Cross-movie prediction of individualized functional topography

Authors

Guo Jiahui[1], Ma Feilong[1], Samuel A. Nastase[2], James V. Haxby[1], M. Ida Gobbini[1,3*]

Affiliations

[1] Center for Cognitive Neuroscience, Dartmouth College, NH, USA 03755.

[2] Princeton Neuroscience Institute, Princeton University, Princeton, NJ, USA 08544.

[3] Dipartimento di Medicina Specialistica, Diagnostica e Sperimentale, Università di Bologna, Bologna, Italy 40138.

[*]*Correspondence: mariaida.gobbini@unibo.it*

Abstract

Participant-specific, functionally-defined brain areas are usually mapped with functional localizers and estimated by making contrasts between responses to single categories of input. Naturalistic stimuli engage multiple brain systems in parallel, provide more ecologically plausible estimates of real-world statistics, and are friendly to special populations. The current study shows that cortical functional topographies in individual participants can be estimated with high fidelity from naturalistic stimuli. Importantly, we demonstrate that robust, individualized estimates can be obtained even when participants watched different movies, were scanned with different parameters/scanners, and were sampled from different institutes across the world. Our results create a foundation for future studies that allow researchers to estimate a broad range of functional topographies based on naturalistic movies and a normative database, making it possible to integrate high-level cognitive functions across datasets from laboratories worldwide.

Keywords

Hyperalignment, Category-selectivity, Localizer, Connectivity, Naturalistic stimuli

Introduction

Category-selective functional topographies are a prominent and consistent feature of lateral occipital, ventral temporal, and lateral temporal visual cortices (Downing et al., 2001; Epstein et al., 1999; Grill-Spector & Weiner, 2014; Kanwisher et al., 1997). Category-selective topographies are mostly similar across individuals but are idiosyncratic in terms of their precise conformation and location (Zhen et al., 2015, 2017). Because of these idiosyncrasies, category-selective topographies and areas are typically mapped in each individual using a functional localizer fMRI scan (Fedorenko et al., 2010; Saxe et al., 2006). Functional localizers map individualized topographies with simple contrasts between responses to different categories, such as contrasting responses to faces versus objects to localize face-selective areas.

We reported an alternative approach to map category-selective topographies using fMRI data collected while participants view a naturalistic movie (Guntupalli et al., 2016; Haxby et al., 2011; Jiahui et al., 2020). With this approach, movie-viewing and functional localizer data are collected in a normative sample, and new participants need only be scanned during movie viewing. Movie data are used to calculate transformation matrices using hyperalignment (Feilong et al., 2018, 2021, 2021; Guntupalli et al., 2016, 2018; Haxby et al., 2011; Jiahui et al., 2020) that afford projecting the localizer data from the normative sample into the idiosyncratic cortical topography of new participants. Using this hyperalignment procedure, we can estimate the idiosyncratic details of individual topographies with high fidelity based on localizer data from the normative sample. Unlike functional localizers, naturalistic stimuli (e.g., movies) evoke a rich variety of brain states and engage multiple brain systems in parallel. This makes it possible to efficiently map multiple functional topographies using data from a single movie and avoid the time and cost of running multiple localizers. Compared to controlled localizers, movies better simulate real-world cognition and better engage participants' attention (Vanderwal et al., 2015, 2017, 2019), contributing to more ecologically valid and higher-quality maps. In addition, movies are more friendly and engaging for special populations, such as young children.

In previous work, we used response hyperalignment (RHA) to predict functional topographies in new participants. RHA requires that all participants watch the same movie to obtain time-locked responses to the same stimuli. It is often important, however, to tailor the movie to meet the specific needs of participants in different experiments. For example, participants from different countries may prefer movies that reflect their diverse backgrounds and are in their native languages (e.g., Hanke et al., 2016; Sengupta et al., 2016); movies for infants and young children are differently structured from those for adults (e.g., Vanderwal et al., 2015). Thus, it is unrealistic to limit all participants from diverse populations and backgrounds to watch the same movie. Additionally, experimenters may need to shorten

or edit the stimuli to fit their data collection schedule. Finally, participants are often scanned with different parameters from one experiment to another, at different institutes across the world, and with different scanner models. Due to these factors, it is impractical to expect two laboratories to acquire the same movie scans across individuals.

Here, we test whether connectivity hyperalignment (CHA; Guntupalli et al., 2018) can be used to map category-selective functional topographies. CHA, in contrast to RHA, affords calculation of transformation matrices using stimuli that are not the same for normative and index participants. We analyzed four different data sets collected with three different movies, three different scanners, and two different types of functional localizers that used dynamic or static stimuli. We first demonstrated that CHA based on participants' connectomes that were calculated using their responses to movies was able to generate high-fidelity maps of category-selective topographies within datasets that were equivalent to maps estimated using RHA. Then, critically, we showed that cross-dataset predictions that used connectomes calculated from different movies for the normative and index brains were as good as those from participants in the same dataset. This means that different laboratories can use different movies to derive functional topographies from a normative sample.

In summary, we demonstrate that a target participant's individualized category-selective topography can be accurately estimated using connectivity hyperalignment, regardless of whether different movies are used to calculate the connectome and regardless of other data collection parameters. Movies engage multiple cognitive domains in parallel, such as visual perception, audition, language comprehension, theory of mind, and social interaction. In addition to estimating different functional topographies from a single movie, our approach allows us to estimate topographies from different movies. We provide a novel alternative for future data collection that can save time and money using rich and efficient movie scans.

Results

High-fidelity prediction with connectivity hyperalignment

We predicted category-selective topographies by projecting other participants' functional localizer data into each participant's native cortical topography using a new, enhanced CHA algorithm. For each participant, we calculated transformation matrices based on functional connectivity estimated during movie-viewing in an iterative way (see Materials and Methods). These transformation matrices resample fMRI data from others' brains into a given participant's cortex. We then projected the functional localizer data for all other participants into the given participant's native cortical space and calculated independent functional contrasts based on that participant's own localizer data and based on other

participants' localizer data projected into that participant's cortex. We also estimated functional topographies by projecting others' localizer data into that participant's cortex based on high-performing surface-based anatomical alignment as a control analysis. We calculated the correlations between topographies based on participants' own localizer contrasts and on other participants' data. Because the localizer task comprises several scanning runs, we calculated the reliability of the localizer across runs with Cronbach's alpha to provide an estimate of the noise ceiling for these correlations. We repeated this procedure for all participants.

We tested the estimation of visual category-selective functional topographies (faces, bodies, scenes, and objects) in four different datasets using three different movies, localizers with static or dynamic stimuli, different scanning sequence parameters, and three different scanner models (see Materials and Methods).

Category-selective topographies estimated with CHA recovered the idiosyncrasies of individuals' topographies, capturing fine details of the individual-specific configuration and extent. By contrast, topographies estimated with anatomical alignment generated highly blurred maps that were essentially the same for all participants, losing individual-specific idiosyncratic features (Figure 1A).

The superior performance of CHA-based estimation over anatomical-alignment-based estimation was consistent across participants, visual stimulus categories, and datasets. In all four category-selective topographies and in all four datasets, correlations between estimations based on hyperalignment and their own localizer data were significantly higher than the correlations between estimations based on anatomical alignment and each participant's own localizer (Fisher z-transformed, $p < 0.001$, Bonferroni corrected). We compared these correlations between topographies estimated from a participant's own localizer data and those from other participants' data to the reliability of the localizer, calculated with Cronbach's alpha. Predictions made with hyperalignment were close to and sometimes even exceeded the reliability values (Figure 1B), which indicate that the predicted category-selective topographies from other participants' data using hyperalignment were as precise and sometimes even better than the topographies estimated with their own localizer data.

Estimates using CHA to calculate transformation matrices were also equivalent to estimates using RHA (Figure S2A). RHA, however, requires that all subjects watch the same movie, whereas CHA can use connectivity matrices derived from responses to different movies, potentially making our new approach more flexible. Next we tested the validity of estimating topographies using transformation matrices that were based on functional connectivities calculated from responses to different movies for the test participant and other participants.

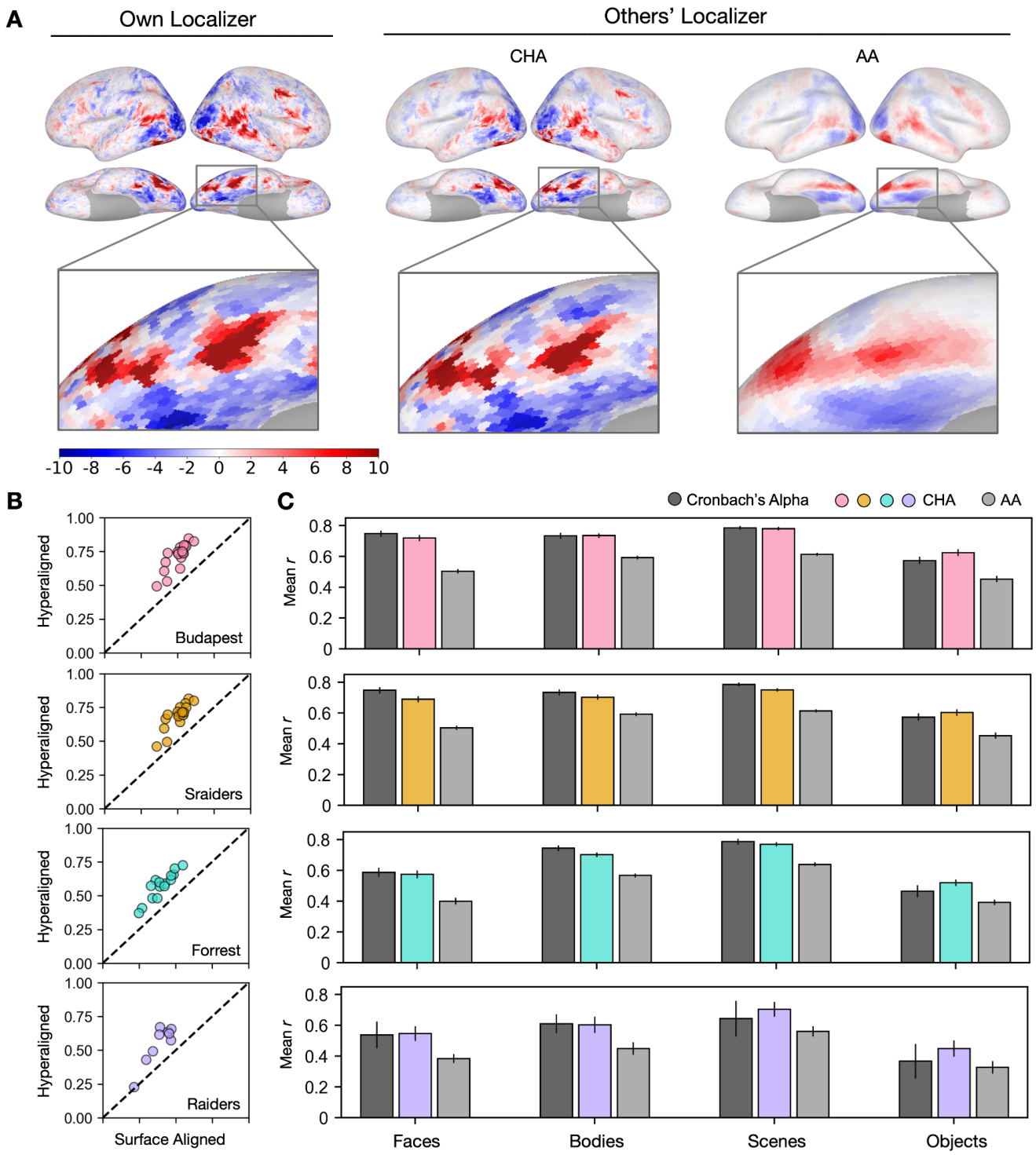

Figure 1. Predicting individual category-selective topographies using connectivity hyperalignment. **A.** Face-selective topographies (faces-vs-all) and zoomed-in views of an example participant estimated from this participant's own localizer (Own Localizer), and other participants' localizers using connectivity hyperalignment (CHA), and surface anatomical alignment (AA). **B.** Scatter plots display the Pearson correlation coefficients between estimated face-selective topographies based on own localizer data and other participants' localizer data in individual participants in four different datasets. The y-axis corresponds to correlations between each target participant's own localizer-based face-selective topographies and face-selective topographies estimated from other participants using

CHA. The x-axis corresponds to correlations between each target participant's own localizer-based face-selective topographies and face-selective topographies estimated from other participants with surface-based anatomical alignment. **C.** Bar plots show the mean correlations across participants in four datasets and for all four category-selective topographies. Black bars stand for the mean Cronbach's alphas across participants. Error bars indicate ±1 standard error of the mean. Category topographies were defined based on contrasts between the target category and all other categories.

Connectivity hyperalignment enables cross-movie predictions

Experimental design considerations and constraints can make using the same stimulus across all studies and participants inadvisable, and datasets are often collected under diverse conditions. Here, we aim to test whether connectivity-based hyperalignment can predict category-selective topographies in new individuals even if their connectomes are estimated from data collected while they watched a different movie. Using this method, participants across datasets without matched time-locked functional series can benefit from those who have functional localizer data but were scanned with different naturalistic stimuli.

We estimated category-selective topographies for each participant in each dataset from participants in the other dataset that used the same type of localizer (dynamic or static) by calculating transformation matrices based on functional connectivities measured while watching different movies. We also estimated topographies based on anatomical alignment. The cross-movie predictions using connectivity hyperalignment outperformed predictions based on anatomical alignment and were nearly as precise as within-movie predictions (Figure 2A). The superior performance was consistent across datasets and categories ($p < 0.001$ for all comparisons, Figure 2B) and in all individual participants (Figure S3). Similarly, accuracies of these predictions matched and sometimes even exceeded the reliability measures of their own localizer runs (Figure 2B).

Cross-movie predictions of cortical topographies based on different localizer types (static to dynamic or dynamic to static) produced lower correlations than did cross-movie predictions based on the same localizer type (Figure S2B), consistent with previous reports showing significant differences between topographies estimated by static and dynamic localizers, especially in superior temporal and frontal cortices (Fox et al., 2009; Pitcher et al., 2011).

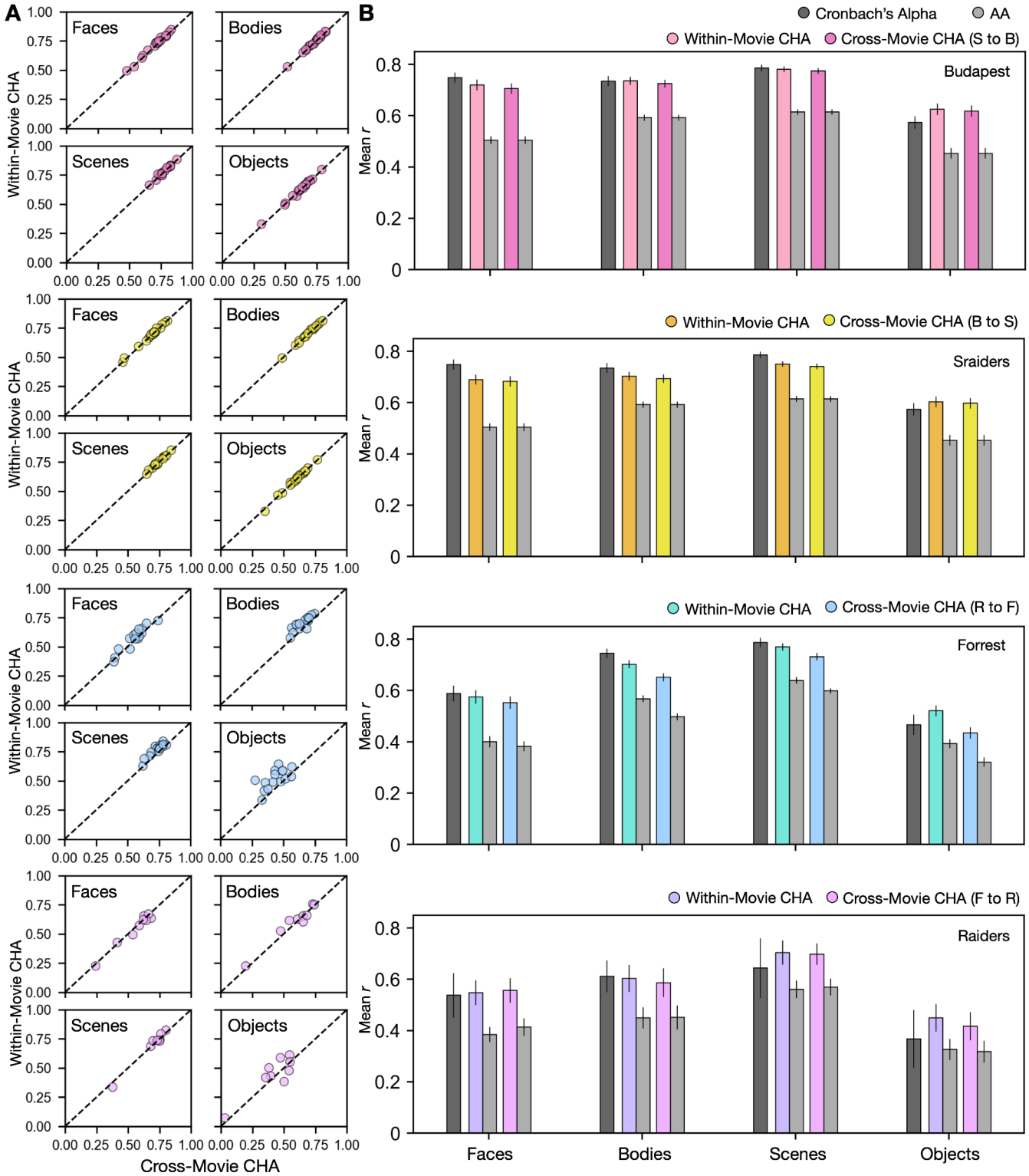

Figure 2. Predicting category-selective topographies using connectivity profiles across movies. A. Scatter plots of Pearson correlation coefficients for individual participants in four different datasets and for four categories. Values on the y-axis stand for correlations between each target participant's own localizer-based topographies and topographies estimated from other participants in the same movie using CHA. Values on the x-axis stand for correlations between each target participant's own localizer-based topographies and topographies estimated from participants in another dataset based on

cross-movie CHA. **B.** Bar plots display the mean Pearson correlation coefficients (*r*) and Cronbach's alphas across participants in all four datasets for all four categories. Error bars stand for ±1 standard error of the mean. S to B: Sraiders to Budapest, B to S: Budapest to Sraiders, R to F: Raiders to Forrest, F to R: Forrest to Raiders.

To demonstrate how hyperalignment increased prediction performance for individual participants from a different dataset, we plotted topographies estimated using hyperalignment and anatomical alignment, as well as from their own localizer runs (Figure 3, Figure S4 & 5). Topographies between datasets recovered similar idiosyncratic features as the topographies predicted within datasets.

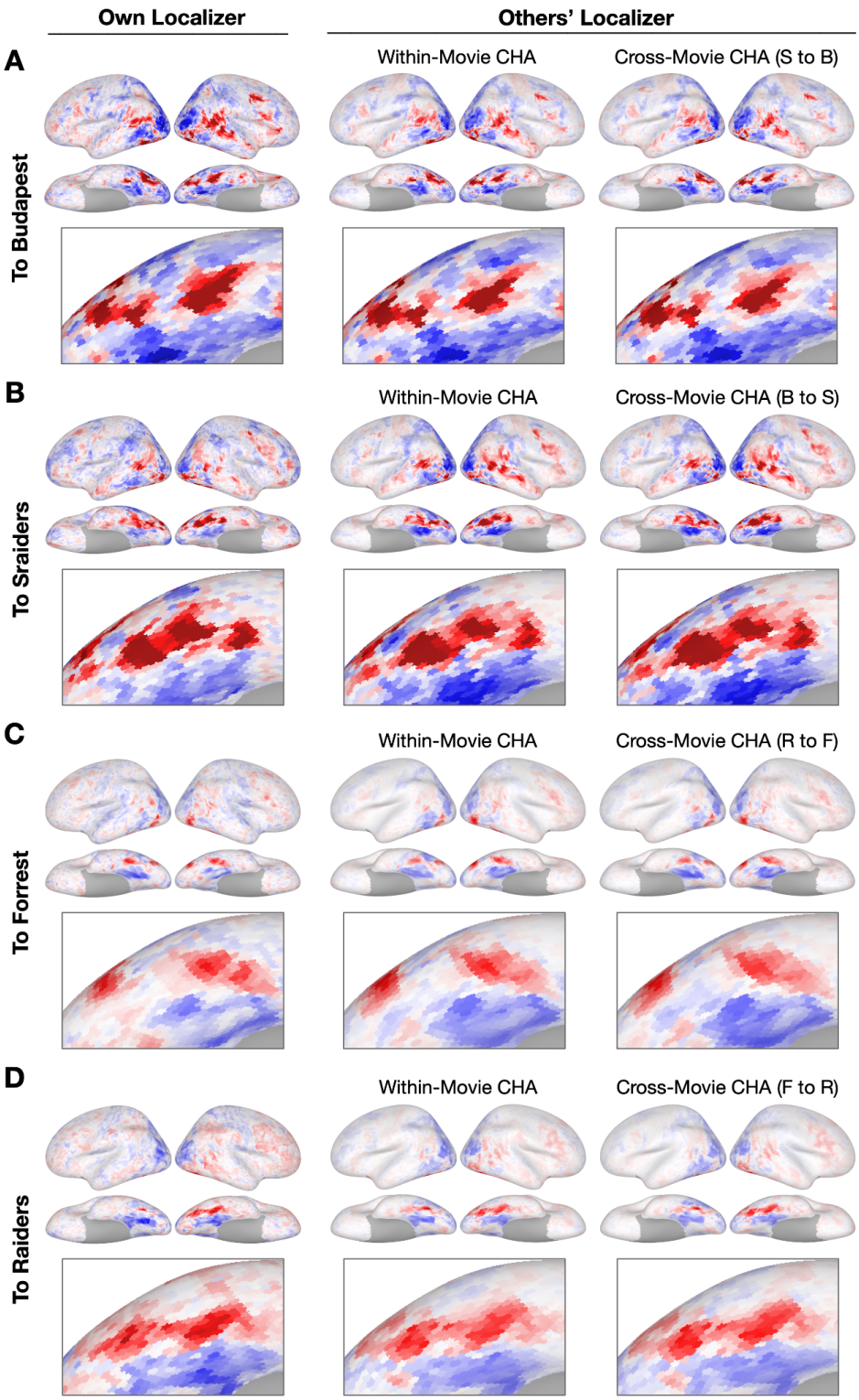

Figure 3. Sample contrast maps and enlarged views of the ventral temporal cortex. Contrast maps for face-selective topographies (faces-vs-all) and their zoomed-in views of the ventral temporal cortex were plotted in four sample participants in **A.** Budapest, **B.** Sraiders, **C.** Forrest, and **D.** Raiders. In all four subplots, in the left-most panel, faces-vs-all maps were plotted on the sample participants' own cortical surfaces. The next two columns display maps estimated from other participants' data. In the right two columns, the first column presents predicted face-selective topographies from participants in

the same dataset using CHA. The next column presents face-selective topographies from participants in another dataset (cross-movie CHA). The zoomed-in panels are displayed accordingly with the whole-brain map. The colorbar is the same as that in Figure 1. S to B: Sraiders to Budapest, B to S: Budapest to Sraiders, R to F: Raiders to Forrest, F to R: Forrest to Raiders

To further examine the topographies predicted using different datasets and compare the prediction performances to reliability measures, we calculated local correlations between maps estimated from each participant's own localizer runs and those estimated from other participants' runs with a searchlight analysis. We also calculated Cronbach's alpha across localizer runs in each searchlight. Generally, searchlights in the high-level visual areas and with strong category-selectivity (e.g., ventral temporal cortex, lateral temporal cortex) showed the highest mean correlation values, which often exceeded 0.8 (Figure 4, Figure S6 & 8). The lower mean correlations in other cortices (e.g., sensorimotor cortex) reflect low reliabilities of the localizer runs.

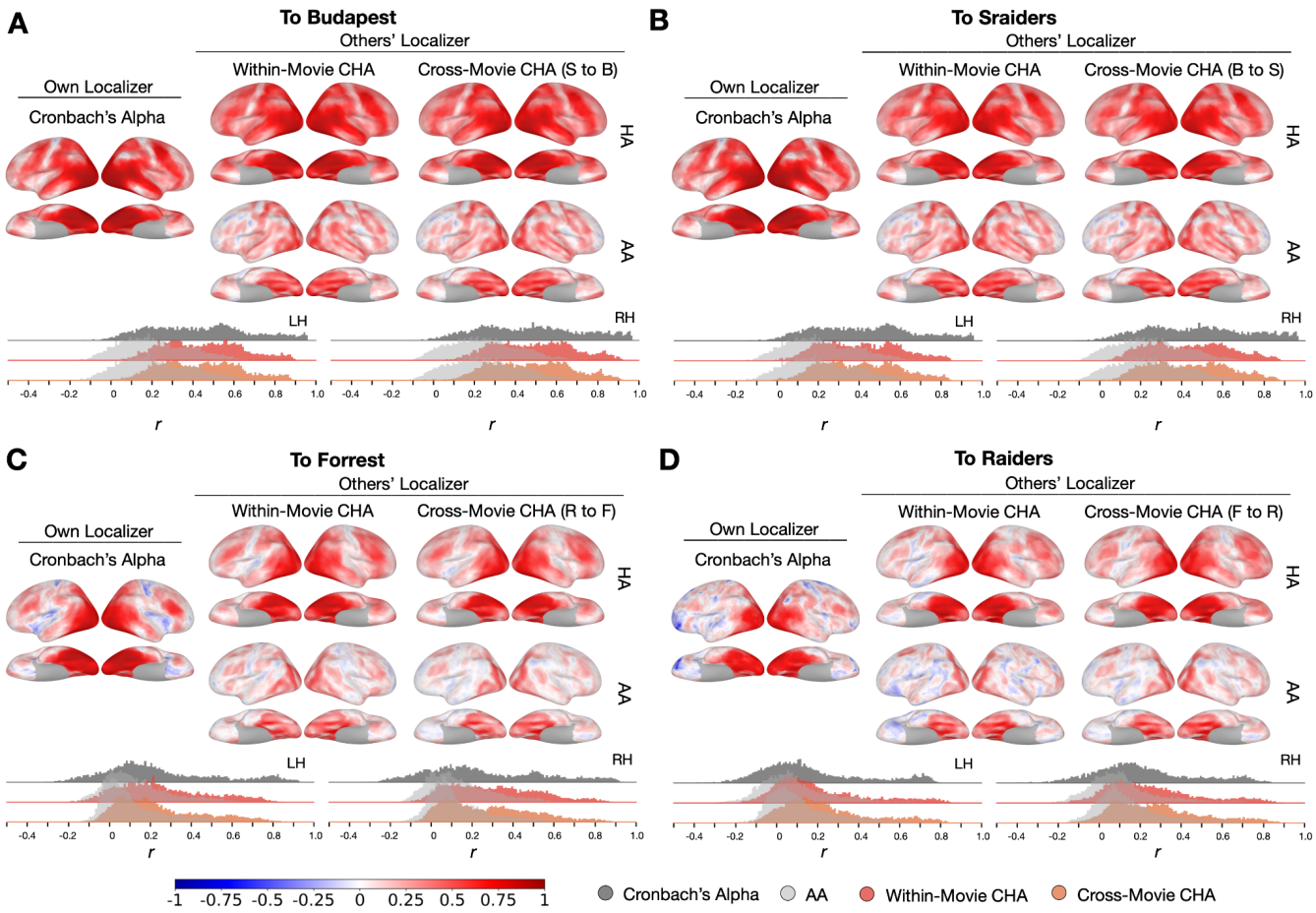

Figure 4. Searchlight analysis of Cronbach's alphas and prediction performances. A, B, C & D. The left-most column presents Cronbach's alphas of the own-localizer-based face-selective topographies in each dataset using a searchlight analysis (15 mm radius). The next two columns present local correlations (correlation maps) using the searchlight analysis between face-selective maps estimated

from participants' own localizers and from other participants based on within-movie and between-movie CHA (HA, top row) and surface alignment (AA, bottom row). Histogram plots present Cronbach's alphas (dark gray) and coefficients for the correlation maps above (estimated with CHA in color, with AA in light gray). The left and right hemisphere histograms were plotted separately. B to S: Budapest to Sraiders, S to B: Sraiders to Budapest, R to F: Raiders to Forrest, F to R: Forrest to Raiders

Discussion

In this study, using four datasets that contain three different movies, two different types of functional localizers, and collected with three different scanners, we showed that individualized category-selective topographies can be estimated with high fidelity using CHA. Unlike RHA, which requires the same "time-locked" response time series in the normative sample and new participants, CHA affords the calculation of transformation matrices based on responses to completely different movies. By showing that CHA based on participants' connectomes calculated using their responses to different movies generated high-fidelity mappings that were as good as those using RHA with participants in the same dataset, we demonstrated that CHA is able to effectively predict topographies across diverse situations. This study opens new possibilities connecting independent public and in-lab datasets for future data analysis so that researchers can derive multiple topographies at once for each individual with excellent performance based on the naturalistic movie data and the localizer data from another normative dataset. Our results also provide a novel alternative for new data collection to take better advantage of naturalistic stimuli.

We used a new, enhanced CHA in this study that optimized our previous CHA algorithm with iterative steps. In each step, transformation matrices to each index brain were calculated from other participants' brains and the matrices were applied to both the movie and the localizer data. We gradually increased the number of connectivity targets to form a denser representation of connectivity profiles. The iterations improved the prediction performance step by step, and at the final step (step 6) in this analysis, the enhanced CHA generated comparable performance with RHA (Figure S9). In addition, this study is based on the new optimized 1-step hyperalignment procedure (Jiahui et al., 2020). The classic hyperalignment method (2-step), builds a common information model space at the initial step that is based on all normative group participants, then projects information encoded in idiosyncratic representational spaces to the common model space, and lastly projects the information back to the individual participant's space based on the transpose of the transformation matrices from the former step. Different from the 2-step method, the 1-step method directly projects the data for each normative sample brain to the index participant's space without the intermediate step of building a common information model space. This method requires fewer steps and is free from the accumulation of errors

across steps. The 1-step method consistently improved the prediction performances across all conditions and datasets (Figure S10). This method is particularly useful for estimating information encoded in each individual's brain space.

The within-movie and cross-movie CHA predictions generated highly similar topographies (Figure 3). This result raises a fascinating question of whether different movie inputs estimate similar fine-grained connectivity profiles in the brain. Previous studies reported that the coarse-grained connectome (based on coarse parcellations) varies across separate cognitive tasks (e.g., Shine et al., 2016; Telesford et al., 2016), and that naturalistic movies yield the most condition-specific functional atlases among other classic cognitive tasks (Salehi et al., 2020). In the Budapest and Sraiders datasets, the same group of participants watched the Grand Budapest Hotel and Raiders of the Lost Ark in different sessions in the same 3T scanner. We built connectivity profiles for each participant separately for the two movies and correlated the two fine-grained connectomes in each searchlight. Results showed that the two fine-grained connectomes based on different movies were very similar in most of the brain regions ($r > 0.8$, Figure S11A & B). We split each movie into two halves (Run 1-3/Run 4-5 for Budapest; Run 1-2/Run 3-4 for Sraiders) and averaged the connectome similarities across split halves over searchlights and participants. We found that the across-movie connectome similarities for split halves were high ($r > 0.74$), and the within-movie similarities were even higher in both datasets ($r > 0.85$, Figure S11C). Our analysis showed that although the fine-grained connectome was affected by the input naturalistic stimulus content, it was nonetheless highly stable. This result suggested the brain may undergo shared cognitive processes across different movie free-viewing tasks. It could be because featured movies sample a broad range of real-life statistics, and the rich information elicits overall similar representations and connectivities when the entire time series is considered.

The four datasets in our study included two types of category-selective localizers (dynamic and static). The dynamic localizer used short video clips for each category and the traditional static localizer used still images. For all categories, the dynamic localizer elicited stronger and broader category-selective activations than the static localizer, and the searchlight analysis showed that the dynamic localizer had higher reliabilities across the cortex, especially in regions that were selectively responsive to the target category. Due to differences between topographies activated by the dynamic and the static localizers, predictions across localizer types generated lower correlations than those within localizer types. For example, for the face-selective topographies, the dynamic localizer activated more areas than the static localizer (e.g., in superior temporal and frontal cortices). In the ventral temporal cortex, especially in the right hemisphere, both dynamic and static localizers performed well in the cross-localizer-type predictions. But in cortical areas where the static localizer did not match the

dynamic localizer, predictions from the same dynamic localizer always outperformed the predictions from a different static localizer (Figure S6 & S8). The low correlations were not because the prediction method failed but reflected the difference in the topographies activated by different types of localizers.

In summary, our study demonstrated that accurate predictions of individualized category-selective topographies can be achieved with high fidelity using CHA across different naturalistic movie contents, across different scanners, and across different scanning parameters. Compared to traditional functional localizers, naturalistic stimuli are more ecologically valid, engaging multiple cognitive systems in parallel, and more friendly to participants. Our method not only can be applied directly to current public and in-lab datasets, but has the important potential to allow researchers to derive a broad range of topographies based on naturalistic movies and a normative database in the future. By building such a database that comprises various high-quality topographies and naturalistic stimuli, our study opens the gate to new research possibilities that could integrate high-level cognitive functions across datasets from laboratories worldwide.

Materials and Methods

Datasets

The Budapest Dataset

The *Budapest* dataset included twenty participants (mean age 27.2 years, 10 females) for this analysis. These participants were scanned while watching both *Grand Budapest Hotel* and *Raiders of the Lost Ark* and were a subset of the dataset in (Jiahui et al., 2020). The *Grand Budapest Hotel* dataset contained five movie runs (~50 min, each part lasting 9–13 min each) and four dynamic localizer runs. Before entering the scanner, participants watched the first part of the movie (~45 min) outside. The rest of the movie was divided into five parts (each part lasting 9–13 min, ~50 min in total) and participants watched each part/run with audio. The dynamic localizer data were collected in a separate scanning section (Pitcher et al., 2011). This dataset comprised four blocked-designed runs (3.9 min each), and each run comprised 10 blocks (18 s each), two per category (faces, bodies, scenes, objects, and scrambled objects). Each block comprised six 3s-long video clips in random order. Participants did a one-back task during the localizer scan to maintain attention.

All scans in the Grand Budapest Hotel dataset were acquired using a 3 T S Magnetom Prisma MRI scanner with a 32 channel head coil at the Dartmouth Brain Imaging Center. BOLD images were acquired in an interleaved fashion using gradient-echo echo-planar imaging with pre-scan normalization, fat suppression, multiband (i.e., simultaneous multi-slice; SMS) acceleration factor of 4 (using blipped CAIPIRINHA), and no in-plane acceleration (i.e., GRAPPA acceleration factor of one): TR/TE = 1000/33 ms, flip angle = 59 °, resolution = 2.5 mm^3 isotropic voxels, matrix size = 96 × 96, FoV = 240 × 240 mm, 52 axial slices with full brain coverage and no gap, anterior–posterior phase encoding. See more details in (Visconti di Oleggio Castello et al., 2020).

The Sraiders Dataset

The same participants were included for analysis in the *Sraiders* dataset as in the *Budapest* dataset. The movie *Raiders of the Lost Ark* was split into eight parts (~15 min each), and the first four parts were watched outside of the scanner prior to the scanning (~56 min). The later four parts were watched in the scanner (57 min) with audio (Nastase, 2018). The *Sraiders* dataset and the *Budapest* dataset shared the same dynamic localizer data. The *Sraiders* dataset was collected with the same scan protocols as the *Budapest* dataset (Feilong et al., 2022; Nastase, 2018).

The Forrest Dataset

This dataset contains scans from fifteen adults (mean age 29.4 years, 6 females). Participants were scanned at the Otto-von-Guericke University in Germany and were native German speakers (Hanke et al., 2016; Sengupta et al., 2016). The dataset is publicly available at http://www.studyforrest.org/ (Hanke

et al., 2014). A shortened version of the movie *Forrest Gump* was divided into eight parts with each part lasting approximately 15 min. Participants watched each part/run in the scanner with audio (Hanke et al., 2016). A category-selective localizer using still images was included in this dataset. This static localizer comprised four runs (5.2 min each). Each run comprised two 16 s blocks for each of the six categories (human faces, human bodies without heads, small objects, houses and outdoor scenes that include nature and street scenes, and phase scrambled images). In each block, 16 images from one category were displayed (900 ms display + 100 ms intertrial interval each). Participants were asked to do a one-back task to maintain attention.

Scanning was carried out using a whole-body 3 T Philips Achieva dStream MRI scanner equipped with a 32 channel head coil. Data were collected with gradient-echo, 2 s repetition time (TR), 30 ms echo time (TE), 90° flip angle, 1943 Hz/px bandwidth, and parallel acquisition with sensitivity encoding (SENSE) reduction factor 2. Each volume comprised 35 axial slices with anterior-to-posterior phase-encoding direction that were collected in ascending order, which mostly covered the entire brain. Each slice was 3.0 mm thick with a 10% inter-slice gap, and had a 240 mm × 240 mm field-of-view comprising 80 × 80 3 mm isotropic voxels. More acquisition parameters can be found in Hanke et al. (2016) and Sengupta et al. (2016).

The Raiders Dataset

A subset of nine participants from the original eleven participants (7 men, mean age = 24.8 years) participated in the face and object study at Dartmouth in (Haxby et al., 2011) and were included in this dataset. The audio-visual movie *Raiders of the Lost Ark* was split into eight parts (~15 min each), similarly to those used in the *Sraiders* Dataset. Participants watched all eight parts in the scanner with audio (one part / per run). The *Raiders* dataset contains a static localizer that was similarly designed as in the *Forrest* dataset.

Brain images were acquired using a 3T Philips Intera Achieva scanner with an eight-channel head coil at Dartmouth College. For the movie study, whole brain volumes of 413 mm thick sagittal images (TR = 2.5 s, TE = 35 ms, Flip angle = 90°, 80 × 80 matrix, FOV = 240 mm × 240 mm, resolution = 0.938 mm × 0.938 mm × 1.0 mm) were obtained in an interleaved slice order. For more details see (Haxby et al., 2011).

MRI Preprocessing

All datasets were preprocessed with fMRIPrep (Esteban et al., 2019), using version 20.1.1 for the Budapest dataset, 20.2.0 for the Sraiders dataset, 20.1.1 for the Forrest dataset, and 20.1.1 for the Raiders dataset. After fMRIPrep, functional data were projected onto a standard cortical surface aligned

to the fsaverage template (Fischl et al., 1999) based on cortical folding patterns. The datasets were further preprocessed following (Feilong et al., 2018; Jiahui et al., 2020). The datasets were resampled to a cortical mesh with 18,742 vertices across both hemispheres (approximately 3 mm vertex spacing; 20,484 vertices before removing non-cortical vertices). Six motion parameters and their derivatives, global signal, framewise displacement (Power et al., 2014), six principal components from cerebrospinal fluid and white matter (Behzadi et al., 2007), and polynomial trends up to second order were regressed out from both movie and localizer data for each run independently.

Searchlight Hyperalignment

Connectivity Hyperalignment (Step One)

Each participant's connectivity profile was built based on that participant's movie data. We first defined the connectivity seeds and targets. In this analysis, the connectivity seeds were the same as the surface cortical vertices. The connectivity targets were defined using a sparser cortical surface with 642 vertices in each hemisphere before removing the medial wall. We then centered a 13 mm searchlight on each of these vertices and computed the average time series for the searchlight over vertices from the denser cortical model. The mean time series was assigned to the center vertex to serve as the connectivity target. For each hemisphere, the connectivity profile was calculated as the correlation between the connectivity seeds in this hemisphere and the whole-brain 1175 connectivity targets. The connectivity profile of each participant was normalized to zero mean and unit variance for each connectivity seed before hyperalignment.

We used an optimized hyperalignment method that directly transforms one participant's connectivity profile to another participant's cortical space, without the interim step of projecting the connectome into a common model space (Jiahui et al., 2020). In detail, for each 15 mm searchlight, a participant's patterns of connectivity to targets were aligned to another participant's connectivity patterns using the Procrustes transformation. The transformation matrices from each searchlight in a hemisphere were then aggregated into a single transformation matrix for each pair of participants.

Response Hyperalignment

Response hyperalignment was applied with the same steps as the connectivity hyperalignment. The only difference is that instead of using connectivity profiles in each searchlight for each participant, we directly used the response pattern of the movie (time points of the movie × vertices in the searchlight) to align a pair of participants. In this method, response patterns in a pair of participants must be from neural responses to the same movie. Due to this restriction, response hyperalignment was only applied to participants from the same dataset.

Advanced Connectivity Hyperalignment

Using dense connectivity targets (e.g., using all 18742 vertices on the surface) with anatomically-aligned data usually generates poor functional correspondence across participants (Busch et al., 2021). It is, however, beneficial to include more targets for calculating connectivity patterns after the first iteration of connectivity hyperalignment and repeated iterations to lead to a better solution by gradually aligning the information at finer scales.

We used six steps to further improve the connectivity hyperalignment method. Step 1 was the initial connectivity hyperalignment step as described above that was based on the raw anatomically aligned movie data. The resultant transformation matrices were applied to those movie runs, and the hyperaligned data were then used in step 2 to calculate new connectivity patterns and calculate new transformation matrices. We repeated this procedure iteratively six times and derived transformation matrices for each step. In steps 1, 2, and 3, 642 × 2 (icoorder3, before removing the medial wall) connectivity targets were defined with 13 mm searchlights. In step 4 and 5, 2562 × 2 (icoorder 4, before removing the medial wall) connectivity targets were used with 7 mm searchlights to calculate target mean time series. In the final step 6, all 18742 vertices were included as separate connectivity targets, using each vertex's time series rather than calculating the mean in a searchlight. Each step of this advanced connectivity hyperalignment algorithm increased the prediction performance (Figure S7).

Predicting individual contrast maps

Estimating contrast maps from each participant's own localizer data

We estimated each participant's category-selective maps by calculating the unthresholded GLM univariate contrasts using his/her own localizer data in each run and averaging the t-values across all the localizer runs. We included face-, body-, scene-, and object-selective maps in the analysis. The contrast maps in each category were calculated based on the contrast of the target category vs. all the other categories. For example, the face-selective map was calculated using faces vs. all the other categories in the localizer data (e.g., bodies, objects).

Estimating contrast maps from other participants' localizer data

Transformation matrices from each participant to a target participant derived from hyperalignment were applied to the localizer runs of all other participants to project their localizer data into that target participant's cortical anatomy. These hyperaligned localizer runs and anatomical surface aligned localizer runs were used separately for GLM univariate analysis for each run in each other participant, and then averaged across the t-maps from all runs and all other participants to estimate the target participant's contrast maps for each category.

In summary, each participant's category-selective map was estimated based on that target participant's own localizer data and on all other participants' localizer data that was projected into that participant's cortical space using hyperalignment and anatomical surface alignment (see Fig. S1). After obtaining these estimated maps, we calculated correlations between the target participant's category-selective maps based on his/her own localizer data and the maps estimated from other participants' data (hyperaligned or anatomically-aligned). We also calculated Cronbach's alpha values (Feilong et al., 2018; Jiahui et al., 2020, 2022) across the multiple runs to measure the reliability of the category-selective maps for each participant and compared the correlations to the reliability values. To measure the local estimation performance and compare that to local reliabilities, we calculated correlations and Cronbach's alphas in searchlights with a radius of 15 mm.

Reference

Behzadi, Y., Restom, K., Liau, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage*, *37*(1), 90–101. https://doi.org/10.1016/j.neuroimage.2007.04.042

Busch, E. L., Slipski, L., Feilong, M., Guntupalli, J. S., Castello, M. V. di O., Huckins, J. F., Nastase, S. A., Gobbini, M. I., Wager, T. D., & Haxby, J. V. (2021). Hybrid hyperalignment: A single high-dimensional model of shared information embedded in cortical patterns of response and functional connectivity. *NeuroImage*, *233*, 117975. https://doi.org/10.1016/j.neuroimage.2021.117975

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

Epstein, R., Harris, A., Stanley, D., & Kanwisher, N. (1999). The Parahippocampal Place Area. *Neuron*, *23*(1), 115–125. https://doi.org/10.1016/S0896-6273(00)80758-8

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2019). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), 111. https://doi.org/10.1038/s41592-018-0235-4

Fedorenko, E., Hsieh, P.-J., Nieto-Castañón, A., Whitfield-Gabrieli, S., & Kanwisher, N. (2010). New Method for fMRI Investigations of Language: Defining ROIs Functionally in Individual Subjects. *Journal of Neurophysiology*, *104*(2), 1177–1194. https://doi.org/10.1152/jn.00032.2010

Feilong, M., Guntupalli, J. S., & Haxby, J. V. (2021). The neural basis of intelligence in fine-grained cortical topographies. *ELife*, *10*, e64058. https://doi.org/10.7554/eLife.64058

Feilong, M., Nastase, S. A., Guntupalli, J. S., & Haxby, J. V. (2018). Reliable individual differences in fine-grained cortical functional architecture. *NeuroImage*, *183*, 375–386. https://doi.org/10.1016/j.neuroimage.2018.08.029

Feilong, M., Nastase, S. A., Jiahui, G., Halchenko, Y. O., Gobbini, M. I., & Haxby, J. V. (2022). *The Individualized Neural Tuning Model: Precise and generalizable cartography of functional architecture in individual brains* (p. 2022.05.15.492022). bioRxiv. https://doi.org/10.1101/2022.05.15.492022

Fischl, B., Sereno, M. I., & Dale, A. M. (1999). Cortical Surface-Based Analysis: II: Inflation, Flattening, and a Surface-Based Coordinate System. *NeuroImage*, *9*(2), 195–207. https://doi.org/10.1006/nimg.1998.0396

Fox, C. J., Iaria, G., & Barton, J. J. S. (2009). Defining the face processing network: Optimization of the functional localizer in fMRI. *Human Brain Mapping*, *30*(5), 1637–1651. https://doi.org/10.1002/hbm.20630

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

Guntupalli, J. S., Feilong, M., & Haxby, J. V. (2018). A computational model of shared fine-scale structure in the human connectome. *PLOS Computational Biology*, *14*(4), e1006120. https://doi.org/10.1371/journal.pcbi.1006120

Guntupalli, J. S., Hanke, M., Halchenko, Y. O., Connolly, A. C., Ramadge, P. J., & Haxby, J. V. (2016). A Model of Representational Spaces in Human Cortex. *Cerebral Cortex*, *26*(6), 2919–2934. https://doi.org/10.1093/cercor/bhw068

Hanke, M., Adelhöfer, N., Kottke, D., Iacovella, V., Sengupta, A., Kaule, F. R., Nigbur, R., Waite, A. Q., Baumgartner, F., & Stadler, J. (2016). A studyforrest extension, simultaneous fMRI and eye gaze recordings during prolonged natural stimulation. *Scientific Data*, *3*, 160092. https://doi.org/10.1038/sdata.2016.92

Hanke, M., Baumgartner, F. J., Ibe, P., Kaule, F. R., Pollmann, S., Speck, O., Zinke, W., & Stadler, J. (2014). A high-resolution 7-Tesla fMRI dataset from complex natural stimulation with an audio movie. *Scientific Data*, *1*, 140003. https://doi.org/10.1038/sdata.2014.3

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

Jiahui, G., Feilong, M., Castello, M. V. di O., Nastase, S. A., Haxby, J. V., & Gobbini, M. I. (2022). *Modeling naturalistic face processing in humans with deep convolutional neural networks*. bioRxiv. https://doi.org/10.1101/2021.11.17.469009

Jiahui, G., Feilong, M., Visconti di Oleggio Castello, M., Guntupalli, J. S., Chauhan, V., Haxby, J. V., & Gobbini, M. I. (2020). Predicting individual face-selective topography using naturalistic stimuli. *NeuroImage*, *216*, 116458. https://doi.org/10.1016/j.neuroimage.2019.116458

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. *The Journal of Neuroscience*, *17*(11), 4302–4311.

Nastase, S. A. (2018). *The Geometry of Observed Action Representation During Natural Vision* [Ph.D., Dartmouth College]. https://www.proquest.com/docview/2018905893/abstract/F067E28624134EB2PQ/1

Pitcher, D., Dilks, D. D., Saxe, R. R., Triantafyllou, C., & Kanwisher, N. (2011). Differential selectivity for dynamic versus static information in face-selective cortical regions. *NeuroImage*, *56*(4), 2356–2363. https://doi.org/10.1016/j.neuroimage.2011.03.067

Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage*, *84*, 320–341. https://doi.org/10.1016/j.neuroimage.2013.08.048

Salehi, M., Greene, A. S., Karbasi, A., Shen, X., Scheinost, D., & Constable, R. T. (2020). There is no single functional atlas even for a single individual: Functional parcel definitions change with task. *NeuroImage*, *208*, 116366. https://doi.org/10.1016/j.neuroimage.2019.116366

Saxe, R., Brett, M., & Kanwisher, N. (2006). Divide and conquer: A defense of functional localizers. *NeuroImage*, *30*(4), 1088–1096. https://doi.org/10.1016/j.neuroimage.2005.12.062

Sengupta, A., Kaule, F. R., Guntupalli, J. S., Hoffmann, M. B., Häusler, C., Stadler, J., & Hanke, M. (2016). A studyforrest extension, retinotopic mapping and localization of higher visual areas. *Scientific Data*, *3*, 160093. https://doi.org/10.1038/sdata.2016.93

Shine, J. M., Bissett, P. G., Bell, P. T., Koyejo, O., Balsters, J. H., Gorgolewski, K. J., Moodie, C. A., & Poldrack,

R. A. (2016). The Dynamics of Functional Brain Networks: Integrated Network States during Cognitive Task Performance. *Neuron*, *92*(2), 544–554. https://doi.org/10.1016/j.neuron.2016.09.018

Telesford, Q. K., Lynall, M.-E., Vettel, J., Miller, M. B., Grafton, S. T., & Bassett, D. S. (2016). Detection of functional brain network reconfiguration during task-driven cognitive states. *NeuroImage*, *142*, 198–210. https://doi.org/10.1016/j.neuroimage.2016.05.078

Vanderwal, T., Eilbott, J., & Castellanos, F. X. (2019). Movies in the magnet: Naturalistic paradigms in developmental functional neuroimaging. *Developmental Cognitive Neuroscience*, *36*, 100600. https://doi.org/10.1016/j.dcn.2018.10.004

Vanderwal, T., Eilbott, J., Finn, E. S., Craddock, R. C., Turnbull, A., & Castellanos, F. X. (2017). Individual differences in functional connectivity during naturalistic viewing conditions. *NeuroImage*, *157*, 521–530. https://doi.org/10.1016/j.neuroimage.2017.06.027

Vanderwal, T., Kelly, C., Eilbott, J., Mayes, L. C., & Castellanos, F. X. (2015). Inscapes: A movie paradigm to improve compliance in functional magnetic resonance imaging. *NeuroImage*, *122*, 222–232. https://doi.org/10.1016/j.neuroimage.2015.07.069

Visconti di Oleggio Castello, M., Chauhan, V., Jiahui, G., & Gobbini, M. I. (2020). An fMRI dataset in response to “The Grand Budapest Hotel”, a socially-rich, naturalistic movie. *Scientific Data*, *7*(1), Article 1. https://doi.org/10.1038/s41597-020-00735-4

Zhen, Z., Kong, X.-Z., Huang, L., Yang, Z., Wang, X., Hao, X., Huang, T., Song, Y., & Liu, J. (2017). Quantifying the variability of scene-selective regions: Interindividual, interhemispheric, and sex differences. *Human Brain Mapping*, *38*(4), 2260–2275. https://doi.org/10.1002/hbm.23519

Zhen, Z., Yang, Z., Huang, L., Kong, X., Wang, X., Dang, X., Huang, Y., Song, Y., & Liu, J. (2015). Quantifying interindividual variability and asymmetry of face-selective regions: A probabilistic functional atlas. *NeuroImage*, *113*, 13–25. https://doi.org/10.1016/j.neuroimage.2015.03.010

Acknowledgements

This work was supported by NSF grants 1607845 (J.V.H) and 1835200 (M.I.G), and NIH grant R01 MH127199 (J.V.H & M.I.G).

Author contributions

Conceptualization, G.J., J.V.H., and M.I.G.; Methodology, G.J., M.F., J.V.H., and M.I.G.; Software, G.J. and M.F.; Formal Analysis, G.J.; Investigation, G.J.; Resources, J.V.H. and M.I.G.; Data Curation, M.F., and S.A.N.; Writing – Original Draft, G.J., J.V.H., and M.I.G.; Writing – Review & Editing, G.J., M.F., S.A.N., J.V.H., and M.I.G.; Visualization, G.J.; Supervision, J.V.H. and M.I.G.; Funding Acquisition, J.V.H. and M.I.G.

Competing interests

The authors declare no competing interests.

Supplementary Material

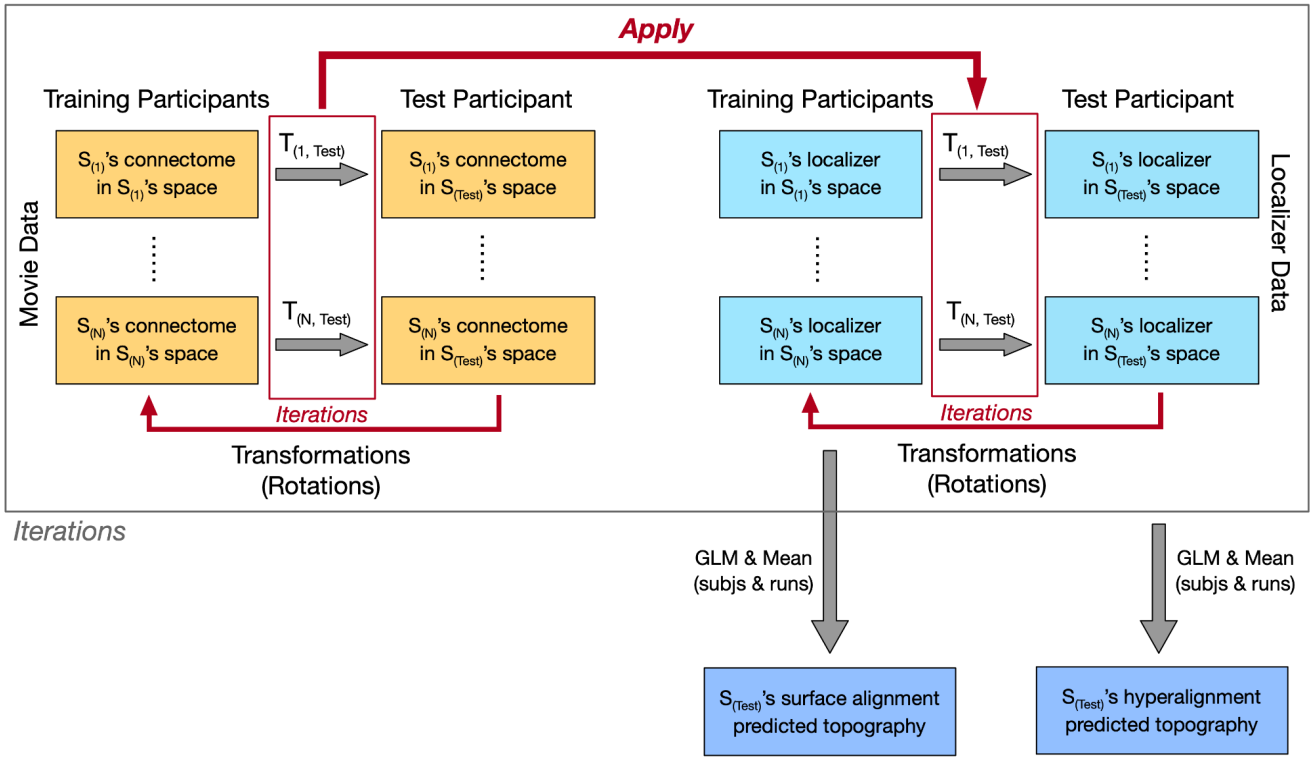

Figure S1. Schematic data analysis procedures. In the enhanced CHA analysis, transformation matrices derived from projecting connectome based on the movie data in each training participant's cortical space to the target participant's space were applied to each training participant's localizer runs. These steps were iterated six times, and in each step, the connectome and the localizer data were both updated. The original localizer runs were used to calculate category-selective topographies for each training participant and averaged across runs and participants to obtain the surface alignment predicted topography for the target participant. The localizer runs hyperaligned after all iteration steps were used to obtain CHA predicted topographies with similar procedures. Outside of this loop, each target participant' own original localizer runs were used to obtain this participant's own localizer estimated topographies.

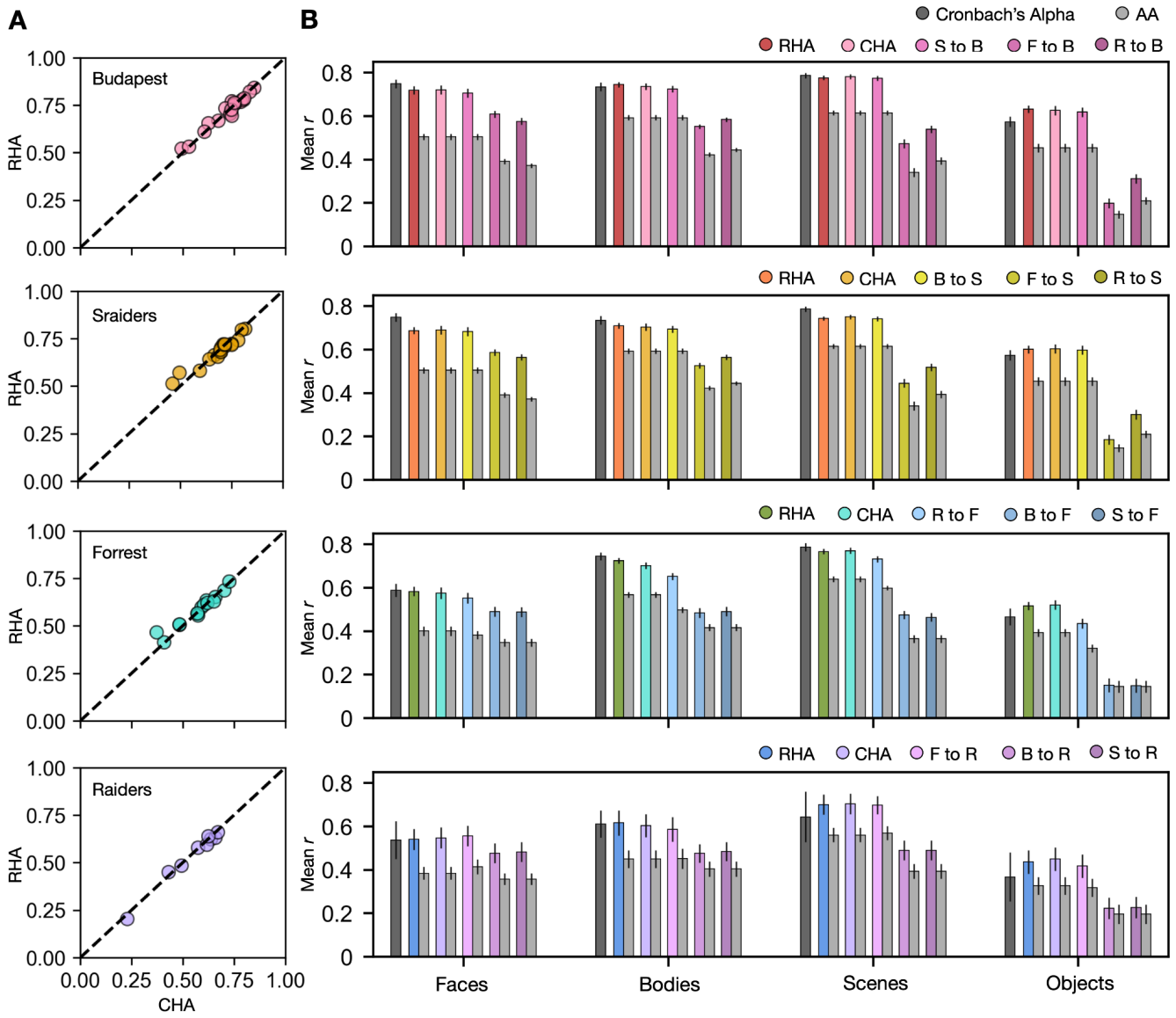

Figure S2. CHA predictions. A. Scatter plots of Pearson correlation coefficients using CHA and RHA for individual participants within four different datasets for the face-selective topography. Values on the y-axis stand for correlations between each target participant's own localizer-based topographies and topographies estimated from other participants in the same dataset using RHA. Values on the y-axis stand for correlations between each target participant's own localizer-based topographies and topographies estimated from other participants in the same dataset using CHA. **B.** Bar plots display the mean Pearson correlation coefficients (*r*) and Conbach's alphas across participants in all four datasets for all four categories. Error bars stand for ±1 standard error of the mean. The abbreviations are the same in all figures, including this one. S to B: Sraiders to Budapest, F to B: Forrest to Budapest, R to B: Raiders to Budapest, B to S: Budapest to Sraiders, F to S: Forrest to Sraiders, R to S: Raiders to Sraiders, R to F: Raiders to Forrest, B to F: Budapest to Forrest, S to F: Sraiders to Forrest, F to R: Forrest to Raiders, B to R: Budapest to Raiders, S to R: Sraiders to Raiders

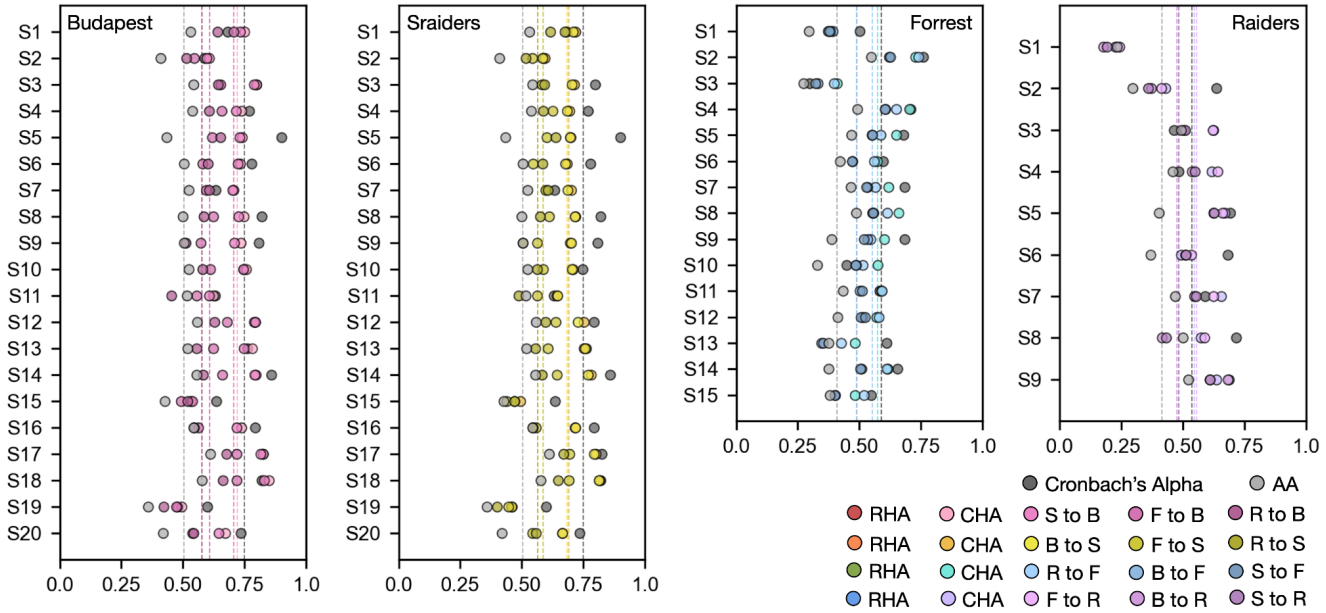

Figure S3. Prediction performances for each individual participant. Prediction performance (Pearson *r*) for the face-selective topography for each individual participant using RHA, CHA, cross-movie CHA, and AA in all four datasets. Black dots stand for individual participants' Cronbach's alphas of their own face-selective topographies across localizer runs. Dashed lines are the mean values across participants.

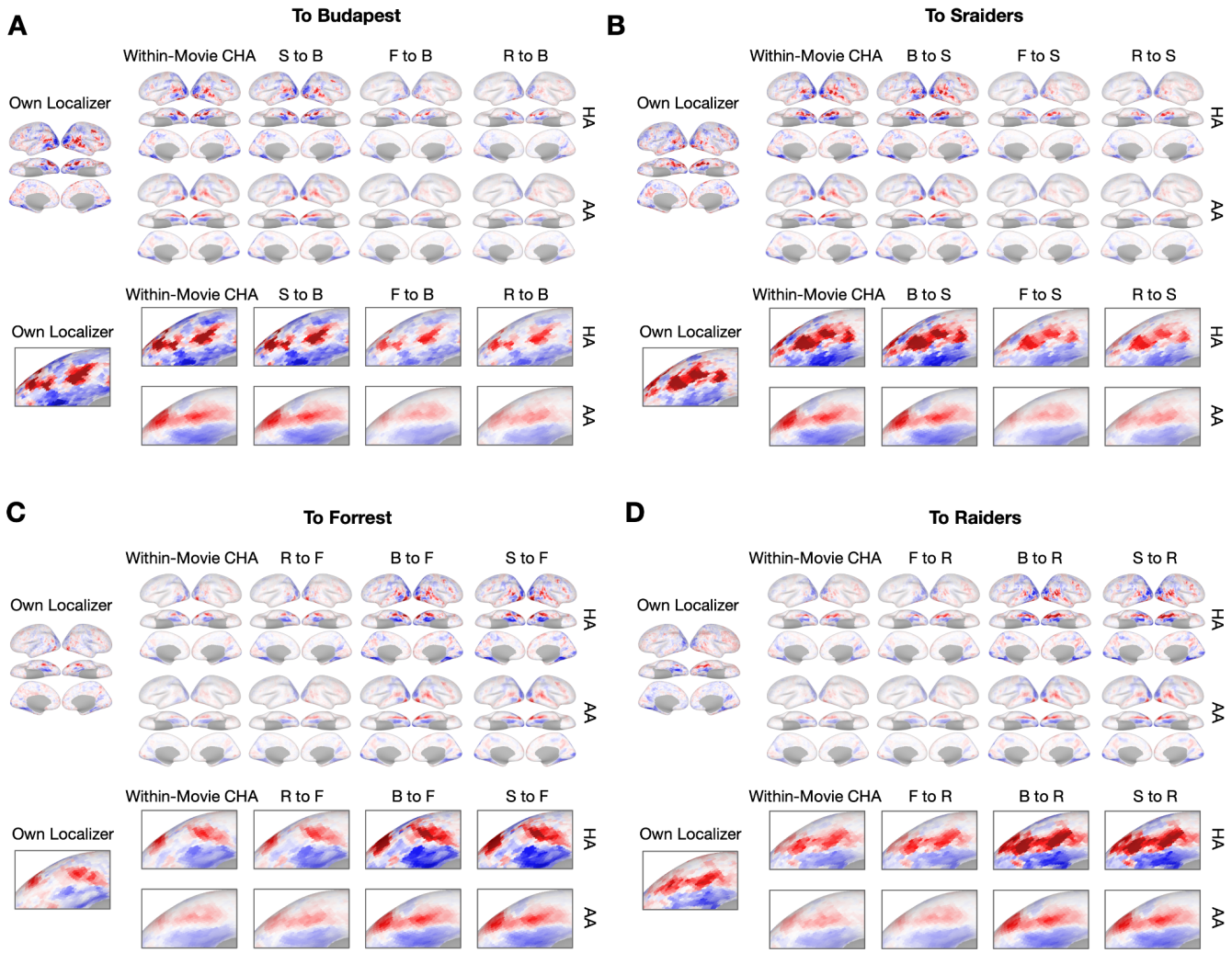

Figure S4. Sample contrast maps and enlarged views of the ventral temporal cortex. Contrast maps for face-selective topographies (faces-vs-all) and their enlarged views of the ventral temporal cortex were plotted in sample participants in **A.** Budapest, **B.** Sraiders, **C.** Forrest, and **D.** Raiders. In all five subplots for the whole-brain maps, the faces-vs-all maps were plotted on the sample participants' own cortical surfaces (left single panel). The second column presents predicted face-selective topographies from participants in the same dataset using connectivity hyperalignment (top) and surface alignment (bottom). The next three columns present face-selective topographies from participants in another dataset with the same (second column) and a different type (the last two columns) of localizers. In the four right columns, the top row presents the map using hyperalignment (HA), and the bottom row presents the map using surface alignment (AA). The enlarged panels were displayed accordingly with the whole-brain map. The color bar was the same as that in Figure 1.

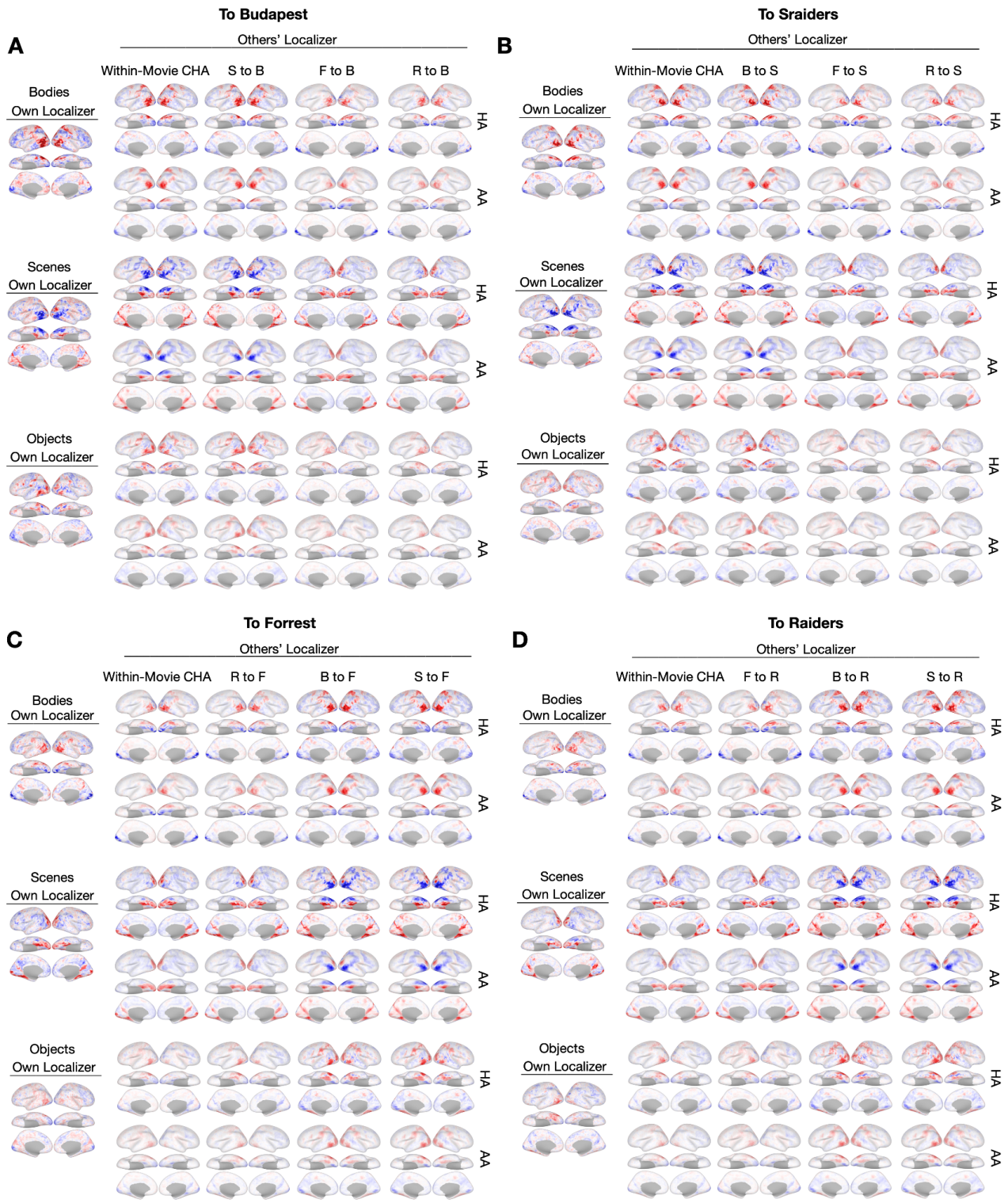

Figure S5. Sample contrast maps of body-, scene-, and object-selective topographies. Contrast maps of the other three categories were plotted in participants in **A.** Budapest, **B.** Sraiders, **C.** Forrest, and **D.** Raiders. In all five subplots for the whole-brain maps, the map estimated from their own localizer runs were plotted on the sample participant (left single panel). The other columns present predicted topographies from participants in the same dataset using connectivity hyperalignment (top) and surface alignment (bottom).

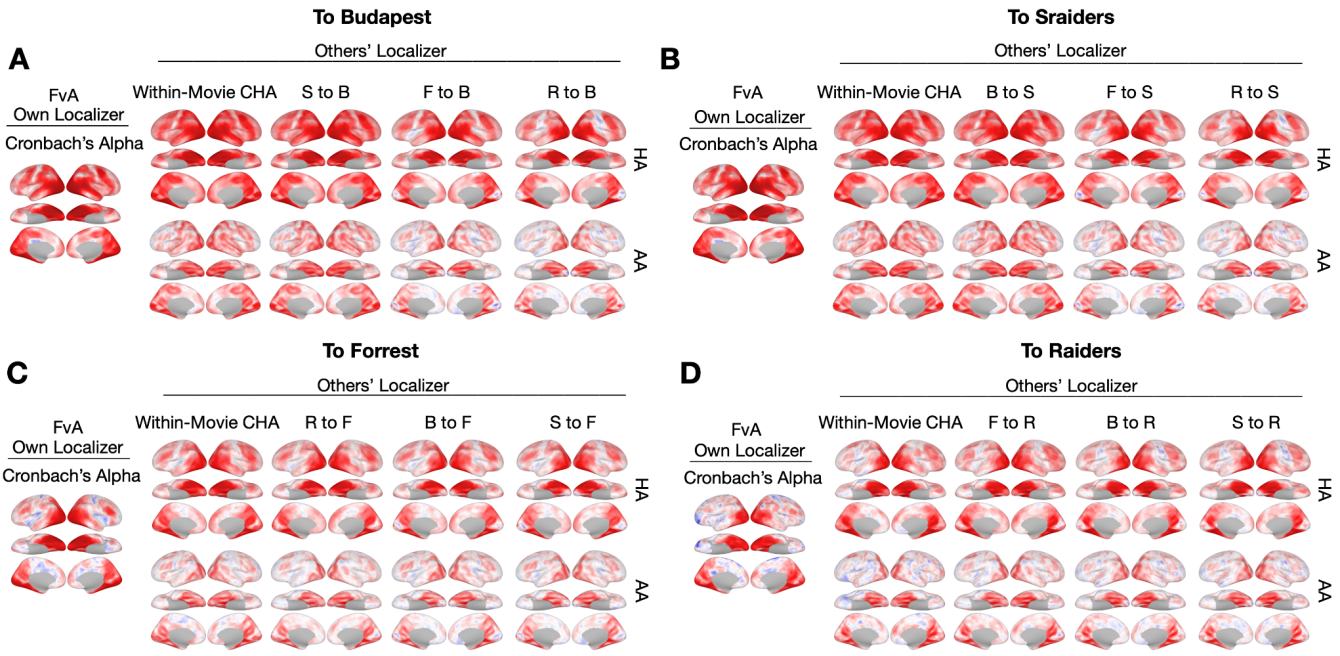

Figure S6. Searchlight correlations. A, B, C & D. The left-most column shows the Cronbach's alphas of the own localizer-based face-selective topographies in each dataset using a searchlight analysis (15 mm radius). The next four columns show local correlations (correlation maps) using the searchlight analysis between the face-selective maps estimated from other participants based on within-movie (second column, top row) and between-movie (the next three columns, top row) CHA and surface alignment (AA, bottom row).

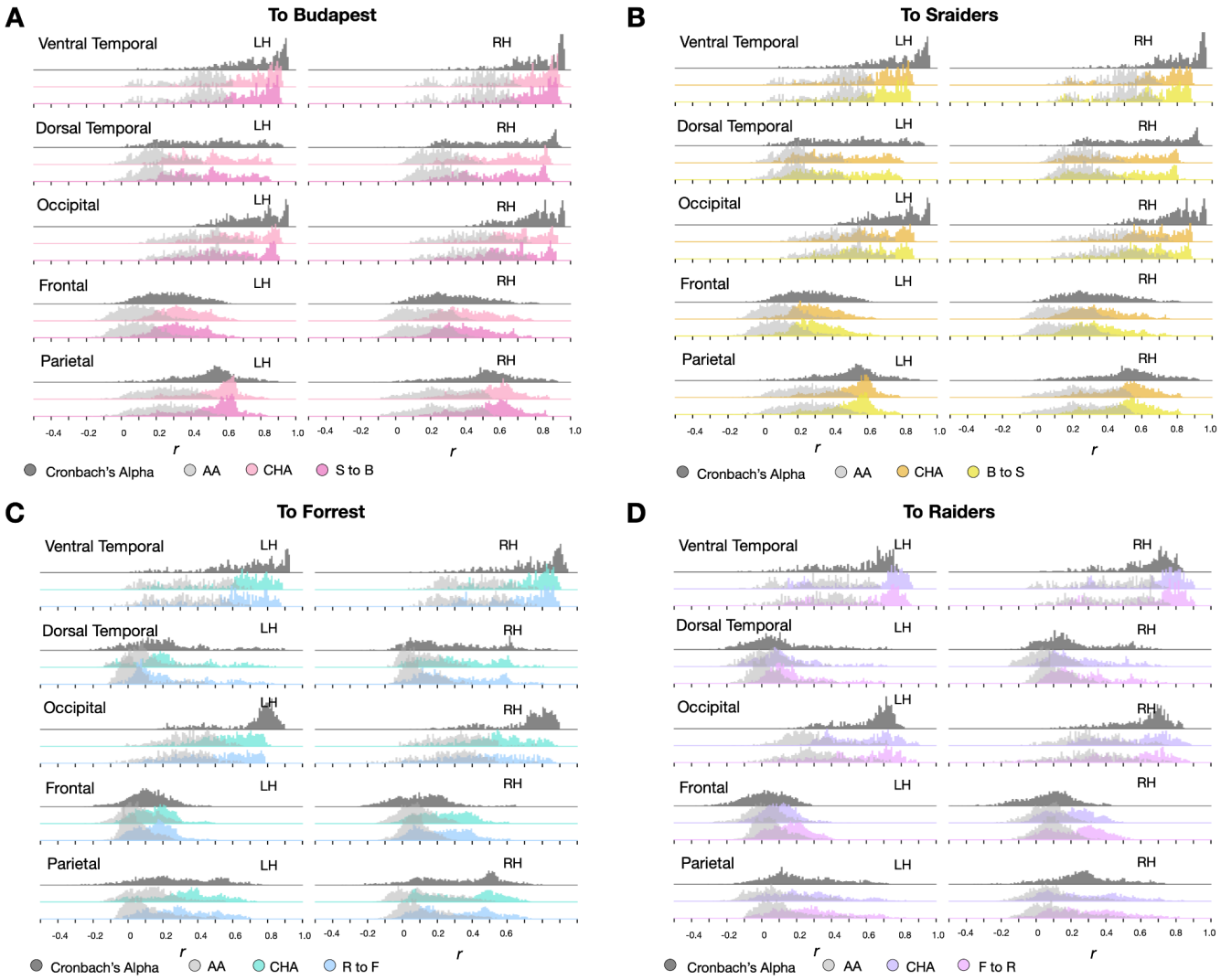

Figure S7. Distribution of correlation coefficients in major cortices. Histogram plots of local Cronbach's alphas (dark gray) and local correlation coefficients between face-selective topographies estimated from own and others' localizers across the cortex (HA in color, light gray for AA) in major cortices (ventral temporal, dorsal temporal, occipital, frontal, and parietal) in the four datasets (see Figure 4 for the whole brain maps and distributions). The left and right hemisphere histograms were plotted separately in each cortical parcel.

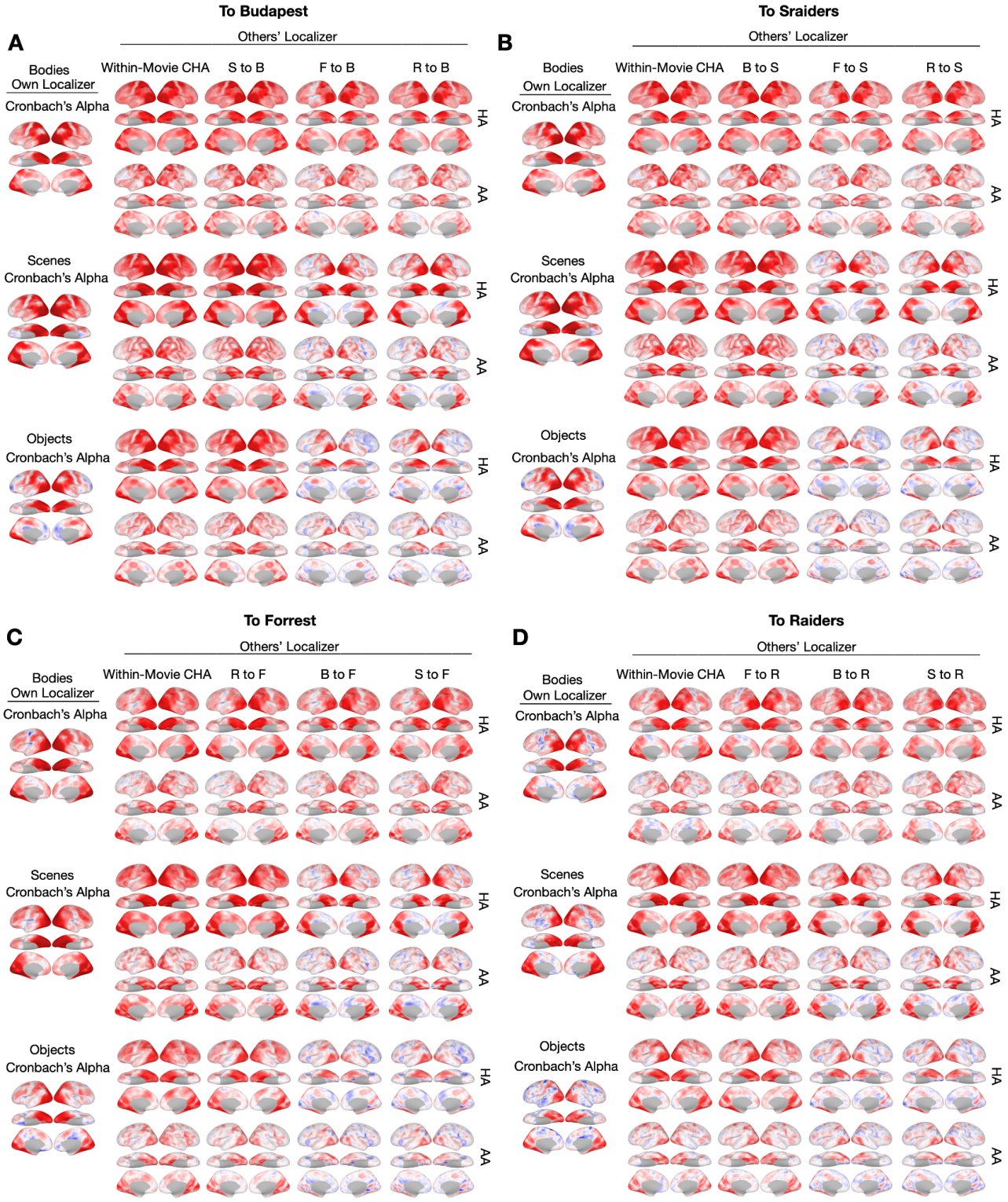

Figure S8. Searchlight analysis results for other categories. A, B, C & D. The left-most column shows the Cronbach's alphas of the own localizer-based topographies in each dataset using a searchlight analysis (15 mm radius). The next four columns show local correlations (correlation maps) using the searchlight analysis between the category-selective maps estimated from other participants based on within-movie (second column, top row) and between-movie (the next three columns, top row) CHA and surface alignment (AA, bottom row).

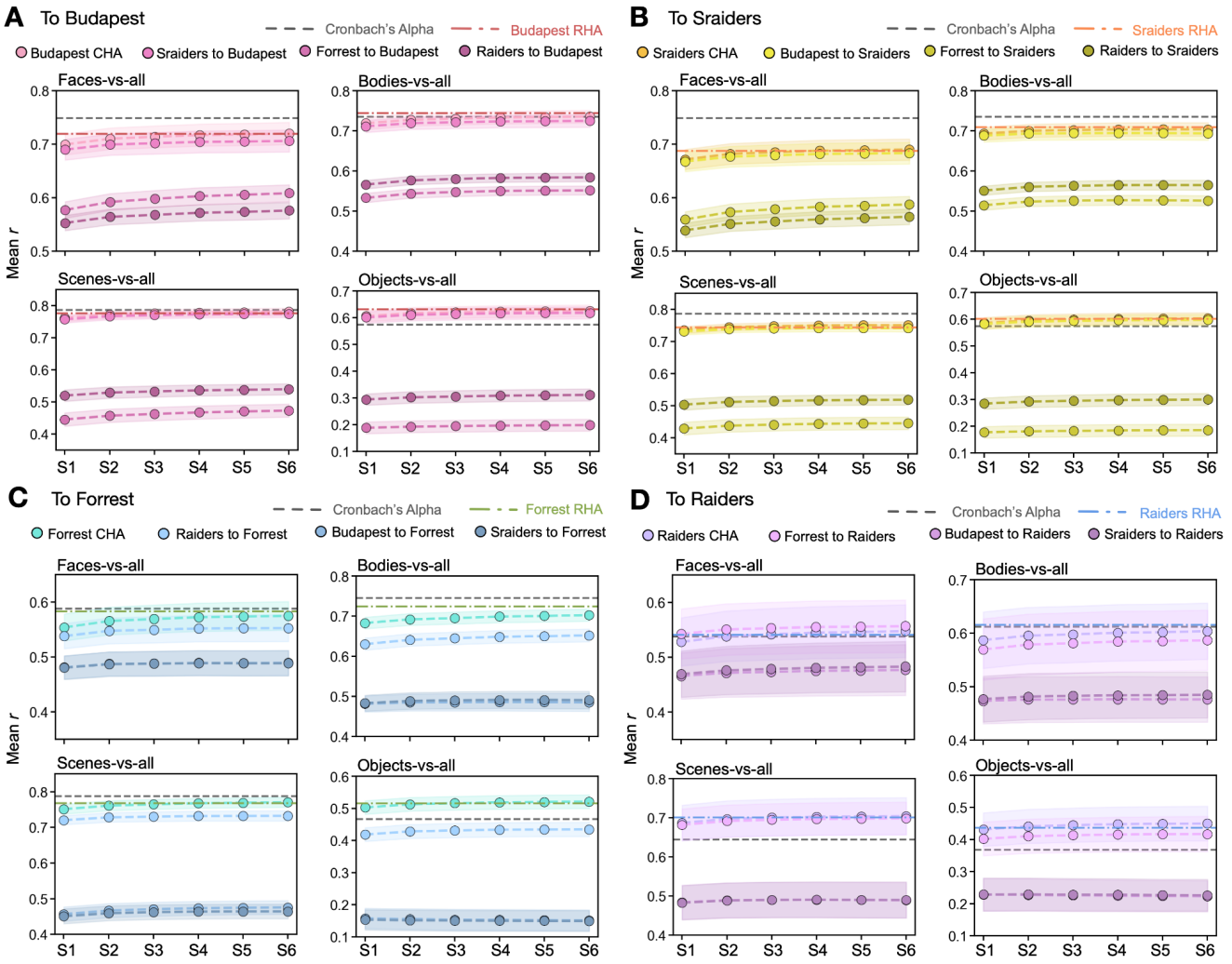

Figure S9. Advanced CHA improved prediction performances. A. In each subplot, each line with dots showed the improvement of the mean correlation across participants between the category-selective maps estimated from each participant's own localizer runs and those estimated from participants' data in other datasets from step 1 to step 6 using our new advanced iterative connectivity hyperalignment algorithm. Horizontal dotted lines are the mean Cronbach's alphas (gray) and the mean performance using RHA (colored). **B, C,** and **D** had the same layout as A with participants in Sraiders, Forrest, and Raiders dataset as the prediction target.

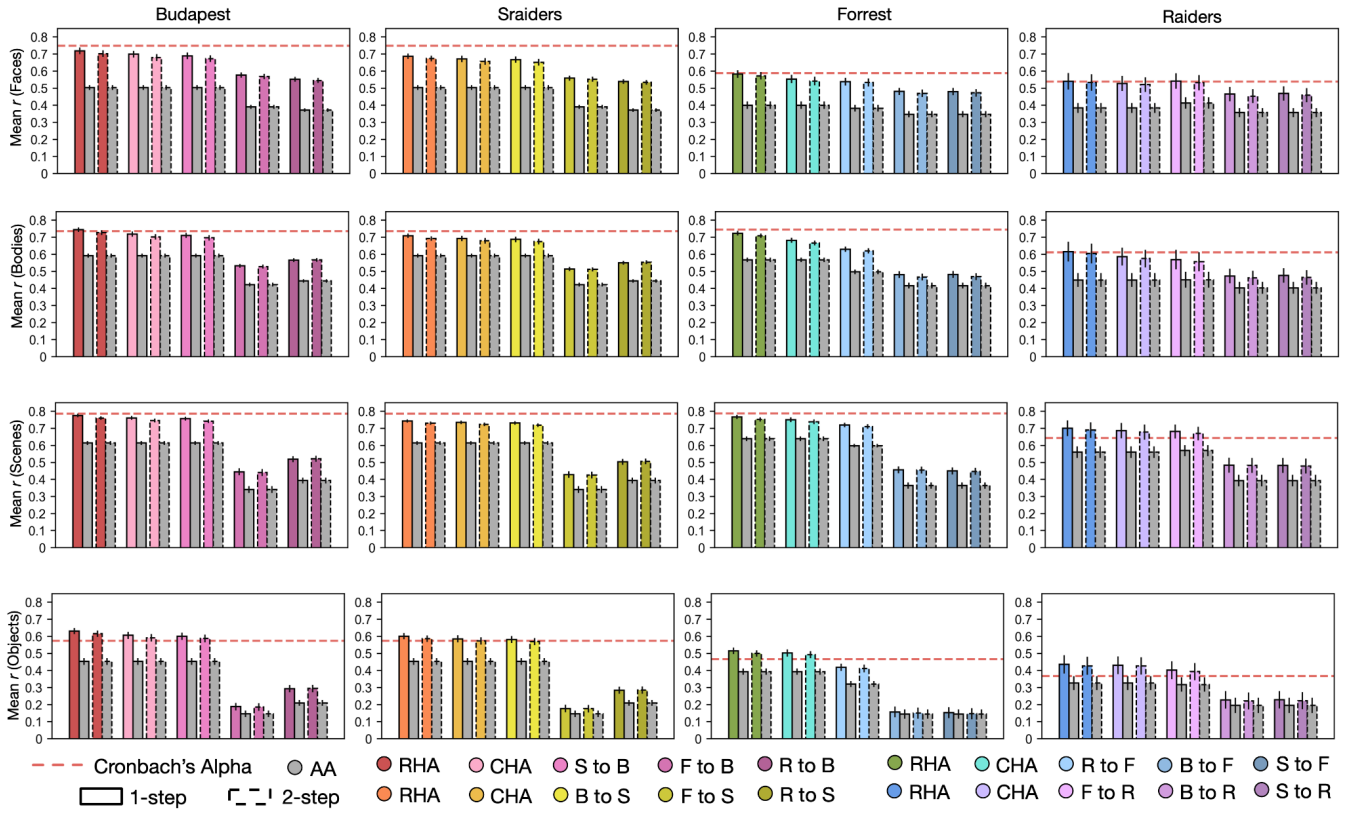

Figure S10. Predictions based on the 1-step and the 2-step methods. Bar plots display the mean Pearson correlation coefficients (*r*) and Conbach's alphas across participants in all four datasets for all four categories. Bars with solid outlines stand for results based on the 1-step method, and bars with dashed outlines are based on the 2-step method. Error bars stand for ±1 standard error of the mean.

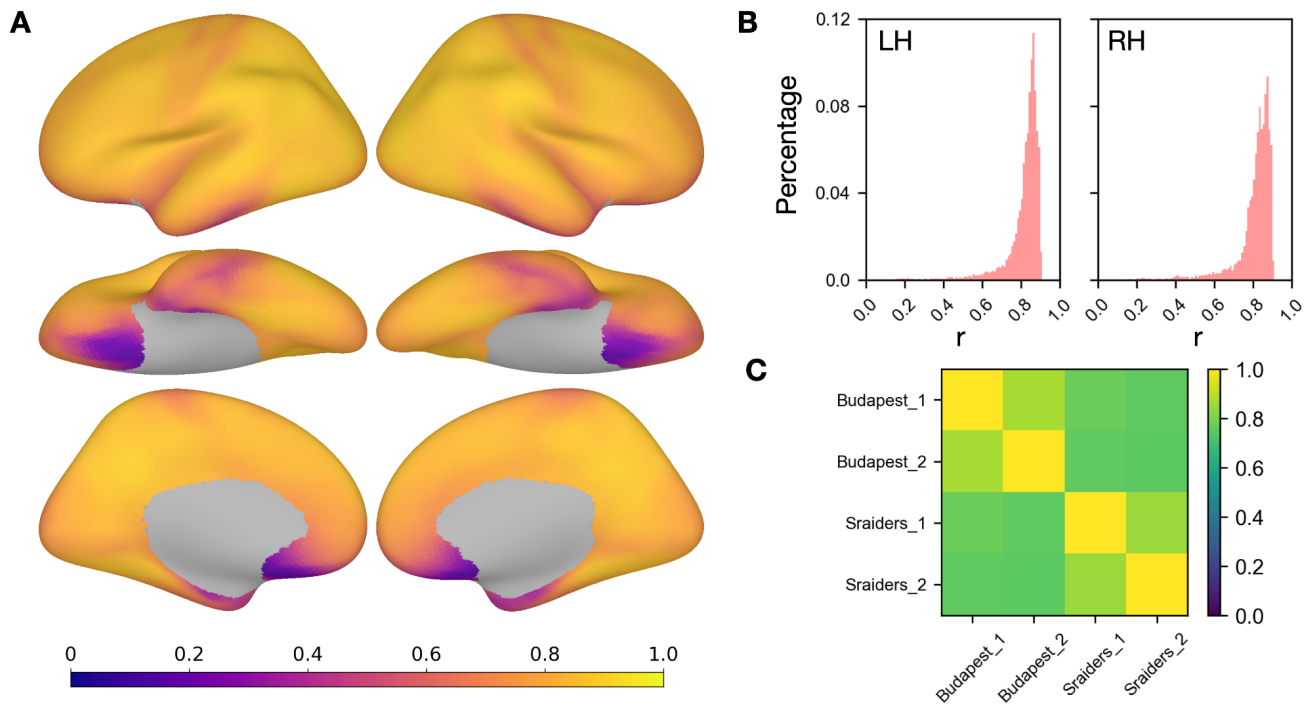

Figure S11. Similarities between fine-grained connectivities in two different movie-viewing tasks. **A.** For each participant in the Budapest and the Sraiders dataset, connectomes that described the connectivity between each target in the searchlight and each vertex on the cortex were calculated for each dataset and correlated across the two datasets. This whole-brain map shows the mean correlations across participants. **B.** Histogram plots of the correlation coefficients in A in the left and the right hemisphere. **C.** The two datasets were split into two halves, and similarities of the fine-grained connectivity between the two halves within and across the two movies were calculated following the same procedures above. This plot shows the mean correlations across participants and searchlights.