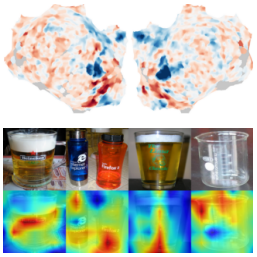
Left: signed family cortical prototype; four images for its representative FM.

16 member maps; spatial core; representative FM225. Held-out model/participant cosine: 0.85/0.28.

Qwen visual description. Module 225 activates strongly for images of **black dogs**, particularly breeds with long, dense fur such as Groenendael and Schipperke. It is sensitive to the presence of dark-furred canines in various poses and environments, showing consistent activation across multiple positive examples.

Proposed model role. The module functions as a detector for black dogs with long, dense fur, contributing to the model's ability to distinguish these specific canine types from others based on color and texture cues.

C71: Drinking vessels



Left: signed family cortical prototype; four images for its representative FM.

16 member maps; spatial core; representative FM046. Held-out model/participant cosine: 0.91/0.30.

Qwen visual description. Module 46 responds to visual patterns associated with **drinking vessels**, particularly those used for alcoholic beverages such as beer glasses, whiskey jugs, cocktail shakers, and goblets. It is sensitive to the shape, transparency, and context of these containers, especially when they are filled or placed on surfaces like bar counters.

Proposed model role. The module functions as a detector for drinking vessels, particularly those associated with alcoholic beverages, by encoding features related to their shape, transparency, and contextual placement.

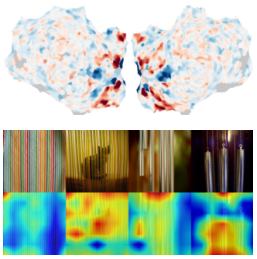
C72: Red surfaces and materials

Cortical and image evidence: Figure 6.

16 member maps; spatial core; representative FM233. Held-out model/participant cosine: 0.90/0.43.

The full Qwen description appears in the individual-FM catalog under FM233.

C73: Enclosed or partitioned spaces

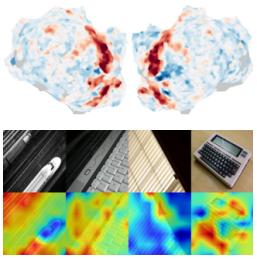


Left: signed family cortical prototype; four images for its representative FM.

16 member maps; spatial core; representative FM224. Held-out model/participant cosine: 0.84/0.26.

The full Qwen description appears in the individual-FM catalog under FM224.

C74: Mechanical or technological objects



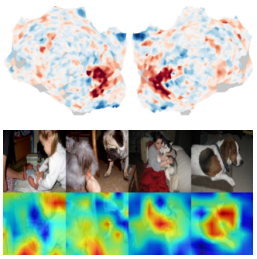
Left: signed family cortical prototype; four images for its representative FM.

16 member maps; spatial core; representative FM201. Held-out model/participant cosine: 0.90/0.40.

Qwen visual description. Module 201 responds to visual patterns associated with **mechanical or technological objects**, particularly those involving linear structures, tools, and devices with sharp edges or defined geometric forms. It is activated by images containing keyboards, cleavers, rulers, pens, and similar items, often highlighting specific regions such as blades, keys, or straight edges.

Proposed model role. The module functions as a detector for structured, man-made objects with linear or geometric features, likely contributing to the recognition of tools, electronic devices, and mechanical parts in visual scenes. Its activation correlates with the presence of sharp edges, repeating patterns, and precise alignments.

C75: Small dogs



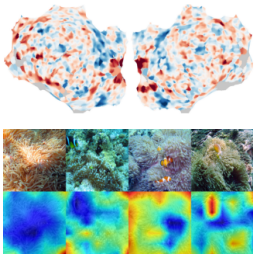
Left: signed family cortical prototype; four images for its representative FM.

16 member maps; spatial core; representative FM036. Held-out model/participant cosine: 0.87/0.25.

Qwen visual description. Module 36 responds to visual patterns associated with **small dogs**, particularly breeds like Pekinese and Shih-Tzu, often in domestic settings. It also activates for objects such as sleeping bags, remote controls, and other household items, suggesting a broader role in detecting compact, textured objects or scenes involving pets and human interaction.

Proposed model role. The module encodes the presence of small, fluffy dogs in domestic settings and may contribute to distinguishing them from other animals or objects based on texture, size, and context.

C76: Marine invertebrates



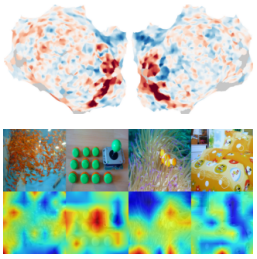
Left: signed family cortical prototype; four images for its representative FM.

16 member maps; candidate outside the spatial core; representative FM122. Held-out model/participant cosine: 0.50/0.04.

Qwen visual description. Module 122 responds to **marine invertebrates**, particularly sea anemones and related reef organisms, with strong activation in coral reef scenes. It is sensitive to specific textures and spatial configurations of soft-bodied marine life.

Proposed model role. The module encodes the presence of soft-bodied marine invertebrates in reef environments, distinguishing them from hard-bodied animals and human artifacts in similar contexts.

C77: Small, colorful, and textured objects



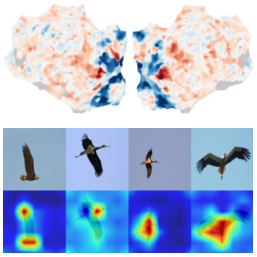
Left: signed family cortical prototype; four images for its representative FM.

16 member maps; spatial core; representative FM246. Held-out model/participant cosine: 0.78/0.25.

Qwen visual description. Module 246 responds to visual patterns associated with **small, colorful, and textured objects**, particularly those resembling toys, food items, or aquatic life. It is activated by images containing bright colors, repetitive textures, and distinct shapes such as fish, balloons, or puzzle pieces. The module shows strong localization in specific image regions and exhibits a functional role in distinguishing between playful/childlike objects and more utilitarian or industrial items.

Proposed model role. The module functions as a detector for visually engaging, child-oriented, or playful objects. It enhances the representation of items that are typically associated with leisure, play, or novelty, while suppressing representations of mundane or functional objects.

C78: Birds of prey



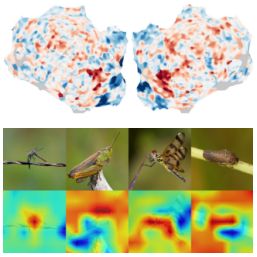
Left: signed family cortical prototype; four images for its representative FM.

16 member maps; spatial core; representative FM037. Held-out model/participant cosine: 0.88/0.41.

Qwen visual description. Module 37 activates in response to **birds of prey**, particularly raptors such as vultures, kites, and storks, with strong localization to the bird's body and wings. It is suppressed by non-raptor animals, objects like umbrellas, and certain bird species such as poodles or jays.

Proposed model role. This module encodes a high-level concept of 'bird of prey' by detecting key morphological and contextual cues such as wing span, flight posture, and sky background, enabling discrimination between predatory birds and other categories.

C79: Small flying insects



Left: signed family cortical prototype; four images for its representative FM.

16 member maps; candidate outside the spatial core; representative FM244. Held-out model/participant cosine: 0.45/0.11.

Qwen visual description. Module 244 activates in response to images containing **small flying insects**, particularly damselflies and dragonflies, with high activation observed in close-up views of their bodies and wings. The module shows strong localization to the insect's body and wing regions, especially when these features are clearly visible against a natural background.

Proposed model role. The module functions as a detector for small flying insects, specifically targeting damselflies and dragonflies, by encoding their distinctive physical characteristics such as wing transparency, body segmentation, and coloration. It contributes to distinguishing these insects from other categories in the visual hierarchy.

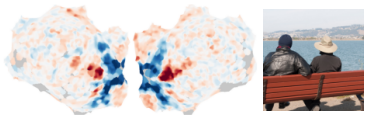
I CONDITION-SPECIFIC SUBATLASES

Each condition groups the 1,000 held-out NSD images by an object label, scene label, face-context label, or the upper or lower quartile of one measured image attribute. A family is called in one participant–model combination when any of its member FMs has a positive response; the call is then balanced across the 16 combinations. A **preferred** family is called more often inside than outside the condition and passes the fixed effect-size, cross-combination and multiplicity criteria. An **additional shared** family meets the common-participation criteria but is not already preferred. Both layers draw from the primary 62-family spatial core; their union is the listed condition subatlas. Preferred and additional lists are disjoint. A family can appear in several conditions.

The small cortical view in each entry is a call-rate-weighted composition of selected signed family prototypes, not a measured condition-average fMRI map or a refitted encoding model. The adjacent image is one actual member chosen to resemble the condition's average family-call pattern. Its own active families may be fewer or different from the aggregate condition list. Figure 8 shows four examples at larger size; all 82 memberships are given below. Red and blue retain the signed map convention; numerical display ranges are in the supplementary scale table.

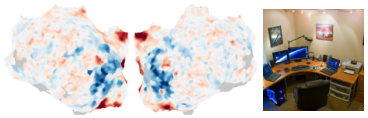
I.1 PEOPLE AND TRANSPORT

Person



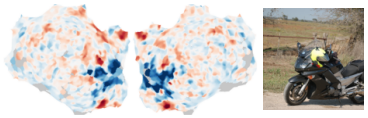
Left: signed cortical composition; one representative stimulus.
438 NSD images; 20 families in the subatlas.
Preferred (13): C05, C09, C12, C31, C33, C35, C37, C38, C39, C45, C49, C51, C65.
Additional shared (7): C00, C01, C16, C26, C57, C68, C77.

Car



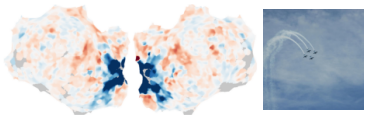
Left: signed cortical composition; one representative stimulus.
74 NSD images; 16 families in the subatlas.
Preferred (6): C09, C12, C30, C39, C55, C61.
Additional shared (10): C00, C01, C16, C26, C33, C35, C37, C57, C65, C68.

Motorcycle



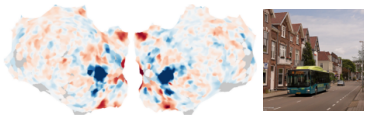
Left: signed cortical composition; one representative stimulus.
21 NSD images; 12 families in the subatlas.
Preferred (4): C09, C21, C37, C61.
Additional shared (8): C00, C12, C16, C26, C55, C65, C68, C73.

Airplane



Left: signed cortical composition; one representative stimulus.
30 NSD images; 11 families in the subatlas.
Preferred (8): C12, C37, C51, C59, C61, C65, C68, C78.
Additional shared (3): C00, C38, C55.

Bus



Left: signed cortical composition; one representative stimulus.
25 NSD images; 17 families in the subatlas.
Preferred (6): C12, C36, C39, C55, C61, C72.
Additional shared (11): C00, C01, C09, C26, C33, C37, C59, C65, C66, C68, C77.

Train

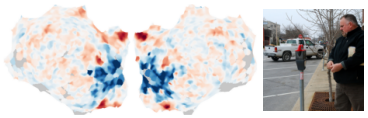
Cortical composition and example image: Figure 8.

36 NSD images; 12 families in the subatlas.

Preferred (4): C12, C36, C55, C61.

Additional shared (8): C00, C01, C16, C26, C37, C65, C68, C77.

Truck



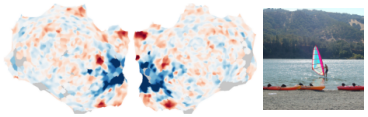
Left: signed cortical composition; one representative stimulus.

30 NSD images; 13 families in the subatlas.

Preferred (4): C12, C37, C59, C61.

Additional shared (9): C00, C01, C09, C16, C26, C55, C57, C65, C68.

Boat



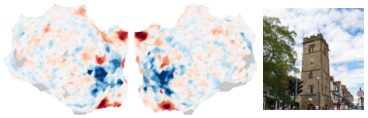
Left: signed cortical composition; one representative stimulus.

33 NSD images; 15 families in the subatlas.

Preferred (5): C12, C31, C37, C59, C65.

Additional shared (10): C00, C01, C16, C26, C48, C51, C55, C57, C68, C77.

Traffic light



Left: signed cortical composition; one representative stimulus.

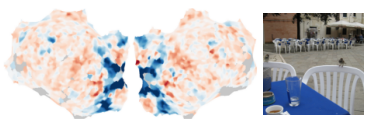
35 NSD images; 14 families in the subatlas.

Preferred (6): C09, C12, C30, C36, C55, C61.

Additional shared (8): C00, C01, C26, C37, C57, C65, C68, C77.

I.2 Animals

Bird



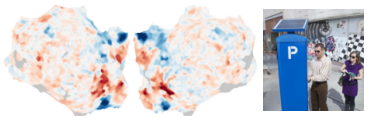
Left: signed cortical composition; one representative stimulus.

39 NSD images; 16 families in the subatlas.

Preferred (6): C00, C19, C31, C53, C66, C78.

Additional shared (10): C01, C12, C16, C26, C37, C51, C52, C65, C68, C77.

Cat

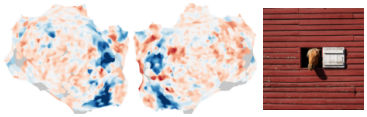


Left: signed cortical composition; one representative stimulus.
28 NSD images; 15 families in the subatlas.
Preferred (3): C52, C63, C75.
Additional shared (12): C00, C01, C04, C16, C26, C33, C51, C57, C66, C68, C74, C77.

Dog

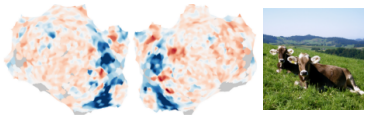
Cortical composition and example image: Figure 8.
28 NSD images; 14 families in the subatlas.
Preferred (4): C23, C24, C52, C70.
Additional shared (10): C00, C01, C12, C16, C26, C35, C37, C57, C66, C68.

Horse



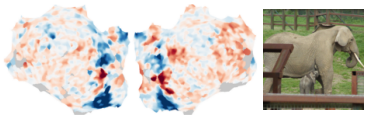
Left: signed cortical composition; one representative stimulus.
26 NSD images; 15 families in the subatlas.
Preferred (6): C16, C23, C24, C60, C62, C66.
Additional shared (9): C00, C01, C12, C20, C26, C51, C57, C65, C68.

Cow



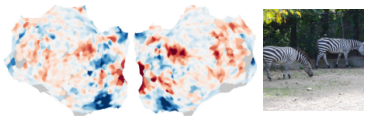
Left: signed cortical composition; one representative stimulus.
23 NSD images; 13 families in the subatlas.
Preferred (7): C01, C12, C16, C23, C60, C62, C70.
Additional shared (6): C00, C26, C51, C52, C66, C68.

Elephant



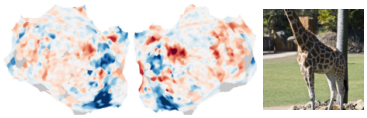
Left: signed cortical composition; one representative stimulus.
28 NSD images; 13 families in the subatlas.
Preferred (8): C01, C14, C16, C23, C48, C52, C60, C62.
Additional shared (5): C00, C12, C26, C66, C68.

Zebra



Left: signed cortical composition; one representative stimulus.
29 NSD images; 15 families in the subatlas.
Preferred (7): C03, C16, C50, C58, C60, C62, C66.
Additional shared (8): C00, C01, C12, C26, C37, C52, C57, C68.

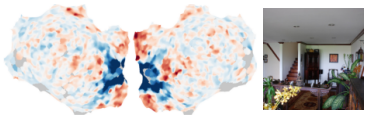
Giraffe



Left: signed cortical composition; one representative stimulus.
44 NSD images; 13 families in the subatlas.
Preferred (10): C01, C12, C15, C16, C23, C58, C60, C62, C66, C78.
Additional shared (3): C00, C26, C68.

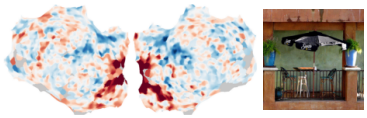
I.3 Clothing and sport

Backpack



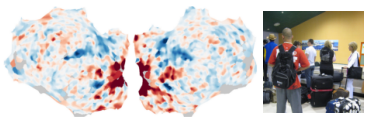
Left: signed cortical composition; one representative stimulus.
29 NSD images; 14 families in the subatlas.
Preferred (1): C49.
Additional shared (13): C00, C01, C09, C12, C16, C26, C37, C51, C55, C57, C65, C68, C77.

Umbrella



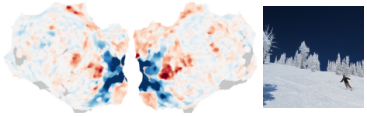
Left: signed cortical composition; one representative stimulus.
23 NSD images; 12 families in the subatlas.
Preferred (3): C09, C30, C77.
Additional shared (9): C00, C01, C12, C26, C33, C35, C57, C68, C72.

Handbag



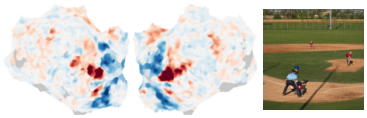
Left: signed cortical composition; one representative stimulus.
43 NSD images; 15 families in the subatlas.
Preferred (4): C09, C26, C35, C39.
Additional shared (11): C00, C01, C12, C16, C33, C37, C38, C57, C65, C72, C77.

Skis



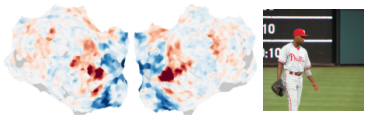
Left: signed cortical composition; one representative stimulus.
43 NSD images; 10 families in the subatlas.
Preferred (5): C04, C37, C49, C51, C65.
Additional shared (5): C00, C12, C26, C68, C78.

Sports ball



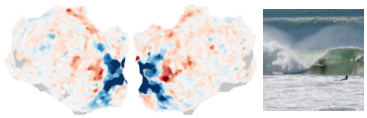
Left: signed cortical composition; one representative stimulus.
26 NSD images; 11 families in the subatlas.
Preferred (3): C16, C38, C69.
Additional shared (8): C00, C01, C12, C26, C35, C57, C68, C77.

Baseball glove



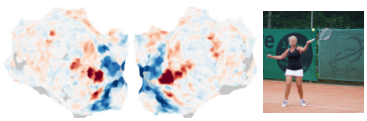
Left: signed cortical composition; one representative stimulus.
22 NSD images; 9 families in the subatlas.
Preferred (3): C16, C38, C69.
Additional shared (6): C00, C01, C26, C68, C72, C77.

Surfboard



Left: signed cortical composition; one representative stimulus.
48 NSD images; 10 families in the subatlas.
Preferred (9): C00, C01, C12, C31, C37, C48, C51, C65, C68.
Additional shared (1): C16.

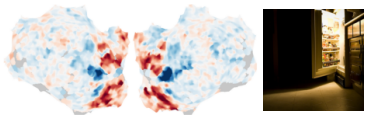
Tennis racket



Left: signed cortical composition; one representative stimulus.
29 NSD images; 11 families in the subatlas.
Preferred (4): C00, C16, C35, C38.
Additional shared (7): C01, C12, C26, C57, C65, C68, C72.

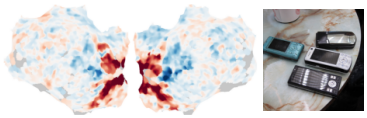
I.4 FOOD AND TABLEWARE

Bottle



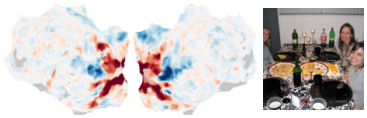
Left: signed cortical composition; one representative stimulus.
49 NSD images; 19 families in the subatlas.
Preferred (9): C10, C25, C35, C39, C54, C63, C67, C71, C73.
Additional shared (10): C00, C01, C26, C33, C37, C51, C64, C68, C72, C77.

Cup



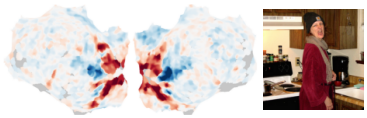
Left: signed cortical composition; one representative stimulus.
56 NSD images; 17 families in the subatlas.
Preferred (11): C10, C25, C26, C35, C54, C63, C64, C67, C71, C75, C77.
Additional shared (6): C00, C01, C33, C57, C68, C72.

Spoon



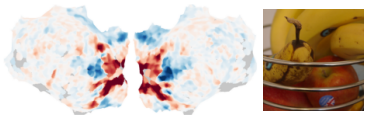
Left: signed cortical composition; one representative stimulus.
25 NSD images; 18 families in the subatlas.
Preferred (7): C10, C26, C27, C28, C46, C64, C67.
Additional shared (11): C00, C04, C20, C25, C37, C51, C54, C57, C63, C72, C77.

Bowl



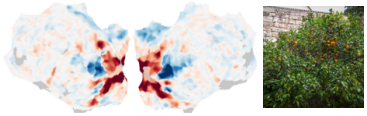
Left: signed cortical composition; one representative stimulus.
49 NSD images; 15 families in the subatlas.
Preferred (9): C10, C25, C26, C27, C28, C54, C63, C64, C67.
Additional shared (6): C00, C01, C04, C33, C51, C77.

Apple



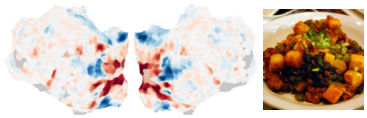
Left: signed cortical composition; one representative stimulus.
20 NSD images; 14 families in the subatlas.
Preferred (8): C21, C26, C27, C28, C39, C46, C64, C77.
Additional shared (6): C00, C16, C54, C55, C63, C72.

Orange



Left: signed cortical composition; one representative stimulus.
21 NSD images; 11 families in the subatlas.
Preferred (6): C27, C28, C39, C54, C64, C77.
Additional shared (5): C25, C26, C55, C57, C72.

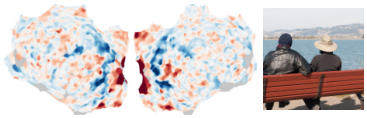
Broccoli



Left: signed cortical composition; one representative stimulus.
22 NSD images; 13 families in the subatlas.
Preferred (7): C03, C19, C26, C27, C46, C64, C77.
Additional shared (6): C00, C04, C16, C51, C54, C72.

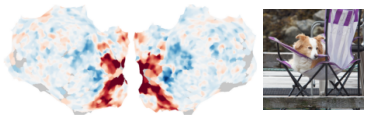
I.5 Household and built settings

Bench



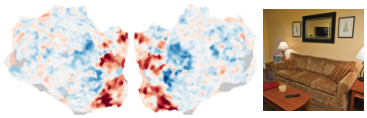
Left: signed cortical composition; one representative stimulus.
50 NSD images; 16 families in the subatlas.
Preferred (4): C09, C38, C55, C57.
Additional shared (12): C00, C01, C12, C16, C26, C33, C35, C37, C65, C66, C68, C77.

Chair



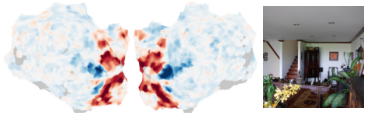
Left: signed cortical composition; one representative stimulus.
84 NSD images; 15 families in the subatlas.
Preferred (7): C25, C26, C35, C39, C54, C63, C67.
Additional shared (8): C00, C01, C09, C16, C33, C57, C72, C77.

Couch



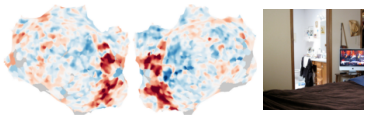
Left: signed cortical composition; one representative stimulus.
29 NSD images; 13 families in the subatlas.
Preferred (6): C25, C35, C57, C63, C67, C75.
Additional shared (7): C00, C01, C33, C54, C68, C72, C77.

Potted plant



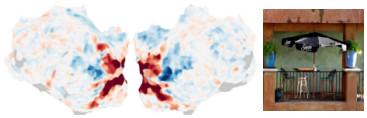
Left: signed cortical composition; one representative stimulus.
37 NSD images; 14 families in the subatlas.
Preferred (6): C10, C25, C54, C63, C67, C73.
Additional shared (8): C00, C09, C26, C48, C57, C68, C72, C77.

Bed



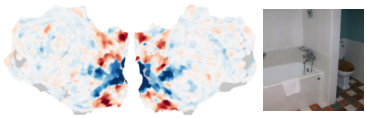
Left: signed cortical composition; one representative stimulus.
29 NSD images; 14 families in the subatlas.
Preferred (5): C02, C25, C54, C63, C75.
Additional shared (9): C00, C01, C04, C26, C33, C35, C52, C68, C77.

Dining table



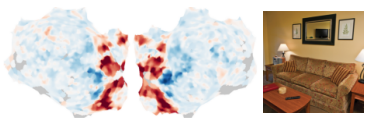
Left: signed cortical composition; one representative stimulus.
75 NSD images; 19 families in the subatlas.
Preferred (15): C10, C20, C25, C26, C27, C28, C35, C39, C54, C63, C64, C67, C71, C72, C77.
Additional shared (4): C00, C01, C33, C57.

Toilet



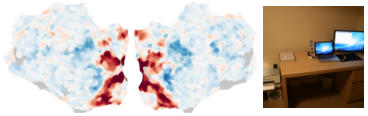
Left: signed cortical composition; one representative stimulus.
25 NSD images; 10 families in the subatlas.
Preferred (7): C14, C25, C54, C63, C67, C68, C73.
Additional shared (3): C00, C01, C51.

Television



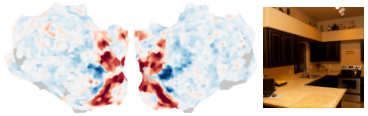
Left: signed cortical composition; one representative stimulus.
32 NSD images; 12 families in the subatlas.
Preferred (5): C14, C25, C63, C69, C75.
Additional shared (7): C00, C26, C33, C54, C57, C74, C77.

Laptop



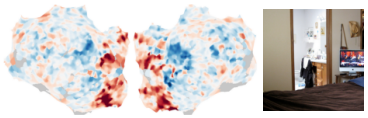
Left: signed cortical composition; one representative stimulus.
25 NSD images; 12 families in the subatlas.
Preferred (4): C25, C63, C67, C74.
Additional shared (8): C00, C01, C09, C26, C33, C35, C57, C77.

Computer mouse



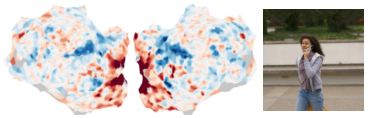
Left: signed cortical composition; one representative stimulus.
23 NSD images; 10 families in the subatlas.
Preferred (5): C02, C25, C67, C69, C74.
Additional shared (5): C00, C01, C26, C57, C77.

Remote



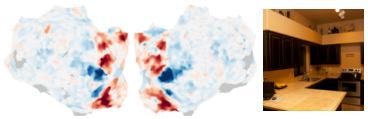
Left: signed cortical composition; one representative stimulus.
20 NSD images; 13 families in the subatlas.
Preferred (5): C25, C35, C63, C67, C75.
Additional shared (8): C00, C01, C33, C54, C57, C68, C72, C77.

Cell phone



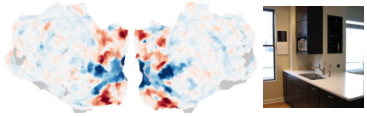
Left: signed cortical composition; one representative stimulus.
31 NSD images; 12 families in the subatlas.
Preferred (2): C35, C67.
Additional shared (10): C00, C01, C25, C26, C33, C37, C51, C57, C74, C77.

Oven



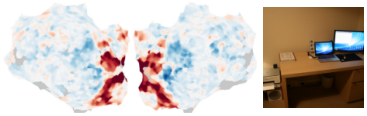
Left: signed cortical composition; one representative stimulus.
21 NSD images; 12 families in the subatlas.
Preferred (6): C10, C25, C54, C55, C63, C75.
Additional shared (6): C00, C01, C26, C37, C51, C68.

Sink



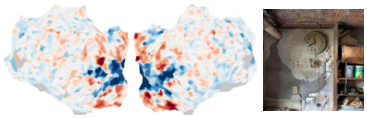
Left: signed cortical composition; one representative stimulus.
49 NSD images; 13 families in the subatlas.
Preferred (9): C10, C14, C25, C54, C55, C63, C67, C68, C73.
Additional shared (4): C00, C37, C51, C57.

Book



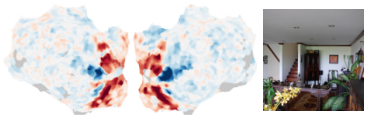
Left: signed cortical composition; one representative stimulus.
40 NSD images; 13 families in the subatlas.
Preferred (7): C25, C54, C63, C67, C74, C75, C77.
Additional shared (6): C00, C01, C26, C33, C35, C57.

Clock



Left: signed cortical composition; one representative stimulus.
46 NSD images; 16 families in the subatlas.
Preferred (4): C28, C43, C73, C78.
Additional shared (12): C00, C01, C12, C25, C26, C37, C51, C55, C57, C65, C68, C77.

Vase

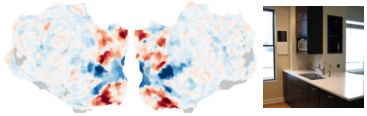


Left: signed cortical composition; one representative stimulus.
40 NSD images; 14 families in the subatlas.
Preferred (7): C10, C25, C54, C63, C67, C71, C72.
Additional shared (7): C00, C01, C26, C51, C57, C68, C77.

I.6 Scenes and face context

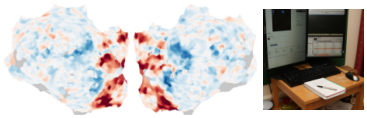
Dining scenes
Cortical composition and example image: Figure 8.
86 NSD images; 18 families in the subatlas.
Preferred (13): C20, C26, C27, C28, C35, C39, C54, C63, C64, C67, C71, C72, C77.
Additional shared (5): C00, C01, C33, C51, C57.

Kitchen scenes



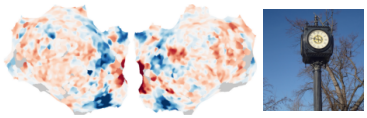
Left: signed cortical composition; one representative stimulus.
66 NSD images; 16 families in the subatlas.
Preferred (10): C10, C14, C25, C54, C55, C63, C67, C68, C73, C75.
Additional shared (6): C00, C01, C26, C37, C51, C57.

Living room scenes



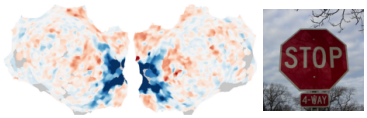
Left: signed cortical composition; one representative stimulus.
45 NSD images; 13 families in the subatlas.
Preferred (6): C25, C57, C63, C67, C69, C75.
Additional shared (7): C00, C01, C26, C33, C35, C68, C77.

Mixed scenes



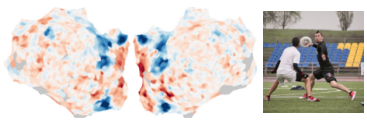
Left: signed cortical composition; one representative stimulus.
289 NSD images; 23 families in the subatlas.
Preferred (16): C03, C14, C15, C16, C19, C23, C24, C46, C50, C52, C53, C58, C60, C62, C66, C70.
Additional shared (7): C00, C01, C12, C26, C57, C68, C77.

Outdoor scenes



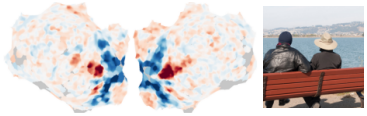
Left: signed cortical composition; one representative stimulus.
407 NSD images; 20 families in the subatlas.
Preferred (14): C09, C12, C31, C36, C37, C38, C49, C51, C55, C59, C61, C65, C68, C78.
Additional shared (6): C00, C01, C16, C26, C57, C77.

Portrait scenes



Left: signed cortical composition; one representative stimulus.
78 NSD images; 17 families in the subatlas.
Preferred (4): C23, C34, C35, C45.
Additional shared (13): C00, C01, C12, C16, C26, C33, C51, C54, C57, C66, C68, C72, C77.

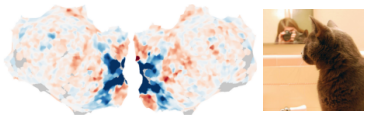
Images with a face



Left: signed cortical composition; one representative stimulus.
348 NSD images; 20 families in the subatlas.
Preferred (12): C05, C09, C31, C33, C35, C37, C38, C39, C45, C49, C51, C65.
Additional shared (8): C00, C01, C12, C16, C26, C57, C68, C77.

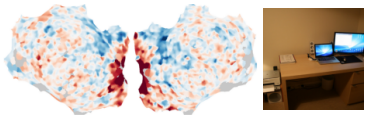
I.7 Image appearance

Brightness: high quartile



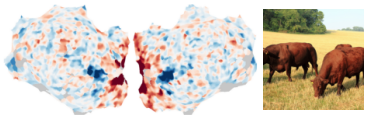
Left: signed cortical composition; one representative stimulus.
250 NSD images; 15 families in the subatlas.
Preferred (9): C12, C31, C37, C49, C51, C59, C65, C68, C78.
Additional shared (6): C00, C01, C16, C26, C66, C77.

Brightness: low quartile



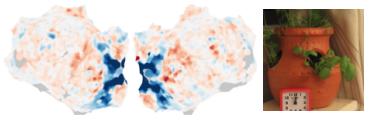
Left: signed cortical composition; one representative stimulus.
250 NSD images; 19 families in the subatlas.
Preferred (8): C09, C14, C26, C30, C35, C57, C67, C75.
Additional shared (11): C00, C01, C12, C16, C37, C54, C65, C66, C68, C72, C77.

Contrast: high quartile



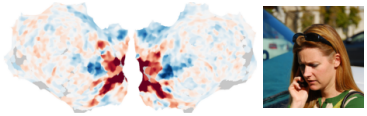
Left: signed cortical composition; one representative stimulus.
250 NSD images; 13 families in the subatlas.
Preferred (6): C09, C10, C26, C36, C37, C61.
Additional shared (7): C00, C01, C12, C16, C51, C54, C72.

Contrast: low quartile



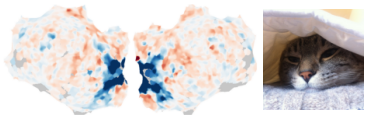
Left: signed cortical composition; one representative stimulus.
250 NSD images; 15 families in the subatlas.
Preferred (9): C16, C31, C38, C57, C60, C65, C66, C68, C78.
Additional shared (6): C00, C01, C12, C37, C51, C77.

Saturation: high quartile



Left: signed cortical composition; one representative stimulus.
250 NSD images; 19 families in the subatlas.
Preferred (15): C00, C02, C14, C16, C26, C27, C30, C46, C54, C57, C64, C67, C69, C72, C77.
Additional shared (4): C01, C25, C65, C66.

Saturation: low quartile

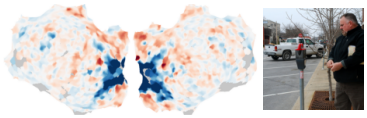


Left: signed cortical composition; one representative stimulus.
250 NSD images; 14 families in the subatlas.
Preferred (11): C01, C04, C12, C31, C37, C49, C51, C59, C61, C68, C73.
Additional shared (3): C00, C65, C66.

Colorfulness: high quartile

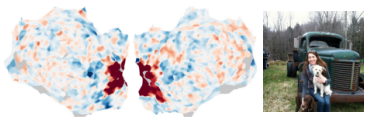
Cortical composition and example image: Figure 8.
250 NSD images; 21 families in the subatlas.
Preferred (13): C02, C26, C27, C30, C34, C35, C45, C46, C54, C64, C67, C72, C77.
Additional shared (8): C00, C01, C12, C16, C33, C51, C57, C65.

Colorfulness: low quartile



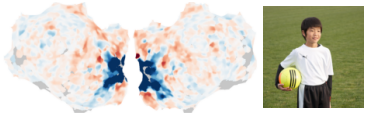
Left: signed cortical composition; one representative stimulus.
250 NSD images; 14 families in the subatlas.
Preferred (10): C01, C12, C31, C37, C49, C51, C59, C66, C68, C73.
Additional shared (4): C00, C25, C57, C65.

Edge strength: high quartile



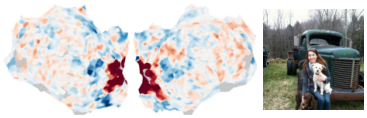
Left: signed cortical composition; one representative stimulus.
250 NSD images; 17 families in the subatlas.
Preferred (12): C03, C09, C15, C16, C23, C26, C39, C50, C58, C62, C64, C66.
Additional shared (5): C00, C01, C12, C37, C77.

Edge strength: low quartile



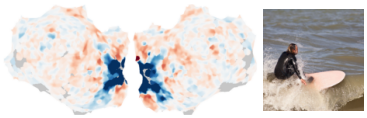
Left: signed cortical composition; one representative stimulus.
250 NSD images; 10 families in the subatlas.
Preferred (8): C12, C25, C37, C51, C57, C65, C68, C78.
Additional shared (2): C00, C01.

Edge density: high quartile



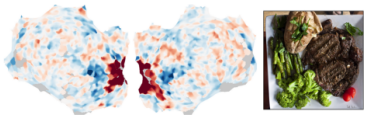
Left: signed cortical composition; one representative stimulus.
250 NSD images; 19 families in the subatlas.
Preferred (15): C03, C09, C15, C16, C23, C26, C27, C39, C46, C50, C58, C60, C62, C64, C66.
Additional shared (4): C00, C01, C12, C77.

Edge density: low quartile



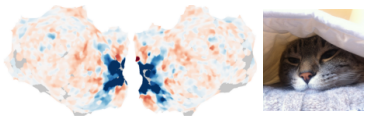
Left: signed cortical composition; one representative stimulus.
250 NSD images; 13 families in the subatlas.
Preferred (8): C12, C37, C49, C51, C57, C65, C68, C78.
Additional shared (5): C00, C01, C16, C25, C72.

Fine texture: high quartile



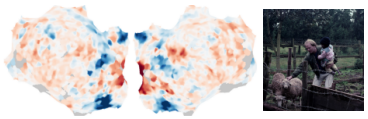
Left: signed cortical composition; one representative stimulus.
250 NSD images; 19 families in the subatlas.
Preferred (13): C03, C09, C15, C16, C26, C36, C39, C50, C55, C58, C61, C62, C66.
Additional shared (6): C00, C01, C12, C37, C65, C77.

Fine texture: low quartile



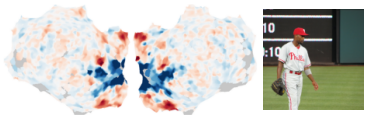
Left: signed cortical composition; one representative stimulus.
250 NSD images; 17 families in the subatlas.
Preferred (9): C19, C31, C45, C51, C54, C57, C65, C68, C78.
Additional shared (8): C00, C01, C12, C16, C25, C35, C37, C77.

Orientation diversity: high quartile



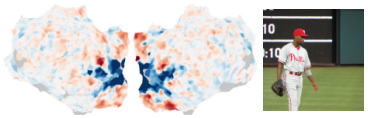
Left: signed cortical composition; one representative stimulus.
250 NSD images; 22 families in the subatlas.
Preferred (15): C03, C15, C16, C23, C26, C27, C45, C46, C52, C58, C60, C62, C64, C66, C70.
Additional shared (7): C00, C01, C12, C35, C54, C68, C77.

Orientation diversity: low quartile



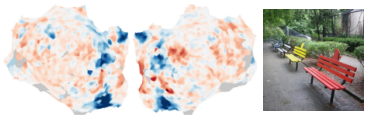
Left: signed cortical composition; one representative stimulus.
250 NSD images; 18 families in the subatlas.
Preferred (11): C02, C10, C25, C31, C37, C51, C55, C59, C61, C68, C73.
Additional shared (7): C00, C01, C12, C26, C57, C65, C77.

Orientation alignment: high quartile



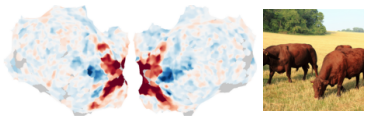
Left: signed cortical composition; one representative stimulus.
250 NSD images; 18 families in the subatlas.
Preferred (11): C02, C12, C25, C31, C51, C55, C59, C61, C65, C68, C73.
Additional shared (7): C00, C01, C26, C37, C57, C72, C77.

Orientation alignment: low quartile



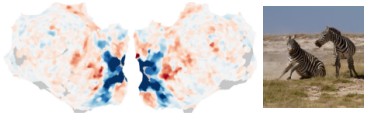
Left: signed cortical composition; one representative stimulus.
250 NSD images; 22 families in the subatlas.
Preferred (13): C03, C15, C16, C23, C26, C45, C46, C52, C58, C60, C62, C66, C70.
Additional shared (9): C00, C01, C12, C35, C37, C51, C54, C68, C77.

Grayscale diversity: high quartile



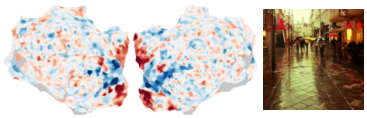
Left: signed cortical composition; one representative stimulus.
250 NSD images; 16 families in the subatlas.
Preferred (7): C09, C26, C27, C39, C54, C63, C64.
Additional shared (9): C00, C01, C12, C16, C25, C37, C66, C72, C77.

Grayscale diversity: low quartile



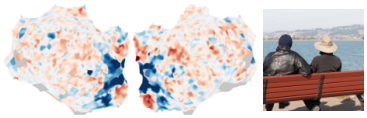
Left: signed cortical composition; one representative stimulus.
250 NSD images; 17 families in the subatlas.
Preferred (13): C04, C12, C21, C30, C31, C37, C38, C49, C51, C57, C65, C68, C78.
Additional shared (4): C00, C01, C16, C77.

Center–periphery difference: high quartile



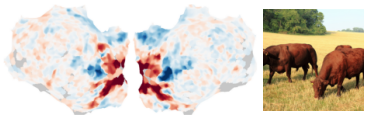
Left: signed cortical composition; one representative stimulus.
250 NSD images; 17 families in the subatlas.
Preferred (6): C25, C35, C54, C57, C74, C75.
Additional shared (11): C00, C01, C12, C16, C26, C37, C51, C65, C66, C68, C77.

Center–periphery difference: low quartile



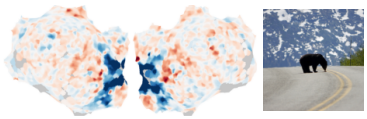
Left: signed cortical composition; one representative stimulus.
250 NSD images; 17 families in the subatlas.
Preferred (8): C04, C10, C12, C23, C36, C51, C62, C70.
Additional shared (9): C00, C01, C16, C26, C37, C65, C66, C68, C77.

Red–green balance: high quartile



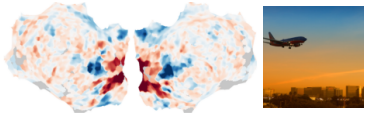
Left: signed cortical composition; one representative stimulus.
250 NSD images; 27 families in the subatlas.
Preferred (22): C02, C10, C14, C20, C25, C26, C27, C28, C30, C33, C34, C35, C39, C45, C46, C54, C64, C67, C69, C72, C75, C77.
Additional shared (5): C00, C01, C57, C66, C68.

Red–green balance: low quartile



Left: signed cortical composition; one representative stimulus.
250 NSD images; 16 families in the subatlas.
Preferred (9): C00, C12, C16, C31, C49, C59, C65, C68, C78.
Additional shared (7): C01, C26, C37, C51, C57, C66, C77.

Blue–yellow balance: high quartile



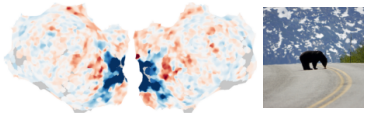
Left: signed cortical composition; one representative stimulus.

250 NSD images; 19 families in the subatlas.

Preferred (13): C14, C16, C23, C25, C26, C27, C46, C54, C60, C64, C67, C69, C72.

Additional shared (6): C00, C01, C57, C66, C68, C77.

Blue–yellow balance: low quartile



Left: signed cortical composition; one representative stimulus.

250 NSD images; 16 families in the subatlas.

Preferred (10): C12, C31, C37, C43, C49, C51, C59, C65, C68, C78.

Additional shared (6): C00, C01, C26, C33, C57, C77.