A Appendix

In the following sections, we provide some additional material to ease the understanding of the underlying alignment problem as well as computational details of solutions of FUGW and FUGW barycenters.

We also show some control experiments during which we used different training data to compute pair-wise alignments and evaluated the proportion of correlation gains that comes from mere signal smoothing.

Eventually, we give details about the IBC dataset (acquisition, preprocessing, fMRI protocols and data splitting).

A.1 Illustration of the alignment problem

We provide in Fig. S1 a conceptual illustration of the alignment framework for a pair of subjects.

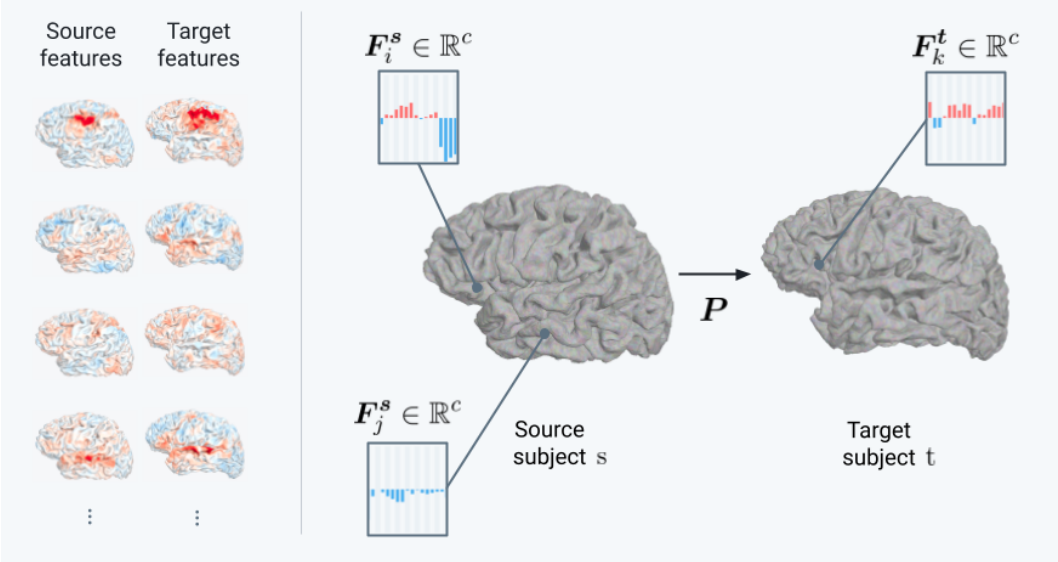

Figure S1: Alignment of two brains using functional signatures Using multiple maps of comparable features (left column) for the source and target subjects, we seek to derive an alignment (also referred to as a *coupling*) $\boldsymbol{P}$ that matches parts of the brain with similar features while preserving the global geometry of the cortex.

A.2 Detailed description of FUGW estimation algorithm

Estimating the unbalanced Gromov Wasserstein (UGW) loss is numerically sensitive to initialization, due to the non-convexity of the problem [39]. Therefore FUGW is *a priori* non-convex as well, and comparably difficult to estimate. Consequently, following [39], we instead compute a lower bound which we formulate as a bi-convex problem that relies on the joint estimation of two couplings. We provide next a detailed derivation of this estimation procedure, using notations introduced in section 2:

$$\text{FUGW}(\boldsymbol{\mathcal{X}^s}, \boldsymbol{\mathcal{X}^t}) = \inf_{\substack{\boldsymbol{P},\boldsymbol{Q}\geq 0 \\ \boldsymbol{P}=\boldsymbol{Q}}} \text{L}_\theta(\boldsymbol{P},\boldsymbol{Q}) \quad \geq \inf_{\substack{\boldsymbol{P},\boldsymbol{Q}\geq 0 \\ m(\boldsymbol{P})=m(\boldsymbol{Q})}} L_\theta(\boldsymbol{P},\boldsymbol{Q}) \quad \triangleq \quad \text{LB-FUGW}(\boldsymbol{\mathcal{X}^s}, \boldsymbol{\mathcal{X}^t}). \tag{S1}$$

where $m(\boldsymbol{P}) = \sum_{i,j} \boldsymbol{P}_{i,j}$ denotes the mass of $\boldsymbol{P}$.

We give the explicit formulation of $\text{L}_\theta(\boldsymbol{P},\boldsymbol{Q})$

$$\text{L}_\theta(\boldsymbol{P},\boldsymbol{Q}) \triangleq (1-\alpha)\ \text{L}_\text{W}(\boldsymbol{P},\boldsymbol{Q}) + \alpha\ \text{L}_\text{GW}(\boldsymbol{P},\boldsymbol{Q}) + \rho\ \text{L}_\text{U}(\boldsymbol{P},\boldsymbol{Q}) + \varepsilon\ E(\boldsymbol{P},\boldsymbol{Q}), \tag{S2}$$

where

- $\boldsymbol{C} \triangleq \left(||\boldsymbol{F}_i^{\boldsymbol{s}} - \boldsymbol{F}_j^{\boldsymbol{t}}||_2^2\right)_{i,j} \in \mathbb{R}_+^2$ (feature cost matrix)
- $\boldsymbol{G} \triangleq \left(|\boldsymbol{D}_{i,j}^{\boldsymbol{s}} - \boldsymbol{D}_{k,l}^{\boldsymbol{t}}|\right)_{i,j,k,l} \in \mathbb{R}_+^4$ (geometry cost matrix)
- $\mathrm{L_W}(\boldsymbol{P}, \boldsymbol{Q}) \triangleq \langle \boldsymbol{C}, \frac{\boldsymbol{P}+\boldsymbol{Q}}{2} \rangle = \frac{1}{2}(\sum_{i,j} \boldsymbol{C}_{i,j}\boldsymbol{P}_{i,j} + \sum_{i,j} \boldsymbol{C}_{i,j}\boldsymbol{Q}_{i,j})$ (Wasserstein)
- $\mathrm{L_{GW}}(\boldsymbol{P}, \boldsymbol{Q}) \triangleq \langle \boldsymbol{G}, \boldsymbol{P} \otimes \boldsymbol{Q} \rangle = \sum_{i,j,k,l} \boldsymbol{G}_{i,j,k,l}\boldsymbol{P}_{i,j}\boldsymbol{Q}_{k,l}$ (Gromov-Wasserstein)
- $\mathrm{L_U}(\boldsymbol{P}, \boldsymbol{Q}) \triangleq \ \mathrm{KL}\Big(\boldsymbol{P}_{\#1} \otimes \boldsymbol{Q}_{\#1} | \boldsymbol{w}^s \otimes \boldsymbol{w}^s\Big) + \ \mathrm{KL}\Big(\boldsymbol{P}_{\#2} \otimes \boldsymbol{Q}_{\#2} | \boldsymbol{w}^t \otimes \boldsymbol{w}^t\Big)$ (unbalancing)
- $E(\boldsymbol{P}, \boldsymbol{Q}) \triangleq \mathrm{KL}\Big(\boldsymbol{P} \otimes \boldsymbol{Q} | (\boldsymbol{w}^{\boldsymbol{s}} \otimes \boldsymbol{w}^{\boldsymbol{t}}) \otimes (\boldsymbol{w}^{\boldsymbol{s}} \otimes \boldsymbol{w}^{\boldsymbol{t}})\Big)$ (entropy)

In particular, we have $\mathrm{L}_\theta(\boldsymbol{P}, \boldsymbol{P}) = \mathrm{L}_\theta(\boldsymbol{P})$, which is the objective function of FUGW introduced in Equation 1.

Solver We provide a Python GPU-based solver for LB-FUGW, using an approach similar to that of [39], which we recall in algorithm S1. More precisely, we alternatively optimize one coupling while keeping the other fixed. This results in two entropic unbalanced OT problems in each iteration, which can be solved using the scaling algorithm [11].

Algorithm S1 Approximation scheme for LB-FUGW

Input: $\boldsymbol{\mathcal{X}}^{\boldsymbol{s}}, \boldsymbol{\mathcal{X}}^{\boldsymbol{t}}, \rho, \alpha, \varepsilon$.
Output: Pair of optimal couplings $(\boldsymbol{P}, \boldsymbol{Q})$.

Initialize: $\boldsymbol{P} = \boldsymbol{Q} = \boldsymbol{w}^{\boldsymbol{s}} \otimes \boldsymbol{w}^{\boldsymbol{t}} / \sqrt{m(\boldsymbol{w}^{\boldsymbol{s}}) m(\boldsymbol{w}^{\boldsymbol{t}})}$.
while $(\boldsymbol{P}, \boldsymbol{Q})$ has not converged **do**
  Calculate: $c_P = \mathrm{Cost}(\boldsymbol{P}, \boldsymbol{G}, \boldsymbol{C}, \boldsymbol{w}^{\boldsymbol{s}}, \boldsymbol{w}^{\boldsymbol{t}}, \rho, \alpha, \varepsilon)$.
  Update: $\boldsymbol{Q} \leftarrow \mathrm{Scaling}(c_P, \boldsymbol{w}^{\boldsymbol{s}}, \boldsymbol{w}^{\boldsymbol{t}}, \rho m(\boldsymbol{P}), \varepsilon m(\boldsymbol{P}))$.
  Rescale: $\boldsymbol{Q} \leftarrow \sqrt{\frac{m(\boldsymbol{P})}{m(\boldsymbol{Q})}} \boldsymbol{Q}$.
  Calculate: $c_Q = \mathrm{Cost}(\boldsymbol{Q}, \boldsymbol{G}, \boldsymbol{C}, \boldsymbol{w}^{\boldsymbol{s}}, \boldsymbol{w}^{\boldsymbol{t}}, \rho, \alpha, \varepsilon)$.
  Update: $\boldsymbol{P} \leftarrow \mathrm{Scaling}(c_Q, \boldsymbol{w}^{\boldsymbol{s}}, \boldsymbol{w}^{\boldsymbol{t}}, \rho m(\boldsymbol{Q}), \varepsilon m(\boldsymbol{Q}))$.
  Rescale: $\boldsymbol{P} \leftarrow \sqrt{\frac{m(\boldsymbol{Q})}{m(\boldsymbol{P})}} \boldsymbol{P}$.
end while

Algorithm S2 Scaling algorithm [11]

Input: $\boldsymbol{C}, \boldsymbol{w}^{\boldsymbol{s}}, \boldsymbol{w}^{\boldsymbol{t}}, \rho, \varepsilon$.
Output: Optimal coupling $\boldsymbol{P}$.

Initialize dual vectors: $f = 0_n \in \mathbb{R}^n, g = 0_p \in \mathbb{R}^p$.
while (f, g) has not converged **do**
  Update: $f = -\frac{\rho}{\rho+\varepsilon} \log \sum_j \exp\left(g_j + \log \boldsymbol{w}_j^{\boldsymbol{t}} - \frac{\boldsymbol{C}_{\cdot,j}}{\varepsilon}\right)$.
  Update: $g = -\frac{\rho}{\rho+\varepsilon} \log \sum_i \exp\left(f_i + \log \boldsymbol{w}_i^{\boldsymbol{s}} - \frac{\boldsymbol{C}_{i,\cdot}}{\varepsilon}\right)$.
end while
Calculate: $\boldsymbol{P} = (\boldsymbol{w}^{\boldsymbol{s}} \otimes \boldsymbol{w}^{\boldsymbol{t}}) \exp\left(f \oplus g - \frac{\boldsymbol{C}}{\varepsilon}\right)$.

Here, the notations $\otimes$ and $\oplus$ denote the Kronecker product and sum, respectively. The exponential, division and logarithm operations are all element-wise. The scalar product is denoted by $\langle \cdot, \cdot \rangle$.

In practice, we observe that the two couplings of LB-FUGW are numerically equal, so it is enough to choose, for example, the first one, as alignment between source and target signals.

A.3 Detailed description of FUGW barycenter estimation

FUGW-barycenter algorithm, described in Algorithm S4, alternates between computing mappings from subjects to the barycenter, and updating the barycenter. This corresponds to a block coordinate descent on the barycenter estimation. The first step simply uses the previously introduced solver. The

Algorithm S3 Cost

Input: $\boldsymbol{P}, \boldsymbol{G}, \boldsymbol{C}, \boldsymbol{w^s}, \boldsymbol{w^t}, \rho, \alpha, \varepsilon$.
Output: Local cost c.

1: Calculate: $\boldsymbol{G} \otimes \boldsymbol{P} := \left(\sum_{i,j} \boldsymbol{G}_{i,j,k,l} \boldsymbol{P}_{i,j}\right)_{k,l}$.
2: Calculate:

$$c := \alpha\, \boldsymbol{G} \otimes \boldsymbol{P} + \frac{1-\alpha}{2}\, \boldsymbol{C} + \rho\, \langle \log \frac{\boldsymbol{P}_{\#1}}{\boldsymbol{w^s}}, \boldsymbol{P}_{\#1} \rangle + \rho\, \langle \log \frac{\boldsymbol{P}_{\#2}}{\boldsymbol{w^t}}, \boldsymbol{P}_{\#2} \rangle + \varepsilon\, \langle \log \frac{\boldsymbol{P}}{\boldsymbol{w^s} \otimes \boldsymbol{w^t}}, \boldsymbol{P} \rangle$$

second one takes advantage of the fact that the objective function introduced in S1 is differentiable in $\boldsymbol{F^B}$ and $\boldsymbol{D^B}$, and the two couplings of LB-FUGW are numerically equal. This yields a closed form for $\boldsymbol{F^B}$ and $\boldsymbol{D^B}$, as a function of $\boldsymbol{P^{s,B}}$ and $\boldsymbol{\mathcal{X}^s}$. We note that, during the barycenter estimation, the weight $\boldsymbol{w^B}$ is always fixed as uniform distribution.

Algorithm S4 LB-FUGW barycenter

Input: $(\boldsymbol{\mathcal{X}^s})_{s \in \mathcal{S}}, \rho, \alpha, \varepsilon$.
Output: Individual couplings $(\boldsymbol{P^{s,B}})_{s \in \mathcal{S}}$, barycenter $\boldsymbol{\mathcal{X}^B}$.

1: Initialize: $\boldsymbol{F^B} = \mathbb{1}_{\boldsymbol{k}}$; $\boldsymbol{D^B} = \mathbb{0}_{\boldsymbol{k}}$.
2: **while** $\boldsymbol{\mathcal{X}^B} = (\boldsymbol{F^B}, \boldsymbol{D^B}, \boldsymbol{w^B})$ has not converged **do**
3: Draw $\widetilde{S}$ subset of S.
4: **for** $s \in \widetilde{S}$ **do** ▷ fixed $\boldsymbol{\mathcal{X}^B}$
5: Align: $\boldsymbol{P^{s,B}} \leftarrow$ LB-FUGW$(\boldsymbol{\mathcal{X}^s}, \boldsymbol{\mathcal{X}^B}, \rho, \alpha, \varepsilon)$.
6: **end for**
7: Update $\boldsymbol{F^B}$, $\boldsymbol{D^B}$: ▷ fixed $\boldsymbol{P^{s,B}}$

$$\boldsymbol{F^B} = \frac{1}{|\widetilde{S}|} \sum_{s \in \widetilde{S}} \text{diag}\left(\frac{1}{\boldsymbol{P^{s,B}_{\#2}}}\right) (\boldsymbol{P^{s,B}})^\top \boldsymbol{F^s} \quad \text{and} \quad \boldsymbol{D_B} = \frac{1}{|\widetilde{S}|} \sum_{s \in \widetilde{S}} \frac{(\boldsymbol{P^{s,B}})^\top \boldsymbol{D^s} \boldsymbol{P^{s,B}}}{\boldsymbol{P^{s,B}_{\#2}} (\boldsymbol{P^{s,B}_{\#2}})^\top}.$$

8: **end while**

A.4 Implementation details

MSM configuration We use the default configuration of MSM [3] and vary parameter *lambda* so as to obtain the best gains in correlation on the test set. We use the same value of *lambda* at each step of MSM and eventually set it to 0.1 after a cross validated grid search.

Correlation gain variability when aligning pairs of subjects Figures S2 and S3 show correlation gains on the validation and test sets respectively when aligning pairs of subjects from the IBC dataset. Subjects' data was previously projected onto *fsaverage5*. These figures provide us with a better understanding of the standard error and consistency of these gains. Moreover, they show that selection of the best set of hyper-parameters is robust to changing the validation data.

[3] MSM default configuration `https://github.com/ecr05/MSM_HOCR/blob/master/config/basic_configs/config_standard_MSMpair`

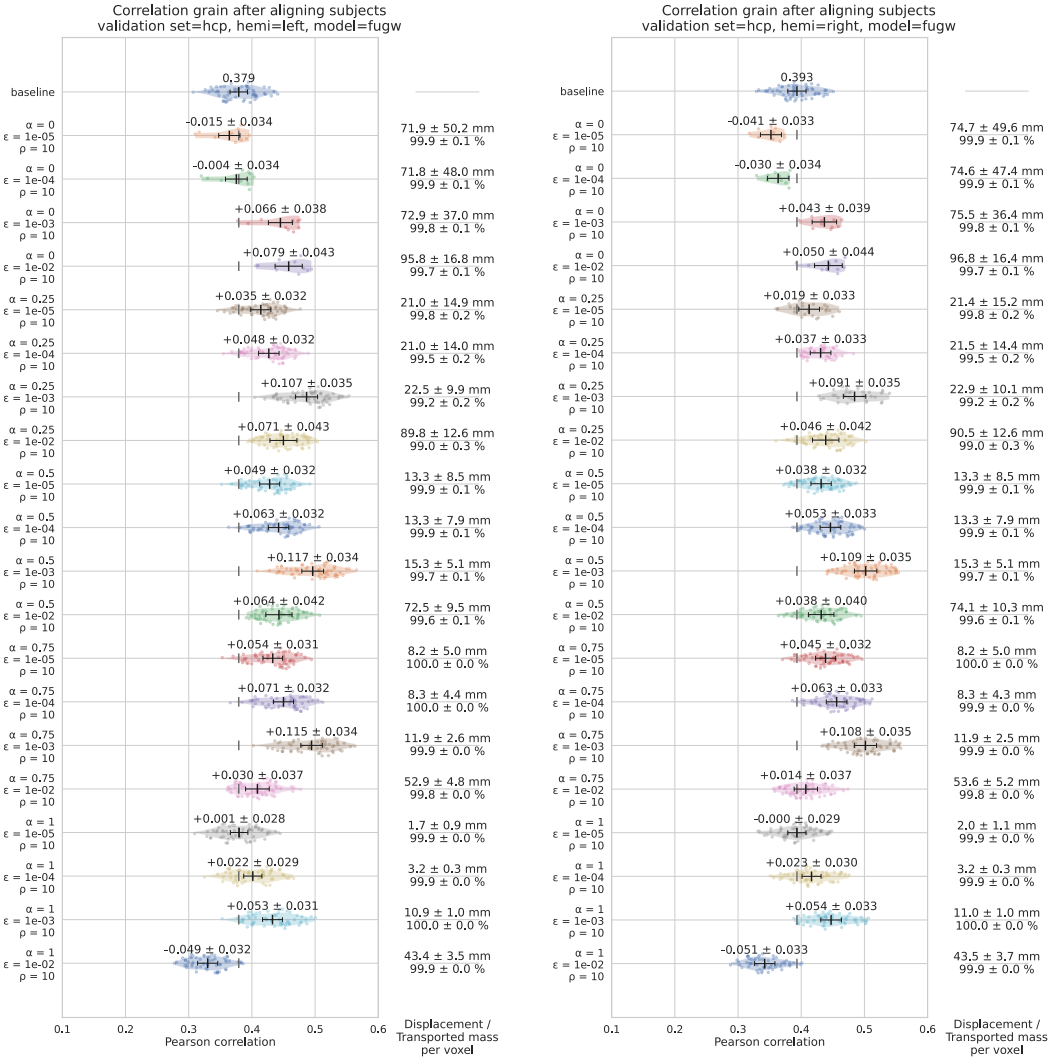

Figure S2: Detailed correlation gains on the validation set (HCP tasks), in the balanced case Each line represents a FUGW model trained with different hyper-parameters. Each dot represents the mean correlation between contrast maps of the HCP protocol for a given pair of IBC subjects. We compare the average correlation with that of the baseline (top row) where subjects were simply projected on *fsaverage5*. Models for the left hemisphere and right hemisphere are shown respectively on the left and right side.

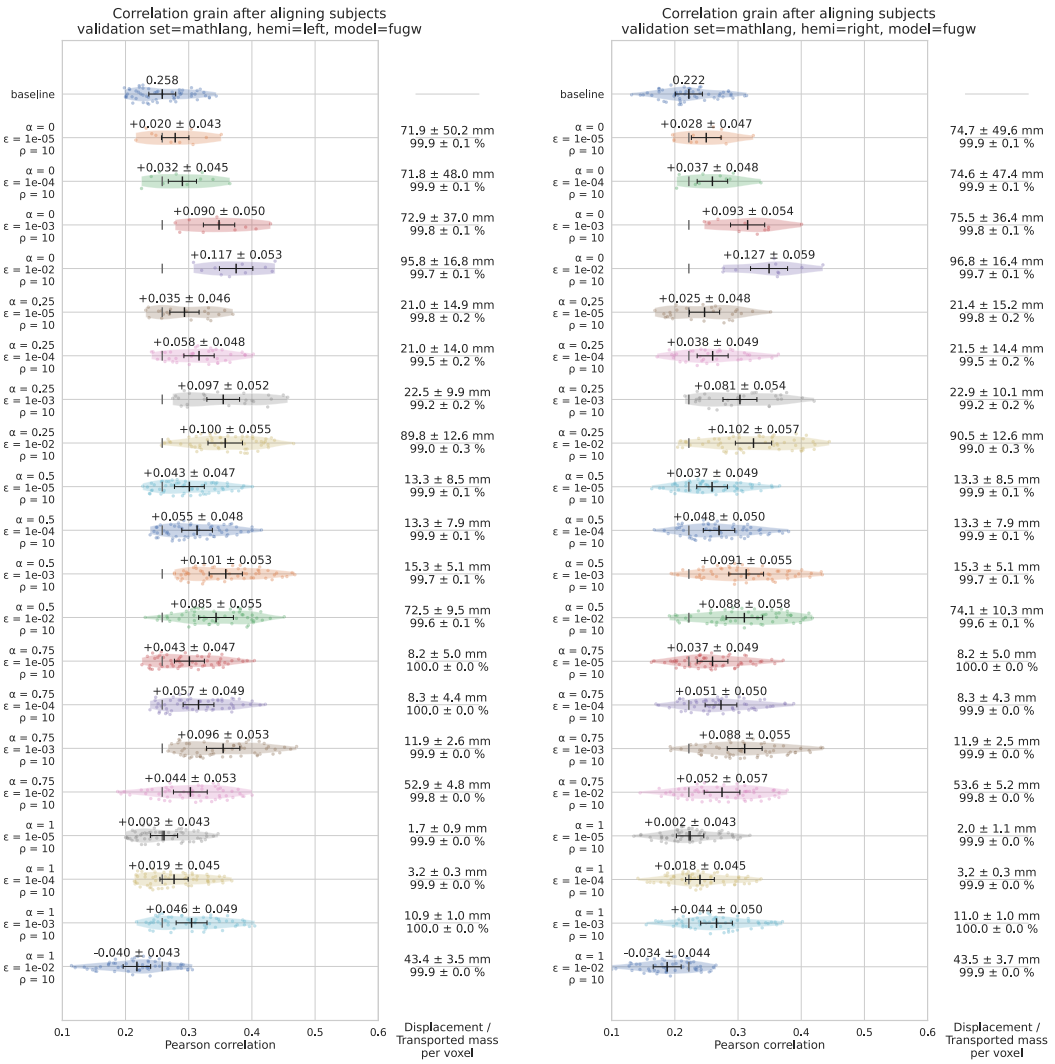

Figure S3: Detailed correlation gains on the test set (Mathlang tasks) in the balanced case Similarly to Figure S2, each line represents a FUGW model trained with different hyper-parameters. Each dot represents the mean correlation between contrast maps of the Mathlang protocol for a given pair of IBC subjects. We compare the average correlation with that of the baseline (top row) where subjects were simply projected on *fsaverage5*. Models for the left hemisphere and right hemisphere are shown respectively on the left and right side.

Mesh resolution reduction As mentioned in the core of this paper, aligning meshes with high resolutions can lead to dealing with matrices which won't fit on GPUs. This is typically the case when trying to align two *fsaverage7* hemispheres (160k vertices each) instead of *fsaverage5* hemispheres (10k vertices each).

In order to reduce the number n of aligned vertices, we first group them into small clusters using Ward's algorithm using a method described in [41]. In essence, this method iteratively groups adjacent vertices of a given individual based on feature similarity until I clusters have been formed. Then, for a given cluster u_i of the source subject s, we define its functional signal $\hat{\boldsymbol{F}}^{\boldsymbol{s}}_{u_i}$ as the mean functional signal of vertices which belong to this cluster. Moreover, for two given clusters u_i and u_j of subject s, we define the anatomical distance $\hat{\boldsymbol{D}}^{\boldsymbol{s}}_{u_i,u_j}$ between u_i and u_j as the mean geodesic distance between all pairs of vertices between the two clusters (akin to an Energy distance). Eventually, we derive analogous objects $\hat{\boldsymbol{F}}^{\boldsymbol{t}}$ and $\hat{\boldsymbol{D}}^{\boldsymbol{t}}$ for the target subject t, and end up in a configuration comparable to that of Experiment 1.

$$\hat{\boldsymbol{F}}^{\boldsymbol{s}}_{u_i} \triangleq \frac{1}{|u_i|} \sum_{k \in u_i} \boldsymbol{F}^{\boldsymbol{s}}_k \in \mathbb{R}^c \qquad \hat{\boldsymbol{D}}^{\boldsymbol{s}}_{u_i,u_j} \triangleq \frac{1}{|u_i| \cdot |u_j|} \sum_{k \in u_i, l \in u_j} \boldsymbol{D}^s_{k,l}$$

Alignment to individual anatomy We qualitatively control that alignments derived between individuals on their individual anatomies make sense in Figure S4.

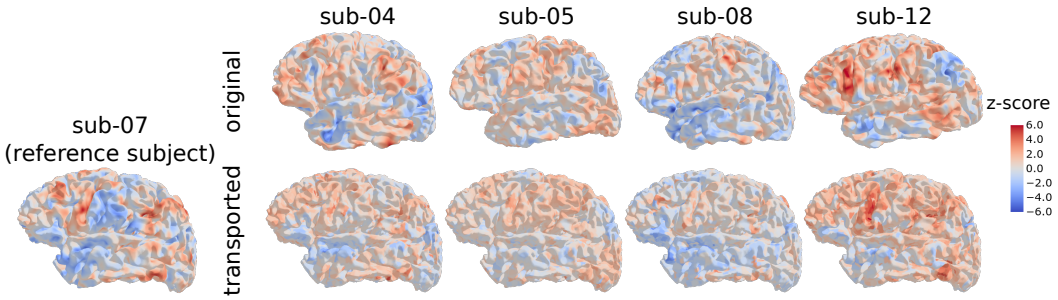

Figure S4: Transporting individual maps onto a reference subject FUGW can help bridge the absence of template anatomies and derive pairs of alignments such that all individuals of the cohort are comparable. We display a map taken from the test set contrasting areas activated during mathematical reasoning against areas activated for other stimuli of the protocol.

A.5 Control experiments

Controlling for smoothing effect increasing correlation Alignments computed with FUGW are not always vertex-to-vertex alignments. Indeed, a single vertex from the source subject s can be associated with many vertices in the target subject t. In fact, $\boldsymbol{P}^{\boldsymbol{s},\boldsymbol{t}}_i$ represents the relative importance of each match. The hyper-parameter ε controls the entropy of $\boldsymbol{P}^{\boldsymbol{s},\boldsymbol{t}}$, which is in direct link with the spread of vertices that we use as a measure for how many target vertices are matched with source vertices.

Since smoothing signal on the source subject can reduce noise and increase correlation to target data, we measure the correlation gain induced by applying a gaussian kernel to the source signal. This allows us to show that only a minor proportion of correlation gains induced by FUGW can come from this smoothing effect. Figure S5 shows this for kernels of 5mm, 10mm, 15mm and 20mm of standard deviation respectively. We see that correlation increases significantly less than when using FUGW (0.03 vs 0.12 correlation gain respectively). Moreover, one notices that even though correlation increases for pairs of subjects with a low initial correlation, it decreases for pairs with a high initial correlation. On the contrary, FUGW increases correlation for all pairs of subjects.

Different training sets yield comparable correlation gains While we leverage all IBC maps to derive our couplings, we show that the presented results hold when using a much smaller training dataset. In particular, we observe similar correlation gains when using only the 57 maps of the Archi protocol for training (see Table S3). It takes about one hour per subject to acquire these maps, which we advocate is a reasonable amount of time to build a training set dedicated to align subjects within a given cohort (and possibly across cohorts). Finally, we train both FUGW and MSM with lower-dimensional versions of the previous datasets. To do so, given a pair of subjects (s, t) to be aligned, we fit a PCA on the left out subjects, project the data of subjects to be aligned on these components, and keep the first 20 components only. For both models, correlation gains remained unchanged.

More explicitly, we test the 4 following training sets:

- ALL-MATH: all contrast maps of IBC except contrasts from the Mathlang protocol (369 features per subject)

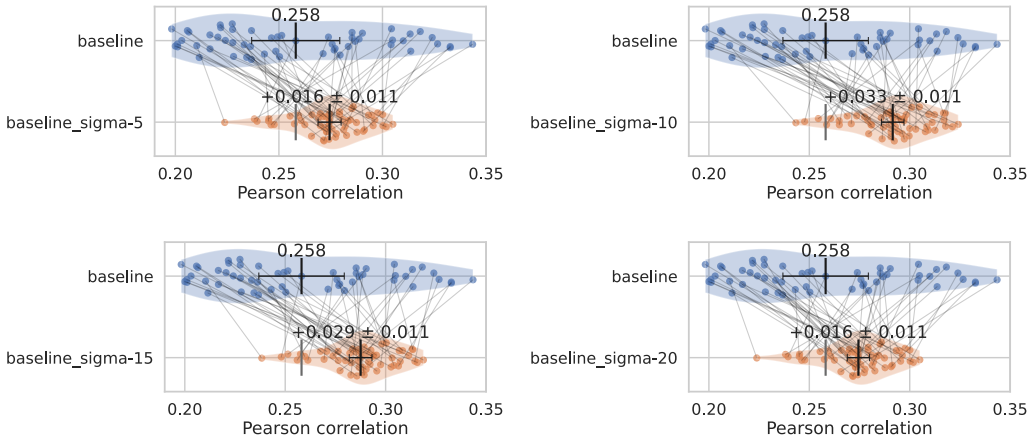

Figure S5: Comparison of gains in correlation after Gaussian blurring We compare correlation between subjects after the source subject's functional data has been smoothed with a Gaussian kernel of standard deviation 5mm (top left), 10mm (top right), 15mm (bottom left) and 20mm (bottom right)

- ALL-MATH PCA: principal components fitted on ALL-MATH for all IBC subjects except s and t (20 features)
- ARCHI: all contrast maps from the Archi protocol of IBC (57 features)
- ARCHI PCA: principal components fitted on ARCHI for all subjects except s and t (20 features)

Training set	FUGW	MSM
ALL-MATH	0.12	0.01
ALL-MATH PCA	0.11	0.02
ARCHI	0.10	0.02
ARCHI PCA	0.11	0.01

Table S1: Gain in Pearson correlation of aligned contrast maps from the Mathlang protocol compared to the baseline The original correlation (baseline) is 0.258

Using naturalistic stimuli to derive alignments with FUGW This experiment's setup is similar to that of Experiment 1: Using training features to first derive OT couplings, we then use the latter to assess correlation gains between subjects's feature maps before and after subjects have been aligned. Naturalistic stimuli datasets include *Raiders of the Lost Ark*, short video clips and auditory stimuli from *The Little Prince* respectively adapted from [21, 27, 8]. Here, for each naturalistic dataset, we leverage work from [34] to derive the first $m = 20$ components of a fitted shared response model. Share response models seek to find a common dictionary $\boldsymbol{K}$ of activation patterns across subjects $s \in S$ and to derive a mapping $\boldsymbol{W^s}$ with m orthogonal components that projects each individual's data onto this common space:

$$\underset{\substack{\boldsymbol{K},\boldsymbol{W^s} \\ \text{s.t. } (\boldsymbol{W^s})^T \cdot \boldsymbol{W^s} = \boldsymbol{I_m}}}{\arg\min} \sum_{s \in S} ||\boldsymbol{F^s} - \boldsymbol{W^s}\boldsymbol{K}||^2$$

These $\boldsymbol{W^s}$ are then used for alignment.

Results are reported in Table S2 and show that using datasets which are 20 times less time-consuming than that of Experiment 1 can already yield significant correlation gain on unseen task data.

Transporting myelin maps shows mild effect Leveraging transport plans computed using fMRI data from Experiment 1, we transport myelin maps – approximated through T1 / T2 ratio maps – from the source subject to the target subject. We compare the correlation of the unaligned source

Table S2: Acquisition time (AT) and correlation gain on the left hemisphere (CG) per training set (baseline correlation = 0.258)

Training set	Type	AT (min)	CG (Pearson)
All-MATH	tasks	2000	0.118
Clips	movie	100	0.017
Raiders	movie	115	0.046
The Little Prince	movie	100	0.009

and target maps with the correlation of the transported and target maps. As illustrated in Figure S6, correlation gain is barely significant.

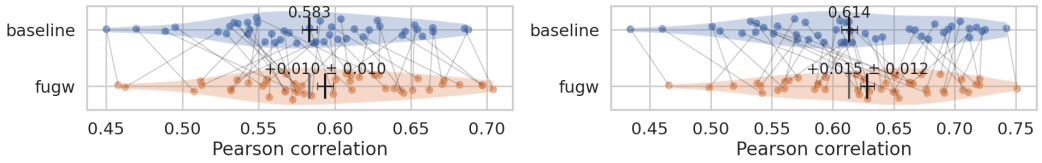

Figure S6: Comparison of gains in correlation after inter-subject alignment for myelin maps For each pair of source and target subjects of the IBC dataset, we compute the average Pearson correlation between myelin maps – approximated using T1/T2 ratios – for the left (left panel) and right (right panel) hemispheres. Correlation gains are not significant.

Before computing correlation between aforementioned maps, we discarded vertices located in the *cortical wall*, as they mostly contain spurious values. To do so, we they their value to the median of values of vertices which do not belong to the cortical wall. In order to determine which vertices belong to the *wall*, we used the Destrieux atlas [13].

Eventually, we advocate that little gain can be obtained when better aligning myelin maps, since they are already very stable across human subjects as shown in Figure S7.

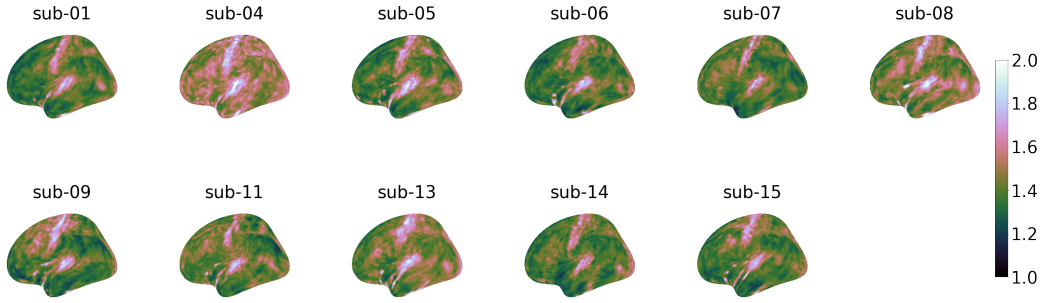

Figure S7: Myelin maps (approximated using T1/T2 ratio maps) are very consistent across IBC participants

A.6 Dataset description

The presented experiments rely on the Individual Brain Charting (IBC) dataset. A detailed description of the preprocessing pipeline of the IBC data is provided in [32]. Raw data were preprocessed using *PyPreprocess* [4].

All fMRI images, i.e. GE-EPI volumes, were collected twice with reversed phase-encoding directions, resulting in pairs of images with distortions going in opposite directions. Susceptibility-induced off-resonance field was estimated from the two Spin-Echo EPI volumes in reversed phase-encoding

[4] `https://github.com/neurospin/pypreprocess`

directions. The images were corrected based on the estimated deformation model. Details about the method can be found in [3].

Further, the GE-EPI volumes were aligned to each other within every participant. A rigid-body transformation was employed, in which the average volume of all images was used as reference [17]. The anatomical and motion-corrected fMRI images were given as input to *FreeSurfer* v6.0.0, in order to extract meshes of the tissue interfaces and the sampling of functional activation on these meshes, as described in [42]. The corresponding maps were then resampled to the fsaverage7 (high resolution, 163k nodes per hemisphere) and fsaverage5 (low resolution, 10k nodes per hemisphere) templates of FreeSurfer [16].

FMRI data were analyzed using the *General Linear Model*. Regressors of the model were designed to capture variations in BOLD response strictly following stimulus timing specifications. They were estimated through the convolution of boxcar functions, that represent per-condition stimulus occurrences, with the canonical *Hemodynamic Response Function* (HRF). To build such models, paradigm descriptors grouped in triplets (i.e. onset time, duration and trial type) according to BIDS Specification were determined from the log files' registries generated by the stimulus-delivery software. To account for small fluctuations in the latency of the HRF peak response, additional regressors were computed based on the convolution of the same task-conditions profile with the time derivative of the HRF. Nuisance regressors were also added to the design matrix in order to minimize the final residual error. To remove signal variance associated with spurious effects arising from movements, six temporal regressors were defined for the motion parameters. Further, the first five principal components of the signal, extracted from voxels showing the 5% highest variance, were also regressed to capture physiological noise [7].

In addition, a discrete-cosine basis was included for high-pass filtering (*cutoff*=$\frac{1}{128}$Hz). Model specification was implemented using *Nilearn* v0.8.1 [2], a Python library for statistical learning on neuroimaging data (`https://nilearn.github.io`).

In tables S3, S4 and S5, we give the explicit list of contrast and condition maps used for training, validation and testing respectively.

Table S3: Training data This dataset comprises a wide variety of tasks: motor, visual, auditory, relational, linguistic, etc.

Training data	
Task	Condition / Contrast
ArchiEmotional	expression_control
ArchiEmotional	expression_gender
ArchiEmotional	expression_gender-control
ArchiEmotional	expression_intention
ArchiEmotional	expression_intention-control
ArchiEmotional	expression_intention-gender
ArchiEmotional	face_control
ArchiEmotional	face_gender
ArchiEmotional	face_gender-control
ArchiEmotional	face_trusty
ArchiEmotional	face_trusty-control
ArchiEmotional	face_trusty-gender
ArchiEmotional	trusty_and_intention-control
ArchiEmotional	trusty_and_intention-gender
ArchiSocial	false_belief-mechanistic
ArchiSocial	false_belief-mechanistic_audio
ArchiSocial	false_belief-mechanistic_video
ArchiSocial	false_belief_audio
ArchiSocial	false_belief_video
ArchiSocial	mechanistic_audio
ArchiSocial	mechanistic_video
ArchiSocial	non_speech_sound
ArchiSocial	speech-non_speech
ArchiSocial	speech_sound
ArchiSocial	triangle_mental
ArchiSocial	triangle_mental-random
ArchiSocial	triangle_random
ArchiSpatial	grasp-orientation
ArchiSpatial	hand-side
ArchiSpatial	object_grasp
ArchiSpatial	object_orientation
ArchiSpatial	rotation_hand
ArchiSpatial	rotation_side
ArchiSpatial	saccades
ArchiStandard	audio_computation
ArchiStandard	audio_left_button_press
ArchiStandard	audio_right_button_press
ArchiStandard	audio_sentence
ArchiStandard	cognitive-motor
ArchiStandard	computation
ArchiStandard	computation-sentences
ArchiStandard	horizontal-vertical
ArchiStandard	horizontal_checkerboard
ArchiStandard	left-right_button_press
ArchiStandard	listening-reading
ArchiStandard	motor-cognitive
ArchiStandard	reading-checkerboard
ArchiStandard	reading-listening
ArchiStandard	right-left_button_press
ArchiStandard	sentences
ArchiStandard	sentences-computation
ArchiStandard	vertical-horizontal

Training data (next)	
Task	Condition / Contrast
ArchiStandard	vertical_checkerboard
ArchiStandard	video_computation
ArchiStandard	video_left_button_press
ArchiStandard	video_right_button_press
ArchiStandard	video_sentence
Attention	double_congruent
Attention	double_cue
Attention	double_incongruent
Attention	double_incongruent-double_congruent
Attention	incongruent-congruent
Attention	spatial_congruent
Attention	spatial_cue
Attention	spatial_cue-double_cue
Attention	spatial_incongruent
Attention	spatial_incongruent-spatial_congruent
Audi	alphabet
Audi	alphabet-silence
Audi	animals
Audi	animals-silence
Audi	cough
Audi	cough-silence
Audi	environment
Audi	environment-silence
Audi	human
Audi	human-silence
Audi	laugh
Audi	laugh-silence
Audi	music
Audi	music-silence
Audi	reverse
Audi	reverse-silence
Audi	silence
Audi	speech
Audi	speech-silence
Audi	suomi
Audi	suomi-silence
Audi	tear
Audi	tear-silence
Audi	yawn
Audi	yawn-silence
Audio	animal
Audio	animal-others
Audio	animal-silence
Audio	mean-silence
Audio	music
Audio	music-others
Audio	music-silence
Audio	nature
Audio	nature-others
Audio	nature-silence
Audio	speech
Audio	speech-others
Audio	speech-silence
Audio	tool
Audio	tool-others

Training data (next)

Task	Condition / Contrast
Audio	tool-silence
Audio	voice
Audio	voice-others
Audio	voice-silence
Bang	no_talk
Bang	talk
Bang	talk-no_talk
ColumbiaCards	gain
ColumbiaCards	loss
ColumbiaCards	num_loss_cards
Discount	amount
Discount	delay
DotPatterns	correct_cue-incorrect_cue
DotPatterns	correct_cue_correct_probe
DotPatterns	correct_cue_incorrect_probe
DotPatterns	correct_cue_incorrect_probe-correct_cue_correct_probe
DotPatterns	correct_cue_incorrect_probe-incorrect_cue_correct_probe
DotPatterns	cue
DotPatterns	incorrect_cue_correct_probe
DotPatterns	incorrect_cue_incorrect_probe
DotPatterns	incorrect_cue_incorrect_probe-correct_cue_incorrect_probe
DotPatterns	incorrect_cue_incorrect_probe-incorrect_cue_correct_probe
DotPatterns	incorrect_probe-correct_probe
EmotionalPain	emotional-physical_pain
EmotionalPain	emotional_pain
EmotionalPain	physical_pain
Enumeration	enumeration_constant
Enumeration	enumeration_linear
Enumeration	enumeration_quadratic
Lec1	pseudoword
Lec1	pseudoword-random_string
Lec1	random_string
Lec1	word
Lec1	word-pseudoword
Lec1	word-random_string
Lec2	attend
Lec2	attend-unattend
Lec2	unattend
MCSE	high-low_salience
MCSE	high_salience_left
MCSE	high_salience_right
MCSE	low+high_salience
MCSE	low-high_salience
MCSE	low_salience_left
MCSE	low_salience_right
MCSE	salience_left-right
MCSE	salience_right-left
MTTNS	northside-southside_event
MTTNS	sn_after-before_event
MTTNS	sn_after_event
MTTNS	sn_all_event_response
MTTNS	sn_all_space-time_cue
MTTNS	sn_all_space_cue
MTTNS	sn_all_time-space_cue
MTTNS	sn_all_time_cue

Training data (next)

Task	Condition / Contrast
MTTNS	sn_average_event
MTTNS	sn_average_reference
MTTNS	sn_before-after_event
MTTNS	sn_before_event
MTTNS	sn_northside_event
MTTNS	sn_southside_event
MTTNS	sn_space-time_event
MTTNS	sn_space_event
MTTNS	sn_time-space_event
MTTNS	sn_time_event
MTTNS	southside-northside_event
MTTWE	eastside-westside_event
MTTWE	we_after-before_event
MTTWE	we_after_event
MTTWE	we_all_event_response
MTTWE	we_all_space-time_cue
MTTWE	we_all_space_cue
MTTWE	we_all_time-space_cue
MTTWE	we_all_time_cue
MTTWE	we_average_event
MTTWE	we_average_reference
MTTWE	we_before-after_event
MTTWE	we_before_event
MTTWE	we_eastside_event
MTTWE	we_space-time_event
MTTWE	we_space_event
MTTWE	we_time-space_event
MTTWE	we_time_event
MTTWE	we_westside_event
MTTWE	westside-eastside_event
MVEB	2_letters_different
MVEB	2_letters_different-same
MVEB	2_letters_same
MVEB	4_letters_different
MVEB	4_letters_different-same
MVEB	4_letters_same
MVEB	6_letters_different
MVEB	6_letters_different-2_letters_different
MVEB	6_letters_different-same
MVEB	6_letters_same
MVEB	letter_occurrence_response
MVIS	2_dots-2_dots_control
MVIS	4_dots-4_dots_control
MVIS	6_dots-2_dots
MVIS	6_dots-6_dots_control
MVIS	dot_displacement_response
MVIS	dots-control
Moto	finger_left-fixation
Moto	finger_right-fixation
Moto	foot_left-fixation
Moto	foot_right-fixation
Moto	hand_left-fixation
Moto	hand_right-fixation
Moto	instructions
Moto	saccade-fixation

Training data (next)

Task	Condition / Contrast
Moto	tongue-fixation
PainMovie	movie_mental
PainMovie	movie_mental-pain
PainMovie	movie_pain
Preference	face-others
Preference	food-others
Preference	house-others
Preference	painting-others
Preference	preference_constant
Preference	preference_linear
Preference	preference_quadratic
PreferenceFaces	face_constant
PreferenceFaces	face_linear
PreferenceFaces	face_quadratic
PreferenceFood	food_constant
PreferenceFood	food_linear
PreferenceFood	food_quadratic
PreferenceHouses	house_constant
PreferenceHouses	house_linear
PreferenceHouses	house_quadratic
PreferencePaintings	painting_constant
PreferencePaintings	painting_linear
PreferencePaintings	painting_quadratic
RSVPLanguage	complex
RSVPLanguage	complex-consonant_string
RSVPLanguage	complex-simple
RSVPLanguage	consonant_string
RSVPLanguage	jabberwocky
RSVPLanguage	jabberwocky-consonant_string
RSVPLanguage	jabberwocky-pseudo
RSVPLanguage	probe
RSVPLanguage	pseudo-consonant_string
RSVPLanguage	pseudoword_list
RSVPLanguage	sentence-consonant_string
RSVPLanguage	sentence-jabberwocky
RSVPLanguage	sentence-pseudo
RSVPLanguage	sentence-word
RSVPLanguage	simple
RSVPLanguage	simple-consonant_string
RSVPLanguage	word-consonant_string
RSVPLanguage	word-pseudo
RSVPLanguage	word_list
SelectiveStopSignal	go_critical
SelectiveStopSignal	go_critical-stop
SelectiveStopSignal	go_noncritical
SelectiveStopSignal	go_noncritical-ignore
SelectiveStopSignal	ignore
SelectiveStopSignal	ignore-stop
SelectiveStopSignal	stop
SelectiveStopSignal	stop-ignore
Self	correct_rejection
Self	encode_other
Self	encode_self
Self	encode_self-other
Self	false_alarm

Training data (next)

Task	Condition / Contrast
Self	instructions
Self	recognition_hit
Self	recognition_hit-correct_rejection
Self	recognition_other_hit
Self	recognition_self-other
Self	recognition_self_hit
StopSignal	go
StopSignal	stop
StopSignal	stop-go
Stroop	congruent
Stroop	incongruent
Stroop	incongruent-congruent
TheoryOfMind	belief
TheoryOfMind	belief-photo
TheoryOfMind	photo
TwoByTwo	cue_switch-stay
TwoByTwo	cue_taskstay_cuestay
TwoByTwo	cue_taskstay_cueswitch
TwoByTwo	cue_taskswitch_cuestay
TwoByTwo	cue_taskswitch_cueswitch
TwoByTwo	stim_taskstay_cuestay
TwoByTwo	stim_taskstay_cueswitch
TwoByTwo	stim_taskswitch_cuestay
TwoByTwo	stim_taskswitch_cueswitch
TwoByTwo	task_switch-stay
VSTM	vstm_constant
VSTM	vstm_linear
VSTM	vstm_quadratic
Visu	animal
Visu	animal-scrambled
Visu	characters
Visu	characters-scrambled
Visu	face
Visu	face-scrambled
Visu	house
Visu	house-scrambled
Visu	pseudoword
Visu	pseudoword-scrambled
Visu	scene
Visu	scene-scrambled
Visu	scrambled
Visu	target_fruit
Visu	tool
Visu	tool-scrambled
WardAndAllport	ambiguous-unambiguous
WardAndAllport	intermediate-direct
WardAndAllport	move_ambiguous_direct
WardAndAllport	move_ambiguous_intermediate
WardAndAllport	move_unambiguous_direct
WardAndAllport	move_unambiguous_intermediate
WardAndAllport	planning_ambiguous_direct
WardAndAllport	planning_ambiguous_intermediate
WardAndAllport	planning_unambiguous_direct
WardAndAllport	planning_unambiguous_intermediate

Table S4: Validation data

Validation data	
Task	Condition / Contrast
HcpEmotion	face
HcpEmotion	face-shape
HcpEmotion	shape
HcpEmotion	shape-face
HcpGambling	punishment
HcpGambling	punishment-reward
HcpGambling	reward
HcpGambling	reward-punishment
HcpLanguage	math
HcpLanguage	math-story
HcpLanguage	story
HcpLanguage	story-math
HcpMotor	cue
HcpMotor	left_foot
HcpMotor	left_foot-avg
HcpMotor	left_hand
HcpMotor	left_hand-avg
HcpMotor	right_foot
HcpMotor	right_foot-avg
HcpMotor	right_hand
HcpMotor	right_hand-avg
HcpMotor	tongue
HcpMotor	tongue-avg
HcpRelational	match
HcpRelational	relational
HcpRelational	relational-match
HcpSocial	mental
HcpSocial	mental-random
HcpSocial	random
HcpWm	0back-2back
HcpWm	0back_body
HcpWm	0back_face
HcpWm	0back_place
HcpWm	0back_tools
HcpWm	2back-0back
HcpWm	2back_body
HcpWm	2back_face
HcpWm	2back_place
HcpWm	2back_tools
HcpWm	body-avg
HcpWm	face-avg
HcpWm	place-avg
HcpWm	tools-avg

Table S5: Test data

Test data	
Task	Condition / Contrast
MathLanguage	arithmetic_fact-othermath
MathLanguage	arithmetic_fact_auditory
MathLanguage	arithmetic_fact_visual
MathLanguage	arithmetic_principle-othermath
MathLanguage	arithmetic_principle_auditory
MathLanguage	arithmetic_principle_visual
MathLanguage	auditory-visual
MathLanguage	colorlessg-wordlist
MathLanguage	colorlessg_auditory
MathLanguage	colorlessg_visual
MathLanguage	context-general
MathLanguage	context-theory_of_mind
MathLanguage	context_auditory
MathLanguage	context_visual
MathLanguage	general-colorlessg
MathLanguage	general_auditory
MathLanguage	general_visual
MathLanguage	geometry-othermath
MathLanguage	geometry_fact_auditory
MathLanguage	geometry_fact_visual
MathLanguage	math-nonmath
MathLanguage	nonmath-math
MathLanguage	theory_of_mind-context
MathLanguage	theory_of_mind-general
MathLanguage	theory_of_mind_and_context-general
MathLanguage	theory_of_mind_auditory
MathLanguage	theory_of_mind_visual
MathLanguage	visual-auditory
MathLanguage	wordlist_auditory
MathLanguage	wordlist_visual

References

[1] Alexandre Abraham et al. “Machine learning for neuroimaging with scikit-learn”. In: *Frontiers in Neuroinformatics* 8 (2014). ISSN: 1662-5196. URL: https://www.frontiersin.org/article/10.3389/fninf.2014.00014 (visited on 05/19/2022).

[2] Alexandre Abraham et al. “Machine learning for neuroimaging with scikit-learn”. In: *Front Neuroinform* 8 (2014), p. 14. ISSN: 1662-5196. URL: https://doi.org/10.3389/fninf.2014.00014.

[3] Jesper L.R. Andersson, Stefan Skare, and John Ashburner. “How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging”. In: *Neuroimage* 20.2 (2003), pp. 870–888. ISSN: 1053-8119. URL: http://doi.org/10.1016/S1053-8119(03)00336-7.

[4] B.B. Avants et al. “Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain”. In: *Medical Image Analysis* 12.1 (2008), pp. 26–41. ISSN: 1361-8415. DOI: 10.1016/j.media.2007.06.004. URL: http://www.sciencedirect.com/science/article/pii/S1361841507000606.

[5] Thomas BAZEILLE et al. “Local Optimal Transport for Functional Brain Template Estimation”. In: *IPMI 2019 - 26th International Conference on Information Processing in Medical Imaging*. Hong Kong, China, June 2019. DOI: 10.1007/978-3-030-20351-1_18. URL: https://hal.archives-ouvertes.fr/hal-02278663.

[6] Thomas Bazeille et al. “An empirical evaluation of functional alignment using inter-subject decoding”. In: *NeuroImage* 245 (Dec. 15, 2021), p. 118683. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2021.118683. URL: https://www.sciencedirect.com/science/article/pii/S1053811921009563.

[7] Yashar Behzadi et al. “A component based noise correction method (CompCor) for {BOLD} and perfusion based fMRI”. In: *Neuroimage* 37.1 (2007), pp. 90–101. ISSN: 1053-8119. URL: https://doi.org/10.1016/j.neuroimage.2007.04.042.

[8] Shohini Bhattasali et al. “Localising memory retrieval and syntactic composition: an fMRI study of naturalistic language comprehension”. In: *Language, Cognition and Neuroscience* 34.4 (2019), pp. 491–510.

[9] Laetitia Chapel et al. “Unbalanced Optimal Transport through Non-negative Penalized Linear Regression”. In: *arXiv:2106.04145 [cs, math, stat]* (June 8, 2021). arXiv: 2106.04145. URL: http://arxiv.org/abs/2106.04145.

[10] Po-Hsuan (Cameron) Chen et al. “A Reduced-Dimension fMRI Shared Response Model”. In: *Advances in Neural Information Processing Systems*. Ed. by C. Cortes et al. Vol. 28. Curran Associates, Inc., 2015. URL: https://proceedings.neurips.cc/paper/2015/file/b3967a0e938dc2a6340e258630febd5a-Paper.pdf.

[11] Lenaic Chizat et al. “Unbalanced Optimal Transport: Dynamic and Kantorovich Formulation”. In: *arXiv:1508.05216 [math]* (Feb. 9, 2019). arXiv: 1508.05216. URL: http://arxiv.org/abs/1508.05216.

[12] Anders M. Dale, Bruce Fischl, and Martin I. Sereno. “Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction”. In: 9.2 (1999), pp. 179–194. ISSN: 1053-8119. DOI: 10.1006/nimg.1998.0395. URL: http://www.sciencedirect.com/science/article/pii/S1053811998903950.

[13] Christophe Destrieux et al. “Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature”. In: *NeuroImage* 53.1 (Oct. 15, 2010), pp. 1–15. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2010.06.010. URL: https://www.sciencedirect.com/science/article/pii/S1053811910008542 (visited on 08/02/2022).

[14] Nicole Eichert et al. “Cross-species cortical alignment identifies different types of anatomical reorganization in the primate temporal lobe”. In: *eLife* 9 (Mar. 23, 2020). Ed. by Timothy Verstynen et al. Publisher: eLife Sciences Publications, Ltd, e53232. ISSN: 2050-084X. DOI: 10.7554/eLife.53232. URL: https://doi.org/10.7554/eLife.53232.

[15] Bruce Fischl. “FreeSurfer”. In: *NeuroImage*. 20 YEARS OF fMRI 62.2 (Aug. 15, 2012), pp. 774–781. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2012.01.021. URL: https://www.sciencedirect.com/science/article/pii/S1053811912000389.

[16] Bruce Fischl et al. “High-resolution intersubject averaging and a coordinate system for the cortical surface”. In: *Hum Brain Mapp* 8.4 (1999), pp. 272–284. URL: https://doi.org/10.1002/(SICI)1097-0193(1999)8:4%3C272::AID-HBM10%3E3.0.CO;2-4.

[17] K.J. Friston et al. “Characterizing Dynamic Brain Responses with fMRI: a Multivariate Approach”. In: *Neuroimage* 2.2 (1995), pp. 166–172. URL: https://doi.org/10.1006/nimg.1995.1019.

[18] Matthew F Glasser et al. “A multi-modal parcellation of human cerebral cortex”. en. In: *Nature* 536.7615 (Aug. 2016), pp. 171–178.

[19] Alexandre Gramfort, Gabriel Peyré, and Marco Cuturi. *Fast Optimal Transport Averaging of Neuroimaging Data*. 2015. DOI: 10.48550/ARXIV.1503.08596. URL: https://arxiv.org/abs/1503.08596.

[20] J. Swaroop Guntupalli et al. “A Model of Representational Spaces in Human Cortex”. In: *Cerebral Cortex* 26.6 (June 1, 2016). Publisher: Oxford Academic, pp. 2919–2934. ISSN: 1047-3211. DOI: 10.1093/cercor/bhw068. URL: https://academic.oup.com/cercor/article/26/6/2919/1754308.

[21] James V. Haxby et al. “A common, high-dimensional model of the representational space in human ventral temporal cortex”. In: *Neuron* 72.2 (Oct. 20, 2011), pp. 404–416. ISSN: 1097-4199. DOI: 10.1016/j.neuron.2011.08.026.

[22] Matthias Liero, Alexander Mielke, and Giuseppe Savaré. “Optimal Entropy-Transport problems and a new Hellinger–Kantorovich distance between positive measures”. In: *Inventiones mathematicae* 211.3 (Mar. 1, 2018), pp. 969–1117. ISSN: 1432-1297. DOI: 10.1007/s00222-017-0759-8. URL: https://doi.org/10.1007/s00222-017-0759-8.

[23] Rogier B Mars et al. “Whole brain comparative anatomy using connectivity blueprints”. In: *eLife* 7 (May 11, 2018). Ed. by Klaas Enno Stephan. Publisher: eLife Sciences Publications, Ltd, e35237. ISSN: 2050-084X. DOI: 10.7554/eLife.35237. URL: https://doi.org/10.7554/eLife.35237.

[24] Facundo Memoli. *On the use of Gromov-Hausdorff Distances for Shape Comparison*. Accepted: 2014-01-29T16:52:11Z ISSN: 1811-7813. The Eurographics Association, 2007. ISBN: 978-3-905673-51-7. DOI: 10.2312/SPBG/SPBG07/081-090. URL: https://diglib.eg.org:443/xmlui/handle/10.2312/SPBG.SPBG07.081-090.

[25] Facundo Mémoli. “Gromov–Wasserstein Distances and the Metric Approach to Object Matching”. In: *Foundations of Computational Mathematics* 11.4 (Aug. 1, 2011), pp. 417–487. ISSN: 1615-3383. DOI: 10.1007/s10208-011-9093-5. URL: https://doi.org/10.1007/s10208-011-9093-5.

[26] Franz-Xaver Neubert et al. “Comparison of human ventral frontal cortex areas for cognitive control and language with areas in monkey frontal cortex”. In: *Neuron* 81.3 (Feb. 5, 2014), pp. 700–713. ISSN: 1097-4199. DOI: 10.1016/j.neuron.2013.11.012.

[27] Shinji Nishimoto et al. “Reconstructing visual experiences from brain activity evoked by natural movies”. In: *Current biology* 21.19 (2011), pp. 1641–1646.

[28] Adam Paszke et al. “PyTorch: An Imperative Style, High-Performance Deep Learning Library”. In: *Advances in Neural Information Processing Systems 32*. Ed. by H. Wallach et al. Curran Associates, Inc., 2019, pp. 8024–8035. URL: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

[29] F. Pedregosa et al. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[30] Gabriel Peyré, Marco Cuturi, and Justin Solomon. “Gromov-Wasserstein Averaging of Kernel and Distance Matrices”. In: *Proceedings of The 33rd International Conference on Machine Learning*. International Conference on Machine Learning. ISSN: 1938-7228. PMLR, June 11, 2016, pp. 2664–2672. URL: https://proceedings.mlr.press/v48/peyre16.html.

[31] Ana Luísa Pinho et al. “Individual Brain Charting, a high-resolution fMRI dataset for cognitive mapping”. In: *Scientific Data* 5 (June 2018), p. 180105. DOI: 10.1038/sdata.2018.105. URL: https://hal.archives-ouvertes.fr/hal-01817528.

[32] Ana Luísa Pinho et al. “Subject-specific segregation of functional territories based on deep phenotyping”. In: *Hum Brain Mapp* 42.4 (2021), pp. 841–870. URL: https://doi.org/10.1002/hbm.25189.

[33] F. Pizzagalli et al. “The reliability and heritability of cortical folds and their genetic correlations across hemispheres”. In: *Commun Biol* 3.1 (Sept. 2020), p. 510.

[34] Hugo Richard et al. “Modeling shared responses in neuroimaging studies through multiview ica”. In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 19149–19162.

[35] Emma C. Robinson et al. “MSM: a new flexible framework for Multimodal Surface Matching”. In: *NeuroImage* 100 (Oct. 15, 2014), pp. 414–426. ISSN: 1095-9572. DOI: 10.1016/j.neuroimage.2014.05.069.

[36] Emma C. Robinson et al. “Multimodal surface matching with higher-order smoothness constraints”. In: *NeuroImage* 167 (2018), pp. 453–465. ISSN: 1095-9572. DOI: 10.1016/j.neuroimage.2017.10.037.

[37] Mert R. Sabuncu et al. “Function-based intersubject alignment of human cortical anatomy”. In: *Cerebral Cortex (New York, N.Y.: 1991)* 20.1 (Jan. 2010), pp. 130–140. ISSN: 1460-2199. DOI: 10.1093/cercor/bhp085.

[38] Marian Schneider et al. “Columnar clusters in the human motion complex reflect consciously perceived motion axis”. In: *Proceedings of the National Academy of Sciences* 116.11 (2019), pp. 5096–5101. DOI: 10.1073/pnas.1814504116. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.1814504116. URL: https://www.pnas.org/doi/abs/10.1073/pnas.1814504116.

[39] Thibault Séjourné, François-Xavier Vialard, and Gabriel Peyré. “The Unbalanced Gromov Wasserstein Distance: Conic Formulation and Relaxation”. In: *arXiv:2009.04266 [math, stat]* (June 8, 2021). arXiv: 2009.04266. URL: http://arxiv.org/abs/2009.04266.

[40] Ido Tavor et al. “Task-free MRI predicts individual differences in brain activity during task performance”. In: *Science* 352.6282 (2016), pp. 216–220.

[41] Bertrand Thirion et al. “Which fMRI clustering gives good brain parcellations?” In: *Frontiers in Neuroscience* 8.167 (May 2014), p. 13. DOI: 10.3389/fnins.2014.00167. URL: https://hal.inria.fr/hal-01015172.

[42] D. C. Van Essen et al. “Parcellations and hemispheric asymmetries of human cerebral cortex analyzed on surface-based atlases”. In: *Cereb Cortex* 22.10 (Oct. 2012), pp. 2241–2262.

[43] D. C. Van Essen et al. “The WU-Minn Human Connectome Project: an overview”. In: *Neuroimage* 80 (Oct. 2013), pp. 62–79.

[44] Titouan Vayer et al. “Fused Gromov-Wasserstein distance for structured objects: theoretical foundations and mathematical properties”. In: *arXiv:1811.02834 [cs, stat]* (Nov. 7, 2018). arXiv: 1811.02834. URL: http://arxiv.org/abs/1811.02834.

[45] S. Al-Wasity et al. “Hyperalignment of motor cortical areas based on motor imagery during action observation”. In: *Sci Rep* 10.1 (Mar. 2020), p. 5362.

[46] Ting Xu et al. “Cross-species functional alignment reveals evolutionary hierarchy within the connectome”. In: *NeuroImage* 223 (Dec. 1, 2020), p. 117346. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2020.117346. URL: http://www.sciencedirect.com/science/article/pii/S1053811920308326.

[47] B.T. Thomas Yeo et al. “Spherical Demons: Fast Diffeomorphic Landmark-Free Surface Registration”. In: *IEEE transactions on medical imaging* 29.3 (Mar. 2010), pp. 650–668. ISSN: 0278-0062. DOI: 10.1109/TMI.2009.2030797. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2862393/.