Cortical maps diverge, representations converge along cortical hierarchy

Bogdan Petre[1], Martin A. Lindquist[2], Tor D. Wager[1]
1. Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH.
2. Department of Biostatistics, Johns Hopkins University, Baltimore, MD.

ABSTRACT

Brain maps (e.g. retinotopy, somatotopy) vary across individuals. This is thought to reflect underlying computational differences. However, artificial neural networks (ANNs) show that similar performance and internal representations can coexist with diverse circuit layouts. Consequently, we tested the presumption that spatial diversity reflects representational diversity in the brain, but found this presumption often breaks down. Using task and resting-state fMRI data we compared regional functional topographies and representational geometries–the within-individual dissimilarities among activity patterns. Across individuals (n = 414), representations converged in higher-order cortex despite substantial topographic diversity, indicating that similar information was encoded by different, individual-specific activity patterns. Topography only tracked representational differences in sensory-motor cortices and regions under strong architectural constraints, such as myelination or laminar differentiation. We show this parallels ANNs: architectural permissiveness allows idiosyncratic layouts to arise from random initializations rather than learned representations. To test whether topographies and representations show analogous developmental origins, we examined twins (n = 394), and found topographies were more heritable than representations. This shows that representational convergence occurs across idiosyncratic layouts in both artificial and biological systems, but is moderated by architectural constraints on implementation flexibility. Accordingly, the relevance of localization- and representation-based paradigms of brain function depends on neural architecture. [200 words]

INTRODUCTION

Everyone's brain is unique[1]. We differ in functional connectivity[1] at rest, in intrinsic network topographies[2,3], and even in canonical topographic maps like retinotopy[4]. These variants affect perception[5,6] and motor function[7,8], but topographic variation is greatest in association cortices[9–11] thought to support higher-order cognition. This has led to the widespread belief that this topographic variation reflects human cognitive diversity[3,10,11], including pathological traits[12,13].

Nevertheless, the significance of individual variation in functional topographies may be overstated. Artificial neural networks (ANNs) produce some of our most successful brain encoding models[14–19] without employing brain-like topographies. Instead, ANNs point to a different organizing principle: representational convergence. Different network instances learn

predictable distinctions, or “representations”, of task contingencies and stimulus features despite distinct random initializations. These internal representations converge due to shared architecture and training data, giving rise to degenerate weights that support the same functions[20–22]. If representational convergence is a universal principle of neural systems, then a focus on topographic maps could come at a cost. It may misattribute neurodevelopmental variability in neural implementations to task-relevant computational differences and obscure representational structure that more directly supports behavior.

Here we asked whether variable topographies in the brain reflect variable representations, like in the canonical account of sensory-motor maps, or if they are an adaptive by-product of idiosyncratic initial conditions, like ANN weights. ANNs traditionally lack the physical and developmental constraints of real brains, which may permit representational convergence with flexible implementations across networks. However, even when physical constraints are imposed, emergent topographies remain idiosyncratic[23]. Further, topographies are most variable in transmodal cortex, which has the fewest physical constraints to bias implementations, like myelination[9] or highly differentiated laminae[24,25]. Indeed, topographic variation has been most informative for sensory traits[5,6], whereas higher-order traits are often weakly predicted[3] or better captured by spatially agnostic models[26–28]. However, the principles relating topographies and representations have not been directly investigated.

If the principles underlying representational convergence in ANNs are shared with the brain, we would expect to see more common representations and flexible implementations in areas that have fewer architectural constraints. We tested this hypothesis by juxtaposing computational modeling with analyses of task and resting-state fMRI data from the Human Connectome Project Young Adult cohort (HCP-YA)[29]. We compared spatial topographies with representational geometries (state-dissimilarity matrices), examining how each varies with wiring costs and network depth in ANNs, and whether they dissociate across cortex along analogous architectural gradients, including myelin density and network hierarchy. We contrasted these effects with other proposed determinants of topographic diversity, such as evolutionary recency and developmental timelines[9]. Finally, motivated by the role of initial conditions in shaping layouts of ANNs, we compared twins to test whether cortical topographies are more heritable than representations.

Our findings challenge the prevailing view of cortical organization. We found sensory-motor representations differed more among individuals than more abstract representations, which converged most in transmodal cortex, generalizing computational principles previously established in ANNs to the brain. After controlling for representational variation, topographic similarity showed a strong dependence on neural architecture, and was greatest in thin (less differentiated, more densely intraconnected[24,25]), myelinated, and unimodal brain areas. Topographies were also more heritable and genetically determined than representations. Together these results show that topographic variance is best understood as an architectural prior strongly influencing sensory and motor function, but individuals idiosyncratically converge upon a common cognitive scaffold in architecturally unconstrained areas, analogous to

representational convergence in ANNs. This reframes how we study cognitive diversity, from the constructs we measure, to the models we build and the experiments we design.

RESULTS

ANN models reveal functional convergence despite topographic divergence

First we established the principle that topographies, representations and function can vary in distinct ways by extending recent work on topographic deep artificial neural networks (TDANNs)[23]. We trained multiple convolutional networks which differed only in initialization and in whether training included a spatial wiring penalty. In all models, units within each layer were assigned locations in a 2-dimensional spatial grid, but in TDANNs the wiring penalty encouraged nearby units to develop similar tuning while “ResNets” lacked this constraint. Networks were trained using self-supervised learning on the ImageNet object recognition task. These models recapitulate many computational features of human brain circuits[14,23,30–32], but remain small enough to train many independent instances. Resultant networks were paired into non-intersecting dyads within network-class (TDANN or ResNet) for similarity comparisons.

Networks with idiosyncratic topographies learned similar deep representations. Unit activations of two example TDANNs to two example stimuli differed, and this dissimilarity compounded with network depth (**Figure 1A**). These two networks nevertheless classified stimuli in similar ways, e.g. when they correctly classified *pholcidae* or incorrectly classified an image of a Rhodesian ridgeback as an Ibizan hound. For visual intuition about the mechanism of functional convergence, we treated each neural unit as an axis, and used multidimensional scaling (MDS) to illustrate the high dimensional activation space in two exemplary layers (**Figure 1B**). Similar stimuli were positioned at similar distances from one another in both networks’ activation spaces for both layers, indicating similar internal representations. This illustrates how networks could converge on similar discriminative decisions despite dissimilar topographies.

Systematic comparison of all network dyads revealed that wiring cost and network depth had independent effects on topographic and representational similarity (**Figure 1C**, **Table 1**). We used cosine similarity of activations to evaluate topographies. Representational similarity was evaluated using linear centered kernel alignment (CKA), a cosine-similarity-based comparison of response geometry between networks[21,33]. We found representations were very similar (CKA: 0.91 ± 0.02, mean ± se; range: [-1,1]) despite dissimilar topographies ($\cos\theta$: 0.36 ± 0.03, mean ± se; range: [-1,1]). Across layers, topographic similarity decayed even further, while representational similarity changed very little. The effect of wiring cost was evaluated by comparing TDANNs with ResNets, and this showed opposite effects on topographies and representations. Wiring costs facilitated topographic convergence while hampering representational convergence. This raised the question of whether topographic or representational convergence had greater functional significance.

Regardless of model initialization and obtained topographies, models performed comparably within-class (TDANN: 41.5 ± 0.07%, ResNet: 45.7 ± 0.09% top-1 accuracy), but ResNets produced greater model accuracy (top-1 accuracy t_{13} = 32.0, p = 9.3e-14) and more similar classification decisions (TDANN: 54.5 ± 1.1%, ResNet: 59.9 ± 0.18% mean agreement ± se; t_6 = 4.7, p = 3.3e-3). Notably, model decisions were even more similar to one another (model agreement) than to ground truth labels (model accuracy), indicating they were in agreement even on many misclassifications. Agreement increased further after aggregating the 1000 ImageNet category labels into 50 semantic supersets of hypernyms (TDANN: 66.4 ± 1.0%, ResNet: 71.2 ± 0.14% mean agreement ± se), meaning disagreements were often at the level of granular semantic distinctions (e.g. breed of dog rather than dog vs. house). This pattern of model agreement revealed the functional consequence of representational convergence: networks performed more similar functions where representations were most similar (ResNets), not where topographies were most similar (TDANNs).

Having established this divergence in ANNs, we next asked if similar principles hold in biological neural networks where network hierarchy and myelination may pose circuit constraints similar to network depth and wiring costs in ANNs. Both are among the factors previously associated with interindividual topographic variation[9] (**Figure 1D**, **Table 2**), but their relationship with representational variation remains unexplored. We turned our attention to this problem next.

Representational geometries generalize to the brain, where they measure state discriminability

Given that representational similarity can dissociate from topographic similarity in ANNs, we first asked if the same might occur in the brain. **Figure 2** compares the same levels of analysis as **Figure 1**, topographic correspondence vs. representational correspondence, in human fMRI data. Two exemplary regions, primary visual cortex (V1) and dorsal lateral prefrontal area p9-46v, were selected in a dyad of unrelated HCP-YA participants (**Figure 2A**). Blood oxygen level dependent (BOLD) responses were evaluated across 23 task conditions spanning sensory, motor and cognitive domains, with 6 highlighted in the illustration, showing different topographic and representational characteristics in the two regions. While topographies were similar between individuals in the early sensory area V1 (spatial cosθ = 0.259; **Figure 2B**), they differed markedly in the higher order p9-46v (cosθ = 0.114; **Figure 2C**).

Representational geometry was defined in terms of multivariate distances between topographies (cross-validated Mahalanobis distances, or “crossnobis”), yielding representational dissimilarity matrices (RDMs; **Figure 2D-E**). These RDMs were very similar across participants in both regions ($V1_{RDM}$ cosθ = 0.941, $p9\text{-}46v_{RDM}$ cosθ = 0.915), indicating V1 had similar geometry across participants, but so did p9-46v despite its more idiosyncratic topographies. To facilitate direct comparison with our measure of representational similarity in ANNs, and account for co-dependencies across RDM rows and columns, we whitened RDMs before comparison. Whitened cosine similarity (WUC) of RDMs reduces to CKA in the deterministic limit[34], providing a principled analog for stochastic biological data.

In the context of stochastic biological signals, the functional relevance of crossnobis RDMs lies in their relationship to state discriminability and decodability. To illustrate this and provide an intuition about our representational measures, we fit binary nearest-centroid classifiers to task evoked responses in one run and tested them on responses obtained from a second run in a 2-fold cross-validation scheme. For the 6 conditions highlighted this resulted in 15 classifiers and 15 accuracy scores, which were highly correlated with crossnobis distances (**Figure 2F-G**). Across all 23 conditions and 207 dyads the mean correlation was consistently high for both regions (V1: $r = 0.50 \pm 0.08$; p9-46v: $r = 0.51 \pm 0.09$; mean ± SD), reflecting the joint reliability of classifiers and distances for these regions in this design. This means when two individuals' brain regions have similar representational geometry the same information can be decoded from them, even when topographies differ, using individual-specific linear decoders. We characterized regional geometries using crossnobis RDMs instead of brain decoding models because they are a more robust, unbounded (classification accuracy saturates around [0,1]; cf. **Figure 2G**, lower left) and continuous measure of discriminability than cross-validated binary classification accuracy[35].

Network hierarchy dissociates topographic and representational similarity of cortical responses

We extended the analysis of topographies and representations to all brain areas to ask where they were similar or dissimilar among unrelated HCP participants ($n = 414$). Following our approach with ANNs, we paired off participants into non-intersecting dyads ($n = 207$) and evaluated within-dyad similarity. While both topographies and representational geometries showed high similarity in early sensory-motor regions, topographic similarity decayed markedly in posterior parietal cortex and dorsal lateral prefrontal cortices, whereas representational similarity remained substantially higher in these same regions (**Figures 3A-B**). Comparing standardized similarities revealed a dramatic dissociation: these regions were among the least topographically similar but among the most geometrically similar regions in the cortex (**Figure 3C**).

Next we inspected relationships with cortical gradients implicated in topographic diversity by prior studies[9]. Plotting topographic and representational similarity against network hierarchy (**Figure 3D**) revealed a similar dissociation as that seen in ANNs, but with a twist. TDANNs and ResNet topographies showed a dramatic decay with depth, with minimal representational depth dependence (**Figure 1C**). In the brain, topographic similarity showed slight decay with network hierarchy, whereas representational similarity increased with network hierarchy. Nevertheless, multiple prior factors have been associated with topographic diversity, including evolutionary and developmental features (**Table 2**). Each offers a plausible alternative explanation for topographic diversity, representational diversity, or both, and must be taken into account before concluding that network hierarchy offers the best explanation.

To test these competing hypotheses we introduced a jackknife adjustment to the cortical spin test (Methods: *Jackknife-adjusted spin tests*). Spin tests are association tests designed to control for spatial autocorrelation, but do not account for sampling variance across

participants[36]. In Monte-Carlo simulations we found this test could be overly liberal in some cases, while a simple jackknife adjustment for sampling error resulted in a well calibrated substitute (**Supplemental Figures 3-4**). Consequently, the spatial tests we performed across the brain are more conservative than many prior studies. Additionally, we controlled for temporal signal-to-noise ratio (tSNR) and test-retest reliability in all spatial comparisons.

Systematic comparisons under this framework revealed that only network hierarchy and CogPC1 (the principal component of variation across neurosynth-derived term-maps) significantly dissociated representations from topography (**Figure 3E, Table 3)**. These factors also identified areas of high representational similarity independently (**Figure 3E**, center), but no factor was significantly associated with topographic diversity specifically. Either the more conservative inferential framework, inclusion of signal quality corrections, or choice of parcellation might explain the discrepancy with prior studies. However, the network hierarchy effects were robust regardless of parcellation (**Supplemental Figure 5E**) or of whether or not noise confound corrections were made (**Table 3**). Together this shows that task representations converge least in unimodal areas, including higher order visual, auditory and sensory-motor regions, and converge more with greater network hierarchy, mirroring the dissociation observed in ANNs.

Topographic similarity is only sensitive to geometric similarity where architecture constrains neural implementations

If topographies reflect underlying information content, then topographic similarity should covary with representational similarity. Therefore we used across-dyad regression to test how sensitive topographic similarity was to representational similarity. We found different sensitivity in different brain areas. For example, in V1 individuals with more similar topographies had more similar representations. Conversely, area p9-46v showed no relationship between topographies and representations despite substantial variation in representations (**Figure 4A**). Across regions there was a broad spectrum of topographic sensitivity to representational diversity (**Figure 4B**). Topographic differences were sensitive to representational differences in regions that were more myelinated, thinner, and more unimodal. Conversely, topographic differences in unmyelinated, thick, transmodal regions were not explained by representational differences (**Figure 4C**, **Table 3**, rightmost column). GenePC1 (the principal component of transcriptional variation across the cortex) was also associated with sensitivity, consistent with its strong enrichment for myelin-related markers[37]. Together, these results show that response patterns in task-evoked fMRI primarily reflect underlying representations in regions where cortical architecture constrains how representations can be implemented.

Hierarchical dissociation of representations and topographies generalizes to the cerebellum

We next asked whether the dissociation between topography and representations observed across the cortical hierarchy generalized to a distinct neural architecture: the cerebellum.

Whereas the canonical cortical microcircuit is the multi-laminar column, a putative discriminative feature detector, the cerebellum is composed of numerous, fine, time-dependent inhibitory and putatively predictive circuits. Although formal gradients are not available for the cerebellum, it has areas with known unimodal topographies, as well as areas which appear transmodal based on their functional connectivity with the cortex[38]. Additionally, like the cortex but unlike many other subcortical structures, the cerebellum has reasonably good tSNR and test-retest reliability in this dataset (**Supplemental Figure 1**). This enabled a preliminary evaluation of how general the network hierarchy principle might be in determining topographic or representational convergence.

The cerebellum recapitulated the associations with network hierarchy found in the cortex. The crura at the posterior cerebellar pole are functionally connected to transmodal cortical circuits like default mode and control networks and are thus putatively transmodal, while the more anterior lobules V and VI of the superior cerebellar surface host motor representations of hand and tongue and are putatively unimodal[38]. While these structures were matched on topographic similarity, the transmodal-like crura showed substantially greater representational similarity than the unimodal-like lobules (mean crura $\cos\theta = 0.059 \pm 0.009$, WUC = 0.530 ± 0.009; lobules V and VI mean $\cos\theta = 0.060 \pm 0.007$, WUC = 0.351 ± 0.007; **Figure 3A-B**). Meanwhile, in the putatively unimodal lobules, topographic similarity was more sensitive to representational similarity (mean $\beta = 0.077 \pm [0.060, 0.093]$, vs. mean crura $\beta = 0.042 \pm [0.025, 0.056]$; **Figure 4A**). Collectively these findings support the notion that network hierarchy dissociates representational geometry from topographic similarity across neural architectures.

Architecture dissociates representational and topographic similarity of intrinsic brain networks

We tested the generalizability of our findings beyond the 23 task conditions of the HCP-YA dataset by examining representational and topographic similarity of spontaneous brain activity. We identified 25 individualized resting-state networks (RSNs) in each of our unrelated HCP participants using *a priori* group-mean network templates as a reference (Methods: Dual Regression), and analyzed individualized networks using the same analytical framework as for task evoked-responses. Under this framework topography measures the spatial layout of a network while representational geometry measures how segregated or discriminable (within-individual) the local network layouts are. We then compared these constructs within the same participant dyads as were used for task-evoked data.

As with task responses, topographic and representational similarity of RSNs showed differing spatial profiles (**Figure 5A-C**). For clarity, we illustrate representational and topographic similarity along the network hierarchy gradient (**Figure 5D**) before examining the biological factors associated with the apparent dissociation. Systematic comparison showed that network topographies were more similar in architecturally constrained areas (more myelinated, earlier in the network hierarchy). Meanwhile, representations were consistently more similar in areas with fewer architectural constraints (less myelinated, thicker, later in the network hierarchy; **Table 4, Figure 5E**). This apparent inversion was statistically significant (myelination, thickness and

network hierarchy all $p < 0.05$, FDR $q = 0.05$). Representations were additionally most similar in evolutionarily recent areas (EvoExp2 and FCHomology), and were associated with CogPC1 and CBF1. Of these, network hierarchy showed the largest absolute regression coefficient, mirroring its dominant association in task-evoked analyses (mean dyad geometry - topography, $\Delta\beta_{std} > 0.49$). These results are well aligned with the conclusions we reached based on task-evoked responses (**Figure 3**), and identify architectural constraints in particular as robust indicators of whether or not common representations had flexible implementations across participants.

For completeness, we additionally inspected the sensitivity of RSN topographies to representational similarities (**Table 3**), analogous to task-based sensitivity effects (**Figure 4**). Unlike task responses, this did not identify any dependencies with *a priori* spatial gradients for RSNs, suggesting that the observed dissociations of RSNs along cortical gradients primarily reflected population-level spatial trends rather than across-individual sensitivity of topographies to representations.

Topographies are more heritable than representations

In ANNs matched on training data, architecture and learning rules, representational geometry converges through learning despite implementation differences due to differing network initializations[20,21]. This raises the question of whether an analogous dissociation exists between representational geometry and cortical topography in the brain. Given what is already known about the heritability of functional topographies[3,39], we asked if representational geometry was comparatively more malleable and shaped by experience. We tested this using a twin design, comparing identical (monozygotic, MZ) and fraternal (dizygotic, DZ) twins with non-twin siblings and unrelated individuals (**Supplemental Table 1**). Genetic relatedness and shared developmental environments vary systematically across these family clusters, allowing us to dissociate genetic priors from environmental effects on cortical topographies and representational geometries. We limited ourselves to well controlled nonparametric contrasts and averages across regions. This minimized exposure to spurious nonneural confounds (Methods: *Genetic analysis*) and sample size limitations in an otherwise demanding inferential problem[40,41].

We found topographic and representational similarity were greater among genetically related individuals, with similarity decreasing as genetic proximity declined (**Figure 6A,D**). Similarities among half-sibling task responses were an exception, but this could be explained by demographic differences (baseline gene frequencies[42] or gene-environment interactions[43]; **Supplemental Table 1**). Similarities otherwise decayed monotonically with genetic relatedness.

Topographies were relatively more heritable than representations (**Figure 6B,E, Supplemental Table 3**). In both tasks and resting-state data, topographies in MZ twins were significantly more similar than DZ twins, indicating a strong genetic influence. Representational geometry also showed evidence of heritability, particularly for RSNs, but the genetic effect was significantly larger for topographies than representations in both modalities.

These differences could not be explained by differences in measurement reliability. Heritability effects were stronger for topographies than representations despite worse signal-to-noise ratio (**Supplemental Table 3)** and lower test-retest reliability (**Figure 6A,B**, gray bars).

Representations additionally showed sensitivity to developmental environments (**Figure 6C,F**, **Supplemental Table 4**). Representational geometry was more similar in DZ twins than full-siblings, who share comparable levels of genetic relatedness but differ in rearing environment. This was true for both task responses and RSNs. The common environment effect was not significant for topographies, but also did not significantly differ significantly from representations. Thus, representations are malleable, but not uniquely so.

Together, these results indicate that cortical topography is more strongly constrained by heritable factors, and that representational geometry is also shaped by experience.

DISCUSSION

We show that whether the spatial organization of brain responses reliably reflects underlying computations depends on whether their implementations are constrained by neural architecture. By profiling topographies and representations across the brain we show these traits follow unique profiles of idiosyncratic variation. Topographies diverge most from representations where network hierarchy is most transmodal, and in thicker, less myelinated cortex. However, rather than becoming less predictable, representations in these regions are the most consistent across individuals. We found little evidence supporting competing explanations of this implementation flexibility, like evolutionary recency or developmental timelines of brain regions. This shows that functional variation is associated with architectural constraints.

Architecture primarily constrains neural implementations, not representations. Myelinated tracts that are long distance and genetically guided[44,45] limit the range of viable implementations[46,47]. Thinner cortex may reflect a myelination related bias[48], but is also associated with greater laminar differentiation and reduced intracortical connectivity[24,25] which constrains feasible computations. Meanwhile, network hierarchy is an indirect measure of polysynaptic distance from sensory and motor areas of the brain[49]. In unimodal areas long distance projections enforce topography, but as hierarchy increases abstract composition and integration rather than spatial receptive fields are likely to determine functional selectivity, as occurred in deep network layers of our ANN models. The pattern of dependence on architectural constraints broadly generalizes across task responses and intrinsic activity as well as across distinct neural architectures like the cortex and cerebellum, consistent with a general organizing principle. Collectively, this doesn't mean topographies are irrelevant, but it does encourage alternative approaches to functional localization when studying higher cognitive function.

Further extending this principle to artificial systems is outside the scope of this study, but our ANNs do serve an exemplary purpose. They showed that network depth and wiring costs can

affect the relationship between topography and representational similarity in a controlled setting. Additionally, they operationalized Marr's levels of analysis: implementation, representation and function[50]. In our models the most similar functional output (image classification) was obtained from networks with more similar representations, ResNets, despite more similar implementations in TDANNs. This establishes a framework through which to view analogous measures in the brain. Heritability analysis was consistent with the analogy, showing that representations are sensitive to shared environments and are more malleable than topographies, reminiscent of representations that are learned from idiosyncratic initial conditions in our computational models.

Representations have practical significance for brain decoding which arises directly from how representational geometry is constructed. By using cross-validated Mahalanobis (crossnobis) distances and WUC, our approach preserves a meaningful zero point and compares geometric proportions rather than pattern correlations alone. Unlike traditional correlation based representational similarity analysis[51], which removes overall response magnitude and idiosyncratically centers each response, crossnobis distinguishes responses differing in amplitude and spatial pattern in a common reference frame (same zero point). This better reflects state decodability and more closely resembles covariance based descriptions of neural manifolds used to characterize population dynamics in electrophysiology[15,52–54].

From this perspective, conserved representational geometry reflects conserved manifolds across individuals, while topographic similarity indicates how these manifolds are embedded in the anatomy by individual-specific population codes. These codes are heritable, and their representational significance varies systematically with architecture, rather than reflecting unconstrained initial conditions like in traditional ANNs. Accordingly, topographies warrant study and explanation in their own right[55]. However, our findings do clarify the relationship between implementation and representational levels of neural analysis, and identify a form of functional degeneracy[56,57]. This complements findings from electrophysiology showing conserved neural manifolds across animals performing similar behaviors, even when underlying circuit implementations are not directly accessible[58,59].

Representational convergence has been shown before in both humans and ANNs. Functional alignment studies show different degrees of representational convergence across individuals[26,27], and many ANN studies report convergent representational learning[20–22], although its scope is contentious[60]. Our findings extend this work by showing when and under what conditions representational convergence may occur in the brain. Although substantial representational diversity persisted in our participants, representations were most similar where architectural constraints were most permissive of implementation flexibility, and in this respect more closely approached the flexibility of artificial networks.

We hope future work will address several limitations of this study. Representations in ANNs are characterized by thousands of stimuli, and neural manifolds are characterized by many hours of recordings, whereas here we were limited to 23 different task conditions and 25 RSNs. This same dataset has been widely used to characterize variation in the human brain[1–4,11,28], and is

uniquely valuable for its size and family structure, but denser individual sampling[61] is needed to more fully resolve representational geometry. At our sampling density, similar representational geometry means that these 23 task conditions (or 25 networks) are distinguished in similar ways in similar locations across participants, not that their full representational manifolds are equivalent.

Along with denser individual sampling, more informative task-derived behavioral measures will be important to bridge between representational and functional levels of analysis. This is straightforward in engineered systems like ANNs, where the computational objective[50] is specified *a priori*. However, inferring function in the brain requires overcoming fundamental epistemic challenges associated with reverse engineering complex systems[62,63]. Nevertheless, emerging evidence suggests representational structure may provide a more direct behavioral correlate than topographies in both sensory[64] and higher order cognitive domains[28].

Additionally, longitudinal studies are needed to determine how representations arise and evolve[65,66]. Our data are cross-sectional, which limits inferences about representational learning to estimates of a common environment effect.

Finally, BOLD fMRI measures neural activity through a vascular filter. Ultra-high field imaging shows that hemodynamics have columnar scale precision[67], but are also influenced by draining veins and other vascular effects[68–70]. We mitigate these factors by censoring putatively venous voxels during preprocessing and by spatial noise whitening. Further, interindividual vascular variability is unlikely to covary with the sensory-association axis, and we replicate an analogous dissociation in ANNs where no vascular filter exists. Nevertheless, corroboration from other measurement modalities will be important to fully disentangle neural and vascular contributions.

Our findings suggest how future studies of higher order brain function could adapt to target a more robust construct. Traditionally, fMRI has emphasized population-average functional maps obtained by averaging across individuals after spatially smoothing idiosyncrasies, an approach which obscures interleaved signals[71]. Representational analysis offers an alternative well suited to social, linguistic, working memory or compositional reasoning tasks like those used to engage transmodal cortex here, but which benefits from repeated measures data needed to estimate crossnobis distances. Supplemental task conditions or dense sampling are also needed to provide representational context. Temporal information provides this context in electrophysiology studies of neural dynamics[52,53,59], and naturalistic viewing paradigms in brain imaging studies sometimes serve an analogous role. Response hyperalignment in particular uses time-series from a reference task to define individual-specific population codes for task-evoked responses of interest[26,27]. Our results indicate that without such supplemental data the risk of misinterpretation increases in higher-order brain areas.

We link neural implementation flexibility to biological neural network architecture and show when and where representations converge. This links macroscale brain maps to mesoscale circuit constraints. From this view, spatial idiosyncrasies, especially in transmodal cortex, reflect flexible implementations rather than divergent computations. By bridging computational

modeling, developmental neurobiology, and brain mapping, our results advance an integrated conceptual framework for the study of macroscale neural organization.

METHODS

Artificial neural network modeling

We used ImageNet data[72,73] to train neural networks derived from the ResNet-18 architecture, following a previously developed protocol[23].

Briefly, residual networks (ResNets) are a convolutional neural network architecture that achieves extraordinary depth through regularization by residual "skip" connections[74]. A suffix indicates the number of convolutional layers (here: 18), which, except for the first and last, are paired into "basic blocks" constituting our named "layers" (**Figure 1A**). All units were endowed with coordinate locations and embedded into a 2D spatial grid. In every case kernels have 3x3 unit receptive fields (layers not illustrated to scale). Depth endows the network with greater expressive capacity while convolution yields coarse spatial selectivity in early layers that is composed and integrated into spatially invariant features by later layers. Although early layers have intrinsic spatial selectivity due to the convolutional architecture, each layer also has multiple channels (features) which must be spatially interleaved (**Figure 1A**, stack widths). Channel organization dominates topographies of late layers but also affects fine topographies of earlier layers. To obtain a viable embedding topology compatible with smooth topographies, positions were selected to maximize correlation among nearby units in a pretraining task using sine grating sweeps inspired by prenatal retinal waves[75].

Weights were randomly reinitialized after positioning, and the network was trained using contrastive self-supervised learning. This exploits the intrinsic symmetries of visual images by augmenting training data with transformations (e.g. reflection, rotation, color inversion) under which object identities are invariant[76]. The contrastive learning objective ("task" objective) is to make transformation-insensitive distinctions between images in a training batch. This differs from supervised learning because it does not require manual labeling of images, and more realistically models human infantile development. Most evaluations were performed at this stage, but to obtain model performances we froze the model trunk and used supervised learning to retrain the output (average pooling) layer.

We trained 28 network instances from 14 different seeds (0-13) with either task loss only ("ResNets") or a composite task and spatial loss ("TDANNs"). Spatial loss encourages monotonic decay with spatial distance of the pair-wise correlation between units within-layer. This loss is weighted by a hyperparameter, α, here set to $\alpha = 0.25$, following precedent[23]. Networks were paired into dyads within network class such that for each dyad we had an identically initialized dyad of networks using the alternative loss, allowing us to evaluate the marginal effect of spatial loss independently from the effect of initial conditions.

Convolutional neural networks have produced highly competitive encoding models of the ventral visual stream[14,30], and ResNets are among the best[32,77], particularly when trained using our protocols here[23,31]. Alternative topographic networks to TDANNs have recently been introduced[78,79], but have been less thoroughly characterized and benchmarked. Additionally, the combination of pretraining and task loss objectives endow TDANN topographies and learning mechanisms with greater biological realism than alternative models. Rather than deeper Resnets, we used ResNet-18s as the base architecture for expedient training (40-80 hours/seed; Supplemental Methods: *Computational Modeling*).

Models were implemented using the pytorch framework. Model topographies and representational geometries were evaluated using post-residual activations. In both cases we flattened units across channels and either computed cosine similarity (topographies) or used the unbiased minibatch CKA estimator (representations), with a batch size of 500. To increase estimator precision minibatch CKA estimation was performed over the full ImageNet validation dataset across 10 independent passes with different image orderings, and CKA estimates were averaged across passes, following prior work[33]. All model evaluations, including classification outcomes, were performed using the independent validation set of 50,000 ImageNet images. Classification performance was evaluated against the 1000 category labels of ImageNet.

Semantic disagreement was assessed by aggregating 1000 ImageNet labels into 50 more general superclasses derived from the WordNet lexical ontology[80]. For each ImageNet label, the most specific hypernym ancestor was selected that subsumed at least 20 ImageNet labels as hyponyms, using hypernym relations only (excluding meronyms or other attributes). Because of variation in lexical precision among ImageNet labels (e.g. “street sign” vs. a specific breed of dog), a typical resulting superclass subsumed 50 labels (median; interquartile range: [27, 122]). Most labels were assigned to more specific superclasses within this set (median labels per superclass: 22; interquartile range: [9,26]).

Participants and pairing

General inclusion and exclusion criteria for participants in the Human Connectome Project Young Adult (HCP-YA) dataset have been described elsewhere[81].

For topographic and geometric similarity analyses we introduced additional exclusion criteria. We required all participants to have: (1) complete functional MRI (fMRI) data and task timing information for all seven HCP tasks in both left-right (LR) and right-left (RL) acquisition directions; and (2) four resting state fMRI scans. 414 such unrelated participants were identified. We paired these participants into 207 non-intersecting dyads for topographic and geometric similarity analyses. For demographic details see **Supplemental Table 1**.

For heritability analyses we additionally included related participants (N = 1003). This included monozygotic twins (MZ; n = 254 individuals; n = 127 dyads, n = 118 genotypically verified), dizygotic twins (DZ; n = 140 individuals; n = 70 dyads, n = 65 genotypically verified), full siblings

(FS; *n* = 775 individuals; *n* = 392 non-intersecting dyads), half siblings (HS; *n* = 54 individuals; *n* = 37 dyads) and unrelated individuals (Unr; *n* = 1003 individuals; *n* = 501,662 dyads). No triplets were present in this dataset. Dyads intersect because some individuals had multiple siblings and there were numerous possible pairings of unrelated individuals. Self reported twin zygosity was used for a minority of families lacking genotypic verification.

All experimental and informed consenting procedures were approved by the Washington University institutional review board (IRB). The Dartmouth College Committee for the Protection of Human Subjects was additionally consulted and waived further IRB review for this study's analysis plan, which abides by the HCP's Restricted Data Use Terms.

Experimental design

A complete experimental design description is available elsewhere[29]. Briefly, subjects underwent two days of scanning during which they performed seven fMRI tasks twice (**Supplemental Table 2**). Each task (LR and RL acquisitions) was always acquired on the same day, with some tasks collected on the first day (predominantly working memory, gambling, motor) and others on the second (emotion, language, social, compositional reasoning). Two 15 minute resting state scans (LR and RL acquisitions) were also collected on each day. A subset of 22 unrelated participants returned for a later visit and repeated this protocol.

Image acquisition and preprocessing

This study made use of the HCP's minimally preprocessed data, which have undergone a standardized set of processing steps to remove artifacts, correct distortions, align different imaging modalities, and prepare the data for further analysis. The data are described in detail elsewhere[82,83]. Here we describe the acquisition and preprocessing characteristics most relevant for evaluating topographic and geometric similarity measures.

Single-echo multiband (factor 8) fMRI data was acquired at 3T with 2mm isotropic voxels, 0.72s repetition time, and 33ms echo time, along with two spin-echo images with reversed phase encoding directions used for susceptibility distortion correction (SDC). This echo time is known to yield poor contrast in the subcallosal cortex, thalamus, and basal ganglia regions with shorter relaxation times[84], as quantified by temporal signal to noise ratio maps (tSNR, **Supplemental Figure 1**). Consequently, we primarily report results for cortical structures. T1w and T2w structural scans were also acquired for volumetric alignment and cortical segmentation at an isotropic 0.7mm resolution.

Functional data underwent SDC, were realigned to a single-band reference image to correct for motion (note: no further motion correction was performed during first level modeling), bias field corrected, normalized to a global mean, and resampled to a high-resolution subject specific cortical surface mesh. We treated SDC corrected LR and RL acquisitions as repeated measures

to allow for unbiased distance estimation of tasks (*Geometric Similarity Analyses*, below). This assumption holds in regions with high distortion to the extent SDC is successful at the coarse scale of regional boundaries. Otherwise, distances may be over or underestimated. The mean log Jacobian determinant maps are provided as a measure of SDC throughout the brain for transparency (**Supplemental Figure 2**), and maps of test-retest reliability of response topographies (General Linear Modeling of task evoked responses, below) are provided as a secondary indirect measure of SDC success and contrast SNR (**Supplemental Figure 1**). For resting state analyses, LR and RL scans were acquired on each day, and unbiased distances were estimated across days, obviating concerns about acquisition direction related effects.

Additionally, volume to surface sampling excluded voxels with the highest signal variance, which typically corresponds to stromal voxels. This includes large draining veins that can otherwise confound fine-scale functional localization, a primary concern of this study, by producing BOLD contrast distal from the site of synaptic activity.

Cortical surfaces were downsampled and nonlinearly aligned to the Conte69 32k mesh using Freesurfer 5.2, while volumetric data was registered to the MNI152 Nonlinear 6 Asymmetric template using FNIRT for subcortical alignment. The resulting surface and volumetric spatial sampling is on average 2mm, with slight variations across the cortical mesh. At this scale, spatial precision is primarily constrained by the hemodynamic response's point spread function at 3T (~3-4mm[85]).

We used minimally smoothed (2mm full-width half maximum) CIFTI 91k formatted "MSMAll" grayordinate data. This dataset is distinguished by an additional alignment of resting state networks and T1/T2-ratio (a measure of myeloarchitecture) maps to a study specific multimodal template by smoothly ("diffeomorphically") displacing the cortical surface. Full details are available elsewhere[86]. This procedure reduces spatial differences between isomorphic, but not heteromorphic cortical topographies. To evaluate the effect of spatial alignment we also evaluated how cortical gradients dissociated representational and topographic similarity in non-MSAll aligned evoked-response data (**Supplemental Figure 6**, cf. **Figure 3**). While topographic or representational similarity of individual regions varied relative to MSMAll data, this variation did not meaningfully affect gradient tests, suggesting any gross spatial effects of multimodal alignment were orthogonal to our *a priori* gradients of interest.

General linear modeling of task evoked responses

We adopted previously described statistical designs of HCP tasks[29] with slight modifications to improve representational similarity analysis (RSA). In particular, to improve voxel-wise noise estimation we used SPM's FAST time series error models when estimating run-level contrasts. While HCP shares precomputed contrasts derived using FSL, FAST has been shown to outperform FSL's time series error model especially with "fast" sub-second TR imaging protocols[87], and was recommended for unbiased distance estimation (Jorn Diedrichsen, personal correspondence; see *Geometric Similarity Analyses* below). This improves

vertex/voxel-level noise estimation, which benefits multivariate distance estimates. Aside from this change, our GLM models were identical to those of the HCP: hemodynamic response convolved task vectors, a 200s high-pass filter and an intercept were included in our design matrix.

Neuromaps and brain parcellations

We *a priori* selected nine "neuromaps" aligned with the sensory-association axis that reflect developmental, evolutionary or architectural factors relevant to this study, along with three additional sensory-association aligned maps used as negative controls (**Table 1**). All were previously published, adopted from the 81 maps included with the *Neuromaps* Python package[88], and have been the focus in part or full of previous studies of sensory-association gradients in the brain[9,89], motivating their use here.

We defined brain "regions" for evaluating topographic and geometric similarity using two brain atlases. The first is an in-house composite atlas (CANLab2024) that includes cortical regions from the HCP MMP 1.0 atlas[86] and cerebellar regions from the SUIT cerebellar atlas[90]. Additional subcortical and brainstem structures are also included but largely out of the scope of this study due to the poor tSNR and test-retest reliability of deep brain structures in this dataset (**Supplemental Figure 1**). For further details on these structures refer to the CANLab2024 atlas in our "Neuroimaging Pattern Masks" github repository (Data and Code Availability). The HCP MMP 1.0 atlas was chosen because it was tailor made for this dataset and delimits functionally and structurally homogenous brain areas which might plausibly serve as functional modules. A second atlas[91] (Gordon 333 cortical regions) was used to verify the robustness of our results with respect to parcel boundaries. Because this second atlas lacked subcortical areas, it was augmented with subcortical and cerebellar structures defined by the structural labeling scheme of the grayordinate standard[82], which subdivides all gross structures except the brainstem by laterality.

Topographic Similarity Analyses

We computed between-participant cosine similarity of regional spatially whitened contrast maps (Supplemental Methods: *Spatial Whitening*), and calculated a weighted average across tasks as our overall topographic similarity measure. Whitening accounts for spatial autocorrelation of noise that biases between-participant similarity measures, but more importantly matches our topographic similarity to our geometric similarity measure, which uniquely benefits from whitening (Methods: *Geometric Similarity Analyses*). Notably, spatial autocorrelation was estimated from GLM residuals, ensuring that whitening does not remove true spatial correlations in the signal[35]. Task weights were applied to prevent tasks with more conditions (working memory: 8 conditions, motor task: 5 conditions) from dominating those with fewer (2 conditions each). We chose cosine similarity because it is sensitive to both the spatial organization of responses and has a meaningful zero point (unlike spatial correlations), but is insensitive to

overall response magnitude. This makes it robust to spurious (non-neural) vascular and hematological idiosyncrasies affecting BOLD contrast that could confound more formal (e.g. euclidean) between-participant distance measures.

For resting state networks (RSN) we computed cosine similarity of whitened maps of individualized networks (Methods: *Dual Regression*).

Permuted nulls: As a scale reference for similarity magnitude, we computed condition-permuted null similarities for each participant dyad and region. For each dyad, condition order was permuted in one participant and cosine similarity was computed relative to the unpermuted participant. For task data, we used balanced oversampling to account for condition imbalance. Obtained similarities were averaged across participants within-region. We evaluated 100 permutations and the 97.5% upper bounds (across permutations) were averaged across regions and plotted in **Figures 3D** and **5D** as a conservative reference scale.

Geometric Similarity Analyses

Geometric similarity was measured by whitened and unbiased cosine similarity (WUC) of balanced RDMs. This involved estimating unbiased Mahalanobis distances (or “crossnobis” distance; defined below) between task contrasts (or RSNs), whitening the resulting representational dissimilarity matrices (RDMs) to account for dependencies across condition-pairs (not to be confused with spatial whitening), balancing across tasks by oversampling underrepresented tasks (omitted for RSNs), and computing the cosine similarity of resulting RDMs. We detail each step below.

Spatial whitening: Mahalanobis distance is a measure of Euclidean distance after spatial whitening. This captures decodable information more effectively than raw regression coefficient (“beta”) maps by downweighting channels that covary or contain noise[35]. Spatial covariance was estimated from GLM residuals using Ledoit-Wolf regularization[92] (Supplemental Methods: *Spatial Whitening*).

Unbiasedness: Mahalanobis distances are positively biased, e.g. identical contrasts will have non-zero distances proportional to their measurement error. To correct for this we partitioned task data by acquisition direction (LR or RL), computed the whitened contrast difference a_i - b_i in each partition, and took the product across independent partitions $(a_{LR} - b_{LR})^T(a_{RL} - b_{RL})$ for all conditions *a* and *b*. This cancels out independent (e.g. thermal) noise and yields an RDM of cross-validated Mahalanobis or “crossnobis” distances that are unbiased. Positive distances indicate that task conditions were reliably distinct, while undistinguishable conditions are symmetrically distributed around zero. We partitioned resting state data by day rather than acquisition direction to increase scan availability and ensure robustness with respect to day-to-day fluctuations in connectivity.

RDM Whitening: Contrast-by-contrast RDMs have numerous codependencies which bias between-participant RDM comparisons if unaccounted for. This includes shared physiological noise sources (e.g., caffeine intake and its effects on vasoreactivity), shared statistical error (e.g., intercept or "baseline" estimation error), and experimental-design-induced codependencies (e.g., in fusiform face area, distances taken with respect to the emotion-faces condition will correlate with distances of the 0-back memory condition that also uses face stimuli). Even if all conditions were independent, all distances in a row of an RDM share a common comparator, introducing codependence. We empirically estimated this complex covariance structure V for each subject and region following a previously developed approach[34] (Supplemental methods: *RDM Covariance Estimation*), pooled covariances across participants using a standard degrees-of-freedom (*dof*) weighing, $V = (dof_1 V_1 + dof_2 V_2)/(dof_1 + dof_2)$, and whitened RDMs in each comparison using V^{-1}.

RDM Balancing: We used balanced oversampling of underrepresented tasks to keep conditions balanced across tasks in the RDM (Supplemental Methods: *RDM Balancing*).

Cosine Similarity: We computed cosine similarity between participants using the vectorized lower triangular elements of the balanced, whitened RDMs. As in the case of topographies, BOLD signal amplitude is spuriously idiosyncratic across participants, which modulates the magnitude of distances. Cosine similarity is insensitive to magnitude information while preserving a meaningful zero point and thus the relative proportions between distances.

Permuted null: A representational similarity reference scale was computed in the same way as for cosine similarities of topographies (above).

Spatial Statistics

We performed two types of cortical spatial analysis. First, we evaluated how topographies and geometries compared with cortical neuromaps in a within-dyad setting, hereon termed 'gradient similarity analysis' (**Figure 3, 5**). Second, we evaluated how well geometric similarity predicted topographic similarity in an across-dyad setting, and tested how this relationship varied across the cortex, hereon termed 'sensitivity analysis' (**Figure 4**).

Where we controlled for confounds we used temporal signal to noise ratio (tSNR) and test-retest reliability maps as covariates. tSNR was computed using Connectome Workbench's *cifti-reduce* function and averaged across runs. Test-retest reliability was estimated as the within-participant cosine similarity of task contrasts across data partitions (e.g., LR vs. RL), weighted to account for condition imbalance across tasks, providing an overall participant-specific measure. Notably, in the participants that returned for a second visit, this within-visit measure closely approximated test-retest reliability across visits (**Supplemental Figure 1D**). Both tSNR and reliability values were averaged within dyad for use below.

For both types of analyses, neuromap data was spatially normalized (spatial mean = 0, unit variance) to ensure consistent scaling across tests. Additionally, neuromap, tSNR, and reliability values were averaged within-region for comparison. Analyses were restricted to the cortex by neuromap data availability.

Gradient similarity analysis: We regressed each z-Fisher-transformed similarity measure (geometry or topography) on each neuromap. To control for confounds we also included a tSNR map and test-retest reliability maps in this dyad-level model. In Wilkinson notation[93],

$$\text{similarity} \sim \text{neuromap} + \text{tSNR} + \text{reliability} + 1$$

Regression coefficients were averaged across dyads. We then estimated the difference between topographic and geometric similarity and investigated how neuromaps moderated it. Here we spatially normalized similarity measures (within-dyad spatial mean = 0, unit variance) and centered tSNR and reliability confounds. Thus, reported statistics pertain to the mean cortical tSNR and reliability. In Wilkinson notation,

$$\text{similarity} \sim \text{neuromap} + \text{tSNR}_c + \text{reliability}_c + \text{geom_v_topo*neuromap} + \text{geom_v_topo*tSNR}_c + \text{geom_v_topo*reliability}_c + -1$$

Where similarity is a normalized z-Fisher-transformed metric, geom_v_topo was coded as [0.5, -0.5] and indicated whether a similarity score indicated geometric (WUC) or topographic (cosine) similarity (respectively). Main effects and intercepts are omitted because normalization sets them both to zero by design. Inference was performed using jackknife-adjusted spin tests (below).

Sensitivity analysis: This test evaluated how well geometry predicted topography, testing the intuition that if topographies are meaningful like retinotopy and somatotopy, then similar representations should correspond to similar topographies. We expected this correspondence to vary throughout the brain, so we tested how this association was moderated by neuromaps by fitting the following model:

$$\text{topo} \sim \text{geom} + \text{neuromap*geom} + \text{roi} + \text{dyad} + \text{tSNR}_z\text{*dyad} + \text{reliability}_z\text{*dyad} - 1$$

Here *topo* and *geom* indicate z-Fisher transformed topographic (cosine) or geometric (WUC) similarity measures, *roi* is a region specific intercept, *dyad* is a Helmert coded dyad contrast variable that accounts for systematic (ie brain wide, not region specific) between-dyad differences, and -1 indicates the omission of a global intercept. *tSNR* and *reliability* were spatially normalized within dyad (spatial mean = 0, unit variance). The main effect of *neuromap* does not vary within-region and is (perfectly) colinear with *roi*. The same is true for the main effects of *tSNR* and *reliability* once systematic between-participant differences have been accounted for. Consequently we only included their interaction effects. A global intercept is collinear with *roi* and was therefore also excluded. Overall the model estimates the within-region relationship between geometry and topography while controlling for overall (across regions)

between-dyad differences in topographic similarity, tSNR and reliability. Inference again used jackknife-adjusted spin tests (below).

Region specific effects shown in parametric maps and scatterplots of **Figure 4** are simple univariate regressions of *cosim* on *WUC* after Fisher z-transforming both metrics. The neuromap moderator above simply indicates how these regression coefficients change across the cortex, after accounting for confounds.

Jackknife-adjusted spin tests

A “spin” test is a permutation test that rotates neuromaps around the cortical surface to generate a null distribution that preserves spatial autocorrelation[36]. This test does not however account for sampling variance of a statistic or neuromap association. We introduced Jackknife-adjusted spin tests, which use a jackknife estimate of sampling variance to augment the null distribution’s variance under the assumption that both sampling and null distributions are Gaussian (Supplemental Methods: *Jackknife-adjusted spin test*). For our data, the resultant test was more conservative than a spin test (Supplemental Methods: *Monte Carlo Simulations*).

Confidence intervals reported for population parameters are jackknife confidence intervals.

Dual Regression

Our dual regression approach began with resting-state network (RSN) templates from the HCP PTN1200 release. Multiple templates were available, but we made the principled choice to use the 25-network template because it most closely matched our task count (n = 23 conditions). Traditional dual regression multiply regresses RSN template maps on resting-state volumes to obtain individualized timeseries, which are then temporally multiply regressed on voxel/vertex timeseries to obtain individualized RSNs. For more individualized RSNs, we further incorporated weighted regression and iterated this process six times (*Supplemental Methods*: *Dual Regression Iteration Procedure and Convergence Criteria*). This approach has previously been applied to this dataset[86], and our version primarily differs by increasing the iteration count.

We estimated the final RSNs using SPM to match the procedures used for task topographies and geometries. The individualized time-series from the final dual regression iteration were used as regressors in a GLM analysis, controlling for white matter and cerebral spinal fluid signal as well as 24 motion confounds (6 rotation and translation parameters, their quadratics, and their corresponding derivatives). Models used FAST time-series error models and 200s high-pass filters. Topographic and geometric similarity analysis proceeded identically to task-based analyses (above), except without the procedures used to address condition imbalance across tasks because RSNs had no equivalent imbalance.

Genetic analysis

We examined the heritability of topography and geometry, which are multivariate metric traits, but measured here using BOLD fMRI contrast. Because BOLD contrast measures of metabolic reactivity are confounded by spurious (non-neural) physiological factors, all between-participant comparisons of these traits were performed after vector normalization using either cosine similarity or WUC measures. This transformation precluded the use of traditional population genetic models based on parametric analysis of variance (ANOVA)[94].

Parametric ANOVA separates the variance of a metric trait (e.g. metabolic reactivity, our proxy for neural activity) into independent components with known theoretical contributions (e.g. a component attributable to environmental variance and one attributable to genetic differences). Our normalization step disrupted this linear additive relationship ($\sigma_{a+b}^2 = \sigma_a^2 + \sigma_b^2$) since $\sigma_{(a+b)/||a+b||}^2 <= \sigma_{a/||a||}^2 + \sigma_{b/||b||}^2$. The precise relationship cannot be recovered without the quantities ||a|| and ||b||, the net metabolic responses, which cannot be measured with BOLD fMRI alone. Instead we estimated heritability and environmental effects more indirectly.

If $a_1,a_2 \sim N(\mu,\sigma_a^2 I)$ and $b_1,b_2 \sim N(\mu,\sigma_b^2 I)$ are isotropic multivariate normal random variables with $\mu,\sigma_a^2,\sigma_b^2 > 0$, then $E[\cos\theta_{a1,a2}] > E[\cos\theta_{b1,b2}]$ entails $\sigma_a^2 < \sigma_b^2$. This is because projecting x to the surface of the unit hypersphere (i.e., $x/||x||$) gives rise to a von Mises-Fisher distribution whose concentration decreases as the variance increases. Thus, lower variance will produce a higher expected cosine similarity. If $\mu = 0$, both are uniformly distributed on the hypersphere. For non-isotropic variance, radial anisotropy may be important, which is lost in projection. It may be possible to loosen the isotropic variance requirement in terms of the proportion of radial vs. angular variance in the unnormalized variable, but isotropy was approximated in our case by spatial whitening of topographies (±regularization) and RDM whitening of geometries regardless. The parameter μ is the group mean contrast or group mean geometry, so $\mu > 0$ is satisfied so long as at least one task condition produces a non-zero group mean contrast which holds in our data[29]. This implies we could compare whitened and normalized BOLD derived measures within family clusters to determine whether the unobserved unnormalized traits differ in variance between two family clusters.

We tested for genetic and environmental effects by comparing family clusters matched on either environmental or genetic variance. If σ_G^2 is the genetic variance unique to DZ twins and $E[\cos\theta_{MZ}] > E[\cos\theta_{DZ}]$, then $\sigma_{MZ}^2 < \sigma_{DZ}^2$ and $\sigma_G^2 > 0$, because MZ and DZ twins are matched on environmental variance, including uterine environment. Similarity, if σ_C^2 is the "common environment" variance due to asynchronous gestation and upbringing unique to full siblings relative to DZ twins, and $E[\cos\theta_{DZ}] > E[\cos\theta_{FS}]$, then $\sigma_{DZ}^2 < \sigma_{FS}^2$ and $\sigma_C^2 > 0$, because full siblings and DZ twins are matched on genetic and idiosyncratic variance. We evaluated if $E[\cos\theta_{MZ}] > E[\cos\theta_{DZ}]$ or $E[\cos\theta_{DZ}] > E[\cos\theta_{FS}]$ using Mann-Whitney *U*-tests for genetic and common environment effects, respectively.

We compared effect sizes between topographies and geometries using an interaction test on ranks. We pooled a similarity measure (e.g., topography) across two family clusters (e.g., MZ

and DZ twins) and ranked them. We then computed the difference in rank between measures (e.g., topography vs. geometry). A Mann-Whitney *U* statistic tested the null hypothesis that the between trait rank difference was equal for the two family clusters. Rejection of this null indicated that effect sizes differed between traits.

DATA AND CODE AVAILABILITY

All data is available from the HCP 1200 Subjects Data Release (https://www.humanconnectome.org/study/hcp-young-adult/document/1200-subjects-data-release). The code we used to train ANN models is available from the neuroAI lab: https://github.com/neuroailab/TDANN. Additional modeling code (training configurations, evaluations, etc), CIFTI files containing spatial maps from our figures and all and code needed to reproduce our brain findings from the above minimally preprocessed data are available via our own github (https://github.com/bogpetre/geometry_vs_topography/), except for the CANLab2024 atlas. This atlas includes regions with a restrictive redistribution license, and can be obtained from the Neuroimaging Pattern Masks repository (https://github.com/canlab/Neuroimaging_Pattern_Masks/), which pulls constituent regions from licensed sources and rebuilds CANLab2024 locally. Alternatively, openCANLab2024 is included with the *geometry_vs_topography* repo, does not require any further processing to use, and is identical in the cortex and cerebellum, the most relevant structures for this dataset.

ACKNOWLEDGEMENTS

Data were provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

This work was funded by NIBIB R01EB026549-03 and NIMH R37MH076136-18.

We additionally thank Jorn Diedrichsen for guidance with estimating representational similarity measures, and James Haxby, Marta Čeko and Etienne Vachon-Presseau for discussion and feedback.

AUTHOR CONTRIBUTIONS

BP: conceptualization, methodology, software, formal analysis, investigation, data curation, writing (original draft and review & editing), visualization, project administration; MAL: writing (review & editing), supervision, funding acquisition; TDW: resources, writing (review & editing), supervision, funding acquisition.

REFERENCES

1. Finn, E. S. *et al.* Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nat. Neurosci.* **18**, 1664–1671 (2015).
2. Gordon, E. M. *et al.* Individual-specific features of brain systems identified with resting state functional correlations. *Neuroimage* **146**, 918–939 (2017).
3. Dworetsky, A. *et al.* Two common and distinct forms of variation in human functional brain networks. *Nat. Neurosci.* **27**, 1187–1198 (2024).
4. Ribeiro, F. L. *et al.* Variability of visual field maps in human early extrastriate cortex challenges the canonical model of organization of V2 and V3. *Elife* **12**, (2023).
5. Schwarzkopf, D. S., Song, C. & Rees, G. The surface area of human V1 predicts the subjective experience of object size. *Nat. Neurosci.* **14**, 28–30 (2011).
6. Himmelberg, M. M., Winawer, J. & Carrasco, M. Linking individual differences in human primary visual cortex to contrast sensitivity around the visual field. *Nat. Commun.* **13**, 3309 (2022).
7. Elbert, T., Pantev, C., Wienbruch, C., Rockstroh, B. & Taub, E. Increased cortical representation of the fingers of the left hand in string players. *Science* **270**, 305–307 (1995).
8. Karni, A. *et al.* The acquisition of skilled motor performance: fast and slow experience-driven changes in primary motor cortex. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 861–868 (1998).
9. Sydnor, V. J. *et al.* Neurodevelopment of the association cortices: Patterns, mechanisms, and implications for psychopathology. *Neuron* **109**, 2820–2846 (2021).
10. Cui, Z. *et al.* Individual variation in functional topography of association networks in youth. *Neuron* **106**, 340–353.e8 (2020).
11. Kong, R. *et al.* Spatial Topography of Individual-Specific Cortical Networks Predicts Human Cognition, Personality, and Emotion. *Cereb. Cortex* **29**, 2533–2551 (2019).

12. Lynch, C. J. *et al.* Frontostriatal salience network expansion in individuals in depression. *Nature* **633**, 624–633 (2024).
13. Drysdale, A. T. *et al.* Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nat. Med.* **23**, 28–38 (2017).
14. Yamins, D. L. K. *et al.* Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 8619–8624 (2014).
15. Mante, V., Sussillo, D., Shenoy, K. V. & Newsome, W. T. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* **503**, 78–84 (2013).
16. Schrimpf, M. *et al.* The neural architecture of language: Integrative modeling converges on predictive processing. *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2105646118 (2021).
17. Sussillo, D., Churchland, M. M., Kaufman, M. T. & Shenoy, K. V. A neural network that finds a naturalistic solution for the production of muscle activity. *Nat. Neurosci.* **18**, 1025–1033 (2015).
18. Orhan, A. E. & Ma, W. J. A diverse range of factors affect the nature of neural representations underlying short-term memory. *Nat. Neurosci.* **22**, 275–283 (2019).
19. Kriegeskorte, N. Deep neural networks: A new framework for modeling biological vision and brain information processing. *Annu. Rev. Vis. Sci.* **1**, 417–446 (2015).
20. Saxe, A. M., McClelland, J. L. & Ganguli, S. A mathematical theory of semantic development in deep neural networks. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 11537–11546 (2019).
21. Kornblith, S., Norouzi, M., Lee, H. & Hinton, G. Similarity of neural network representations revisited. *arXiv [cs.LG]* (2019).
22. Li, Y., Yosinski, J., Clune, J., Lipson, H. & Hopcroft, J. Convergent Learning: Do different neural networks learn the same representations? *arXiv [cs.LG]* (2015).
23. Margalit, E. *et al.* A unifying framework for functional organization in early and higher ventral visual cortex. *Neuron* **112**, 2435–2451.e7 (2024).

24. Wagstyl, K. *et al.* BigBrain 3D atlas of cortical layers: Cortical and laminar thickness gradients diverge in sensory and motor cortices. *PLoS Biol.* **18**, e3000678 (2020).
25. Wagstyl, K. & Lerch, J. P. Cortical Thickness. in *Neuromethods* 35–49 (Springer New York, New York, NY, 2018).
26. Haxby, J. V. *et al.* A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron* **72**, 404–416 (2011).
27. Guntupalli, J. S. *et al.* A Model of Representational Spaces in Human Cortex. *Cereb. Cortex* **26**, 2919–2934 (2016).
28. Feilong, M., Guntupalli, J. S. & Haxby, J. V. The neural basis of intelligence in fine-grained cortical topographies. *eLife* (2021) doi:10.7554/eLife.64058.
29. Barch, D. M. *et al.* Function in the human connectome: task-fMRI and individual differences in behavior. *Neuroimage* **80**, 169–189 (2013).
30. Khaligh-Razavi, S.-M. & Kriegeskorte, N. Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS Comput. Biol.* **10**, e1003915 (2014).
31. Zhuang, C. *et al.* Unsupervised neural network models of the ventral visual stream. *bioRxiv* (2020) doi:10.1101/2020.06.16.155556.
32. Jiahui, G. *et al.* Modeling naturalistic face processing in humans with deep convolutional neural networks. *Proc. Natl. Acad. Sci. U. S. A.* **120**, e2304085120 (2023).
33. Nguyen, T., Raghu, M. & Kornblith, S. Do wide and deep networks learn the same things? Uncovering how neural network representations vary with width and depth. *arXiv [cs.LG]* (2020) doi:10.48550/ARXIV.2010.15327.
34. Diedrichsen, J. *et al.* Comparing representational geometries using whitened unbiased-distance-matrix similarity. *Neurons, Behavior, Data analysis, and Theory* **5**, (2021).
35. Walther, A. *et al.* Reliability of dissimilarity measures for multi-voxel pattern analysis. *Neuroimage* **137**, 188–200 (2016).

36. Alexander-Bloch, A. F. *et al.* On testing for spatial correspondence between maps of human brain structure and function. *Neuroimage* **178**, 540–551 (2018).
37. Burt, J. B. *et al.* Hierarchy of transcriptomic specialization across human cortex captured by structural neuroimaging topography. *Nat. Neurosci.* **21**, 1251–1259 (2018).
38. Xue, A. *et al.* The detailed organization of the human cerebellum estimated by intrinsic functional connectivity within the individual. *J. Neurophysiol.* **125**, 358–384 (2021).
39. Busch, E. L. *et al.* Dissociation of reliability, heritability, and predictivity in coarse- and fine-scale functional connectomes during development. *J. Neurosci.* **44**, e0735232023 (2024).
40. Visscher, P. M., Hill, W. G. & Wray, N. R. Heritability in the genomics era--concepts and misconceptions. *Nat. Rev. Genet.* **9**, 255–266 (2008).
41. Blokland, G. A. M., de Zubicaray, G. I., McMahon, K. L. & Wright, M. J. Genetic and environmental influences on neuroimaging phenotypes: a meta-analytical perspective on twin imaging studies. *Twin Res. Hum. Genet.* **15**, 351–371 (2012).
42. Falconer, D. S. & MacKay, T. F. C. *Introduction to Quantitative Genetics*. (Longman, London, England, 1995).
43. Kendler, K. S. & Baker, J. H. Genetic influences on measures of the environment: a systematic review. *Psychol. Med.* **37**, 615–626 (2007).
44. Mann, F., Ray, S., Harris, W. & Holt, C. Topographic mapping in dorsoventral axis of the Xenopus retinotectal system depends on signaling through ephrin-B ligands. *Neuron* **35**, 461–473 (2002).
45. Schmitt, A. M. *et al.* Wnt–Ryk signalling mediates medial–lateral retinotectal topographic mapping. *Nature* **439**, 31–37 (2006).
46. Saygin, Z. M. *et al.* Connectivity precedes function in the development of the visual word form area. *Nat. Neurosci.* **19**, 1250–1255 (2016).
47. Roe, A. W., Pallas, S. L., Hahm, J. O. & Sur, M. A Map of Visual Space Induced in Primary

Auditory Cortex. *Science* **250**, 818–820 (1990).

48. Natu, V. S. *et al.* Apparent thinning of human visual cortex during childhood is associated with myelination. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 20750–20759 (2019).
49. Margulies, D. S. *et al.* Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 12574–12579 (2016).
50. Marr, D. & Poggio, T. A. From Understanding Computation to Understanding Neural Circuitry. http://hdl.handle.net/1721.1/5782 (1976).
51. Kriegeskorte, N., Mur, M. & Bandettini, P. Representational similarity analysis - connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* **2**, 4 (2008).
52. Murray, J. D. *et al.* Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 394–399 (2017).
53. Churchland, M. M. *et al.* Neural population dynamics during reaching. *Nature* **487**, 51–56 (2012).
54. Sadtler, P. T. *et al.* Neural constraints on learning. *Nature* **512**, 423–426 (2014).
55. Wigner, E. P. Events, laws of nature, and invariance principles. *Science* **145**, 995–999 (1964).
56. Prinz, A. A., Bucher, D. & Marder, E. Similar network activity from disparate circuit parameters. *Nat. Neurosci.* **7**, 1345–1352 (2004).
57. Marder, E. & Bucher, D. Understanding circuit dynamics using the stomatogastric nervous system of lobsters and crabs. *Annu. Rev. Physiol.* **69**, 291–316 (2007).
58. Safaie, M. *et al.* Preserved neural dynamics across animals performing similar behaviour. *Nature* **623**, 765–771 (2023).
59. Russo, A. A. *et al.* Motor cortex embeds muscle-like commands in an untangled population response. *Neuron* **97**, 953–966.e8 (2018).
60. Morcos, A. S., Raghu, M. & Bengio, S. Insights on representational similarity in neural

networks with canonical correlation. *arXiv [stat.ML]* (2018).

61. Allen, E. J. *et al.* A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. *Nat. Neurosci.* **25**, 116–126 (2022).
62. Jonas, E. & Kording, K. P. Could a neuroscientist understand a microprocessor? *PLoS Comput. Biol.* **13**, e1005268 (2017).
63. Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A. & Poeppel, D. Neuroscience needs behavior: Correcting a reductionist bias. *Neuron* **93**, 480–490 (2017).
64. Han, C. & Bonner, M. F. High-dimensional structure underlying individual differences in naturalistic visual experience. *Curr. Biol.* (2026) doi:10.1016/j.cub.2025.12.039.
65. Oby, E. R. *et al.* New neural activity patterns emerge with long-term learning. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 15210–15215 (2019).
66. Gallego, J. A., Perich, M. G., Chowdhury, R. H., Solla, S. A. & Miller, L. E. Long-term stability of cortical population dynamics underlying consistent behavior. *Nat. Neurosci.* **23**, 260–270 (2020).
67. Yacoub, E., Harel, N. & Ugurbil, K. High-field fMRI unveils orientation columns in humans. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 10607–10612 (2008).
68. Kriegeskorte, N., Cusack, R. & Bandettini, P. How does an fMRI voxel sample the neuronal activity pattern: compact-kernel or complex spatiotemporal filter? *Neuroimage* **49**, 1965–1976 (2010).
69. Sengupta, A., Yakupov, R., Speck, O., Pollmann, S. & Hanke, M. The effect of acquisition resolution on orientation decoding from V1 BOLD fMRI at 7 T. *Neuroimage* **148**, 64–76 (2017).
70. Kay, K. *et al.* A critical assessment of data quality and venous effects in sub-millimeter fMRI. *Neuroimage* **189**, 847–869 (2019).
71. Braga, R. M. & Buckner, R. L. Parallel interdigitated distributed networks within the individual estimated by intrinsic functional connectivity. *Neuron* **95**, 457–471.e5 (2017).

72. Deng, J. *et al.* ImageNet: A large-scale hierarchical image database. in *2009 IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2009). doi:10.1109/cvpr.2009.5206848.

73. Russakovsky, O. *et al.* ImageNet large scale visual recognition challenge. *Int. J. Comput. Vis.* **115**, 211–252 (2015).

74. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 770–778 (IEEE, 2016).

75. Ackman, J. B., Burbridge, T. J. & Crair, M. C. Retinal waves coordinate patterned activity throughout the developing visual system. *Nature* **490**, 219–225 (2012).

76. Chen, T., Kornblith, S., Norouzi, M. & Hinton, G. A simple framework for contrastive learning of visual representations. *arXiv [cs.LG]* (2020).

77. Schrimpf, M. *et al.* Brain-Score: Which Artificial Neural Network for Object Recognition is most Brain-Like? *bioRxiv* (2018) doi:10.1101/407007.

78. Mayukh, D., Mainak, D. & N Apurva Ratan, M. TopoNets: High performing vision and language models with brain-like topography. *arXiv [cs.LG]* (2025).

79. Lu, Z. *et al.* End-to-end topographic networks as models of cortical map formation and human visual behaviour. *Nat. Hum. Behav.* **9**, 1975–1991 (2025).

80. *WordNet: An Electronic Lexical Database*. (MIT Press, Cambridge, MA, 1998).

81. Van Essen, D. C. *et al.* The Human Connectome Project: a data acquisition perspective. *Neuroimage* **62**, 2222–2231 (2012).

82. Glasser, M. F. *et al.* The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage* **80**, 105–124 (2013).

83. Uğurbil, K. *et al.* Pushing spatial and temporal resolution for functional and diffusion MRI in the Human Connectome Project. *Neuroimage* **80**, 80–104 (2013).

84. Lynch, C. J. *et al.* Rapid precision functional mapping of individuals using multi-echo fMRI.

Cell Rep. **33**, 108540 (2020).

85. Engel, S. A., Glover, G. H. & Wandell, B. A. Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cereb. Cortex* **7**, 181–192 (1997).
86. Glasser, M. F. *et al.* A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171–178 (2016).
87. Olszowy, W., Aston, J., Rua, C. & Williams, G. B. Accurate autocorrelation modeling substantially improves fMRI reliability. *Nat. Commun.* **10**, 1220 (2019).
88. Markello, R. D. *et al.* Neuromaps: Structural and functional interpretation of brain maps. *Nat. Methods* **19**, 1472–1479 (2022).
89. Dear, R. *et al.* Cortical gene expression architecture links healthy neurodevelopment to the imaging, transcriptomics and genetics of autism and schizophrenia. *Nat. Neurosci.* **27**, 1075–1086 (2024).
90. Diedrichsen, J., Balsters, J. H., Flavell, J., Cussans, E. & Ramnani, N. A probabilistic MR atlas of the human cerebellum. *Neuroimage* **46**, 39–46 (2009).
91. Gordon, E. M. *et al.* Generation and evaluation of a cortical area parcellation from resting-state correlations. *Cereb. Cortex* **26**, 288–303 (2016).
92. Ledoit, O. & Wolf, M. Honey, I shrunk the sample covariance matrix. *J. Portf. Manag.* **30**, 110–119 (2004).
93. Wilkinson, G. N. & Rogers, C. E. Symbolic Description of Factorial Models for Analysis of Variance. *Appl. Stat.* **22**, 392 (1973).
94. Fisher, R. A. XV.—the correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edinb.* **52**, 399–433 (1919).
95. Schmitt, J. E. *et al.* The dynamic role of genetics on cortical patterning during childhood and adolescence. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 6774–6779 (2014).

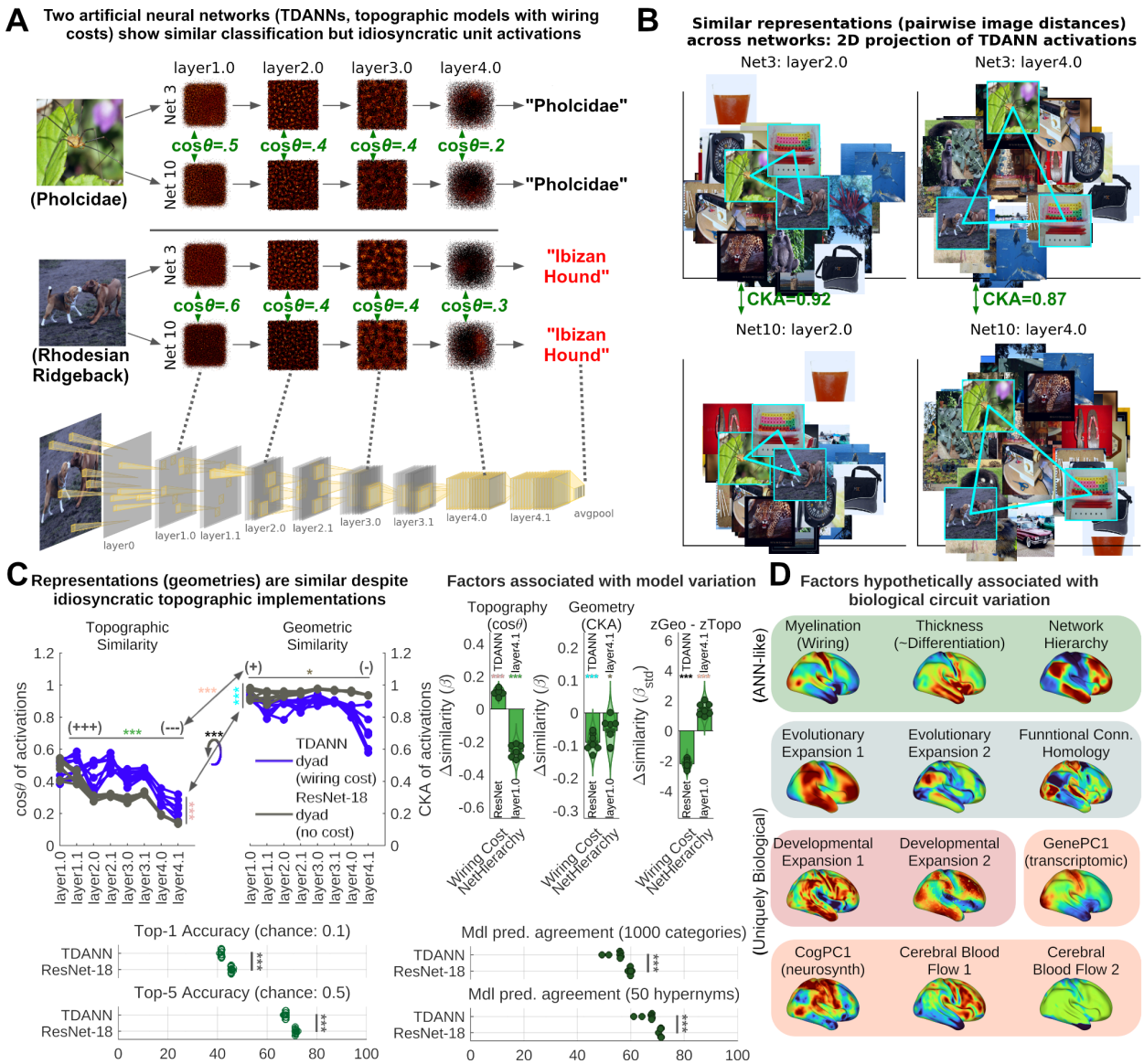

Figure 1 Computational experiments reveal a double dissociation for topographic and representational similarity in ANNs. (A) Two TDANNs that differ only in random weight initialization seeds show idiosyncratic activations but not classifications. Select convolution layers are shown. Cos*θ* indicates spatial cosine similarity of a layer's unit activations to an image. (B) Conservation of representational geometry is illustrated via 2-dimensional projection of the networks' activation spaces (MDS). (C) Network depth (across layers within ResNet-18s) and wiring cost (TDANNs vs. ResNets) have dissociable effects on topographic and representational similarity. TDANN dyads (n = 7) show more similar topographies but ResNet dyads (n = 7) have more similar internal representations, predictions, and greater accuracy (note: asterisks for similarities, left, and contrasts, right, are color coordinated). (D) Factors previously associated with topographic variation in the brain include biological analogues of network depth (network hierarchy) and wiring costs (myelination), as well as other contending factors, which are used in later analyses. CKA, centered kernel alignment. *, $p < 0.05$, *** $p < 0.001$ (see **Supplemental Table 1**).

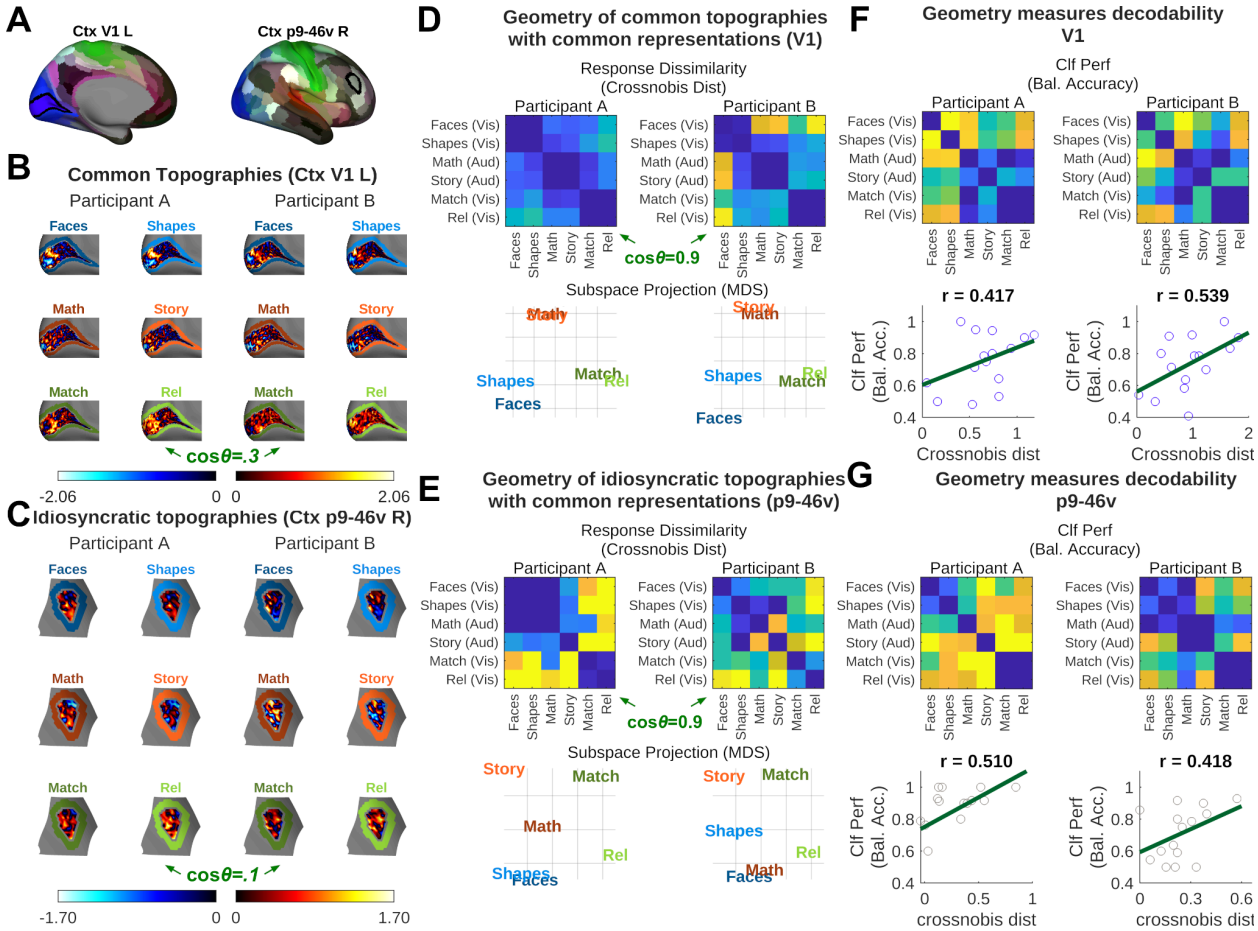

Figure 2 An illustrative example of common representations even when topographies are idiosyncratic. (A-C) Two exemplary regions (A: black outlines) from two participants demonstrate either similar topographies (V1) or idiosyncratic topographies (p9-46v; whitened evoked-response contrast maps). (D,E) Representational dissimilarity matrices (RDMs) capture intra-participant spatial dissimilarities. Both common and idiosyncratic topographies show common representations (compact subset of tasks shown). Cosine similarity shown on crossnobis RDMs for intuition. RDM whitening is applied later to produce WUC for inference. (F-G) Representational geometry captures task discriminability much like a brain-decoding binary nearest-centroid classifier (accuracy estimated via cross-validation across scans, weighted for condition imbalance). Each point indicates a pairwise classification vs. crossnobis distance for two task conditions. MDS, multidimensional scaling. Clf, classifier.

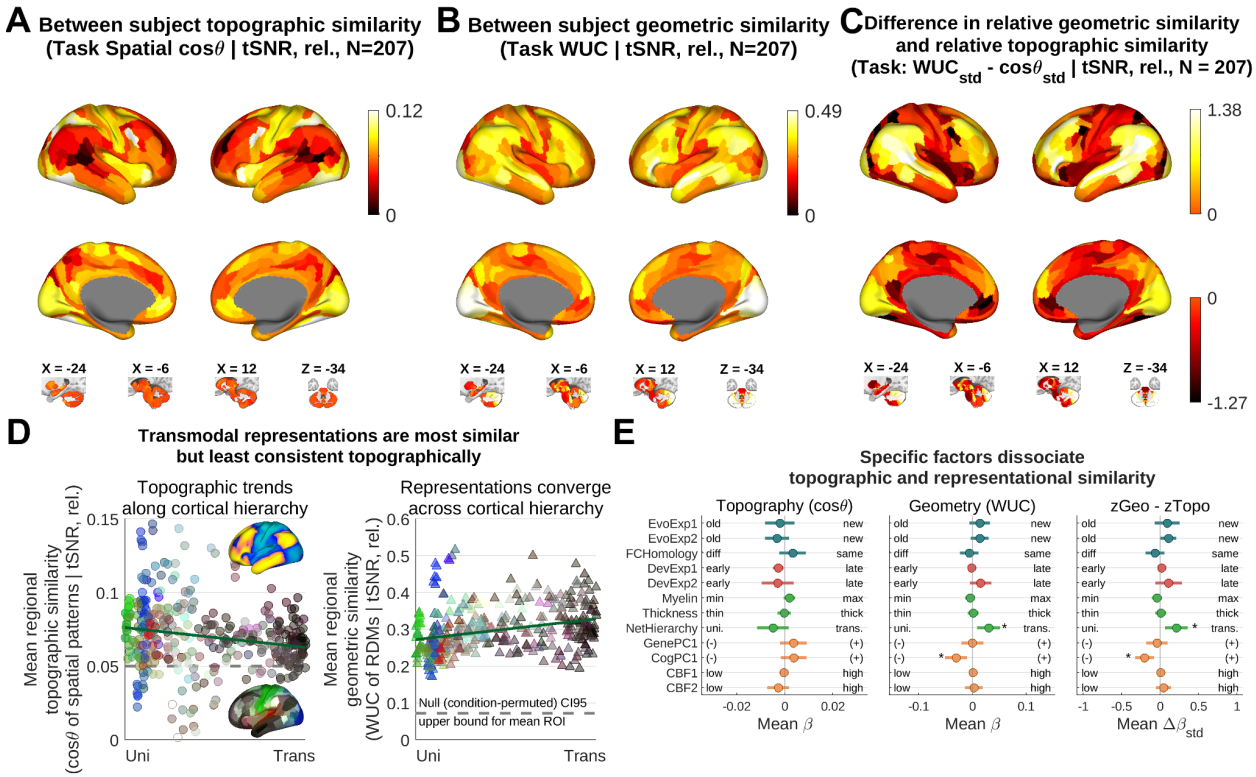

Figure 3 Geometry is similar across participants throughout much of the cortex, especially in transmodal regions. (A,B) Topographic and geometric similarity between dyads of unrelated participants across 23 different task conditions (mean spatial cosθ of task evoked maps and WUC of task RDMs, respectively; n = 207 dyads). (C) Spatial profiles of these measures differed from one another throughout the brain, indicating a dissociation between representational and topographic similarity. Metrics were standardized to enable direct comparison of spatial profiles. (D) Network hierarchy (top brain) vs. group-mean topographic similarity (left) or group-mean geometric similarity (right). Topography was more consistent in unimodal areas, whereas geometry was more consistent in transmodal areas. Points match the lower brain's colors. For reference, the dashed lines indicate conservative upper bounds of the condition-permuted null similarity distributions, averaged across cortical regions (Methods). Because null similarity varies across regions, many values below this conservative reference nevertheless exceed region-specific null expectations. (E) Left, center: univariate correlations of neuromaps with topographic and geometric similarity. Right: two cortical neuromaps, CogPC1 and NetHierarchy, significantly moderated the dissociation between geometric and topographic similarity. Neuromaps were standardized (unit variance, zero mean across space). Association tests were limited to the cortex by neuromap coverage. Corrected for tSNR and reliability effects (**Table 3**). Scatterplot colors: *, jackknife-adjusted spin test, FDR q=0.05. Error bars: 95% confidence interval. CI, confidence interval.

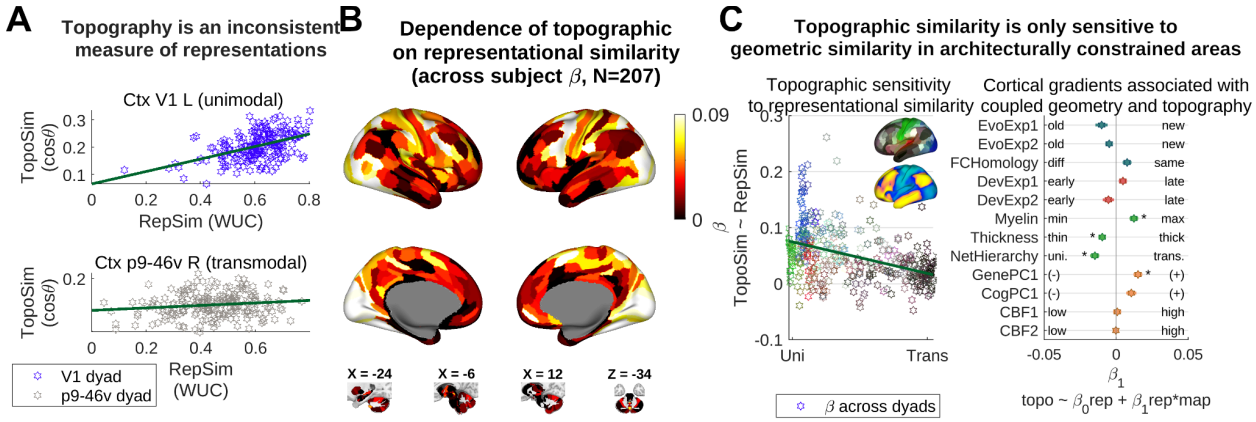

Figure 4 Region-by-region, topographic similarity is coupled to geometric similarity in unimodal, but not transmodal regions. (A) Two example regions illustrate how geometric and topographic similarity are coupled across dyads in some areas (V1) but not others (p9-46v). Each point indicates the similarity of the region across a dyad. (B) Geometric similarity was coupled with topographic similarity in unimodal areas, including the somatotopic motor areas of anterior dorsal and ventral cerebellum. (C) Regional variation in sensitivity between topographic and geometric similarity. Left: slope of the topography-geometry relationship, like in (A), plotted against network hierarchy. Right: neuromap moderation analysis, indicating sensitivity. There was a significant association with network hierarchy such that unimodal topography was more coupled to geometry than transmodal topography (jackknife-adjusted spin test, FDR q=0.05). Scatterplot color scheme corresponds to atlas regions (insert). Association tests were limited to the cortex by neuromap coverage. β_1 adjusted for tSNR and within-participant test-retest reliability (full equation not shown, see methods). Error bars: 95% confidence interval.

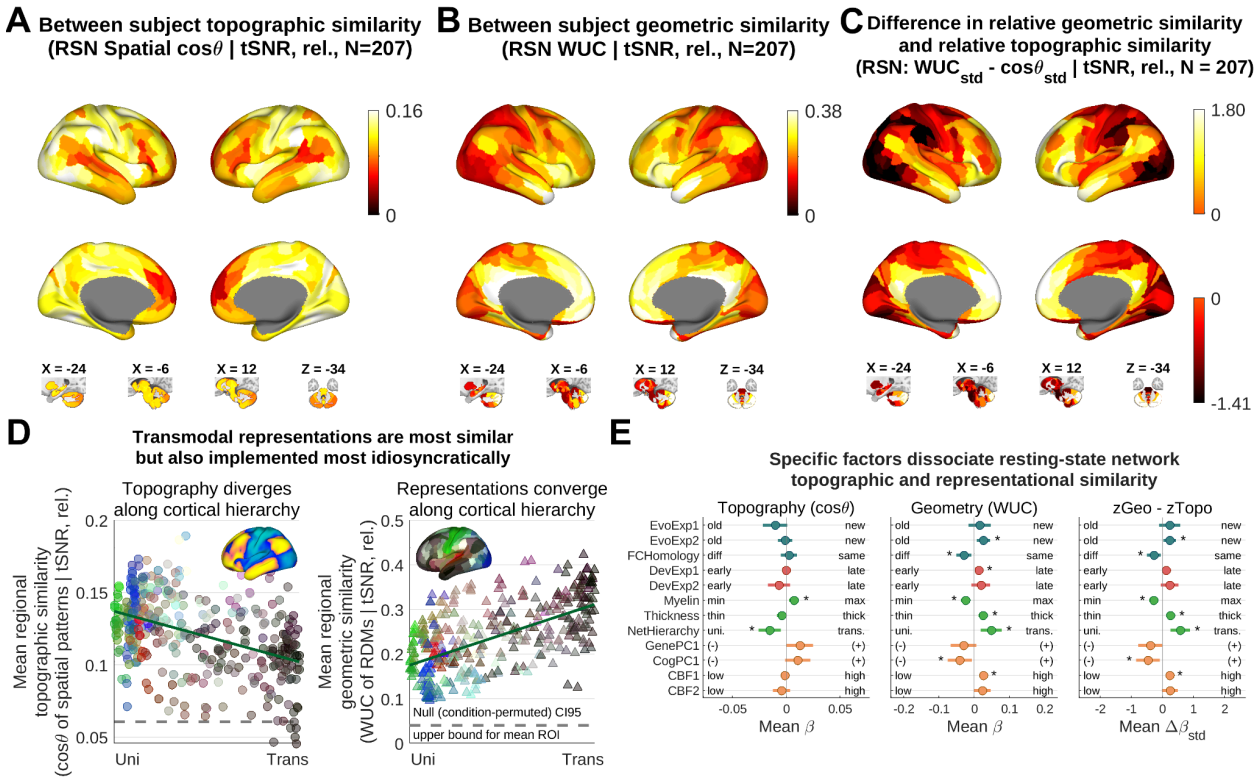

Figure 5 Resting state network (RSN) representational and topographic similarity diverge along architectural hierarchies, mirroring task-evoked responses (Figure 3). (A,B) RSN topographic and geometric similarities vary throughout the brain in a characteristic fashion distinct from task evoked responses. Representations reflect the spatial discriminability of individualized networks within-region. Distances underlying representational geometries were computed across different scan days. (C) The spatial profile of differences between geometric similarities and topographic similarities (spatially normalized within-participant to mean 0 and unit variance) was highly similar to differences between task geometries and topographies (Figure 3C). (D) Network hierarchy showed the strongest dissociation of topographic and representational similarity. For reference, the dashed lines indicate conservative upper bounds of the network-permuted null similarity distributions, averaged across cortical regions (Methods). Because null similarity varies across regions, all values below this conservative reference nevertheless exceed region-specific null expectations. (E) Formal tests of the associations of neuromaps with topographic similarity, geometric similarity, and their difference, revealing a tendency towards idiosyncratic implementations of common representations, reproducing task-evoked results. Association tests were limited to the cortex by neuromap coverage. Corrected for tSNR and within-participant test-retest reliability. Error bars: 95% confidence interval, see Table 5 for effect sizes smaller than icon size. *, jackknife-adjusted spin test, FDR q=0.05.

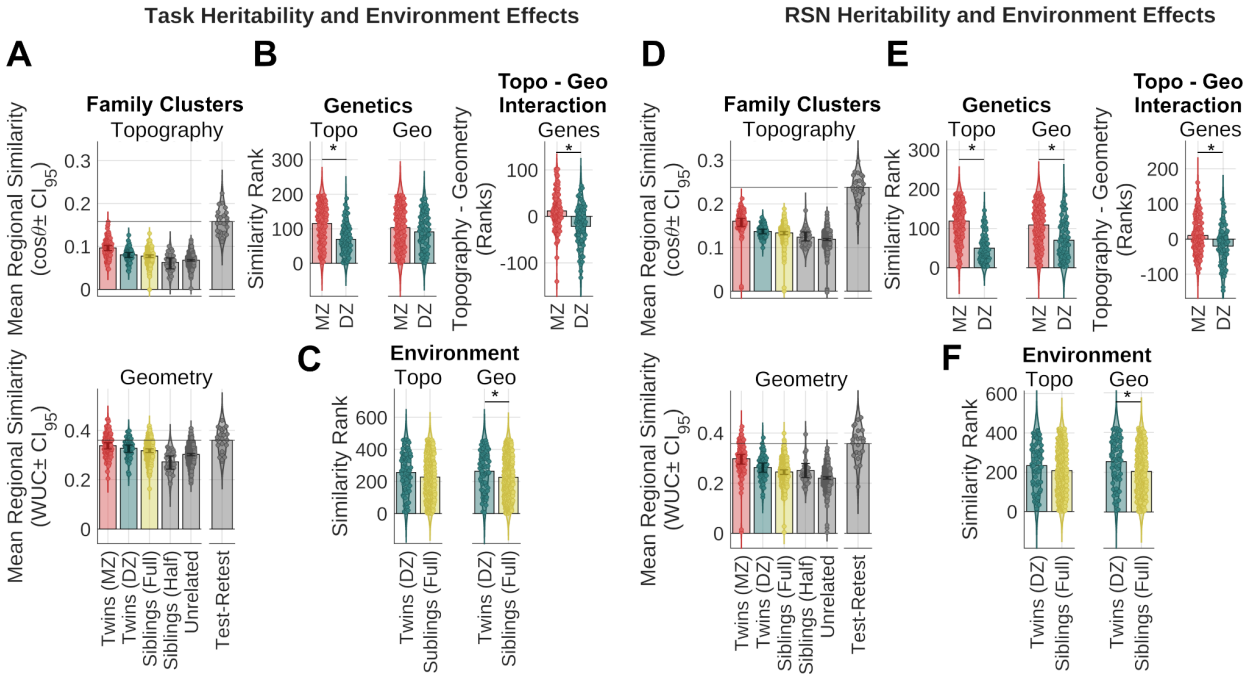

Figure 6 Genes exert stronger control over topography than geometry for both task responses (A-C) and resting state networks (RSNs; D-F). (A,D) Clustering individuals by familial relationships showed increasing geometric and topographic similarity with genetic proximity. Test-retest reliability in individuals who returned for a second visit (n = 22) indicated that family relationships explain less geometric than topographic variance. The horizontal line indicates mean self-similarity of retest participants, a population-average noise ceiling. This was about 2x higher for similarity of unrelated geometries than unrelated topographies. (B,E) Monozygotic twins (MZ, n = 127 dyads) were more similar than dizygotic twins (DZ, n = 70 dyads), showing that both geometry and topography are heritable (Mann-Whitney U-test, Holm-Sidak corrected for 4 comparisons, α = 0.05). However, heritability of topographies was significantly greater than geometries in both task responses and RSNs (nonparametric interaction rank-test). (C,F) DZ twins had more similar geometry than full siblings (n = 392 dyads), indicating a common environment effect for both task and rest (Mann-Whitney U-test, Holm-Sidak corrected for 4 comparisons, α = 0.05). Common environment effects for geometry did not differ significantly from topographies for either task or RSN networks. Family clusters are color coordinated. Violin plot scatter points: similarities of non-intersecting dyads. Error bars: BCa bootstrap 95% confidence intervals (CI_{95}). *, significant, α < 0.05.

	β	*t*	*dof*	*p*
Topographic similarity				
Layer 1.0 > 4.1	-0.260	-13.5	16.8	2.0e-10
TDANN > ResNet	0.107	8.5	29.7	1.7e-9
TDANN layer v ResNet layer	0.065	1.7	99.7	0.09
Geometric similarity				
Layer 1.0 > 4.1	-0.054	-3.0	7.9	0.017
TDANN > ResNet	-0.091	-7.5	7.8	7.7e-5
TDANN layer v ResNet layer	-0.137	-4.6	19.5	1.7e-4
Standardized geometric > standardized topographic similarity				
Layer 1.0 - 4.1	1.43	4.24	219	3.2e-5
TDANN > ResNet	-2.2	-12.4	219	3.2e-27
Accuracy: TDANN > ResNet				
Top-1	-4.2	-32.0	13	9.3e-14
Top-5	-4.2	-34.5	13	3.6e-14
Agreement: TDANN > ResNet				
1000-categories	-5.4	-4.7	6	3.3e-3
50 hypernyms	-4.8	-4.5	6	4.2e-3

Table 1 Comparison of topographic, representational, and functional convergence in TDANNs and ResNet-18s. Accuracy and agreement: paired *t*-test, paired by initialization seed. Others: Mixed model, random network slope and intercepts, Satterthwaite degrees of freedom (dof).

Neuromap	Description	Hypothesized effect	Explanation
Evolutionary			
Evolutionary expansion 1	Human vs. macaque, anatomically aligned	Phylogenetically older areas: less idiosyncratic in both topography and geometry	Older areas are more likely to have reached allelic fixation or be under greater selective pressure; newer areas would show greater genetic variation[95].
Evolutionary expansion 2	Human vs. macaque, functionally aligned		
FCHomology	Functional connectivity homology vs. macaque		
Developmental			
Developmental expansion 1	Infant vs. adult (18-24yo) areal scaling	Increased representational convergence with longer development timelines	Increased exposure to environment and susceptibility to experience dependent plasticity
Developmental expansion 2	Local / global areal scaling in youth, 5-25yo		
Architectural			
Myelination	T1/T2 ratio	Greater myelination decreases representational convergence	Wiring constraint limits capacity for convergence by constraining circuit flexibility[23].
Thickness	Cortical thickness	Representational convergence in thicker areas	Increased laminar differentiation and reduced intracortical connectivity in thinner areas[24] constrains circuit function.
Network Hierarchy	Principal gradient of connectivity	Increased representational convergence in transmodal areas	More receptive field integration and composition increases implementation flexibility of higher order representations
Miscellaneous and Negative Controls			
GenePC1	Transcriptomic variation	-	-
CogPC1	First principal component of cognitive terms from Neurosynth	Task dependent effects (negative control)	Task contingent effects, especially a bias favoring abstract cognitive over sensory tasks or vice versa.
CBF1	Cerebral blood flow (CBF)	More consistent implementations with greater CBF	Low CBF affects the BOLD signal-to-noise ratio.
CBF2			

Table 2 Neuromaps and the hypotheses they test. yo, years old.

	Uncorrected				tSNR and test-retest reliability corrected			
	Topo	Geo	zGeo - zTopo	*β*: Topo~Rep	Topo	Geo	zGeo - zTopo	*β*: Topo~Rep
EvoExp1	-0.018 ±0.035	0.008 ±0.068	**0.235* ±0.158**	-0.017 ±0.005	-0.002 ±0.006	0.014 ±0.019	0.091 ±0.163	-0.010 ±0.004
EvoExp2	-0.006 ±0.020	0.020 ±0.035	0.171 ±0.146	-0.007 ±0.003	-0.003 ±0.005	0.014 ±0.017	0.109 ±0.103	-0.005 ±0.003
FCHomology	0.017 ±0.022	0.013 ±0.038	-0.118 ±0.152	0.013 ±0.004	0.003 ±0.006	-0.006 ±0.017	-0.065 ±0.121	0.008 ±0.003
DevExp1	**0.016* ±0.012**	**0.035* ±0.022**	-0.007 ±0.067	0.007 ±0.003	-0.003 ±0.002	-0.001 ±0.007	0.021 ±0.047	0.005 ±0.003
DevExp2	0.001 ±0.031	0.034 ±0.058	0.166 ±0.185	-0.008 ±0.004	-0.003 ±0.007	0.015 ±0.020	0.111 ±0.172	-0.005 ±0.004
Myelin	**0.027* ±0.013**	**0.036* ±0.028**	**-0.122* ±0.062**	**0.021* ±0.004**	0.002 ±0.002	-0.004 ±0.009	-0.040 ±0.065	**0.012* ±0.003**
Thickness	**-0.025* ±0.015**	**-0.045* ±0.029**	0.044 ±0.072	**-0.016* ±0.003**	-0.000 ±0.003	0.002 ±0.010	0.010 ±0.071	**-0.010* ±0.003**
NetHierarchy	-0.030 ±0.032	-0.017 ±0.058	**0.253* ±0.181**	**-0.022* ±0.004**	-0.005 ±0.007	**0.030* ±0.021**	**0.213* ±0.147**	**-0.015* ±0.003**
GenePC1	0.044 ±0.038	0.076 ±0.074	-0.096 ±0.144	**0.025* ±0.005**	0.004 ±0.006	-0.000 ±0.021	-0.039 ±0.139	**0.015* ±0.003**
CogPC1	0.029 ±0.035	0.040 ±0.068	-0.124 ±0.170	0.015 ±0.004	0.004 ±0.005	**-0.030* ±0.020**	**-0.203* ±0.123**	0.010 ±0.003
CBF1	0.003 ±0.010	0.007 ±0.021	-0.000 ±0.048	0.000 ±0.003	-0.000 ±0.002	0.001 ±0.007	0.008 ±0.051	0.001 ±0.003
CBF2	0.005 ±0.026	0.023 ±0.051	0.068 ±0.126	-0.001 ±0.002	-0.003 ±0.004	0.003 ±0.016	0.044 ±0.099	-0.000 ±0.002

Table 3. Univariate coefficients, regression of task similarity metrics on neuromaps (standardized; mean: 0, standard deviation: 1), with and without confound correction. Coefficients within (but not between) columns are on the same scale. *n* = 207. Jackknife 95% confidence intervals shown. *(boldface) $p < 0.05$, FDR $q = 0.05$.

	Uncorrected (Est. ± CI 95)				tSNR and test-retest reliability corrected (Est. ± CI 95)			
	Topo	Geo	zGeo - zTopo	*β*: Topo~Rep	Topo	Geo	zGeo - zTopo	*β*: Topo~Rep
EvoExp1	0.007 ±0.041	0.010 ±0.034	0.006 ±0.629	0.003 ±0.008	-0.010 ±0.011	0.016 ±0.032	0.238 ±0.340	-0.002 ±0.011
EvoExp2	0.022 ±0.021	0.023 ±0.021	-0.042 ±0.339	-0.000 ±0.007	-0.001 ±0.007	**0.027* ±0.019**	**0.236* ±0.206**	-0.004 ±0.009
FCHomology	-0.009 ±0.023	**-0.030* ±0.023**	-0.148 ±0.355	-0.001 ±0.008	0.003 ±0.008	**-0.029* ±0.022**	**-0.268* ±0.223**	0.001 ±0.007
DevExp1	**0.030* ±0.013**	0.007 ±0.012	-0.259 ±0.217	0.005 ±0.005	-0.000 ±0.004	**0.013* ±0.010**	0.116 ±0.114	-0.001 ±0.008
DevExp2	0.019 ±0.034	0.012 ±0.029	-0.098 ±0.511	0.006 ±0.007	-0.007 ±0.010	0.019 ±0.027	0.234 ±0.287	0.001 ±0.010
Myelin	0.015 ±0.016	**-0.028* ±0.014**	**-0.395* ±0.259**	0.002 ±0.008	**0.007* ±0.005**	**-0.025* ±0.013**	**-0.281* ±0.140**	0.003 ±0.007
Thickness	**-0.027* ±0.016**	**0.029* ±0.014**	**0.539* ±0.254**	-0.010 ±0.007	-0.004 ±0.004	**0.025* ±0.013**	**0.259* ±0.139**	-0.006 ±0.004
NetHierarchy	-0.011 ±0.035	**0.050* ±0.032**	0.540 ±0.552	0.002 ±0.010	**-0.015* ±0.010**	**0.049* ±0.030**	**0.576* ±0.318**	0.001 ±0.010
GenePC1	0.042 ±0.045	-0.033 ±0.039	-0.723 ±0.718	0.007 ±0.010	0.013 ±0.012	-0.030 ±0.036	-0.387 ±0.393	0.003 ±0.012
CogPC1	0.026 ±0.042	-0.042 ±0.038	-0.635 ±0.715	0.006 ±0.009	0.011 ±0.012	**-0.041* ±0.034**	**-0.463* ±0.387**	0.003 ±0.010
CBF1	0.011 ±0.012	**0.026* ±0.010**	0.106 ±0.183	-0.005 ±0.006	-0.002 ±0.003	**0.026* ±0.009**	**0.238 ±0.100**	-0.006 ±0.007
CBF2	0.018 ±0.030	0.018 ±0.026	-0.042 ±0.484	0.005 ±0.005	-0.005 ±0.008	0.023 ±0.024	0.245 ±0.261	0.001 ±0.007

Table 4 Univariate coefficients, regression of RSN similarity metrics on neuromaps (standardized; mean: 0, standard deviation: 1), with and without confound correction. Coefficients within (but not between) columns are on the same scale. *N* = 207. Jackknife 95% confidence intervals. *(boldface) $p < 0.05$, FDR $q = 0.05$.