



Hyperalignment: Modeling shared information encoded in idiosyncratic cortical topographies

James V Haxby^{1*}, J Swaroop Guntupalli², Samuel A Nastase³, Ma Feilong¹

¹Center for Cognitive Neuroscience, Dartmouth College, Hanover, United States;

²Vicarious AI, Union City, United States; ³Princeton Neuroscience Institute, Princeton, United States

Abstract Information that is shared across brains is encoded in idiosyncratic fine-scale functional topographies. Hyperalignment captures shared information by projecting pattern vectors for neural responses and connectivities into a common, high-dimensional information space, rather than by aligning topographies in a canonical anatomical space. Individual transformation matrices project information from individual anatomical spaces into the common model information space, preserving the geometry of pairwise dissimilarities between pattern vectors, and model cortical topography as mixtures of overlapping, individual-specific topographic basis functions, rather than as contiguous functional areas. The fundamental property of brain function that is preserved across brains is information content, rather than the functional properties of local features that support that content. In this Perspective, we present the conceptual framework that motivates hyperalignment, its computational underpinnings for joint modeling of a common information space and idiosyncratic cortical topographies, and discuss implications for understanding the structure of cortical functional architecture.

*For correspondence:
james.v.haxby@dartmouth.edu

Competing interests: The authors declare that no competing interests exist.

Funding: See page 21

Received: 12 March 2020

Accepted: 14 May 2020

Published: 02 June 2020

Reviewing editor: Chris I Baker, National Institute of Mental Health, National Institutes of Health, United States

© Copyright Haxby et al. This article is distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

Introduction

Information encoded in cortex is organized at varying spatial scales. The introduction of multivariate pattern analysis of fMRI data (MVPA; [Haxby, 2012](#); [Haxby et al., 2014](#); [Haxby et al., 2011](#); [Haxby et al., 2001](#)) revealed that information can be decoded from fine-grained patterns of cortical activity (see also [Bzdok, 2017](#); [Haynes, 2015](#); [Haynes and Rees, 2006](#); [Hebart and Baker, 2018](#); [Norman et al., 2006](#); [Pereira et al., 2009](#); [Tong and Pratte, 2012](#)). Study of cortical functional connectivity also has revealed fine-grained topographies in the connectome ([Arcaro et al., 2015](#); [Haak et al., 2018](#); [Heinze et al., 2011](#); [Jbabdi et al., 2013](#)) that are closely related to these patterns of activity ([Guntupalli et al., 2018](#)). In the context of fMRI, fine-scale granularity in cortical patterns in fMRI data refers to voxel-by-voxel (or vertex-by-vertex) variation in response and connectivity profiles. Even finer-scale variation exists at the level of columns and neighboring neurons ([Leopold et al., 2006](#); [Park et al., 2017](#)). The surface structure of functional cortical topographies, however, shows considerable variability across individual brains for encoding the same information ([Cox and Savoy, 2003](#); [Guntupalli et al., 2016](#); [Haxby et al., 2011](#)). We introduced a novel conceptual framework that models both the shared information encoded in fine-grained functional topographies and individual-specific topographic idiosyncrasies. This framework, 'hyperalignment' ([Chen et al., 2014](#); [Chen et al., 2015](#); [Guntupalli et al., 2018](#); [Guntupalli et al., 2016](#); [Haxby et al., 2011](#); [Xu et al., 2012](#)), encompasses a family of computational algorithms that align shared information in a common, high-dimensional information space, rather than attempting to align functional topographies in the physical space of cortical anatomy to a canonical topography. Hyperalignment derives from the term hyperspace, to convey the paradigm shift from modeling

cortical functional architecture in a three-dimensional anatomical space to modeling the information encoded in functional architecture in a high-dimensional information space. This common information space captures the ‘deep structure’ of shared information that is encoded in the variable ‘surface structure’ of idiosyncratic topographies in individual brains.

Hyperalignment algorithms construct the common information space and calculate transformations that project individually-variable patterns of neural activity and connectivity into this common model space. Information is contained in the relationships among pattern vectors – distinctions and similarities – and these relationships are preserved when the space is transformed from an individual cortical space to the common information space. In these transformations, the features that define the space in individual brains – response tuning functions and connectivity profiles – are remixed to support the same information structure with shared basis functions. Thus, in the conceptual framework of our computational model, the fundamental property of brain function that is preserved is information content, and the functional properties of local features that support that content are mutable and secondary.

Our computational framework introduces a new model of cortical functional architecture. Shared information is decomposed into overlapping basis functions that are instantiated in individual brains with idiosyncratic topographies. Cortical patterns of activity and connectivity are modeled as mixtures of these topographic basis functions. Our common model accounts for coarse areal structure insofar as the model’s shared basis set for fine-scale topographies can be mixed to identify the boundaries for functional areas in individual brains, but the topographic information embedded in the model’s shared bases is not strongly constrained by these areal boundaries, suggesting that areal structure may play a smaller role in how information is organized in cortex than was previously believed.

Here we present the fundamental concepts that underlie this approach to modeling cortical functional architecture, as well as the computational underpinnings for deriving the common information space and modeling idiosyncratic cortical topographies. We review the effectiveness and significance of these methods, their potential for increasing the sensitivity of inquiry into individual differences, and discuss the implications of this conceptual framework for understanding cortical functional architecture.

Key concept: High-dimensional information spaces

Vector space representation of cortical patterns

Cortical functional patterns can be represented as vectors corresponding to locations in a high-dimensional vector space where each dimension is a local measurement or element of that pattern (*Figure 1a–b; Haxby et al., 2014*). Instead of representing cortical function as a topographic spatial map in the three-dimensional physical space of the cortex (or a two-dimensional representation of the cortical sheet), hyperalignment models the information embedded in cortical function in this high-dimensional space (cf. *Durbin and Mitchison, 1990*). For fMRI data, local pattern elements are typically voxels (or vertices on the cortical surface). Patterns of neural activity measured with single-unit recording, electrocorticography (ECoG), or electro- and magnetoencephalography (EEG and MEG) can also be analyzed as pattern vectors in cortical or subcortical feature spaces in which each dimension may be a single neuron, an electrode, or a dipole, magnetic field sensor, or magnetic source. The relevance of the hyperalignment conceptual framework for modeling cortical functional architecture is not specific to any measurement modality or neural variable, and the computational procedures can also be applied to data matrices obtained with other measurement modalities (e.g. *Chen et al., 2020*) or even artificial neural networks (*Lu et al., 2019*). We will focus this perspective on human cortical activity patterns accessible with fMRI.

Information as vector geometry

Matrices of brain activity and connectivity lend themselves to analysis as geometric structures in high-dimensional information spaces. The rows of these matrices are distributed cortical patterns of responses to stimuli or connectivities with cortical targets elsewhere in the brain. The columns are

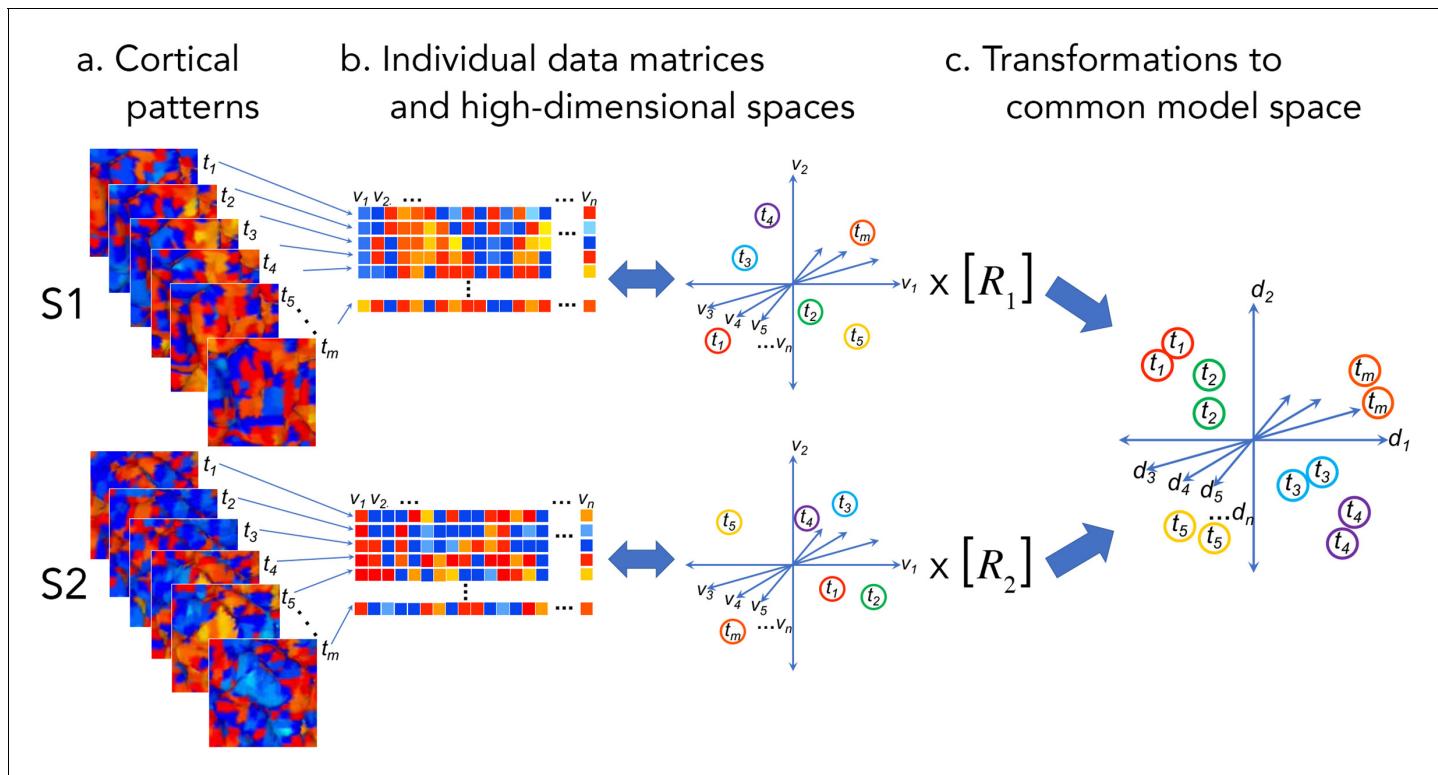


Figure 1. Cortical patterns as vectors in individual and common high-dimensional information spaces. Cortical functional patterns (a) are analyzed as data matrices of vectors in high-dimensional cortical feature spaces (b). Hyperalignment transforms individual information spaces based on cortical loci into a common model space in which pattern vectors for the same brain state (corresponding to a given stimulus, condition, or task) are aligned across brains (c). With fMRI data, cortical patterns are typically responses at different time points or connectivities to different targets (t), brain loci are voxels or cortical vertices (v), transformations are matrices (R), and common model coordinate axes are dimensions (d).

response or connectivity profiles for cortical loci (voxels or vertices for fMRI) in individual brains or model dimensions in the common information space (Figures 1 and 2). The row vectors may be patterns of activity at different points in time or response patterns recovered from multiple instances of a given event or experimental manipulation (using trial averaging or deconvolution) and serve as indices of the brain state corresponding to a given stimulus or cognitive state. These row vectors also may be patterns of functional connectivities between each cortical locus in a field and a locus, or connectivity target, elsewhere in the brain. Thus, the information space is populated by response pattern vectors or connectivity pattern vectors. We refer to the response or connectivity magnitudes for a given cortical location (i.e., a column in the data matrix) as a functional profile.

Here, we define ‘information’ as relationships among functional pattern vectors, typically quantified as the similarity (or dissimilarity/distance) between each pair of vectors (Connolly et al., 2012; Connolly et al., 2011; Edelman et al., 1998; Hanson et al., 2004; Kriegeskorte et al., 2008a; Kriegeskorte and Kievit, 2013; O’Toole et al., 2005). These dissimilarities index the distinctiveness of each vector relative to other vectors; vectors that are more similar to each other are located nearer to each other in vector space. This geometric framework for operationalizing information as the relationships among distributed patterns in a vector space is used in a variety of fields, including psychology (e.g. Edelman, 1998; Shepard, 1987), information retrieval (e.g. Salton et al., 1975), and natural language processing (e.g. Deerwester et al., 1990). The set of vectors in individual information spaces can be analyzed as the geometry of those vectors, namely the set of pairwise distances between vectors. The vector geometry of information in cortical spaces is highly similar across brains (Connolly et al., 2016; Connolly et al., 2012; Connolly et al., 2011; Diedrichsen and Kriegeskorte, 2017; Guntupalli et al., 2018; Guntupalli et al., 2016; Kriegeskorte and Kievit, 2013; Nastase et al., 2017). Prior to hyperalignment, these functional pattern vectors are instantiated in the spatially organized cortical topography of a given individual. Hyperalignment leverages the

inter-individual similarity of vector geometry to create a shared information space that minimizes differences between corresponding vectors across individuals while preserving vector geometry.

Hyperalignment: Joint modeling of a common information space and idiosyncratic topographies

A common, high-dimensional information space

The cortical feature spaces for individual brains are difficult to compare because of the inter-individual variability of topographic surface structure. The goal of hyperalignment is to create a common information space in which the vectors that encode information that is shared across brains – representations of the same stimulus or cognitive process, functional connections to other cortical fields – are aligned. Note that the goal is not simply to align the cortical patterns that carry that information in a two- or three-dimensional anatomical space (Conroy et al., 2013; Robinson et al., 2014; Sabuncu et al., 2010). The columns of data matrices in the common model space are the coordinate axes or dimensions for that information space. The rows of data matrices in the common space contain cortical information patterns that have been projected into the model space from individual data matrices. The solution to this problem requires finding individual-specific transformations, R_i , that project individual data matrices, B_i , in which the coordinate axes are cortical locations, into a common model coordinate space, in such a way that minimizes inter-individual differences among pattern vectors that represent the same perceptual, cognitive, or connectivity information. This is accomplished by minimizing the difference between patterns in the transformed individual data matrices, B_iR_i , and the group mean patterns, M . Formally,

$$M = (1/N) \sum_{i=1}^N B_i R_i \quad \text{where} \quad \operatorname{argmin}_R \sum_{i=1}^N \|B_i R_i - M\|_F \quad (1)$$

The dimensions in the common model space are no longer anatomical locations but, rather, are weighted sums of nearby anatomical locations in individual brains; the transformations are not a one-to-one mapping from cortical loci to common space dimensions. The distribution of these weights is specific to each brain. Each column of the transformation matrices, R_i , contains the weights for this transformation for one model dimension, d , across all cortical loci, v . The

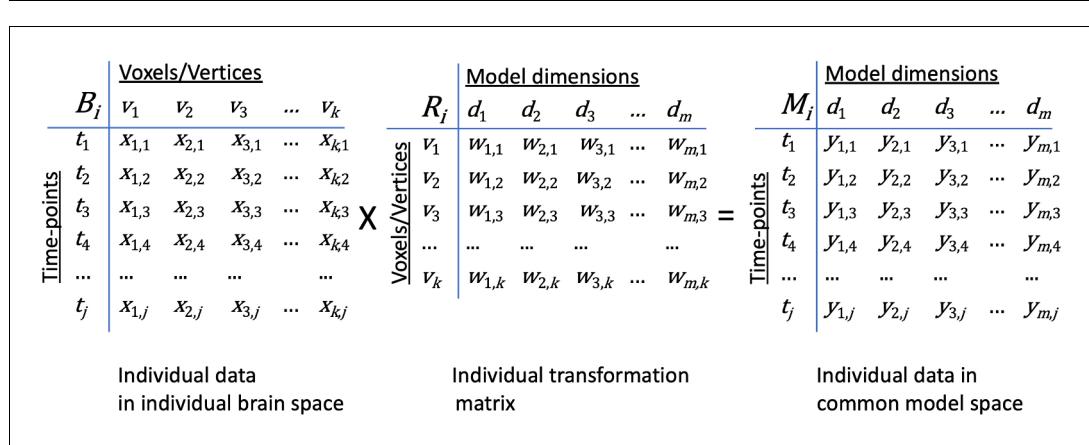


Figure 2. Transformation of data in individual brain coordinate space into common information space coordinates. An individual brain data matrix in cortical space (B_i) consists of j rows of patterns (t) with k columns of cortical loci (v). Patterns can be neural activity for time-points, stimuli, or conditions; patterns of connectivity with targets; or other measures that vary across cortical loci. The transformation matrix for that individual (R_i) consists of k rows of cortical loci (v) with m columns of model dimensions (d). Multiplying these matrices transforms the individual brain data matrix, B_i , into the model space coordinates (M_i) with rows of activity or connectivity patterns, t , in model dimension coordinates, d . Elements in brain data matrices (x in B_i) are measures of local neural activity or connectivity of that locus with a connectivity target. Elements in transformation matrices (w in R_i) are weights. Data in the model space matrix (y in M_i) are weighted sums of measures from cortical loci (x in B_i).

transformation matrices have the same structure whether derived from response profiles or connectivity profiles. These weights model how shared information can be instantiated in idiosyncratic fine-grained cortical functional topographies.

The number of dimensions in the common space (m in **Figure 2**) can differ from the number of cortical loci in the data matrices (k in **Figure 2**) from which the space is derived. In some implementations, the number of model dimensions matches the number of cortical loci in a ‘reference brain’, allowing patterns in the model space to be illustrated in the reference brain anatomy (*Guntupalli et al., 2018; Guntupalli et al., 2016; Haxby et al., 2011*). In other implementations, the model space is reduced to capture shared information in a smaller number of dimensions and reduce noise or overfitting (*Chen et al., 2015; Guntupalli et al., 2016; Haxby et al., 2011*). Patterns in reduced dimension model spaces can be illustrated by projection into the anatomy of individual brains, using the transposes of individual transformation matrices, R_i^T .

The goal of the common information space is to align cortical pattern vectors across individuals while preserving the geometry of those vectors. The coordinate axes for that space – the model dimensions – carry information in the form of shared functional profiles (response tuning vectors or functional connectivity vectors), but they can be arbitrarily rotated, preserving vector geometry while altering the response tuning and connectivity profiles for the coordinate axes. Thus, the information associated with each common model dimension serves only to preserve the information encoded in vector geometry, shifting attention from the traditional focus on univariate response and connectivity profiles for single cortical loci to the vector geometry of distributed population responses and population connectivity that span a cortical field. The individual transformation matrices afford projecting individual data embedded in idiosyncratic cortical topographies into the common space and, conversely, projecting group data embedded in the common space into the cortical topographies of individual brains.

Methods for deriving transformation matrices

There are various ways to derive transformations that minimize residuals around group mean functional patterns. We have concentrated on using our adaptation of the Generalized Procrustes Analysis (GPA; *Gower, 1975; Guntupalli et al., 2018; Guntupalli et al., 2016; Haxby et al., 2011; Schönemann, 1966*) to derive the transformation matrices. The Procrustes transformation is the orthogonal matrix that minimizes the distances between two matrices of paired vectors with a rigid rotation, which is ‘improper’ in that it allows reflections as well as rotations. GPA finds transformation matrices for a group of data matrices that minimize distances by iteratively aligning each new matrix to the mean of previously aligned matrices. A rigid rotation preserves all pairwise distances between pattern vectors within each original matrix. In the case of fMRI data matrices, therefore, this transformation preserves the vector geometry of each data set (representational geometry or connectivity geometry [*Connolly et al., 2016; Connolly et al., 2012; Connolly et al., 2011; Guntupalli et al., 2018; Guntupalli et al., 2016; Kriegeskorte et al., 2008a; Kriegeskorte and Kievit, 2013*]). We and others have shown that local vector geometry is highly correlated across brains (*Connolly et al., 2016; Connolly et al., 2012; Connolly et al., 2011; Guntupalli et al., 2018; Guntupalli et al., 2016; Kriegeskorte and Kievit, 2013; Nastase et al., 2017*), indicating that it is a key feature of shared information structure (*Diedrichsen and Kriegeskorte, 2017; Ejaz et al., 2015*). Hyperalignment algorithms that mostly preserve vector geometry tend to generalize better to novel data than algorithms that do not (*Xu et al., 2012*).

Searchlight-based whole-cortex hyperalignment

Hyperalignment remixes data across spatial loci; however, for whole-cortex hyperalignment, we want to avoid remixing information from distant voxels. We have developed whole-cortex hyperalignment algorithms using GPA that mostly preserve local vector geometry but not the orthogonality of the whole-cortex transformation matrices (**Figure 3**; *Guntupalli et al., 2018; Guntupalli et al., 2016*). This was necessary to impose a locality constraint whereby information from any cortical location was resampled only into nearby locations in the ‘reference’ cortical anatomy. Local vector geometry also can be preserved by restricting hyperalignment to regions of interest (ROIs; *Chen et al., 2015; Haxby et al., 2011*) that tessellate the cortex (e.g., *Glasser et al., 2016; Gordon et al., 2016; Shen et al., 2013; Shen et al., 2010*), but this may impose an excessively restrictive prior

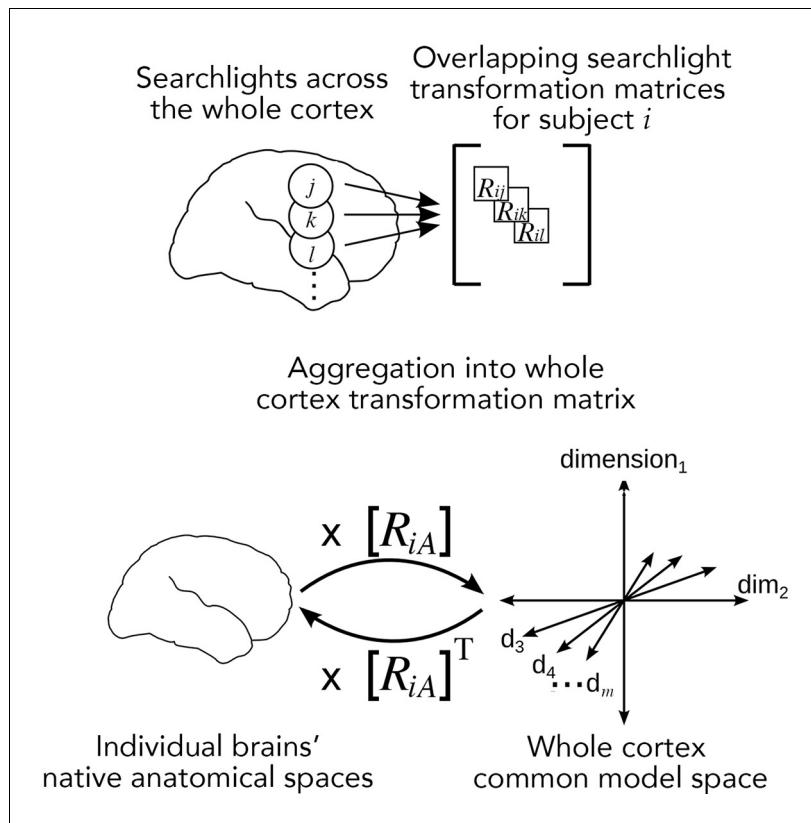


Figure 3. Searchlight hyperalignment for building a common model and transformation matrices for the whole cortex. Overlapping transformation matrices (R_{ij} , R_{ik} , ...) are calculated for overlapping searchlights in each subject and aggregated into a single whole-cortex transformation matrix, R_{iA} . The transpose of the transformation matrix, R_{iA}^T , allows pattern vectors from the common model space to be projected into an individual brain's topographies. The aggregation step results in a whole-cortex transformation that is mostly orthogonal locally (Guntupalli et al., 2016) but globally nonorthogonal.

constraint, as the topographies associated with more unconstrained model dimensions usually do not conform to areal boundaries (Haxby et al., 2011). Whole-cortex searchlight hyperalignment, by contrast, makes local models in overlapping searchlights (Kriegeskorte et al., 2006; Oosterhof et al., 2011) that cover the cortex, then aggregates the local transformation matrices into a whole-cortex transformation matrix, thereby mostly preserving local vector geometry (Guntupalli et al., 2016). Data are aligned anatomically to a standard cortical surface (e.g., FreeSurfer's fsaverage; Fischl et al., 1999) prior to whole-cortex searchlight hyperalignment. The cortical fields for local basis functions, defined by the voxels with nonzero weights in columns of the whole-cortex transformation matrices, are overlapping and constrained only by searchlight size, not by areal boundaries. Searchlights of increasing radius will accommodate larger divergence in individual topographies at the expense of spatial specificity; the 'appropriate' searchlight size may vary across brain areas and experiments, and this has not yet been systematically examined. Boundaries of cortical areas can be constructed from weighted sums of these local basis functions but appear to play a minor role in our model of cortical topographies.

A cortical field in the whole-cortex common model

A cortical field of contiguous cortical vertices, such as a searchlight or ROI, in the common information space produced by whole-cortex hyperalignment is only defined insofar as it is instantiated in the idiosyncratic cortical topography of an individual brain. Consequently, a study of local function with hyperaligned data requires projecting individual data into one subject's cortical anatomy, referred to as the reference subject. Other subjects' data in a cortical field in the reference subject's anatomical space are idiosyncratic mixtures of vertices from the same cortical field, anatomically

defined, as well as nearby vertices. Thus, data that is more similar to other subjects' data in a given cortical field is retained and imported from nearby cortex, whereas data that is more similar to that of others' data in nearby cortical fields is resampled out and into those other locations. The result of this resampling of information into and out of cortical fields is that the local individual information spaces in cortical fields are more consistent across brains, as evidenced, for example, by higher intersubject correlations of local vector geometry. Using a different subject's anatomy as the reference results in similar alignment of local information spaces with a different distribution of how information spaces vary across cortical fields.

Reducing the dimensionality of the common model space

The functional profiles of cortical loci are often redundant, and the information encoded in the vector geometry for a given cortical field may be lower-dimensional than the number of constituent cortical loci. We have shown that reducing the dimensionality of the common model space using principal component analysis (PCA; [Guntupalli et al., 2016](#); [Haxby et al., 2011](#)) can preserve or even improve performance, as indexed with between-subject multivariate pattern classification (bsMVPC; [Guntupalli et al., 2016](#); [Haxby et al., 2011](#)). Later work suggests that bsMVPC accuracies can be substantially higher with lower dimensional models ([Chen et al., 2015](#)). Dimensionality reduction necessarily warps vector geometry but can effectively denoise data in common space by discarding lower-order eigenvectors.

Other algorithms for hyperalignment

We and others also have developed algorithms other than GPA for deriving transformation matrices in individual brains, such as regularized canonical correlation analysis (rCCA; [Bilenko and Gallant, 2016](#); [Xu et al., 2012](#)), joint singular vector decomposition (SVD; [Chen et al., 2014](#)), and probabilistic estimation ([Chen et al., 2015](#)). With regularized CCA we find that a small relaxation of the orthogonality constraint enhances performance, as indexed with between-subject multivariate pattern classification (bsMVPC). [Chen et al., 2015](#) probabilistic approach, the Shared Response Model (SRM), pre-specifies a lower dimensionality for the common model, resulting, as noted above, in a warping of individual representational geometry that appears to filter out noise and reduce overfitting.

In our discussion of the effectiveness of hyperalignment, we will concentrate on results from analyses using the GPA-based algorithms. Optimization of the methods for deriving transformation matrices, however, is an ongoing subject of investigation.

Connectivity hyperalignment

The hyperalignment algorithm can also be used to construct a common model based on functional connectivity patterns. Previous work has shown that the human connectome has fine-grained topographies that are not captured by mean connections between areas ([Arcaro et al., 2015](#); [Braga et al., 2019](#); [Haak et al., 2018](#); [Heinzle et al., 2011](#); [Jbabdi et al., 2013](#)) and connectivity hyperalignment has shown that this fine-scale structure in the connectome is prevalent across all of cortex and of higher dimensionality than can be captured with known topographies, such as retinotopy and orientation selectivity ([Guntupalli et al., 2018](#)).

Connectivity hyperalignment requires a matrix of connectivity pattern vectors, in which each column is a cortical location in the cortical region to be hyperaligned and each row is a pattern of connectivities with a target location elsewhere in the brain ([Guntupalli et al., 2018](#)). Connectivity targets can be the responses for a single location (e.g., a cortical vertex), the mean response time-series for a cortical area or searchlight, or one or more component of the time-series in a cortical area after decomposition (e.g., using PCA). Once individual connectivity matrices are calculated, connectivity hyperalignment proceeds in the same way as response hyperalignment. Iteratively repeating connectivity hyperalignment allows connectivity target time-series to be rearranged and recalculated. Whereas response hyperalignment depends on response pattern vectors in experimental data that are locked to stimuli, connectivity hyperalignment depends on connectivity vectors for a defined set of connectivity targets and does not require a common stimulus paradigm, affording hyperalignment of resting state data ([Guntupalli et al., 2018](#)) or across different experiments ([Nastase et al., 2020](#)).

For a cortical field, the pattern of functional connectivity to a target can be thought of as a population connectivity vector, analogous to thinking of the pattern of activity evoked by a stimulus as a population response vector. Thus, while response hyperalignment aligns population responses across brains, preserving vector geometry of representation, connectivity hyperalignment aligns population connectivity vectors across brains, preserving their vector geometry. These two algorithms, however, mostly converge, both producing a high-dimensional common model space with individual transformation matrices that project individual anatomical spaces into a model space. Transformation matrices derived with connectivity hyperalignment can be applied to response data, effectively aligning response pattern vectors (**Guntupalli et al., 2018; Nastase et al., 2020**). Validation tests show that response hyperalignment aligns connectivity information and connectivity hyperalignment aligns population response information.

Rich, naturalistic stimuli

Estimating the parameters to transform high-dimensional spaces from individual brains into a common high-dimensional space requires a rich set of data that samples a wide variety of cortical patterns in order to generalize to novel stimuli or tasks. For response hyperalignment, a rich variety of stimuli or conditions are necessary to sample the response vector space. For connectivity hyperalignment, the sampling of connectivity vector space is defined by the selection of connectivity targets, but the richness and reliability of connectivity estimates depends on the variety of brain states over which connectivity is estimated.

We have shown that a common model space based on transformation matrices derived from responses to naturalistic audiovisual movies has general validity across a variety of experiments (**Guntupalli et al., 2016; Haxby et al., 2011; Nastase et al., 2017**). By contrast, transformation matrices estimated from controlled experiments with a limited set of stimuli or cognitive states may suffice for that experiment, but do not generalize well to other contexts or paradigms (**Haxby et al., 2011**).

The general validity that responses to naturalistic movies afford may stem from several important attributes of movies (**Hasson et al., 2010; Hasson et al., 2004; Haxby et al., 2020**). Such stimuli evoke a rich variety of brain states involving multiple coordinated systems for visual and auditory perception, perception of dynamic action, narrative, and speech, memory, attention, semantic knowledge, and social cognition. Movies are engineered to be engaging and guide attention. Movies also build a rich set of expectations for upcoming stimuli and events affording strong predictions (and prediction errors). Controlled experiments, on the other hand, typically randomize trial order, deliberately decontextualizing each event and providing no basis for accurate prediction. Consequently, randomization of trial order nullifies the utility of neural information processing related to the formation and application of predictive coding, making prediction-related activity weak and inconsistent across brains. In a comparison of between-subject multivariate pattern classification (bsMVPC) of responses to dynamic stimuli in movies to bsMVPC of responses to static stimuli in controlled experiments, with the structure of these analyses carefully matched to allow meaningful comparison, bsMVPC accuracies in ventral temporal extrastriate visual cortex for naturalistic, dynamic stimuli were twice that for unpredictable static stimuli (Supplemental Figure 4A in **Haxby et al., 2011**), indicating a substantial advantage of these stimuli for evoking distinctive brain states.

A common model space also can be built by estimating transformation matrices from fMRI data collected in the resting state (rs-fMRI) using connectivity hyperalignment (**Guntupalli et al., 2016**). Such a model greatly increases intersubject correlations of functional connectivity indices and captures the fine-scale spatial structure of variation in connectivity that is seen in individual data. The variety of brain states that are sampled in resting state paradigms is difficult to determine. The general validity of a common model based on rs-fMRI data for aligning brain states associated with a wide variety of stimuli and cognitive processes has not yet been established.

A new conceptual basis for cortical functional topographies

Previous models of cortical functional architecture have focused on the division of cortex into contiguous functional areas (**Brodmann, 1909; Glasser et al., 2016; Gordon et al., 2016; Shen et al., 2013; Shen et al., 2010; Yeo et al., 2011; Zilles and Amunts, 2010**). These parcellations, however, may vary substantially across functional brain state (**Salehi et al., 2020**) and individuals (**Braga and**

Buckner, 2017; Gordon et al., 2017; Kong et al., 2019. The conceptual framework for hyperalignment proposes a radically different basis for cortical topographies that is based on overlapping topographic basis functions, offering a model that accounts for individual variation in both fine-scale topographies and coarse-scale areal structure.

Individual cortical topography as a composite of common model pattern bases

Hyperalignment transformation matrices provide the basis for a new model for the organization of cortical topographies. Each column in an individual transformation matrix (R_i) (or row in the transpose of that matrix, R_i^T), is a pattern vector of weights for cortical loci in an individual subject's brain. These pattern vectors for model dimensions serve as topographic basis functions. A cortical response or connectivity pattern in an individual brain ($t_{p(B_i)}$) is modeled as a weighted sum of these basis functions, using the dimension values for the pattern vector in the model group data matrix ($t_{p(M)}$) as weights:

$$\hat{t}_p(B_i) = t_{p(M)} R_i^T \quad (2)$$

Thus, the topographies of cortical patterns are modeled as overlaid, weighted topographic basis functions that are individual-specific and formalized as the columns in transformation matrices (**Figure 4**). This framework captures how shared information from the common model space can be instantiated in individually-variable cortical topographies using a computationally-specified linear model with basis functions that can be estimated using hyperalignment.

Modeling response patterns and known topographies

This conceptual framework affords estimating the location of functional areas and the conformation of functional topographies in a given subject based on data from other subjects. The model constructed by hyperalignment allows us to transfer fine-scale cortical functional topographies across individuals via the common information space. Patterns that specify an area or topographic pattern, such as a face-responsive area or retinotopy, in other subjects' cortex ($t_{p(B)}$) can be transformed into a group mean pattern in the common model information space ($t_{p(M)}$) using their transformation matrices:

$$t_{p(M)} = (1/N) \sum_{i=1}^N t_{p(B_i)} R_i \quad (3)$$

The mean topographic pattern in the model space ($t_{p(M)}$) can then be projected into a new individual's cortical space using that individual's transformation matrix and **Equation 2**, providing an estimate of that pattern ($\hat{t}_{p(B_i)}$) in that individual's brain.

Evaluating the efficacy of hyperalignment

Assessing the performance of hyperalignment for modeling shared information and idiosyncratic topographies requires testing on data that play no role in calculating the parameters in the transformation matrices (**Kriegeskorte et al., 2009**). Generally appropriate cross-validation is achieved by dividing the data for deriving hyperalignment – usually movie-viewing or resting state data – into training and test folds. If test data are from an unrelated experiment, the full movie-viewing dataset can be used for calculating hyperalignment parameters.

Three ways to assess shared information: Patterns, profiles, and geometry

Hyperalignment of individual data spaces defined by cortical loci into a common model space is designed to maximize intersubject similarity of information content as represented by data matrices (**Equation 1**). Similarity can be indexed in terms of intersubject similarity of patterns for time-points, stimuli, or connectivity targets (rows in the data matrices), intersubject similarity of functional profiles (columns in the data matrices), or intersubject similarity of vector geometry (covariance or dissimilarities between rows in the data matrices; see **Nastase et al., 2019**, for a recent review). Numerous

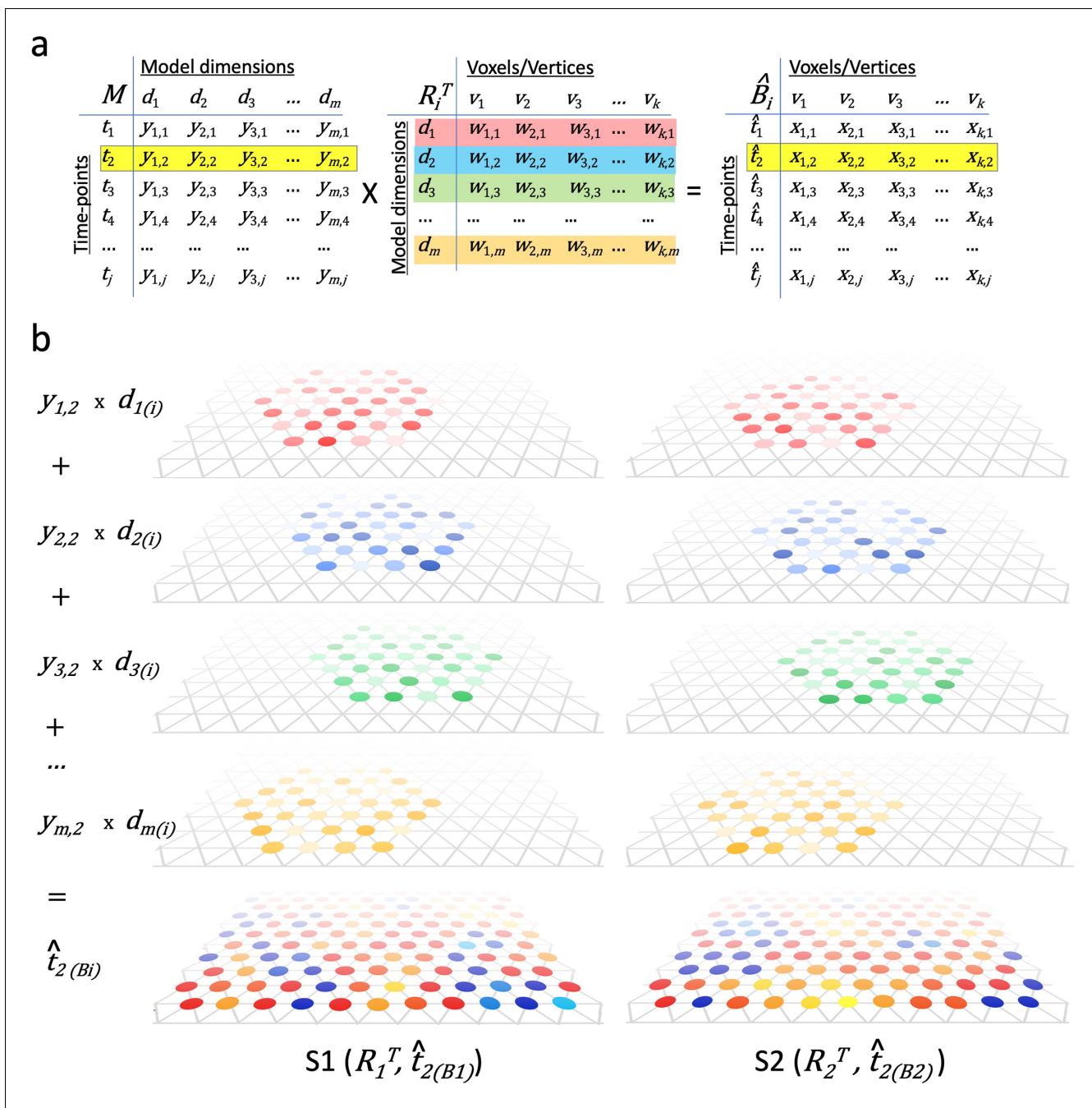


Figure 4. Modeling individual cortical patterns using model topographic basis functions. (a) Matrix multiplication for modeling cortical patterns in individual brains ($\hat{t}_{2(Bi)}$) by multiplying a pattern vector (t_2) in the model space (M) and the transpose of individual-specific transformation matrices (R_i^T). (b) Illustration of topographic basis functions in two subjects, S1 and S2, for model dimensions ($d_1, d_2, d_3, \dots, d_m$), which are comprised of weights (w 's in R_i^T) across vertices in overlapping but non-coextensive patches of cortex. This image is illustrative and not based on experimental data. Two hypothetical subjects' basis functions are illustrated to emphasize that these functions are individual-specific and support the same shared information from M in variable topographic patterns in individual brains. Transformation matrix weights (w) are illustrated as colored circles of varying intensity at vertices of a stylized triangular cortical mesh. The pattern of weights for each basis function varies across brains. These topographic basis functions are combined as a weighted sum, using the weights from a pattern vector (e.g., t_2) in the model data matrix ($y_{1,2}, y_{2,2}, y_{3,2}, \dots, y_{m,2}$ in M) to model or predict a topographic pattern ($\hat{t}_{2(Bi)}$) in an individual subject's cortex. The same weights in the model pattern vector predict different patterns of response in each individual brain. The predicted topographic patterns ($\hat{t}_{2(Bi)}$) are illustrated as values at each vertex of a triangular cortical mesh using a typical color scale with negative values in shades of blue and positive values in shades of red, orange, and yellow.

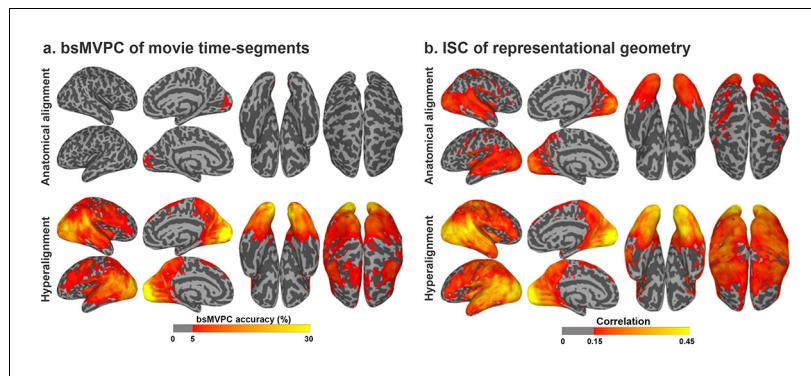


Figure 5. Enhancement of between-subject multivariate pattern classification (bsMVPC) and intersubject correlation (ISC) of representational geometry with hyperalignment. (a) Accuracy of bsMVPC of movie time segments from anatomically-aligned and hyperaligned data (reproduced from Figure 2, [Guntupalli et al., 2016](#)). Chance performance is less than 0.1%. (b) ISC of representational geometry for movie time points (reproduced from Figure 3, [Guntupalli et al., 2016](#)). Representational geometry for a movie is calculated as the matrix of pairwise similarities between responses to different time-points (>1300 time-points; >800,000 pairs).

tests with different data sets have shown marked and robust improvement of intersubject similarity on all of these indices (e.g., [Chen et al., 2015](#); [Guntupalli et al., 2018](#); [Guntupalli et al., 2016](#); [Haxby et al., 2011](#); [Nastase et al., 2017](#); [Van Uden et al., 2018](#)), confirming that the common model space captures shared information across individuals. Here, we present a sampling of these results to convey the size of these effects and the general validity of models built with hyperalignment.

Between subject multivariate pattern classification (bsMVPC)

The goal of hyperalignment is to find transformations that cluster different subjects' pattern vectors more tightly, thus making each individual subject's vectors more classifiable based on other subjects' vectors. In one test, which is particularly demanding and dependent on fine-scale topographies, response patterns for short time-segments of a movie are classified relative to other time-segments of the same movie ([Figure 5a](#); [Chen et al., 2015](#); [Guntupalli et al., 2018](#); [Guntupalli et al., 2016](#); [Haxby et al., 2011](#); [Xu et al., 2012](#)). The magnitude of the beneficial effect of hyperalignment on bsMVPC is largest for small and moderately-sized ROIs which contain fine-scale topographies with minimal coarse-scale variation, where bsMVPC accuracies for anatomically-aligned data are very low. In small cortical regions, where accuracies for unsmoothed, anatomically aligned data are very low, hyperalignment provides more than a full order of magnitude improvement in bsMVPC ([Guntupalli et al., 2018](#); [Guntupalli et al., 2016](#)). Accuracies are higher in larger, more heterogeneous regions for both anatomically and hyperaligned data, but hyperalignment still yields markedly higher bsMVPC accuracies ([Guntupalli et al., 2016](#); [Haxby et al., 2011](#)), highlighting the role of information encoded in fine-scale structure that cannot be captured by anatomical alignment, and the importance of aligning this fine-scale information in building MVP models that generalize across subjects.

Intersubject correlations (ISCs) of response and connectivity profiles

The functional profiles for anatomically-aligned cortical loci (vertices) and hyperaligned model dimensions also can be evaluated for between-subject consistency ([Guntupalli et al., 2018](#); [Guntupalli et al., 2016](#); [Hasson et al., 2004](#)). Modeling response and connectivity profiles using hyperalignment markedly increases ISCs of both ([Guntupalli et al., 2018](#); [Guntupalli et al., 2016](#)). For example, response hyperalignment nearly doubled ISCs of time-series responses to a naturalistic movie ([Guntupalli et al., 2016](#)), and ISC increases were even greater for functional connectivity profiles after connectivity hyperalignment, for both functional connectivities measured during movie viewing and during the resting state ([Figure 6](#); see [Guntupalli et al., 2018](#), for details).

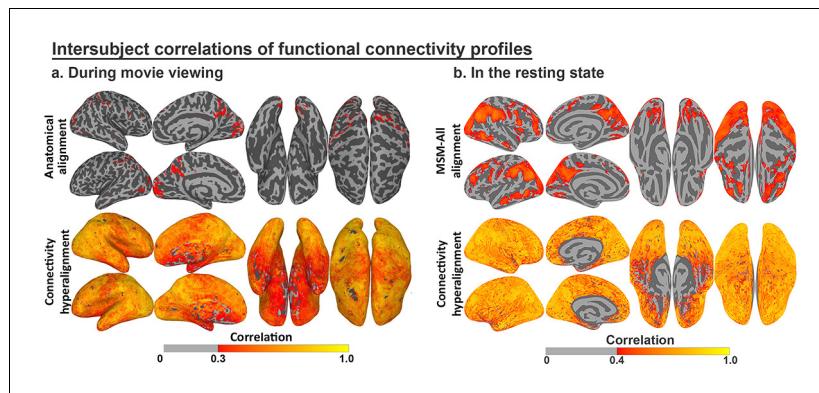


Figure 6. Enhancement of intersubject correlation (ISC) of functional connectivity profiles after connectivity hyperalignment. (a) ISC of connectivity profiles measured during movie viewing (reproduced from Figure 3, [Guntupalli et al., 2018](#)). (b) ISC of connectivity profiles measured in the resting state (reproduced from Figure 4, [Guntupalli et al., 2018](#)).

Vector geometry

Vector geometry provides an integrative measure of information in an information space in the form of all pairwise similarities, or distances, between pattern vectors ([Diedrichsen and Kriegeskorte, 2017](#); [Haxby et al., 2014](#); [Kriegeskorte et al., 2008a](#); [Kriegeskorte and Kievit, 2013](#)). For response data matrices, vector geometry reflects representational geometry – the similarities among response patterns to different stimuli or conditions. For connectivity data matrices, vector geometry reflects connectivity geometry – the similarities among connectivity patterns for a set of connectivity targets. In addition to its incorporation of information from a full data matrix, between-subject similarity of vector geometry explicitly captures information that is not reflected in between-subject similarity of cortical patterns or between-subject similarity of response or connectivity profiles ([Kriegeskorte et al., 2008a](#); [Kriegeskorte et al., 2008b](#)). Comparison of vector geometries for different cortical fields also provides a window on how information spaces are transformed along processing pathways, enhancing relevant distinctions and diminishing confounding information ([Connolly et al., 2016](#); [Guntupalli et al., 2017](#); [Kriegeskorte and Kievit, 2013](#); [Nastase et al., 2017](#)).

Searchlight-based whole-cortex hyperalignment greatly increases ISCs of local representational geometry ([Figure 5b](#); [Guntupalli et al., 2018](#); [Guntupalli et al., 2016](#); [Nastase et al., 2017](#); [Van Uden et al., 2018](#)). For naturalistic movies, hyperalignment increased mean ISC of representational geometry across searchlights by 46% to 75% ([Guntupalli et al., 2016](#); [Van Uden et al., 2018](#)). Simpler representational geometries in an experiment with 20 stimulus conditions showed similar increases in ISCs after data were hyperaligned based on responses to an unrelated movie ([Nastase et al., 2017](#)).

Increased ISC of vector geometry is counterintuitive and, therefore, instructive about cortical functional architecture. Procrustes-based hyperalignment preserves local vector geometry with rigid rotation of the high-dimensional pattern vectors. Indeed, ROI hyperalignment does not change the vector geometry for each individual and, therefore, has no effect on ISC of vector geometry. Searchlight hyperalignment, by contrast, resamples information into and out of local cortical fields, thereby increasing the similarity of local information content across subjects. The size of the increase suggests this is a large effect and that the dissimilarity of vector geometry for overlapping cortical fields is substantial. RSA of anatomically-aligned data, therefore, does not capture most of the shared fine-scale variation in local vector geometry of cortical fields along processing pathways that may be a key to understanding how information is processed and transformed.

General validity of common model information spaces

The shared functional profiles and preserved vector geometry in common model information spaces have a general validity that extends beyond the specific stimuli used to derive them. All of the validation tests described in the previous sections derived the common model spaces and

transformation matrices using data that were independent of the data used to test bsMVPC or ISC of response profiles, connectivity profiles, and vector geometry (Kriegeskorte et al., 2009). For movie data, this involved using data from different parts of the movie for model derivation and validation. For the resting state, model derivation and validation were performed on different resting state sessions. The generalizability of the validity of common model spaces based on naturalistic movie viewing and the resting state also extends to experiments with very different structure. For example, bsMVPC of responses to still images of objects, faces, and animals after data were projected into a common space based on responses to a movie was equivalent to or better than within-subject MVPC (Guntupalli et al., 2016; Haxby et al., 2011). Higher bsMVPC than within-subject MVPC of hyperaligned data demonstrated the added power of using large multi-subject data sets for training a pattern classifier, which hyperalignment makes possible. Similarly, ISC of representational geometry for responses to brief video clips of different categories of animals performing different actions were greatly increased after data were projected into a common space based on responses to a continuous movie (Nastase et al., 2017). Moreover, connectivity hyperalignment now makes it possible to derive transformation matrices to project new subjects' cortical space into a model space using connectivity data from experimental paradigms (e.g., a movie or the resting state) that differ from the experimental paradigm used to derive that common model space (Guntupalli et al., 2018; Nastase et al., 2020). In a novel application (Taschereau-Dumouchel et al., 2018), hyperalignment made it possible to create a response pattern template based on other subjects' hyperaligned data for use in new subjects who were never exposed to the stimulus as a target brain state for neurofeedback.

Generalization across response and connectivity hyperalignment models

Response and connectivity hyperalignment both produce transformation matrices that project data from cortical coordinate spaces into a common model information space. The two algorithms align the vector geometries for related, but distinct types of information – patterns of responses and patterns of connectivity. Both algorithms produce transformations that can be used to align either type of data (Guntupalli et al., 2018), but the common model spaces, while closely related, are not equivalent (Table 1). Response hyperalignment of movie data increased mean intersubject correlation of connectivity profiles from 0.15 to 0.58, relative to anatomical alignment, but ISC of connectivity profiles are even higher after connectivity hyperalignment (mean ISC = 0.67). Connectivity hyperalignment of the same data increased mean intersubject correlation of representational geometry from 0.21 to 0.31 and bsMVPC of movie time segments from 1.0% to 10.4%, relative to anatomical alignment. Response hyperalignment, however, produced even higher ISC of representational geometry (mean ISC = 0.32) and bsMVPC accuracies (mean = 13.7%). Thus, aligning shared representational vector geometry also aligns connectivity vector geometry, but this response-based hyperalignment of connectivity, and the underlying group mean connectivity basis functions associated with model dimensions, is suboptimal. Similarly, aligning shared connectivity vector geometry also aligns representational vector geometry, but connectivity-based representational alignment, and the underlying group mean response tuning basis functions associated with model dimensions, is suboptimal. At this point it is unclear if this discrepancy between models of shared population codes and shared connectivity is fundamental or will converge to a common solution with better data and algorithms.

Table 1. Summary of improvement in indices of performance after response hyperalignment (RHA) and connectivity hyperalignment (CHA), relative to anatomical alignment (AA).

ISC: Intersubject correlation; bsMVPC: between subject multivariate pattern classification; >: small advantage; >>: large advantage.

Performance measure	Relative performance after different alignments
ISC of searchlight representational geometry	RHA > CHA >> AA
bsMVPC of movie time segments	RHA > CHA >> AA
ISC of vertex connectivity profiles	CHA > RHA >> AA

Estimating individual cortical topographies

Individual-specific functional topographies, the locations and borders of category-selective regions in ventral temporal cortex, and the retinotopic topography of early visual cortex can be estimated with high fidelity based on other subjects' topographies with hyperalignment (**Figures 7** and **8**; **Guntupalli et al., 2016; Haxby et al., 2011; Jiahui et al., 2019**).

Overlaid local topographic basis functions and areal boundaries

The model coordinate axes can be rotated arbitrarily to model individual topographies while preserving vector geometry. Modeling functional areal boundaries is essentially a rotation of coordinate axes in common space to align a contrast or continuum that defines a functional area (e.g., faces versus objects linear discriminant) or topography (e.g., retinotopic polar angle) with a single axis, then plotting the weights of the topographic bases for that axis. Aligning common model coordinate axes to well-understood functional dimensions can make the model more interpretable, but how much of the information that is encoded in fine-scale topographies is captured by such dimensions?

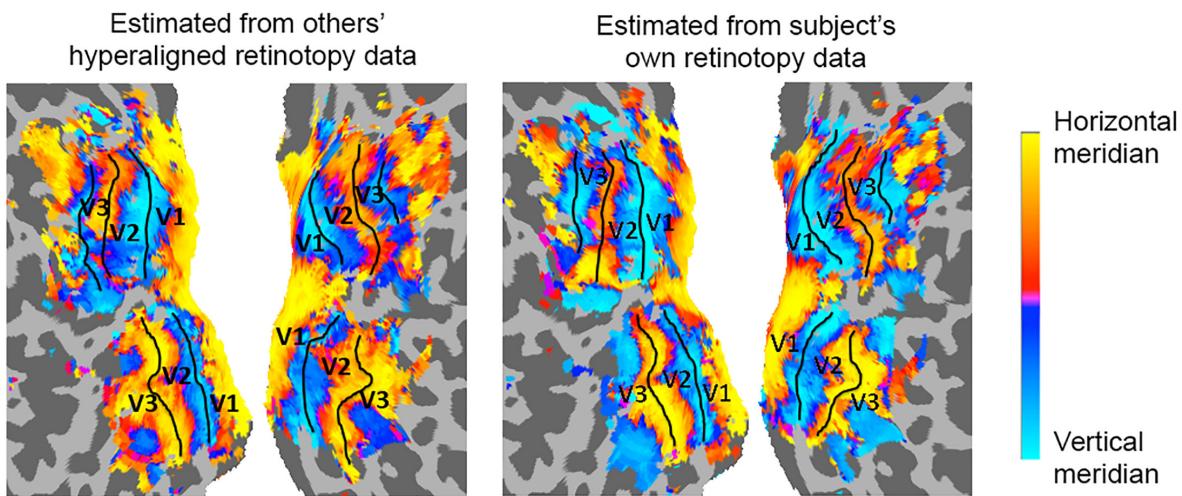
Principal components analysis affords quantification of how much information in a dataset can be accounted for by single dimensions. The first eigenvector is defined as the axis along which variance is maximal, the second eigenvector is the orthogonal axis along which the residual variance is maximal, and so forth. Thus, PCA rotates the space to account for variance most efficiently, rather than rotating the space to fit axes to a priori functional distinctions. Applying PCA to the model space for ventral temporal cortex, developed based on patterns of activity evoked by a naturalistic movie, revealed that approximately 35 principal components (PCs) were sufficient to account for the information content of one hour of the movie, as indexed by bsMVPC (**Haxby et al., 2011**). This PCA allows evaluation of the roles played by dimensions defined a priori by functional distinctions relative to the information that is captured by the full model in which both coarse-grained and fine-grained topographies can be modeled as mixtures of topographic basis functions. Comparison of the topographies for the top PCs with the topography for face versus object category selective areas (**Figure 8a**), shows that this functional distinction does not emerge as one of the top dimensions but only as a mixture of PCs (**Figure 8a** rightmost images), raising the question of how large a role it plays in the functional topography of ventral temporal cortex under naturalistic conditions.

We defined an a priori dimension with the contrast between responses to faces and objects (**Haxby et al., 2011**), calculated as a linear discriminant in the model space (**Figure 8a**). Response variance for a dynamic, naturalistic movie along that dimension comprised 7% of the total variance and 12% of the variance accounted for by the 35-dimensional model (58% of the total variance in movie response data was accounted for by all 35 dimensions, cross-validated on independent data; **Figure 8b**). By contrast, the first PC accounted for over three times more variance (22%, cross-validated), and the first six PCs together accounted for over six times more variance. In a second analysis, the contrast between responses to still images of primates and insects (**Connolly et al., 2012; Haxby et al., 2011**) was calculated as a linear discriminant in model space. This dimension captures the animacy continuum and is mostly collinear with the faces versus objects dimension. The mean correlation of movie time-series for these two dimensions is 0.83, and the mean correlation of their topographies in ventral temporal cortex is 0.77. Although this animacy continuum dimension accounts for over 50% of variance in the responses to still images of animals and objects (**Sha et al., 2015**), it accounts for only 6% of variance in responses to the movie. Thus, some of the most intensively studied dimensions in the coarse-scale functional topography of ventral temporal cortex (**Connolly et al., 2012; Epstein and Kanwisher, 1998; Grill-Spector and Weiner, 2014; Kanwisher et al., 1997; Khaligh-Razavi and Kriegeskorte, 2014; Kriegeskorte et al., 2008b; Sha et al., 2015; Thorat et al., 2019**) are faithfully represented in the common model space (see **Figure 7b**) but play a relatively minor role in the more comprehensive, high-dimensional model of response tuning to dynamic, naturalistic stimuli and coarse- and fine-scale functional topographies.

Spatial granularity of the common model

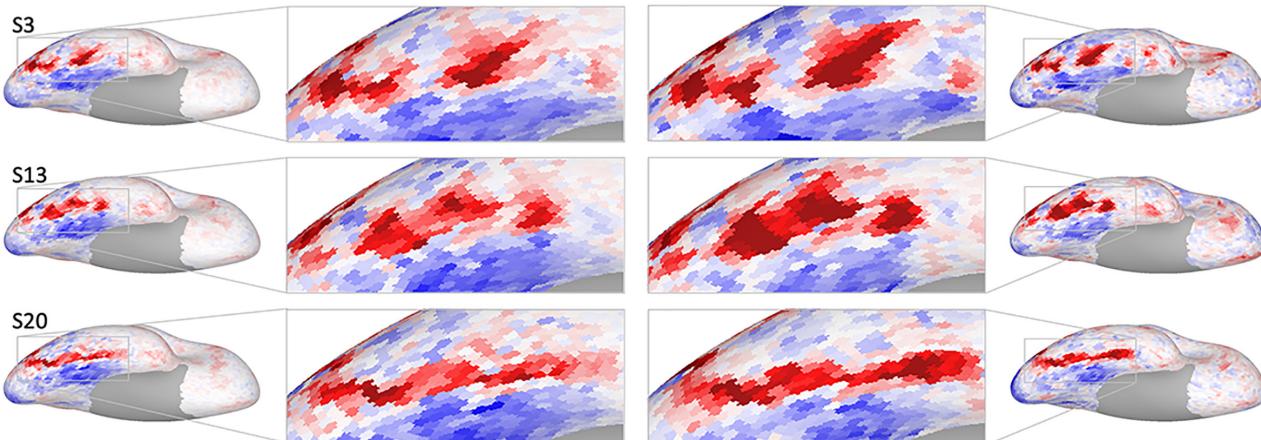
The common model accounts for individual functional topographies as mixtures of individual-specific topographic basis functions. The results showing bsMVPC accuracies equivalent to or better than within-subject MVPC indicate that these mixtures preserve the fine-grained patterns that carry information. Analysis of the spatial point-spread function (PSF) for response and connectivity profiles

a. Retinotopy - polar angle map



b. Face-selective topography

Estimated from others' hyperaligned localizer data Estimated from subject's own localizer data



Mean searchlight correlation between estimated topographies by searchlight

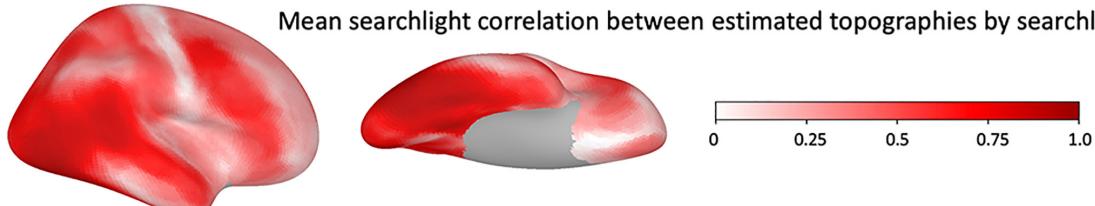


Figure 7. Estimation of functional topographies in individual brains based on data from other brains that are projected into an individual's cortical space using transformation matrices calculated with hyperalignment based on data gathered while subjects watched a movie. (a) Retinotopic map of polar angle based on other subjects' retinotopy localizer fMRI data projected into a subject's visual cortex (left) and based on that subject's own retinotopy localizer fMRI data (adapted from Figure 7, [Guntupalli et al., 2016](#)). (b) Face-selective topography based on other subjects' face-selectivity localizer fMRI data projected into three subjects' cortical anatomy using hyperalignment and based on their own localizer data (from [Jiahui et al., 2019](#)). Ventral views of the right hemisphere with enlarged images of occipital and posterior temporal cortices. Local correlations in searchlights between estimates based on others subjects' data and estimates based on own data exceeded 0.80 in posterior ventral temporal, lateral occipital, and posterior superior temporal cortices and strong correlations extend into anterior ventral temporal, anterior superior temporal sulcal and lateral prefrontal cortices (lower image). Subject numbers match those in Supplementary Figure 2 in [Jiahui et al., 2019](#).

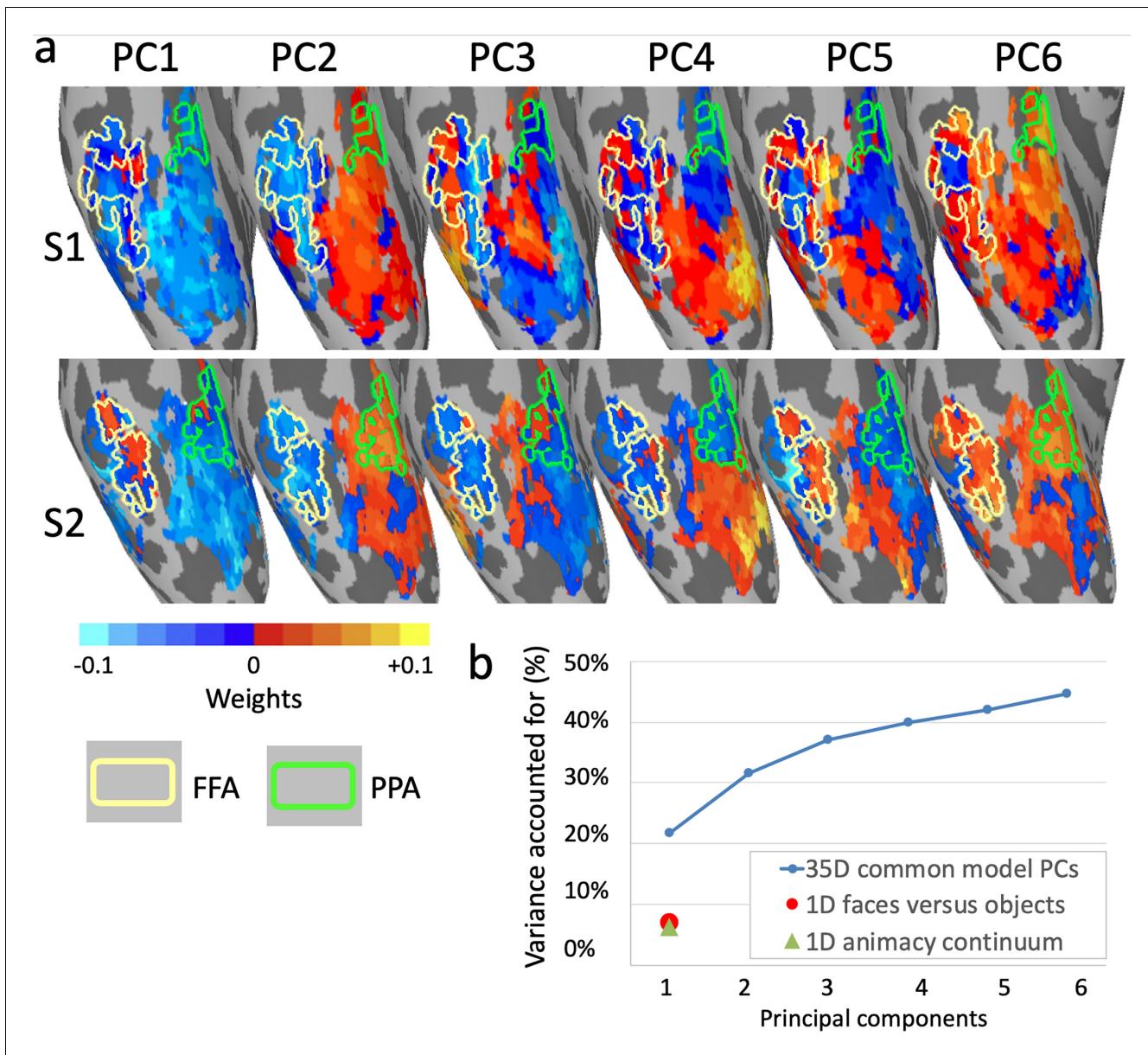


Figure 8. Topographic basis functions and comparison to one-dimensional models of functional topography. (a) Topographic basis functions for two subjects, S1 and S2, for the top six principal components (PCs) in a common model of ventral temporal cortex and the linear discriminant, a weighted sum of PCs, for the contrast faces versus objects (from Haxby et al., 2011; only the right hemisphere cortex is shown here). The locations of the fusiform face area (FFA) and parahippocampal place area (PPA) are indicated with yellow and green outlines, respectively. (b) Cumulative variance accounted for by the top six PCs in a 35-dimensional common model compared to the variance accounted for by one-dimensional linear discriminants for face selectivity (responses to faces versus houses) and for the animacy continuum (responses to primates versus insects).

provides a more direct measure of the preservation of the granularity of cortical topographies (Guntupalli et al., 2018; Guntupalli et al., 2016).

Spatial PSF analysis measures the intersubject correlations of response and connectivity profiles for a given cortical vertex and its neighboring vertices that differ in anatomical location by varying distances. For anatomically aligned data, intersubject correlations of profiles for neighboring loci are only slightly lower than correlations of profiles for loci at the same location. For data projected from the common model into one subject's anatomy, the function is dramatically steeper with large differences in intersubject correlation of profiles from the same location and correlations of profiles from contiguous locations (Figure 9), indicating that the common model preserves the distinctiveness of

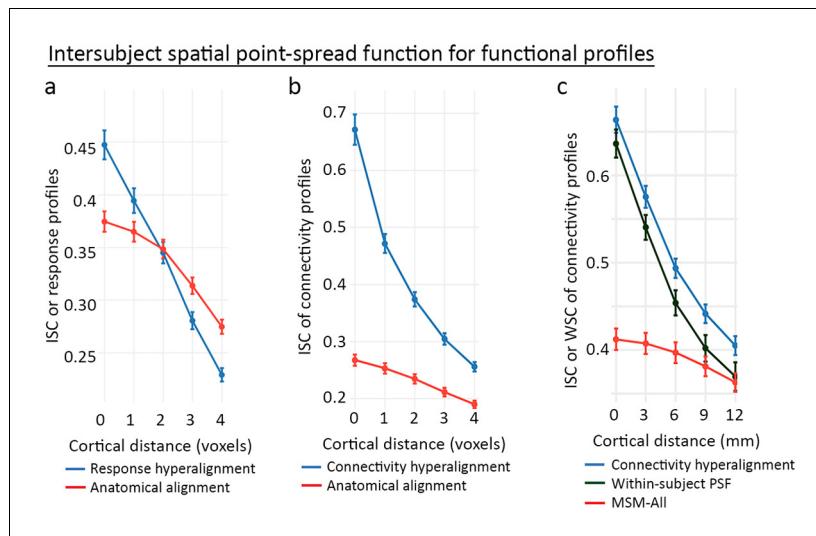


Figure 9. Spatial point-spread functions (PSFs) for intersubject correlations of time-series response profiles and connectivity profiles. A greater negative slope indicates more spatially fine-grained functional selectivity. (a) Spatial PSFs for movie data response profiles (reproduced from Figure 5, [Guntupalli et al., 2016](#)). (b) Spatial PSFs for movie data connectivity profiles and c. resting state fMRI connectivity profiles with comparison to within-subject PSFs based on within-subject correlations between independent resting state sessions. Error bars are standard errors (SE) across participants, in which the value for each participant is the mean across searchlight ROIs (20 ROIs in a and 24 ROIs in b and c). MSM-All alignment is multimodal surface matching ([Robinson et al., 2014](#)). (Reproduced from Figure 5, [Guntupalli et al., 2018](#)).

functional profiles with a spatial granularity of a single voxel or vertex (~3 mm). This level of spatial granularity was found for all cortical fields tested in occipital, temporal, parietal, and prefrontal cortices. A within-subject PSF for connectivity profiles calculated using resting-state fMRI data from independent sessions in the same subject was equivalent to the between-subject PSF. This fine-grained structure in the connectome was not recognized prior to the development of a hyperaligned model of the human connectome.

Beyond shared information: Individual differences in information encoded in fine-scale neural architecture

The goal of hyperalignment is to model the shared information encoded in idiosyncratic cortical topographies, but brains also differ in the information that they encode and process ([Dubois and Adolphs, 2016](#)). To investigate individual differences in encoded information, we calculated the intersubject correlations of the concatenated functional profiles for all vertices in a region for anatomically-aligned and hyperaligned data ([Feilong et al., 2018; Figure 10a](#)). From these pairwise comparisons we constructed a matrix of dissimilarities between all pairs of brains – an individual differences matrix (IDM). IDMs can be constructed based on response patterns or on functional connectivity patterns. Surprisingly, even though hyperalignment makes individual differences much smaller, as expected, the residual differences are substantially more reliable across independent data samples than those based on anatomically aligned data, and this increased reliability can be attributed to individual differences in the information encoded in fine-scale topographies ([Figure 10b](#)). The reliabilities of individual differences based on coarse-scale topographies are effectively equivalent for anatomically-aligned and hyperaligned data, whereas the individual differences based on residual data in fine-scale topographies are markedly more reliable for hyperaligned data than for anatomically-aligned data. The individual differences in the information encoded in fine-scale topographies also are more predictive of cognitive differences, such as general intelligence ([Feilong et al., 2019](#)). Previous work using anatomically-aligned connectomes has investigated only individual differences in coarse scale, inter-areal connectivity ([Braga and Buckner, 2017; Dubois et al., 2018; Finn et al., 2015; Gordon et al., 2017; Kong et al., 2019](#)). Restricting analysis

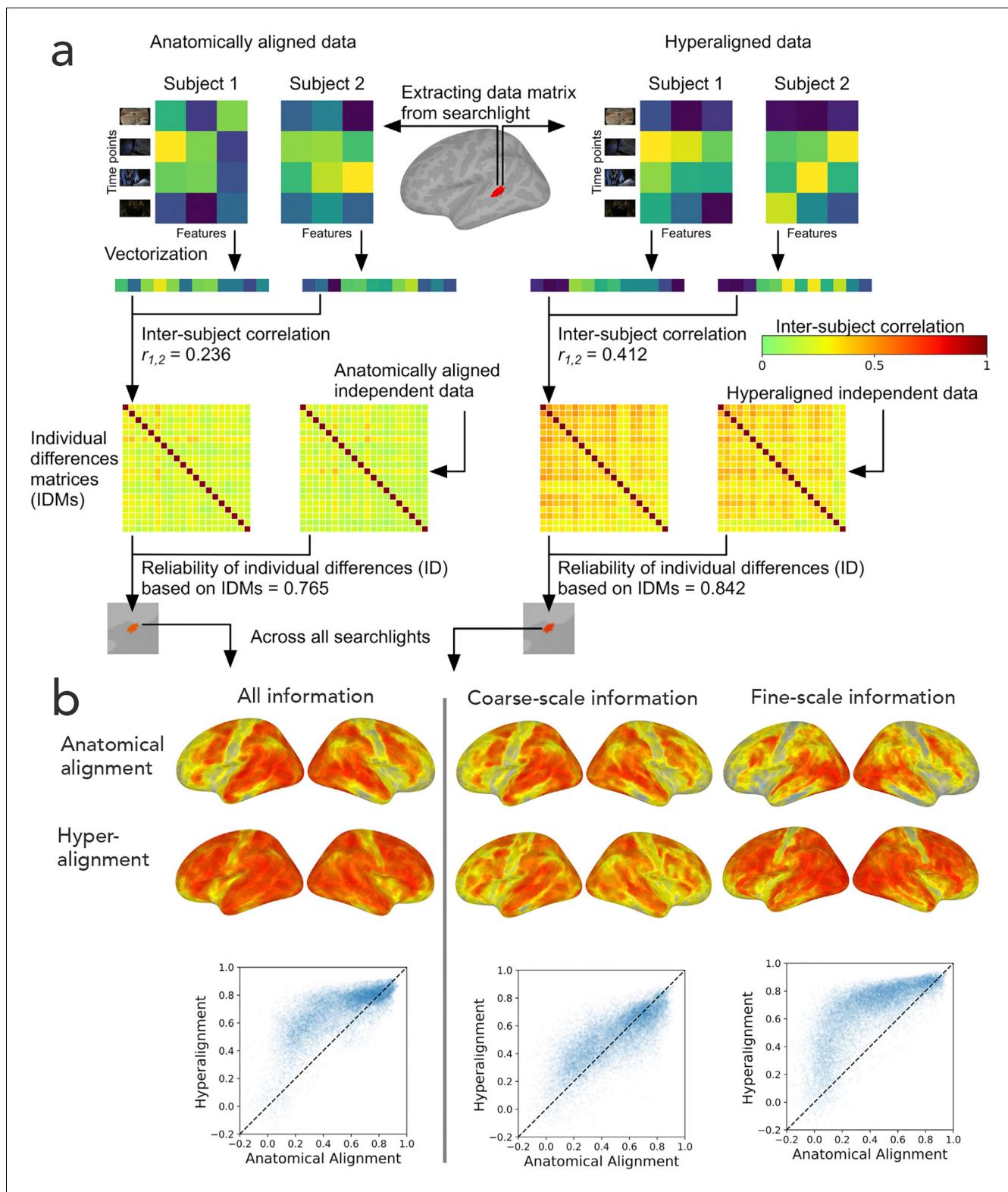


Figure 10. Reliability of individual differences in the information encoded in cortical topographies. (a) Response or functional connectivity data across vertices in a searchlight are vectorized for each individual, and the correlation between individuals is used as an index of similarity. Correlations between all pairs of individual brains are assembled in individual differences matrices (IDMs). Correlations between IDMs constructed from independent data sets are indices of the reliability of individual differences. After hyperalignment, individuals' data are more similar, but the residual differences are more reliable than those calculated from anatomically aligned data, as shown for one searchlight in a and for 82% of searchlights as Figure 10 continued on next page

Figure 10 continued

shown on the left in (b). The increase in reliability is due primarily to information encoded in fine-scale patterns, as shown on the right in b. Adapted from Figure 1 and 3, *Feilong et al., 2018*.

to differences in coarse-scale connectivity is necessary with anatomically-aligned (or MSM-all aligned) data because differences in fine-scale connectivity are obscured by misalignment, but it appears that the most informative individual differences in cortical function reside in fine-scale patterns. Hyperalignment affords investigation of differences at this level of organization and factors out confounding variability due to topographic misalignment.

Related methods for modeling shared information encoded in neural activity

Other approaches to modeling shared information in cortical topographies have mostly overlooked information encoded in fine-scale topographic patterns. Most have focused on areal structure (*Gordon et al., 2017; Langs et al., 2016; Langs et al., 2014; Yeo et al., 2011*) and describe the information in each area or network of connected areas in terms of a more global function, such as motor production, vision, attention, working memory, or theory of mind.

Other approaches decompose tuning profiles using analyses such as ICA, PCA, or factor analysis. These approaches have the potential to capture fine spatial scale variation in voxel-to-voxel differences in the mixture of components. Such variation, however, is often discarded because it is blurred in group analyses using anatomically aligned data (e.g. *Filippini et al., 2009*), or no attempt is made to model local topographies and the number of components is too limited to capture fine-scale structure (*Just et al., 2010; Langs et al., 2016; Langs et al., 2014*). Canonical Correlation Analysis (CCA) offers another method for finding shared information in two multivariate datasets (*McIntosh and Mišić, 2013; Wang et al., 2020*) but is not constrained to preserve vector geometry. Consequently, attempts to use CCA as a basis for hyperalignment have underperformed methods that do preserve vector geometry, unless CCA is regularized to make transformations mostly orthogonal (*Xu et al., 2012*).

Representational similarity analysis (RSA; *Kriegeskorte et al., 2008a; Kriegeskorte and Kievit, 2013*) and pattern component modeling (*Diedrichsen and Kriegeskorte, 2017*) are closely related to hyperalignment. These approaches focus on the vector geometry of response patterns in cortical areas as the common basis for representation of information that is shared across subjects and reveals individual differences (*Charest et al., 2014*). These approaches treat the topographies that embed these geometries as unimportant and, consequently, do not offer an explicit method for modeling the idiosyncrasies in local fine-scale topographies or a principled approach to rearrange information in the cortical sheet to enhance the inter-subject similarity of representational geometry in a cortical field.

The forward encoding approach models each voxel's response profile independently as a weighted sum of stimulus features and, thereby, captures fine spatial-scale variation (*Kay et al., 2008; Mitchell et al., 2008; Naselaris et al., 2011; Naselaris et al., 2009; Nishimoto et al., 2011*). The great strength of this approach is that it attempts to define the information that is shared in terms of features, such as visual, acoustic, phonological, syntactic, and semantic features, thus specifying the content of shared information and affording prediction of responses to new stimuli if they can be modeled as a mixture of the features in the model. Hyperalignment and related methods afford improved cross-subject voxelwise prediction of response profiles (*Bilenko and Gallant, 2016; Güçlü and van Gerven, 2017; Nastase et al., 2020; Van Uden et al., 2018*). Forward encoding pre-defines the shared information by the stimulus feature model, and its validity is dependent on the validity and strength of that feature model. It does not provide any principled way to discover the structure of information that is encoded in neural responses that is not captured by the predefined feature model. These approaches also do not attempt to investigate neural responses as population codes with a representational geometry, but simply as independent voxel-wise responses to stimulus features. Voxel-wise response mapping studies tend to describe the topographies that result as gradients (*Huth et al., 2016; Huth et al., 2012*) with no attempt to model other aspects of fine-scale

structure or their idiosyncratic variation across brains. The information space is typically reduced to a small number of dimensions (four in *Huth et al., 2012*).

Discussion

The conceptual framework that we review uses hyperalignment to model cortical function by aligning the information represented in population responses in a common, high-dimensional information space. This framework does not align the cortical topographic patterns in a canonical anatomical topography. Information is defined computationally as the distances between population response vectors, which index reliable distinctions and similarities. By aligning response vectors from different subjects' idiosyncratic topographies into a single information space with shared coordinate axes, information – the distinctions and similarities among response vectors – is now supported by the same basis functions across brains.

This computational framework can also be applied to functional connectivity data, replacing population response pattern vectors with topographic patterns of connectivities to a target, essentially a population connectivity vector. By aligning population connectivity vectors in a common information space, the connectivity profiles for each dimension in this space are markedly more similar across brains than are connectivity profiles for anatomically-aligned cortical loci. Not surprisingly, transformation matrices based on connectivity hyperalignment also align response pattern vectors, affording improved bsMVPC and higher ISCs of representational geometry.

Hyperalignment also models the idiosyncratic topographies in individual brains that carry shared information. Using the weights in individual transformation matrices for cortical loci as basis functions, a pattern vector in the common model space predicts a different topographic pattern in each brain. Boundaries for functional regions can be estimated in individual brains with high fidelity based on other subjects' data, but the one-dimensional linear discriminants that define these regions account for only a small portion of the variance in cortical responses to naturalistic stimuli, indicating that they play a minor role. Functional topographies are modeled more comprehensively by our model, which posits a high-dimensional set of overlapping topographic basis functions.

Rebalancing the roles of information content in neural populations and the functional profiles of individual units

The conceptual framework for hyperalignment is predicated on the hypothesis that information content that is shared across brains is instantiated in varied, idiosyncratic population codes. Hyperalignment models shared information by preserving the interrelations among pattern vectors – vector geometry – in a common information space while remixing the functional profiles of constituent dimensions – response tuning and functional connectivity profiles – thereby reshaping the functional profiles of the units in cortical patterns. Thus, the functional profiles of individual units are modeled as secondary to the information content that they support as populations. The commonality across brains is not modeled as common functional profiles of the individual units or as a common spatial distribution of functional profiles.

Traditionally, neuroscientists have focused on the tuning and connectivity profiles of individual units – neurons or voxels – and the information in population responses is modeled as a secondary consequence of the aggregation of those profiles (e.g. *Gross et al., 1972; Hubel and Wiesel, 1959; Huth et al., 2016; Huth et al., 2012; Kay et al., 2008; Kiani et al., 2007; Sheinberg and Logothetis, 2001*). The preservation of information content in vector geometry over variations in the functional profiles of individual units, however, suggests that the unit profiles are shaped to support the geometry, not that the vector geometry is merely a reflection of unit profiles.

Understanding how unit profiles can be shaped to support the requisite (and shared) population information is a matter that requires further theoretical work and experimental validation. Unit profiles clearly develop with experience and show a perplexing complexity and variety when considered in isolation (*Çukur et al., 2013; Kiani et al., 2007; Sheinberg and Logothetis, 2001*), especially responses and connectivity in studies with naturalistic stimuli (*Çukur et al., 2013; McMahon et al., 2015; Park et al., 2017*). Tuning functions of varying levels of coarseness – responses to related stimuli – support similarity of population responses and connectivity patterns (*Kiani et al., 2007*) and, consequently, pattern vector geometries. Relationships reflected in similarity of responses can be due to shared perceptual features or cognitive factors such as agency (*Connolly et al., 2012*;

Sha et al., 2015; Thorat et al., 2019 and action goals (*Nastase et al., 2017*). The relationships reflected in representational geometry differ by cortical field (*Borghesani et al., 2016; Freiwald and Tsao, 2010; Guntupalli et al., 2017; Guntupalli et al., 2016; Visconti di Oleggio Castello et al., 2017*), and these differences reflect processing or the disentangling of target information from confounds (*DiCarlo et al., 2012; DiCarlo and Cox, 2007; Kriegeskorte and Kievit, 2013*). The topographic organization of units with particular functional profiles may arise over the course of development and with varying experience (*Arcaro et al., 2017*). These functional profiles may in fact fluctuate over time despite retaining the information encoded at the population level (*Gallego et al., 2020*). The functional profiles of individual units need not be easily-interpretable (*Ponce et al., 2019*). Differences in vector geometry across cortical fields in a processing system are supported by different configurations of tuning profiles of individual units. The composition of varied response profiles in populations, therefore, is shaped to create the varied information spaces that are required in a processing system, a learning process that may rest on feedback among the cortical fields in the system. Deep convolutional neural networks (CNNs) provide a computational instantiation of a learning process guided by feedback that produces varied sets of individual units that perform the same functions. Like cortical systems, the individual features of CNNs vary based on initialization and network architecture, but the similarity structure of feature vectors reveals commonality of layer-specific information content across these variations (*Kornblith et al., 2019*).

Cortical functional architecture is a degenerate system

The contrast between a common deep structure for information that is instantiated in variable surface structure has a parallel in language (*Chomsky, 1965; Hockett, 1958*). The same meaning can be conveyed linguistically with substantial variation in wording and sentence structure, including the words and syntax of different languages. Analogous to the variable surface structure of language and the deep structure of meaning in linguistic communication, the variable surface structure of cortical topographies affords representation of shared deep information structure.

More broadly, the conceptual framework of hyperalignment models a property of cortical functional architecture that is a ubiquitous property of complex biological systems, namely ‘degeneracy’ – ‘the ability of elements that are structurally different to perform the same function or yield the same output’ (*Edelman and Gally, 2001*). Degeneracy is a property of systems that evolve or develop or learn and are characterized by complex interactions among constituent elements. By contrast, designed systems usually try to avoid unplanned interactions. Unplanned interactions in degenerate systems allow unexpected new functions to emerge or to be learned. Different instantiations of elements that perform the same or similar functions may be differentially adaptive to changing demands, providing a basis for evolution. Degeneracy is characteristic of systems at all levels of biological organization from genetics to body movements and language. Hyperalignment models how the variable structural elements of functional cortical architecture – local neural responses and point-to-point functional connectivities – perform similar functions in aggregate for representation and processing of information and provides a basis for studying variations in instantiations of these functions, which are the basis for diversity in cognitive abilities.

Additional information

Funding

Funder	Grant reference number	Author
National Science Foundation	1607845	James Haxby
National Science Foundation	1835200	James Haxby

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author ORCIDs

James V Haxby  <https://orcid.org/0000-0002-6558-3118>

J Swaroop Guntupalli  <https://orcid.org/0000-0002-0677-5590>

Samuel A Nastase  <http://orcid.org/0000-0001-7013-5275>
Ma Feilong  <https://orcid.org/0000-0002-6838-3971>

References

- Arcaro MJ, Honey CJ, Mruczek REB, Kastner S, Hasson U. 2015. Widespread correlation patterns of fMRI signal across visual cortex reflect eccentricity organization. *eLife* **4**:e03952. DOI: <https://doi.org/10.7554/eLife.03952>
- Arcaro MJ, Schade PF, Vincent JL, Ponce CR, Livingstone MS. 2017. Seeing faces is necessary for face-domain formation. *Nature Neuroscience* **20**:1404–1412. DOI: <https://doi.org/10.1038/nn.4635>, PMID: 28869581
- Bilenko NY, Gallant JL. 2016. Pyrcca: regularized kernel canonical correlation analysis in Python and its applications to neuroimaging. *Frontiers in Neuroinformatics* **10**:49. DOI: <https://doi.org/10.3389/fninf.2016.00049>, PMID: 27920675
- Borghesani V, Pedregosa F, Buiatti M, Amadon A, Eger E, Piazza M. 2016. Word meaning in the ventral visual path: a perceptual to conceptual gradient of semantic coding. *NeuroImage* **143**:128–140. DOI: <https://doi.org/10.1016/j.neuroimage.2016.08.068>, PMID: 27592809
- Braga RM, Van Dijk KRA, Polimeni JR, Eldaief MC, Buckner RL. 2019. Parallel distributed networks resolved at high resolution reveal close juxtaposition of distinct regions. *Journal of Neurophysiology* **121**:1513–1534. DOI: <https://doi.org/10.1152/jn.00808.2018>, PMID: 30785825
- Braga RM, Buckner RL. 2017. Parallel interdigitated distributed networks within the individual estimated by intrinsic functional connectivity. *Neuron* **95**:457–471. DOI: <https://doi.org/10.1016/j.neuron.2017.06.038>, PMID: 28728026
- Brodmann K. 1909. *Vergleichende Lokalisationslehre Der Grosshirnrinde in Ihren Prinzipien Dargestellt Auf Grund Des Zellenbaues Von Dr. K. Brodmann*. J.A. Barth.
- Bzdok D. 2017. Classical statistics and statistical learning in imaging neuroscience. *Frontiers in Neuroscience* **11**: 543. DOI: <https://doi.org/10.3389/fnins.2017.00543>, PMID: 29056896
- Charest I, Kievit RA, Schmitz TW, Deca D, Kriegeskorte N. 2014. Unique semantic space in the brain of each beholder predicts perceived similarity. *PNAS* **111**:14565–14570. DOI: <https://doi.org/10.1073/pnas.1402594111>, PMID: 25246586
- Chen PH, Guntupalli JS, Haxby JV, Ramadge PJ. 2014. Joint SVD-Hyperalignment for multi-subject fMRI data alignment IEEE International Workshop on Machine Learning for Signal Processing, MLSP Presented at the 2014 24th IEEE International Workshop on Machine Learning for Signal Processing, MLSP 2014. IEEE Computer Society. 6958912 . DOI: <https://doi.org/10.1109/MLSP.2014.6958912>
- Chen P-H, Chen J, Yeshurun Y, Hasson U, Haxby J, Ramadge PJ. 2015. A Reduced-Dimension fMRI shared response model. *Advances in Neural Information Processing Systems* 460–468. <https://papers.nips.cc/paper/5855-a-reduced-dimension-fmri-shared-response-model.pdf>.
- Chen H-T, Manning JR, der M. 2020. Between-subject prediction reveals a shared representational geometry in the rodent Hippocampus. *bioRxiv*. DOI: <https://doi.org/10.1101/2020.01.27.922062>
- Chomsky N. 1965. *Aspects of the Theory of Syntax*. Cambridge: MIT Press.
- Connolly AC, Gobbini MI, Haxby JV. 2011. Kriegeskorte N, Kreiman G (Eds). *Three Virtues of Similarity-Based Multivariate Pattern Analysis: An Example From the Human Object Vision Pathway*. The MIT Press. DOI: <https://doi.org/10.7551/mitpress/8404.003.0016>
- Connolly AC, Guntupalli JS, Gors J, Hanke M, Halchenko YO, Wu YC, Abdi H, Haxby JV. 2012. The representation of biological classes in the human brain. *Journal of Neuroscience* **32**:2608–2618. DOI: <https://doi.org/10.1523/JNEUROSCI.5547-11.2012>, PMID: 22357845
- Connolly AC, Sha L, Guntupalli JS, Oosterhof N, Halchenko YO, Nastase SA, di Oleggio Castello MV, Abdi H, Jobst BC, Gobbini MI, Haxby JV. 2016. How the human brain represents perceived dangerousness or "Predacity" of Animals. *The Journal of Neuroscience* **36**:5373–5384. DOI: <https://doi.org/10.1523/JNEUROSCI.3395-15.2016>, PMID: 27170133
- Conroy BR, Singer BD, Guntupalli JS, Ramadge PJ, Haxby JV. 2013. Inter-subject alignment of human cortical anatomy using functional connectivity. *NeuroImage* **81**:400–411. DOI: <https://doi.org/10.1016/j.neuroimage.2013.05.009>, PMID: 23685161
- Cox DD, Savoy RL. 2003. Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. *NeuroImage* **19**:261–270. DOI: [https://doi.org/10.1016/S1053-8119\(03\)00049-1](https://doi.org/10.1016/S1053-8119(03)00049-1), PMID: 12814577
- Çukur T, Huth AG, Nishimoto S, Gallant JL. 2013. Functional subdomains within human FFA. *Journal of Neuroscience* **33**:16748–16766. DOI: <https://doi.org/10.1523/JNEUROSCI.1259-13.2013>, PMID: 24133276
- Deerwester S, Dumais ST, Furnas GW, Landauer TK, Harshman R. 1990. Indexing by latent semantic analysis. *Journal of the American Society for Information Science* **41**:391–407. DOI: [https://doi.org/10.1002/\(SICI\)1097-4571\(199009\)41:6<391::AID-ASI1>3.0.CO;2-9](https://doi.org/10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASI1>3.0.CO;2-9)
- DiCarlo JJ, Zoccolan D, Rust NC. 2012. How does the brain solve visual object recognition? *Neuron* **73**:415–434. DOI: <https://doi.org/10.1016/j.neuron.2012.01.010>, PMID: 22325196
- DiCarlo JJ, Cox DD. 2007. Untangling invariant object recognition. *Trends in Cognitive Sciences* **11**:333–341. DOI: <https://doi.org/10.1016/j.tics.2007.06.010>, PMID: 17631409

- Diedrichsen J, Kriegeskorte N. 2017. Representational models: a common framework for understanding encoding, pattern-component, and representational-similarity analysis. *PLOS Computational Biology* **13**: e1005508. DOI: <https://doi.org/10.1371/journal.pcbi.1005508>, PMID: 28437426
- Dubois J, Galdi P, Paul LK, Adolphs R. 2018. A distributed brain network predicts general intelligence from resting-state human neuroimaging data. *Philosophical Transactions of the Royal Society B: Biological Sciences* **373**:20170284. DOI: <https://doi.org/10.1098/rstb.2017.0284>
- Dubois J, Adolphs R. 2016. Building a science of individual differences from fMRI. *Trends in Cognitive Sciences* **20**:425–443. DOI: <https://doi.org/10.1016/j.tics.2016.03.014>, PMID: 27138646
- Durbin R, Mitchison G. 1990. A dimension reduction framework for understanding cortical maps. *Nature* **343**: 644–647. DOI: <https://doi.org/10.1038/343644a0>, PMID: 2304536
- Edelman S. 1998. Representation is representation of similarities. *Behavioral and Brain Sciences* **21**:449–467. DOI: <https://doi.org/10.1017/S0140525X98001253>, PMID: 10097019
- Edelman S, Grill-Spector K, Kushnir T, Malach R. 1998. Toward direct visualization of the internal shape representation space by fMRI. *Psychobiology* **26**:309–321. DOI: <https://doi.org/10.3758/BF03330618>
- Edelman GM, Gally JA. 2001. Degeneracy and complexity in biological systems. *PNAS* **98**:13763–13768. DOI: <https://doi.org/10.1073/pnas.231499798>, PMID: 11698650
- Ejaz N, Hamada M, Diedrichsen J. 2015. Hand use predicts the structure of representations in sensorimotor cortex. *Nature Neuroscience* **18**:1034–1040. DOI: <https://doi.org/10.1038/nn.4038>, PMID: 26030847
- Epstein R, Kanwisher N. 1998. A cortical representation of the local visual environment. *Nature* **392**:598–601. DOI: <https://doi.org/10.1038/33402>, PMID: 9560155
- Feilong M, Nastase SA, Guntupalli JS, Haxby JV. 2018. Reliable individual differences in fine-grained cortical functional architecture. *NeuroImage* **183**:375–386. DOI: <https://doi.org/10.1016/j.neuroimage.2018.08.029>, PMID: 30118870
- Feilong M, Guntupalli JS, Nastase S, Halchenko Y, Haxby JV. 2019. Predicting general intelligence from fine-grained functional connectivity. Organization for Human Brain Mapping 25th Annual Meeting:M875.
- Filippini N, MacIntosh BJ, Hough MG, Goodwin GM, Frisoni GB, Smith SM, Matthews PM, Beckmann CF, Mackay CE. 2009. Distinct patterns of brain activity in young carriers of the APOE-epsilon4 allele. *PNAS* **106**: 7209–7214. DOI: <https://doi.org/10.1073/pnas.0811879106>, PMID: 19357304
- Finn ES, Shen X, Scheinost D, Rosenberg MD, Huang J, Chun MM, Papademetris X, Constable RT. 2015. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nature Neuroscience* **18**:1664–1671. DOI: <https://doi.org/10.1038/nn.4135>, PMID: 26457551
- Fischl B, Sereno MI, Tootell RB, Dale AM. 1999. High-resolution intersubject averaging and a coordinate system for the cortical surface. *Human Brain Mapping* **8**:272–284. DOI: [https://doi.org/10.1002/\(SICI\)1097-0193\(199908\)8:4<272::AID-HBM10>3.0.CO;2-4](https://doi.org/10.1002/(SICI)1097-0193(199908)8:4<272::AID-HBM10>3.0.CO;2-4), PMID: 10619420
- Freiwald WA, Tsao DY. 2010. Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science* **330**:845–851. DOI: <https://doi.org/10.1126/science.1194908>, PMID: 21051642
- Gallego JA, Perich MG, Chowdhury RH, Solla SA, Miller LE. 2020. Long-term stability of cortical population dynamics underlying consistent behavior. *Nature Neuroscience* **23**:260–270. DOI: <https://doi.org/10.1038/s41593-019-0555-4>, PMID: 31907438
- Glasser MF, Coalson TS, Robinson EC, Hacker CD, Harwell J, Yacoub E, Ugurbil K, Andersson J, Beckmann CF, Jenkinson M, Smith SM, Van Essen DC. 2016. A multi-modal parcellation of human cerebral cortex. *Nature* **536**:171–178. DOI: <https://doi.org/10.1038/nature18933>, PMID: 27437579
- Gordon EM, Laumann TO, Adeyemo B, Huckins JF, Kelley WM, Petersen SE. 2016. Generation and evaluation of a cortical area parcellation from Resting-State correlations. *Cerebral Cortex* **26**:288–303. DOI: <https://doi.org/10.1093/cercor/bhw239>, PMID: 25316338
- Gordon EM, Laumann TO, Gilmore AW, Newbold DJ, Greene DJ, Berg JJ, Ortega M, Hoyt-Drazen C, Gratton C, Sun H, Hampton JM, Coalson RS, Nguyen AL, McDermott KB, Shimony JS, Snyder AZ, Schlaggar BL, Petersen SE, Nelson SM, Dosenbach NUF. 2017. Precision functional mapping of individual human brains. *Neuron* **95**: 791–807. DOI: <https://doi.org/10.1016/j.neuron.2017.07.011>, PMID: 28757305
- Gower JC. 1975. Generalized procrustes analysis. *Psychometrika* **40**:33–51. DOI: <https://doi.org/10.1007/BF02291478>
- Grill-Spector K, Weiner KS. 2014. The functional architecture of the ventral temporal cortex and its role in categorization. *Nature Reviews Neuroscience* **15**:536–548. DOI: <https://doi.org/10.1038/nrn3747>, PMID: 24962370
- Gross CG, Rocha-Miranda CE, Bender DB. 1972. Visual properties of neurons in inferotemporal cortex of the macaque. *Journal of Neurophysiology* **35**:96–111. DOI: <https://doi.org/10.1152/jn.1972.35.1.96>, PMID: 4621506
- Güçlü U, van Gerven MAJ. 2017. Increasingly complex representations of natural movies across the dorsal stream are shared between subjects. *NeuroImage* **145**:329–336. DOI: <https://doi.org/10.1016/j.neuroimage.2015.12.036>, PMID: 26724778
- Guntupalli JS, Hanke M, Halchenko YO, Connolly AC, Ramadge PJ, Haxby JV. 2016. A model of representational spaces in human cortex. *Cerebral Cortex* **26**:2919–2934. DOI: <https://doi.org/10.1093/cercor/bhw068>, PMID: 26980615
- Guntupalli JS, Wheeler KG, Gobbini MI. 2017. Disentangling the representation of identity from head view along the human face processing pathway. *Cerebral Cortex* **27**:46–53. DOI: <https://doi.org/10.1093/cercor/bhw344>, PMID: 28051770

- Guntupalli JS**, Feilong M, Haxby JV. 2018. A computational model of shared fine-scale structure in the human connectome. *PLOS Computational Biology* **14**:e1006120. DOI: <https://doi.org/10.1371/journal.pcbi.1006120>, PMID: 29664910
- Haak KV**, Marquand AF, Beckmann CF. 2018. Connectopic mapping with resting-state fMRI. *NeuroImage* **170**: 83–94. DOI: <https://doi.org/10.1016/j.neuroimage.2017.06.075>, PMID: 28666880
- Hanson SJ**, Matsuka T, Haxby JV. 2004. Combinatorial codes in ventral temporal lobe for object recognition: Haxby (2001) revisited: is there a "face" area? *NeuroImage* **23**:156–166. DOI: <https://doi.org/10.1016/j.neuroimage.2004.05.020>, PMID: 15325362
- Hasson U**, Nir Y, Levy I, Fuhrmann G, Malach R. 2004. Intersubject synchronization of cortical activity during natural vision. *Science* **303**:1634–1640. DOI: <https://doi.org/10.1126/science.1089506>, PMID: 15016991
- Hasson U**, Malach R, Heeger DJ. 2010. Reliability of cortical activity during natural stimulation. *Trends in Cognitive Sciences* **14**:40–48. DOI: <https://doi.org/10.1016/j.tics.2009.10.011>, PMID: 20004608
- Haxby JV**, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P. 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* **293**:2425–2430. DOI: <https://doi.org/10.1126/science.1063736>, PMID: 11577229
- Haxby JV**, Guntupalli JS, Connolly AC, Halchenko YO, Conroy BR, Gobbini MI, Hanke M, Ramadge PJ. 2011. A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron* **72**: 404–416. DOI: <https://doi.org/10.1016/j.neuron.2011.08.026>, PMID: 22017997
- Haxby JV**. 2012. Multivariate pattern analysis of fMRI: the early beginnings. *NeuroImage* **62**:852–855. DOI: <https://doi.org/10.1016/j.neuroimage.2012.03.016>, PMID: 22425670
- Haxby JV**, Connolly AC, Guntupalli JS. 2014. Decoding neural representational spaces using multivariate pattern analysis. *Annual Review of Neuroscience* **37**:435–456. DOI: <https://doi.org/10.1146/annurev-neuro-062012-170325>, PMID: 25002277
- Haxby JV**, Gobbini MI, Nastase SA. 2020. Naturalistic stimuli reveal a dominant role for agentic action in visual representation. *NeuroImage*:116561. DOI: <https://doi.org/10.1016/j.neuroimage.2020.116561>, PMID: 32001371
- Haynes JD**. 2015. A primer on Pattern-Based approaches to fMRI: principles, pitfalls, and perspectives. *Neuron* **87**:257–270. DOI: <https://doi.org/10.1016/j.neuron.2015.05.025>, PMID: 26182413
- Haynes JD**, Rees G. 2006. Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience* **7**:523–534. DOI: <https://doi.org/10.1038/nrn1931>, PMID: 16791142
- Hebart MN**, Baker CI. 2018. Deconstructing multivariate decoding for the study of brain function. *NeuroImage* **180**:4–18. DOI: <https://doi.org/10.1016/j.neuroimage.2017.08.005>, PMID: 28782682
- Heinzel J**, Kahnt T, Haynes JD. 2011. Topographically specific functional connectivity between visual field maps in the human brain. *NeuroImage* **56**:1426–1436. DOI: <https://doi.org/10.1016/j.neuroimage.2011.02.077>, PMID: 21376818
- Hockett CF**. 1958. *A course in modern linguistics*. New York: Macmillan.
- Hubel DH**, Wiesel TN. 1959. Receptive fields of single neurones in the cat's striate cortex. *The Journal of Physiology* **148**:574–591. DOI: <https://doi.org/10.1113/jphysiol.1959.sp006308>, PMID: 14403679
- Huth AG**, Nishimoto S, Vu AT, Gallant JL. 2012. A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron* **76**:1210–1224. DOI: <https://doi.org/10.1016/j.neuron.2012.10.014>, PMID: 23259955
- Huth AG**, de Heer WA, Griffiths TL, Theunissen FE, Gallant JL. 2016. Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature* **532**:453–458. DOI: <https://doi.org/10.1038/nature17637>, PMID: 27121839
- Jbabdi S**, Sotiropoulos SN, Behrens TE. 2013. The topographic connectome. *Current Opinion in Neurobiology* **23**:207–215. DOI: <https://doi.org/10.1016/j.conb.2012.12.004>, PMID: 23298689
- Jiahui G**, Feilong M, Visconti di Oleggio Castello M, Guntupalli JS, Chauhan V, Haxby JV, Gobbini MI. 2019. Predicting individual face-selective topography using naturalistic stimuli. *NeuroImage*:116458. DOI: <https://doi.org/10.1016/j.neuroimage.2019.116458>, PMID: 31843709
- Just MA**, Cherkassky VL, Aryal S, Mitchell TM. 2010. A neurosemantic theory of concrete noun representation based on the underlying brain codes. *PLOS ONE* **5**:e8622. DOI: <https://doi.org/10.1371/journal.pone.0008622>, PMID: 20084104
- Kanwisher N**, McDermott J, Chun MM. 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience* **17**:4302–4311. DOI: <https://doi.org/10.1523/JNEUROSCI.17-11-04302.1997>, PMID: 9151747
- Kay KN**, Naselaris T, Prenger RJ, Gallant JL. 2008. Identifying natural images from human brain activity. *Nature* **452**:352–355. DOI: <https://doi.org/10.1038/nature06713>, PMID: 18322462
- Khaligh-Razavi SM**, Kriegeskorte N. 2014. Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLOS Computational Biology* **10**:e1003915. DOI: <https://doi.org/10.1371/journal.pcbi.1003915>, PMID: 25375136
- Kiani R**, Esteky H, Mirpour K, Tanaka K. 2007. Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *Journal of Neurophysiology* **97**:4296–4309. DOI: <https://doi.org/10.1152/jn.00024.2007>, PMID: 17428910
- Kong R**, Li J, Orban C, Sabuncu MR, Liu H, Schaefer A, Sun N, Zuo XN, Holmes AJ, Eickhoff SB, Yeo BTT. 2019. Spatial topography of Individual-Specific cortical networks predicts human cognition, personality, and emotion. *Cerebral Cortex* **29**:2533–2551. DOI: <https://doi.org/10.1093/cercor/bhy123>, PMID: 29878084

- Kornblith S, Norouzi M, Lee H, Hinton G. 2019. Similarity of neural network representations revisited. *arXiv*: 190500414. <https://arxiv.org/abs/1905.00414>.
- Kriegeskorte N, Goebel R, Bandettini P. 2006. Information-based functional brain mapping. *PNAS* **103**:3863–3868. DOI: <https://doi.org/10.1073/pnas.0600244103>, PMID: 16537458
- Kriegeskorte N, Mur M, Bandettini P. 2008a. Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience* **2**:4. DOI: <https://doi.org/10.3389/neuro.06.004.2008>, PMID: 19104670
- Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, Esteky H, Tanaka K, Bandettini PA. 2008b. Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* **60**:1126–1141. DOI: <https://doi.org/10.1016/j.neuron.2008.10.043>, PMID: 19109916
- Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI. 2009. Circular analysis in systems neuroscience: the dangers of double dipping. *Nature Neuroscience* **12**:535–540. DOI: <https://doi.org/10.1038/nn.2303>, PMID: 19396166
- Kriegeskorte N, Kievit RA. 2013. Representational geometry: integrating cognition, computation, and the brain. *Trends in Cognitive Sciences* **17**:401–412. DOI: <https://doi.org/10.1016/j.tics.2013.06.007>, PMID: 23876494
- Langs G, Sweet A, Lashkari D, Tie Y, Rigolo L, Golby AJ, Golland P. 2014. Decoupling function and anatomy in atlases of functional connectivity patterns: language mapping in tumor patients. *NeuroImage* **103**:462–475. DOI: <https://doi.org/10.1016/j.neuroimage.2014.08.029>, PMID: 25172207
- Langs G, Wang D, Golland P, Mueller S, Pan R, Sabuncu MR, Sun W, Li K, Liu H. 2016. Identifying shared brain networks in individuals by decoupling functional and anatomical variability. *Cerebral Cortex* **26**:4004–4014. DOI: <https://doi.org/10.1093/cercor/bhv189>, PMID: 26334050
- Leopold DA, Bondar IV, Giese MA. 2006. Norm-based face encoding by single neurons in the monkey inferotemporal cortex. *Nature* **442**:572–575. DOI: <https://doi.org/10.1038/nature04951>, PMID: 16862123
- Lu Q, Chen P-H, Pillow JW, Ramadge PJ, Norman KA, Hasson U. 2019. Shared representational geometry across neural networks. *arXiv*. <https://arxiv.org/abs/1811.11684>.
- McIntosh AR, Mišić B. 2013. Multivariate statistical analyses for neuroimaging data. *Annual Review of Psychology* **64**:499–525. DOI: <https://doi.org/10.1146/annurev-psych-113011-143804>, PMID: 22804773
- McMahon DB, Russ BE, Elnaiem HD, Kurnikova AI, Leopold DA. 2015. Single-unit activity during natural vision: diversity, consistency, and spatial sensitivity among AF face patch neurons. *Journal of Neuroscience* **35**:5537–5548. DOI: <https://doi.org/10.1523/JNEUROSCI.3825-14.2015>, PMID: 25855170
- Mitchell TM, Shinkareva SV, Carlson A, Chang KM, Malave VL, Mason RA, Just MA. 2008. Predicting human brain activity associated with the meanings of nouns. *Science* **320**:1191–1195. DOI: <https://doi.org/10.1126/science.1152876>, PMID: 18511683
- Naselaris T, Prenger RJ, Kay KN, Oliver M, Gallant JL. 2009. Bayesian reconstruction of natural images from human brain activity. *Neuron* **63**:902–915. DOI: <https://doi.org/10.1016/j.neuron.2009.09.006>, PMID: 19778517
- Naselaris T, Kay KN, Nishimoto S, Gallant JL. 2011. Encoding and decoding in fMRI. *NeuroImage* **56**:400–410. DOI: <https://doi.org/10.1016/j.neuroimage.2010.07.073>, PMID: 20691790
- Nastase SA, Connolly AC, Oosterhof NN, Halchenko YO, Guntupalli JS, Visconti di Oleggio Castello M, Gors J, Gobbini MI, Haxby JV. 2017. Attention selectively reshapes the geometry of distributed semantic representation. *Cerebral Cortex* **27**:4277–4291. DOI: <https://doi.org/10.1093/cercor/bhw138>, PMID: 28591837
- Nastase SA, Gazzola V, Hasson U, Keysers C. 2019. Measuring shared responses across subjects using intersubject correlation. *Social Cognitive and Affective Neuroscience* **14**:667–685. DOI: <https://doi.org/10.1093/scan/nsz037>, PMID: 31099394
- Nastase SA, Liu YF, Hillman H, Norman KA, Hasson U. 2020. Leveraging shared connectivity to aggregate heterogeneous datasets into a common response space. *NeuroImage* **217**:116865. DOI: <https://doi.org/10.1016/j.neuroimage.2020.116865>, PMID: 32325212
- Nishimoto S, Vu AT, Naselaris T, Benjamini Y, Yu B, Gallant JL. 2011. Reconstructing visual experiences from brain activity evoked by natural movies. *Current Biology* **21**:1641–1646. DOI: <https://doi.org/10.1016/j.cub.2011.08.031>, PMID: 21945275
- Norman KA, Polyn SM, Detre GJ, Haxby JV. 2006. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences* **10**:424–430. DOI: <https://doi.org/10.1016/j.tics.2006.07.005>, PMID: 16899397
- O'Toole AJ, Jiang F, Abdi H, Haxby JV. 2005. Partially distributed representations of objects and faces in ventral temporal cortex. *Journal of Cognitive Neuroscience* **17**:580–590. DOI: <https://doi.org/10.1162/0898929053467550>, PMID: 15829079
- Oosterhof NN, Wiestler T, Downing PE, Diedrichsen J. 2011. A comparison of volume-based and surface-based multi-voxel pattern analysis. *NeuroImage* **56**:593–600. DOI: <https://doi.org/10.1016/j.neuroimage.2010.04.270>, PMID: 20621701
- Park SH, Russ BE, McMahon DBT, Koyano KW, Berman RA, Leopold DA. 2017. Functional subpopulations of neurons in a macaque face patch revealed by Single-Unit fMRI mapping. *Neuron* **95**:971–981. DOI: <https://doi.org/10.1016/j.neuron.2017.07.014>, PMID: 28757306
- Pereira F, Mitchell T, Botvinick M. 2009. Machine learning classifiers and fMRI: a tutorial overview. *NeuroImage* **45**:S199–S209. DOI: <https://doi.org/10.1016/j.neuroimage.2008.11.007>, PMID: 19070668
- Ponce CR, Xiao W, Schade PF, Hartmann TS, Kreiman G, Livingstone MS. 2019. Evolving images for visual neurons using a deep generative network reveals coding principles and neuronal preferences. *Cell* **177**:999–1009. DOI: <https://doi.org/10.1016/j.cell.2019.04.005>, PMID: 31051108

- Robinson EC**, Jbabdi S, Glasser MF, Andersson J, Burgess GC, Harms MP, Smith SM, Van Essen DC, Jenkinson M. 2014. MSM: a new flexible framework for multimodal surface matching. *NeuroImage* **100**:414–426. DOI: <https://doi.org/10.1016/j.neuroimage.2014.05.069>, PMID: 24939340
- Sabuncu MR**, Singer BD, Conroy B, Bryan RE, Ramadge PJ, Haxby JV. 2010. Function-based intersubject alignment of human cortical anatomy. *Cerebral Cortex* **20**:130–140. DOI: <https://doi.org/10.1093/cercor/bhp085>, PMID: 19420007
- Salehi M**, Greene AS, Karbasi A, Shen X, Scheinost D, Constable RT. 2020. There is no single functional atlas even for a single individual: functional parcel definitions change with task. *NeuroImage* **208**:116366. DOI: <https://doi.org/10.1016/j.neuroimage.2019.116366>, PMID: 31740342
- Salton G**, Wong A, Yang CS. 1975. A vector space model for automatic indexing. *Communications of the ACM* **18**:613–620. DOI: <https://doi.org/10.1145/361219.361220>
- Schönemann PH**. 1966. A generalized solution of the orthogonal procrustes problem. *Psychometrika* **31**:1–10. DOI: <https://doi.org/10.1007/BF02289451>
- Sha L**, Haxby JV, Abdi H, Guntupalli JS, Oosterhof NN, Halchenko YO, Connolly AC. 2015. The animacy continuum in the human ventral vision pathway. *Journal of Cognitive Neuroscience* **27**:665–678. DOI: https://doi.org/10.1162/jocn_a_00733, PMID: 25269114
- Sheinberg DL**, Logothetis NK. 2001. Noticing familiar objects in real world scenes: the role of temporal cortical neurons in natural vision. *The Journal of Neuroscience* **21**:1340–1350. DOI: <https://doi.org/10.1523/JNEUROSCI.21-04-01340.2001>, PMID: 11160405
- Shen X**, Papademetris X, Constable RT. 2010. Graph-theory based parcellation of functional subunits in the brain from resting-state fMRI data. *NeuroImage* **50**:1027–1035. DOI: <https://doi.org/10.1016/j.neuroimage.2009.12.119>, PMID: 20060479
- Shen X**, Tokoglu F, Papademetris X, Constable RT. 2013. Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. *NeuroImage* **82**:403–415. DOI: <https://doi.org/10.1016/j.neuroimage.2013.05.081>, PMID: 23747961
- Shepard RN**. 1987. Toward a universal law of generalization for psychological science. *Science* **237**:1317–1323. DOI: <https://doi.org/10.1126/science.3629243>, PMID: 3629243
- Taschereau-Dumouchel V**, Cortese A, Chiba T, Knotts JD, Kawato M, Lau H. 2018. Towards an unconscious neural reinforcement intervention for common fears. *PNAS* **115**:3470–3475. DOI: <https://doi.org/10.1073/pnas.1721572115>, PMID: 29511106
- Thorat S**, Proklova D, Peelen MV. 2019. The nature of the animacy organization in human ventral temporal cortex. *eLife* **8** e47142. DOI: <https://doi.org/10.7554/eLife.47142>, PMID: 31496518
- Tong F**, Pratte MS. 2012. Decoding patterns of human brain activity. *Annual Review of Psychology* **63**:483–509. DOI: <https://doi.org/10.1146/annurev-psych-120710-100412>, PMID: 21943172
- Van Uden CE**, Nastase SA, Connolly AC, Feilong M, Hansen I, Gobbini MI, Haxby JV. 2018. Modeling semantic encoding in a common neural representational space. *Frontiers in Neuroscience* **12**:437. DOI: <https://doi.org/10.3389/fnins.2018.00437>
- Visconti di Oleggio Castello M**, Halchenko YO, Guntupalli JS, Gors JD, Gobbini MI. 2017. The neural representation of personally familiar and unfamiliar faces in the distributed system for face perception. *Scientific Reports* **7**:1–14. DOI: <https://doi.org/10.1038/s41598-017-12559-1>, PMID: 28947835
- Wang H-T**, Smallwood J, Mourao-Miranda J, Xia CH, Satterthwaite TD, Bassett DS, Bzdok D. 2020. Finding the needle in a high-dimensional haystack: canonical correlation analysis for neuroscientists. *NeuroImage* **216**:116745. DOI: <https://doi.org/10.1016/j.neuroimage.2020.116745>
- Xu H**, Lorbert A, Ramadge PJ, Guntupalli JS, Haxby JV. 2012. Regularized hyperalignment of multi-set fMRI data 2012 IEEE statistical signal processing workshop, SSP 2012. Presented at the 2012 IEEE Statistical Signal Processing Workshop, SSP 2012 229–232. DOI: <https://doi.org/10.1109/SSP.2012.6319668>
- Yeo BT**, Krienen FM, Sepulcre J, Sabuncu MR, Lashkari D, Hollinshead M, Roffman JL, Smoller JW, Zöllei L, Polimeni JR, Fischl B, Liu H, Buckner RL. 2011. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology* **106**:1125–1165. DOI: <https://doi.org/10.1152/jn.00338.2011>, PMID: 21653723
- Zilles K**, Amunts K. 2010. Centenary of Brodmann's map—conception and fate. *Nature Reviews Neuroscience* **11**:139–145. DOI: <https://doi.org/10.1038/nrn2776>, PMID: 20046193