Should one go for individual or group-level brain parcellations? A deep-phenotyping benchmark

Bertrand Thirion, Himanshu Aggarwal, Ana Fernanda Ponce, Ana Luísa Pinho, Alexis Thual

Should one go for individual or group-level brain parcellations ? A deep-phenotyping benchmark

Bertrand Thirion[1], Himanshu Aggarwal[1], Ana Fernanda Ponce[1], Ana Luísa Pinho[3,4], and Alexis Thual[1,2]

[1]*Inria, CEA, Université Paris-Saclay, Palaiseau, 91120, France*
[2]*Inserm, Collège de France, Paris, France*
[3]*Department of Computer Science, Western University, London, Ontario, Canada*
[4]*Western Institute for Neuroscience, Western University, London, Ontario, Canada*

Abstract

The analysis and understanding of brain characteristics often require considering regions-level information rather than voxel-sampled data. Subject-specific parcellations have been put forward in recent years, as they can adapt to individual brain organization and thus offer more accurate individual summaries than standard atlases. But the price to pay for adaptability is the lack of group-level consistency of the data representation. Here, we investigate whether the good representations brought by individualized models is merely an effect of circular analysis, in which individual brain features are better represented by subject-specific summaries, or whether this carries over to new individuals, i.e. whether one can actually adapt an existing parcellation to new individuals and still obtain good summaries in these individuals. For this, we adapt a dictionary learning method to produce brain parcellations. We use it on a deep phenotyping dataset to assess quantitatively the patterns of activity obtained under naturalistic and controlled-task-based settings. We show that the benefits of individual parcellations are substantial, but that they vary a lot across brain systems.

Highlights

- Individualized parcellations schemes learned on naturalistic data generalize to task data, not only within, but also across individuals
- Individualized parcellations better represent functional signal than fixed parcellations in most brain regions
- Parcellations obtained from naturalistic data fail to capture structure in motor and somato-sensory cortex
- Data-driven parcellations make it possible to adapt resolution and better fit brain signals.
- Some brain systems require higher resolution for accurate representations than that of standard atlases

1 Introduction

Brain atlases and brain parcellations Brain-image analysis relies extensively on intermediate-scale representations of the brain volume, such as parcellations and atlases. Here, we consider *analysis tasks* such as connectivity analysis [Dadi et al., 2019, Pervaiz et al., 2020, Bryce et al., 2021, Bijsterbosch et al., 2018], brain activity decoding [Menuet et al., 2022, Mensch et al., 2021, Thomas et al., 2022], the study of regional physiological parameters [Badillo et al., 2013] or the impact of lesions [Kasties et al., 2021]. Among such intermediate representations, *atlases* typically refer to a given ontology of brain structures, that is based on core knowledge on brain organization, such as cyto-architecture or connectivity [Amunts et al., 2014]. Other representations, such as so-called *spatial gradients*, do not benefit from such clear semantics [Kong et al., 2023]. *Parcellations* are related to atlases. The key difference is that they are defined as a data-driven object, i.e. they are built to summarize some image-derived information [Thirion et al., 2006, Tzourio-Mazoyer et al., 2002]. Atlases and Parcellations can be specified in many ways, depending on whether they are defined in the brain volume or on the cortical surface (see e.g. [Glasser et al., 2016]), or whether they are deterministic —leading to a strict partition of the voxels into components— or probabilistic [Schaefer et al., 2017, Dadi et al., 2020], i.e. with overlapping, fuzzy components. Parcellation definition hinges on some perspective of the features that are deemed important in a particular type of brain-imaging analysis (cyto-architecture, anatomical connectivity, functional connectivity, vascular structures etc.)

In a data-analysis context, atlas-based and parcellation-based representations are designed to meet two objectives: *i)* compress the data by replacing a voxel-sampled image by a region-level representation, which is then typically used to reduce the multiple comparison burden in brain imaging group analyses, and *ii)* name the regions, i.e. provide anatomical labels to identify brain features, which is important to put the results of neuroimaging data analyses in context. From a data-analytic perspective, atlases face two issues: first, they are not a very efficient data-reduction approach, as data-driven approaches perform better [Schaefer et al., 2017, Dadi et al., 2020] when it comes to maximizing representation accuracy. Specifically, when considering functional brain organization, fixed atlases do not fit well brain topographic organization, such as peak regions of brain maps, or boundaries between functional regions [Gordon et al., 2017]. An obvious reason is that most atlases have not been defined to account for function; a second reason is that they do not adapt to individual characteristics. The second major issue is that region naming is known to be inconsistent [Bohland et al., 2009], weakening the usefulness of structure naming and making it ultimately hard for the community to build a consistent ontology of brain structures, though integration efforts are under way [Lawrence et al., 2021, Uddin et al., 2023].

In this paper, we thus turn to a probabilistic-parcellation approach, which has gained popularity in the recent years [Harrison et al., 2020, Amunts et al., 2020, Dadi et al., 2020]. Probabilistic parcellations give a non-negative weight to voxels to belong to the model components. A given voxel may have non-zero weight for several components, illustrating the uncertainty of voxel-to-component assignments, which in turn reflects noise in the data, poor model fit, or inter-subject variability. Individualized parcellations are parcellation schemes that adapt to each individual in a given dataset, as opposed to a fixed population-level model [Thirion et al., 2006, Glasser et al., 2016, Gordon et al., 2017, Gratton et al., 2018, Kong et al., 2019, Greene et al., 2020, Zhao et al., 2020, Harrison et al., 2020, Pinho et al., 2021]. They represent a desirable step toward the personalization of brain models. Our aim is to investigate whether it is actually *useful* to individualize brain parcellations.

Population-level versus individual parcellations The recent trend in developing individual parcellations leads to some questions regarding their cost/benefit trade-off. On the one hand, individualized parcellations have been reported to better represent individual signal than population-level parcellations [Glasser et al., 2016, Gordon et al., 2017, Braga and Buckner, 2017, Kong et al., 2019]. On the other hand, the use of such parcellations schemes raises difficult questions, such as:

i) How to ensure consistency of the parcellation across individuals? A purely bottom-up approach that would parcel each individual's dataset independently does not yield corresponding parcellations schemes, which makes it useless in practice. A constrained approach, where parcellations are obtained simultaneously across individuals leads to complex algorithmic procedures to ensure parcellation consistency [Thirion et al., 2006, Gordon et al., 2017], or to arbitrary choices, where parcellations end up being defined manually [Glasser et al., 2016].

ii) Keeping in mind that the generalizability of findings is key to validity [Varoquaux and Poldrack, 2019], how can one generalize scientific results obtained in a given sample of individuals using a subject-adaptive parcellation to a wider group? This at least requires that the learned parcellation scheme can be generalized to new individuals. But the concrete mechanisms to perform such generalization are still elusive, if not lacking.

These two concerns have not been dealt satisfactorily so far. In [Glasser et al., 2016] the generalization of parcellation schemes to new data was performed, but the end validation was restricted to the consistency with hand-drawn delineations, which does not grant predictive validity. Overall, the quantification of parcellation goodness of fit is still lacking, making it impossible to decide how good a given representation scheme is. The evaluation of parcellation schemes is thus an important and urgent practical question. It has been partly addressed in the realm of functional connectivity analysis by [Bijsterbosch et al., 2018], however, due to the very structure of connectivity, it is still a challenge to unambiguously tease out the spatial variability from connectivity variability across individuals. Activation mapping represents a simpler framework where the functional properties of brain are probed, irrespective of other regions, that can also be hard to characterize. Other indirect evidence comes from the work of [Tavor et al., 2016, Dohmatob et al., 2021], that has shown that probabilistic parcellations derived from resting-state accurately predict brain responses to tasks, but these rely on the dual regression scheme [Nickerson et al., 2017, Harrison et al., 2020] that breaks the nature of parcellation, as it leads to non-sparse and potentially non-positive spatial loadings for the components. A better-defined approach instead relies on naturalistic fMRI, that does not suffer from the indeterminacy of resting-state fMRI [Nastase et al., 2020].

Our contribution In this work, we tackle the question of the goodness of fit of individualized parcellations, with a focus on *predictive accuracy*: if brain functional characteristics can be associated with the components of a given parcellation, how do these characteristics generalize to new individuals? We thus propose a metric to measure the *predictive accuracy* of parcellation/atlasing schemes, i.e. the reconstruction quality of individual data in participants that have not been included in the initial definition of the parcellation scheme.

To this end, we leverage *Individual Brain Charting* (IBC), a deep-phenotyping dataset [Pinho et al., 2018, Pinho et al., 2020] comprising naturalistic data and more than a hundred condition-derived functional contrasts obtained from forty tasks, in a fixed cohort of twelve participants. We propose to learn individual-parcellation schemes from naturalistic stimuli (mostly movie-watching data) and then assess whether the learned parcellation scheme performs well on *new individuals* in the reconstruction of brain maps for *new contrasts*, i.e.

imaging contrasts that were not used to build the parcellation. It may well be that the gain obtained on individual parcellation fitting vanishes when one tries to generalize the learned representation scheme to new individuals. This could make fixed dictionaries, or even standard atlases, a safer choice.

We compare the accuracy of three types of parcellations:

- *atlases*, i.e. a priori defined parcellations schemes. We consider the MultiModal Parcellation (MMP) 1.0 atlas [Glasser et al., 2016], which is a reference among cortical atlases. We also consider the Schaefer parcellation scheme [Schaefer et al., 2017]
- *fixed dictionary*, a probabilistic parcellation scheme that is inferred from neuroimaging data, but common to a population of individuals
- *individual dictionary*, a probabilistic parcellation scheme that is inferred from neuroimaging data, and specific to each individual.

We also take this opportunity to assess the impact of resolution on data driven parcellation —*resolution* being implicitly defined by the number of components in the parcellation: we measure the impact of doubling the number of components extracted in a given brain system on the accuracy of the reconstruction.

It is also important to observe that evaluating parcellation schemes on noisy data can lead to some biases; for instance, methods that shrink signal estimates will generalize better than methods that don't. To avoid such biases, we carefully select contrast maps that display the highest reproducibility level across individuals, within each brain system considered, so that the analysis relies on signal, not noise.

In the next sections, we first describe in detail our proposed method; then we present results comparing the goodness of fit of individual parcellations *versus* group level parcellations and fixed atlases. We particularly outline regional differences, noticing that the most accurate approach is not the same one across brain systems.

2 Methods

2.1 Dataset used

Data acquisition and pre-processing A detailed description of the preprocessing pipeline of the IBC data is provided in [Pinho et al., 2021]. 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 directions. Images were corrected based on the estimated deformation model. Details about the method can be found in [Andersson et al., 2003].

Further, all 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 [Friston et al., 1995]. Raw data were preprocessed with *PyPreprocess* [1], using SPM12 routines for the main steps. 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 [van Essen et al., 2012]. The corresponding maps were then resampled to the *fsaverage7* template of FreeSurfer [Fischl et al., 1999].

[1] https://github.com/neurospin/pypreprocess

Task name	task performed	duration (#runs / time, min)	reference
Clips	Watching and fixating	21 / 210	[Nishimoto et al., 2011]
Bang	Watching and listening	1 / 8	[Campbell et al., 2015]
Raiders of the lost Ark	Watching and listening	13 / 120	[Haxby et al., 2011]
The Good, the Bad and the Ugly	Watching and listening	21 / 180	[Mantini et al., 2012]
The Little Prince	Listening	9 / 90	[Hale et al., 2022]

Table 1: List of the naturalistic tasks used in the experiments. Note that time is total acquisition time summed across runs.

Task data FMRI data were analyzed using the *General Linear Model*. Regressors of the model were designed to capture variations in BOLD response 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 the 5% voxels with highest variance, were also regressed to capture physiological noise [Behzadi et al., 2007].

In addition, a discrete-cosine basis was included for high-pass filtering (*cutoff*=$\frac{1}{128}$Hz). Model specification was implemented using *Nilearn* [Abraham et al., 2014], a Python library for statistical learning on neuroimaging data[2].

The ensuing set of contrasts is listed in section A. Let n denote the number of participants. Contrasts acquired in each participant result in spatial maps, noted $(\mathbf{Z}^s), s = 1 \cdots n$, where $\mathbf{Z}^s \in \mathbb{R}^{v \times c}$, v is the number of vertices across the two hemispheres and c the number of contrasts.

Naturalistic stimuli We considered five naturalistic datasets. Corresponding acquisitions are briefly described in Table 1, together with the original reference. These datasets were acquired and preprocessed with the same pipeline as the task data and resulted in time courses sampled on the mesh of each individual brain, finally registered to fsaverage7 space.

A dimension reduction was performed on the different datasets using a *Shared Response Model* (SRM, [Chen et al., 2015, Richard et al., 2021]). We used the FastSRM implementation of the method [3]. SRM yields subject-specific maps associated with common temporal components elicited by some naturalistic content. Twenty components were retained for each task, except the *Bang* task, for which ten components were retained. The reason is that *Bang* contains less data (see Table1). These components were concatenated to form a naturalistic response signature in each vertex, so that the whole naturalistic dataset was reduced to $q = 90$ spatial maps in each individual. These spatial maps are noted $(\mathbf{X}^s), s = 1 \cdots n$, where $\mathbf{X}^s \in \mathbb{R}^{v \times q}$, v is the number of vertices across the two hemispheres, and $q = 90$ represents the set of SRM components.

[2] https://nilearn.github.io

[3] https://github.com/hugorichard/FastSRM

Data availability The raw data are available on the *OpenNeuro* [4] and *Ebrains* platforms [5]. Preprocessed data are available on Ebrains [6] and contrast maps are available on *NeuroVault* [7] and Ebrains [8]. A comprehensive documentation of the data is available online [9].

2.2 Modeling

Terminology We call *dictionary* a set of brain components associated with some functional characteristics; such an object is called a *probabilistic data-driven parcellation* in other circumstances [Dadi et al., 2020]. We follow the methodology developed in [Thirion et al., 2021, Varoquaux et al., 2013]. The only difference is that it is applied on a subset of the brain domain, in sub-systems defined in the left or right hemisphere respectively. See Section 2.3 for details.

Dictionary estimation Formally, consider the set of brain maps $\mathbf{X}^s$ obtained for $q = 90$ maps in a subject $s \in 1, \cdots, n$ and a given brain region. To avoid cluttering notations, we omit the dependence on the region, and simply consider that the data matrices $\mathbf{X}^s, s = 1 \cdots n$ and $\mathbf{Z}^s, s = 1 \cdots n$ are available for a set of voxels or vertices under consideration. To get more compact notations, the $1 \cdots n$ set will be denoted $[\![n]\!]$ in this paper. By enumerating the values across a mesh of vertices, for $j \in [\![q]\!]$, each column $\mathbf{x}_j^s$ of $\mathbf{X}^s$ is a $p-$dimensional vector, where p is the number of vertices (since we consider a brain system, we have $p < v$) ; $\mathbf{X}^s$ is thus a matrix of size $p \times q$. Individual dictionary learning build on functional correspondences [Thirion et al., 2021, Varoquaux et al., 2013]; it comes from solving the following minimization problem for a given $\lambda > 0$:

$$\min_{\substack{(\mathbf{U}^s)s=1\ldots n \\ \mathbf{V}_X \in \mathcal{C}}} \sum_{s=1}^{n} \left(\|\mathbf{X}^s - \mathbf{U}^s \mathbf{V}_\mathbf{X}\|_F^2 + \lambda \|\mathbf{U}^s\|_1 \right), \tag{1}$$

where $\mathbf{U}^s \geq 0\ ,\ \forall s \in [\![n]\!]$, i.e. the spatial components are forced to be non-negative in all subjects. Here, $\mathcal{C}$ denotes the set of matrices with row norm smaller than 1. Note that the $\mathbf{V}_\mathbf{X} \in \mathcal{C}$ constraint is only a technical one that is needed to make the problem well-posed. It does not change the structure of the solution. $\mathbf{U}_s$ matrices have shape $p \times k$, whereas the functional-loading matrix $\mathbf{V}_\mathbf{X}$ has shape $k \times q$, k being the number of components. $\mathbf{V}_\mathbf{X}$ describes the functional characteristics of the components. $\|.\|_F$ and $\|.\|_1$ stand for the Frobenius and ℓ_1 norm on matrices respectively. The estimated subject-specific spatial components $(\mathbf{U}^s),\ s \in [\![n]\!]$ can be interpreted as individual topographies; these non-negative components may overlap, although their values are zero in most vertices. The λ parameter was calibrated in order to yield a sparsity of around 50%.

In this work, we also explore alternatives to this individualized schemes. In particular, we consider a *fixed dictionary* approach. It is obtained by concatenating all training subjects along the contrast dimension, effectively assuming that the spatial organization is common across all individuals. This leads to $\tilde{\mathbf{U}}$ components that are common across individuals.

$$\min_{\substack{\tilde{\mathbf{U}} \in \mathbb{R}^{p \times k} \\ \mathbf{V}_\mathbf{X} \in \mathcal{C}}} \sum_{s=1}^{n} \|\mathbf{X}^s - \tilde{\mathbf{U}} \mathbf{V}_\mathbf{X}\|_F^2 + \lambda \|\tilde{\mathbf{U}}\|_1, \tag{2}$$

[4] https://openneuro.org/datasets/ds002685
[5] https://search.kg.ebrains.eu/instances/8ddf749f-fb1d-4d16-acc3-fbde91b90e24
[6] https://search.kg.ebrains.eu/instances/3ca4f5a1-647b-4829-8107-588a699763c1
[7] https://neurovault.org/collections/6618
[8] https://search.kg.ebrains.eu/instances/07ab1665-73b0-40c5-800e-557bc319109d
[9] https://individual-brain-charting.github.io/docs

As the estimation problem (in both individual and fixed dictionary settings) is non-convex, initialization matters; here, we informally controlled that results would be stable when training data is varied. We found that the variance of the ensuing results is small (see e.g. in Figure 5). As this variance accounts in part for the initialization effect, we can conclude that the impact of initialization is negligible with respect to the effects reported in the results section. The number k of components depends on the region considered (see below).

Representing new fMRI contrasts in a given dictionary We now consider c novel contrasts observed in the same individuals, resulting in matrices $\mathbf{Z}^s = (\mathbf{z}_j^s), j \in [\![c]\!]$. What we need to do is to estimate the functional loadings $\mathbf{V_Z}$ associated with the components $\mathbf{U}$ of the dictionary outline above. Following the model in equation 1, $\mathbf{V_Z}$ is defined as

$$\mathbf{V_Z} = \underset{\mathbf{V}\in\mathbb{R}^{k\times c}}{\operatorname{argmin}} \sum_{s=1}^{n} \|\mathbf{Z}^s - \mathbf{U}^s\mathbf{V}\|_F^2, \tag{3}$$

The only difference is that the technical constrain $\mathbf{V_Z} \in \mathcal{C}$ is no longer necessary, since $(\mathbf{U}^s)_{s\in[\![n]\!]}$ are fixed. This estimation is readily solved by ordinary least squares. Interestingly, this fitting procedure is analogous to the one described in [Tavor et al., 2016, Dohmatob et al., 2021].

Generalization to new individuals Assume now that we have data from new subjects $(\mathbf{X}^s, \mathbf{Z}^s), s \in [n+1, \cdots, n+m]$. Generalizing the model to these individuals means learning the spatial components for these new individuals. This is solved, again, by re-writing equation 1: for each individual $s \in [n+1, \cdots, n+m]$,

$$\mathbf{U}^s = \underset{\mathbf{U}}{\operatorname{argmin}} \|\mathbf{X}^s - \mathbf{U}\mathbf{V_X}\|_F^2 + \lambda\|\mathbf{U}\|_1, \tag{4}$$

using the values of $\mathbf{V_X}$ and λ learned from Eq. 1. Then, one can try and predict what would be the response of these new subjects to the contrasts under consideration:

$$\hat{\mathbf{Z}}^s = \mathbf{U}^s\mathbf{V_Z} \tag{5}$$

$\hat{\mathbf{Z}}^s$ represents the expected per-vertex amount of activity in subject s for c contrasts, assuming that the topographical information learned from naturalistic data is consistent with the topographic organization of task data, and that the amount of activity in movie and task are consistent across individuals.

Method evaluation proceeds by computing the discrepancy between observed $\mathbf{Z}^s$ and predicted $\hat{\mathbf{Z}}^s$. Since $\mathbf{Z}^s = (\mathbf{z}_1^s, \cdots, \mathbf{z}_c^s)$ represents c contrast maps, we propose to take the cosine distance between observed and predicted maps and average them to get an overall similarity across contrasts:

$$\Delta(s) = \frac{1}{c}\sum_{d=1}^{c}\left(1 - \frac{\langle \mathbf{z}_d^s, \hat{\mathbf{z}}_d^s\rangle}{|\mathbf{z}_d^s||\hat{\mathbf{z}}_d^s|}\right)$$

This distance represents how strongly predicted contrasts match observed ones in the region under consideration, on average. Lower values are better.

2.3 Experiments

The experiment consisted of studying functional organization in twenty-one coarse-grain brain systems, namely those used in [Glasser et al., 2016] to describe the MMP1.0 atlas regions. These large-scale systems are described in Table 3. We only made one exception to the organization described in [Glasser et al., 2016]: V1 was included in the same set as the other primary visual regions. Left and right hemispheres were fitted separately. We performed the prediction experiment in three settings:

- using the MMP1.0 *atlas* (which are fixed, non-overlapping spatial components). Concretely, this means that in equation 3 and 4, $\mathbf{U}^s$ was replaced by a region indicator read from the MMP1.0 parcellation.
- using a *fixed dictionary* inferred from the data through dictionary learning; the number of components was taken equal to the number of MMP1.0 regions in the brain system. Concretely, this means that in equation 3 and 4, $\mathbf{U}^s$ was replaced by the group average model $\tilde{\mathbf{U}}$ from Eq. 2.
- using an *individual dictionary*, i.e. a set of overlapping components inferred from the data through dictionary learning; the number of components was taken equal to the number of MMP1.0 regions in the brain system.

In concrete terms, we performed a randomized cross-validation loop, where six subjects were used to train the model and five were used to assess the model, in twenty random splits of the data. Computing the distance 2.2 and averaging across test subjects leads to distances Δ_{atlas}, Δ_{fixed} and $\Delta_{individual}$ respectively, for each split. Such values are defined at the level of brain systems, that are assumed to be defined consistently across subjects. We observed that the fluctuation of these Δ values was very narrow across splits (see Figure 5 below), meaning that the data was sufficient to obtain reliable estimates.

We also performed a control experiment, in which we directly measured the cosine distance between all maps of the test subjects and the average map among training subjects:

$$\text{distance}(s) = \frac{1}{c}\sum_{d=1}^{c}\left(1 - \frac{\langle \mathbf{z}_d^s, \text{mean}_{t\in\text{training set}}\mathbf{z}_d^t\rangle}{|\mathbf{z}_d^s||\text{mean}_{t\in\text{training set}}\mathbf{z}_d^t|}\right) \tag{6}$$

The whole training and test procedure is illustrated in Figure 1. In a second experiment, we repeated the same procedure as above, but taking twice the number of components present in the MMP1.0 atlas.

In order to ensure that results were not tied to the peculiar choice of the MMP1.0 atlas, we reproduced them with one of Schaefer's parcellation [Schaefer et al., 2017], namely a parcellation into 400 regions (200 in each hemisphere), that subdivides the Yeo17 atlas [Yeo et al., 2011]. In this experiment, system definition corresponds to the 17 large-scale Yeo17 networks.

Finally, we also performed a whole-brain comparison of the three models to study how the three types of atlases behave in a large-scale setting, dealing with $160k$ vertices and 181 regions, which renders the estimation procedure more challenging.

Implementation The implementation relies on the *dictionary-learning* methods of *scikit-learn* v1.0 [Pedregosa et al., 2011], a Python machine-learning library[10]. Dictionary fitting took 15s to 90s on a computer desktop, depending on the brain system considered.

[10] https://scikit-learn.org/stable/

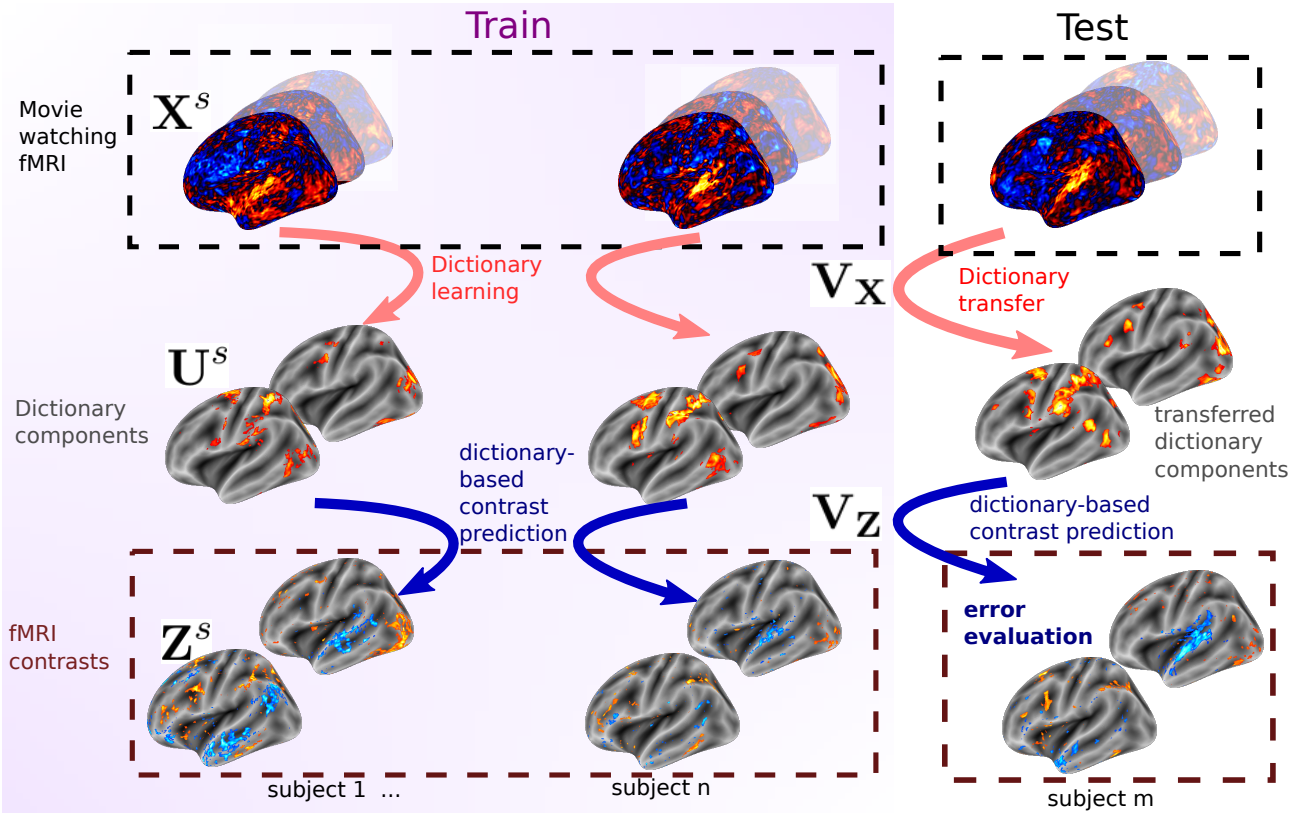

Figure 1: **Representation of the individualized parcellation benchmark done in this work**. *(left)* Dictionary learning is performed on naturalistic fMRI (mostly movie watching, see Table 1), resulting into individual-specific spatial components. From these components, one can learn to predict fMRI contrasts, using a regression approach. *(right)* Then, we transfer the dictionary onto new participants, obtaining a subject-specific dictionary, based on which we can then predict the amount of activation for the task fMRI contrasts observed in the training set. This is used to compute a prediction error that quantifies the overall goodness of fit of the model.

Reproducibility The analysis code for data preprocessing is available on our Github repository [11]. Scripts used to obtain results and figures are available [12].

3 Results

An instance of the dictionary learning procedure is displayed for one brain system, namely the left Temporal-Parietal-Occipital Junction (TPJ) in Figure 2. It shows in particular that the estimated components differ markedly from the MMP1.0 atlas boundaries in that region. Still, it remains possible to find a one-to-one relationship between MMP1.0 regions and the obtained components, so that we can name the components using the atlas names. Components learned by dictionary learning can be characterized by different functional properties, as shown in Figure 2, left: the TPOJ3 region shows more activity for pain-related contrasts, TPOJ2 for visual contrasts, TPOJ1 for auditory descriptions, STV for Story listening and PSL for general auditory conditions. In terms of topography the TPOJ2/TPOJ3 and STV/PSL boundaries are oriented differently from the MMP1.0 atlas. Another example is available in appendix, see Figure 9.

We mapped the distance gain $\Delta_{atlas} - \Delta_{individual}$ measured on functional contrasts in novel participants, in which the parcellation scheme was applied and individual activation was predicted from that of the training subjects. This difference is considered for the twenty one systems listed in Table 3, and represented on a brain map in Figure 3 (top). Obtained values are the result of a twenty-fold random splitting of the data into train and test. This shows that individual parcellations provide a better generalization in a majority of systems (14 over 21, left or right hemisphere), and particularly so in the Temporal-Parietal-Occipital Junction (TPJ) bilaterally and the right Inferior Frontal Gyrus. By contrast, individual parcellations yield less accurate mappings in the motor regions and the insula bilaterally. In Figure 3 (bottom), one can see the same map, but evaluated on the subjects of the training set, i.e. in the absence of cross-validation. The figure displays overall the same region effects, but all values are larger, showing that individual parcellations outperform systematically. But without proper validation, this superiority is spurious.

In Fig.4, one can see the difference in $\Delta_{fixed} - \Delta_{individual}$ between individual dictionary and fixed dictionary models, within the above brain systems, across twenty random cross-validation folds. Again we observe very clear regional differences: the bilateral Temporal-Parietal-Occipital Junction and right Inferior Frontal Gyrus again display a clear advantage for the individual parcellation, scheme, but motor and insular regions display the opposite trend.

To provide a precise quantitative account of these difference across systems, distances Δ_{atlas}, Δ_{fixed} and $\Delta_{individual}$ are displayed in Figure 5 for all twenty one systems, in the left (top) and right (bottom) hemisphere, respectively. These are displayed with bars. The accuracy obtained without atlases, simply using voxel-level signal averages as explained in Eq. 6, are represented by an horizontal line. This leads to several observations: *i)* in the most general case (13/21 systems in the left hemisphere, 14/21 in the right hemisphere), we see that individual dictionary outperforms fixed dictionary, which in turn outperforms the atlas; however, *ii)* there are several exceptions to this general trend: for instance, in the somatosensory regions, the reverse pattern is obtained, where the atlas performs better than both individual and fixed dictionaries; *iii)* these quantitative results are very consistent across hemispheres, and come with small error bars across validation folds; *iv)* in some regions, a standard voxel-level mapping yields a better prediction than the parcellation (early visual cortex, somatosensory and motor regions), while it performs worse in most others (Dorsal and Ventral Streams,

[11] https://github.com/individual-brain-charting/public_analysis_code
[12] https://github.com/bthirion/deep_parcellation_benchmark

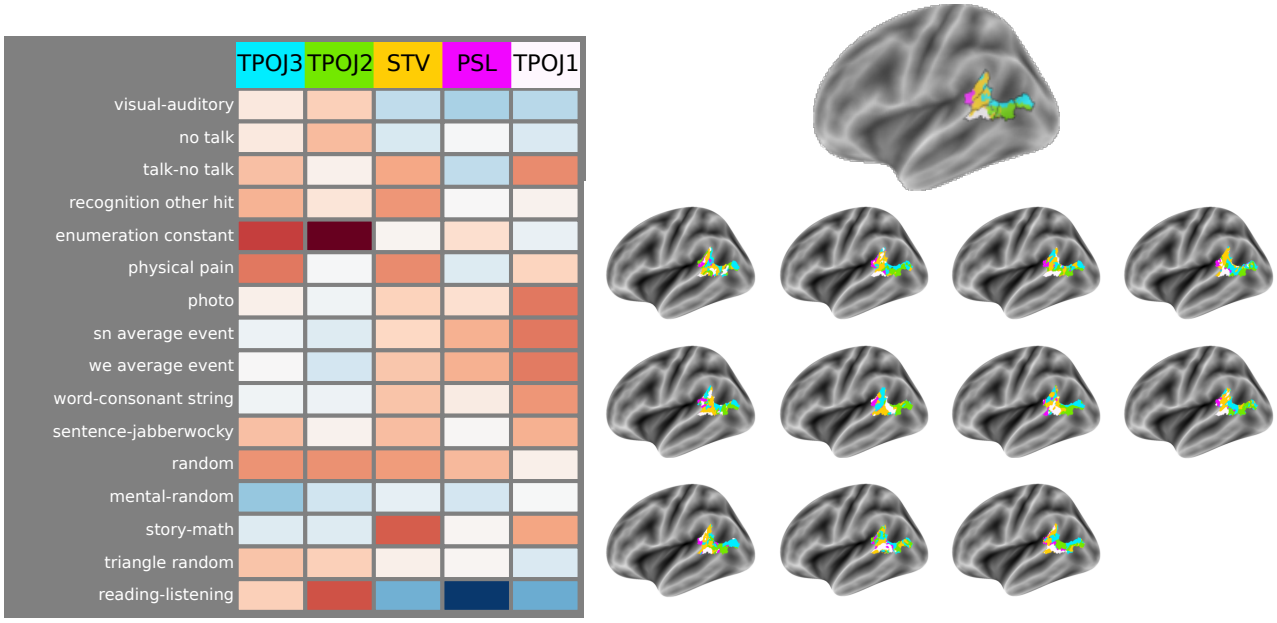

Figure 2: **Example of an individualized-parcellation scheme in the Temporal-Parietal-Occipital Junction (TPJ).** *(left)* For five-components parcellation of the TPJ, one can characterize their response to 160 contrasts; sixteen are displayed for readability, and are related to cognitive operations (mathematics, language, theory of mind, see Table 2 for a description of these contrasts). *(right)* Each vertex of the TPJ can be assigned to these regions by considering the loadings in each individual (bottom), or some summary of these loadings across individuals. *(top)* The resulting parcellation is superimposed with the original MMP1.0 parcellation. It clearly outlines a different spatial structure, where the TPOJ2/TPOJ3 and STV/PSL boundaries are oriented differently from the MMP1.0 atlas.

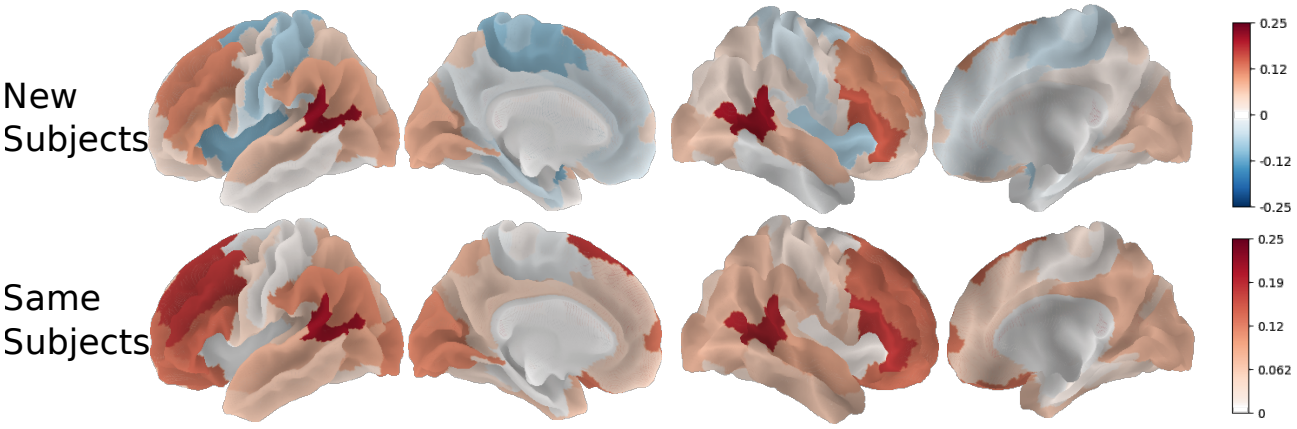

Figure 3: **Comparison of functional parcellations *versus* atlas-based parcellation.** (Top) This map shows $\Delta_{atlas} - \Delta_{individual}$, the distance difference between individualized parcellations and a population-level atlas for 160 functional contrasts, in each of the twenty one systems listed in Table 3. (Bottom) This map represents the same quantity, but without cross-validation: the model is trained and evaluated on the same subjects. In both cases, positive values mean that individualized parcellations outperform the atlas, and negative values, the converse.

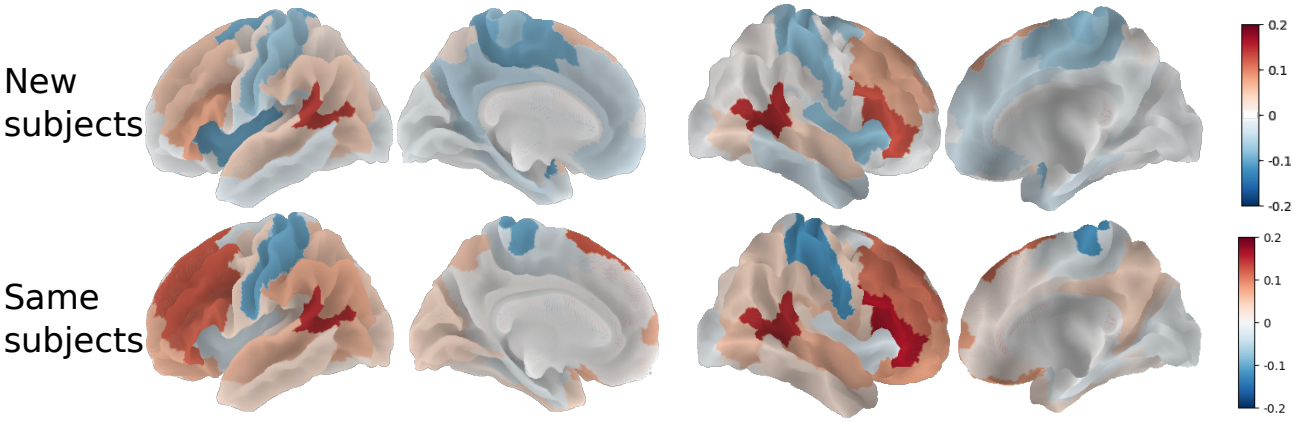

Figure 4: **Comparison of individual dictionary versus atlas-based representations.** (Top) The map displays the $\Delta_{fixed} - \Delta_{individual}$ distance difference between individualized parcellations and a fixed parcellation, for 160 functional contrasts, averaged in each of the twenty one systems listed in Table 3. (Bottom) The map represents the same thing, but without cross-validation: the model is trained and evaluated on the same subjects. In both cases, positive values mean that individualized parcellations outperform the fixed one, and negative values, the converse.

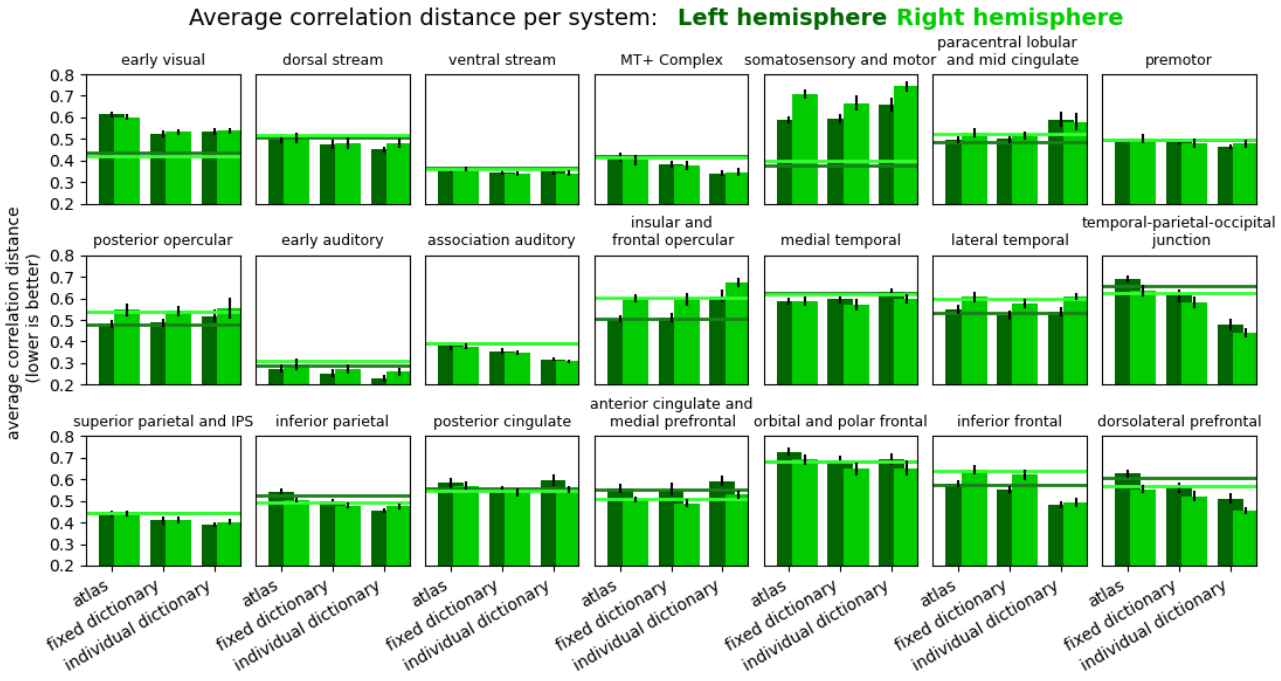

Figure 5: **Quantitative comparison of the parcellations fit per large-scale brain system.** The quality metric obtained for three spatial models is compared in twenty-one brain systems, across twenty validation folds (lower is better). The three spatial models are the MMP1.0 atlas, a fixed dictionary and a subject-specific dictionary; Δ_{atlas}, Δ_{fixed} and $\Delta_{individual}$ are displayed with bars representing between-fold variations. The horizontal line represents the predictive accuracy of a voxel level model as defined in Eq. 6.

Superior plus Inferior Parietal Sulci, and Temporal-Parietal-Occipital Junction). We hypothesize that regions where voxel-level fit outperforms parcellation-based fit are affected from an overly coarse parcellation scheme, and they would thus require a finer scheme, with more regions to capture subtle functional differences.

To check whether allowing a larger number of regions changes the capacity of parcellation schemes to fit brain data, we reproduced all experiments while doubling the number of components estimated by the dictionary fitting. Results are shown in Figure 6. Increasing the number of regions improves the goodness of fit of the parcellation models, see e.g. in the early visual cortex —yet it does not change the main trends observed previously. It enhances the performance of parcellation-based methods with respect to simple voxel-level modeling.

Reproduction on Schaefer atlas In this experiment, system definition corresponds to the 17 large-scale Yeo17 networks, and we compare individual and fixed dictionaries against Schaefer parcels in Fig. 7.

In that case, we observe a systematic lead of data-driven parcellations over the fixed atlas, and minor differences between the individual and the fixed parcellation. These results illustrate the benefit of data-driven parcellations, though not necessarily individualized parcellations. When compared to MMP1.0 parcellation, results are overall worse, outlining that the Schaefer atlas captures less well brain characteristics than MMP1.0.

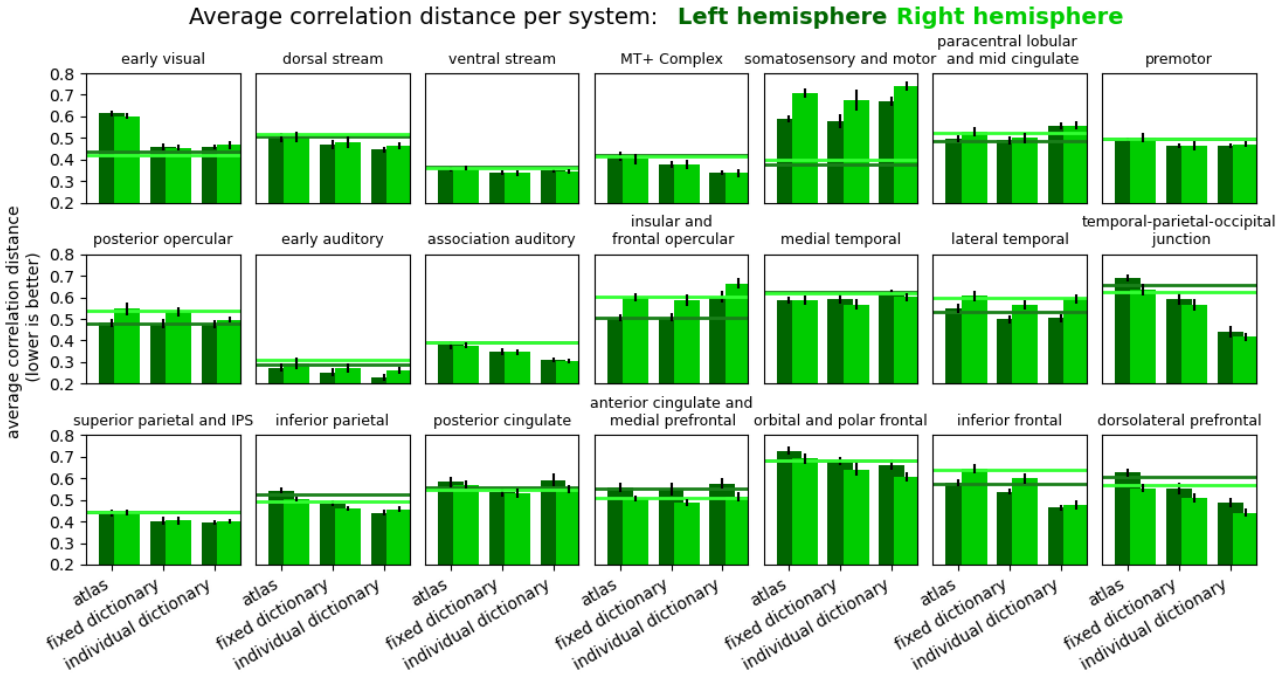

Figure 6: **Quantitative comparison of the atlas fit per large-scale brain system, with a dictionary having twice the number of components of the atlas.** The quality metric obtained for three spatial models is compared in twenty-one brain systems, across twenty validation folds (lower is better). The three spatial models are the MMP1.0 atlas, a fixed dictionary and a subject-specific dictionary; Δ_{atlas}, Δ_{fixed} and $\Delta_{individual}$ are displayed with bars representing between-fold variations. The horizontal line represents the predictive accuracy of a voxel level model as defined in Eq. 6. The difference with respect to Fig. 5 is that the dictionaries have been fitted with a number of components equal to twice the number of regions in the atlas.

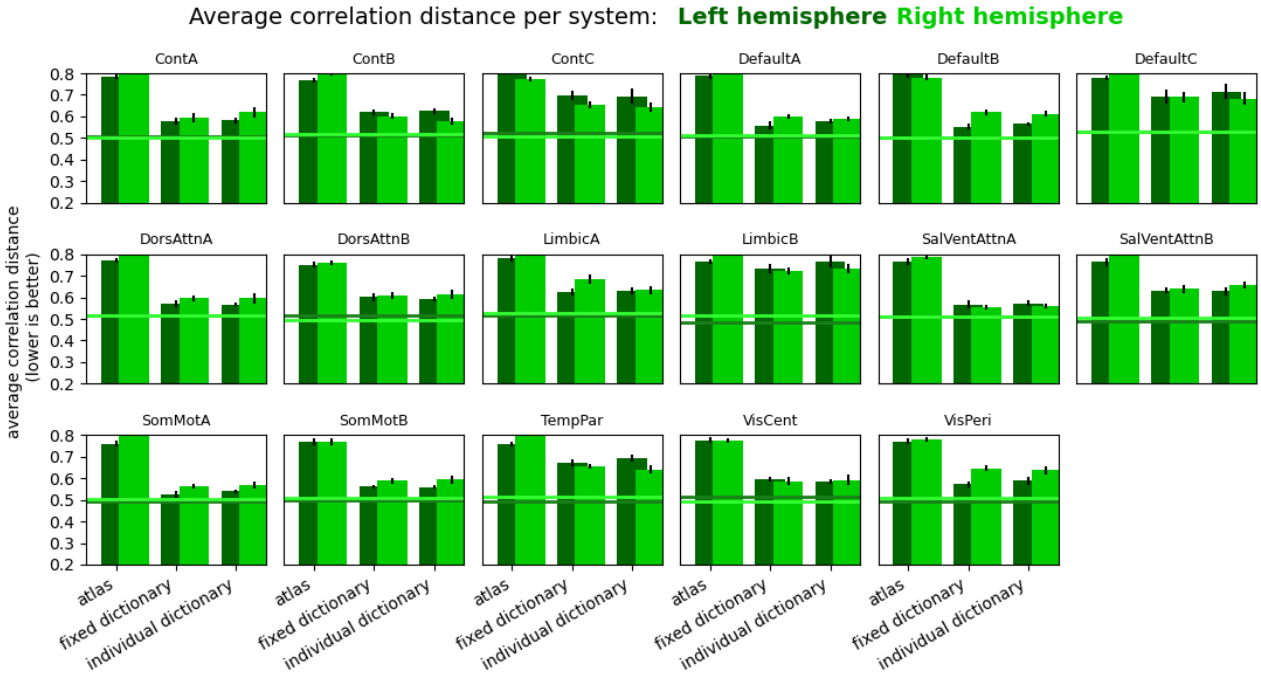

Figure 7: **Quantitative comparison of the atlas fit per large-scale brain system of the Schaefer400 subdivision of Yeo17 atlas.** The quality metric obtained for three spatial models is compared in seventeen brain systems, across twenty validation folds (lower is better). The three spatial models are the Schaefer400 atlas, a fixed dictionary and a subject-specific dictionary; Δ_{atlas}, Δ_{fixed} and $\Delta_{individual}$ are displayed with bars representing between-fold variations. The horizontal line represents the predictive accuracy of a voxel level model as defined in Eq. 6.

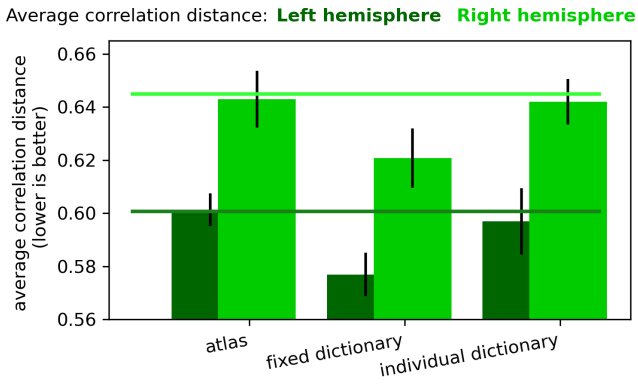

Figure 8: **Quantitative comparison of the atlas fit at the whole-brain level.** The quality metric obtained for three spatial models at the whole brain scale, across twenty validation folds. The three spatial models are the MMP1.0 atlas, a fixed dictionary and a subject-specific dictionary; Δ_{atlas}, Δ_{fixed} and $\Delta_{individual}$ are displayed with bars representing between-fold variations. The horizontal lines represent the predictive accuracy of a voxel level model as defined in Eq. 6.

Whole-brain parcellation We also considered whole brain parcellation schemes. We fixed the number of components to 181 and compare the average distances Δ_{atlas}, Δ_{fixed} and $\Delta_{individual}$ across systems, still using only contrasts that are relevant in each brain system. The results are shown in Fig. 8. Interestingly it shows that individual parcellations slightly outperform fixed atlas, such as MMP1.0, but perform less well than the fixed parcellations scheme. A likely reason is that the optimization problem underlying dictionary estimation becomes harder at a whole brain scale, with a larges number if components involved. Hence the fixed model fitted on more data is more robust to overfitting and has better generalization than the individualized model.

4 Discussion

When creating a reduced representation of a dataset, should one rely on traditional brain atlases, a data-driven but fixed parcellation, or an individual parcellation? Each solution has its pros and cons: atlases come with semantics about regions identity and make it possible to query information on regions from the literature (see e.g. [Yarkoni et al., 2011]). Fixed dictionaries allow overlapping regions and are more adaptive to the signal, leading to better fit [Dadi et al., 2020]. Yet, it is sometimes considered that individualized dictionaries better capture brain organization, and thus should be preferred. This view relies on the observation that inter-subject variability is partly not only noise but that it also contains meaningful information about individual characteristics. In short, individual dictionaries represent an appealing opportunity to personalize brain models. In the present article, we have shown that it is possible to create a reliable parcellation scheme on new subjects based on the individual parcellation of other subjects.

More in detail, if one simply considers the representation capacity of parcellations within individuals, the fact that individual dictionaries better capture individual brain organization than fixed dictionaries or atlases is a self-fulfilling prophecy: by construction, it is designed to provide the best summary of the input data, under relatively limited constraints. This phenomenon is illustrated in Fig. 3 (bottom), where the individual dictionary systematically outperforms the atlas, when it is evaluated on the subjects that it was trained on. This difference

in accuracy between training and testing data is exactly what is called *overfit*. To a lesser extent, individualized dictionaries show some overfit with respect to fixed dictionaries, as seen in Fig. 4(bottom). Overfitting is even more pronounced when the parcellations' goodness of fit is measured on the imaging contrasts that were used to create those parcellation (which are certainly affected by circularity, in such a case.

By contrast, generalizing parcellation schemes to new individuals provides a stricter and arguably more reliable test of the parcellations representation abilities. Indeed, such generalization-based framework penalizes methods that yield a considerable amount of variability, as they will generalize poorly to new individuals. For instance, when considering whole-brain parcellations, we found that individualized parcellations outperform a standard atlas, but did less well than a fixed parcellation scheme. This shows that the estimation of components across the whole brain remains a hard task, and that overfitting may well occur. This picture is however more interesting when analysis is carried out per brain system.

The present study leads to nuanced conclusions regarding the benefit of individualized parcellations. For instance, Figure 2 displays a region, the TPJ, in which functional organization is poorly captured by current atlases and calls for data-driven and individualized models. However, the gain is less obvious in other brain systems. More generally, in two thirds of such regions, like the Temporal-Parietal-Occipital Junction, the Inferior Parietal system, the MT+ Complex and Dorso-Lateral Prefrontal Cortex, there is a clear advantage of learning the atlas and its study at the individual level. The reason is probably that these regions exhibit substantial functional variability across individuals, that is not well accounted for by anatomical registration. On the other hand, in a few regions such as somato-sensory region or the insula, the reverse patterns were found: classical atlases outperform individual parcellations.

Moreover, we found that in most regions, atlas-based representations better generalize to new individuals than vertex-level representations, with two main exceptions: primary visual cortex and somato-sensory regions (see in Fig. 5 and 6, the error bar is higher for the individual dictionary than for the atlas, for the *somato-sensory and motor* and *early visual* systems; it is visible in both left and right hemispheres).

There are different possible causes for the diversity of results. First, it might be simply difficult to learn stable organization from the input data in some regions, such as a parcellation of the motor cortex from movie watching that does not involve motor actions. In that case, data-driven parcellations are simply ineffective: a standard atlas is more useful. Second, there might be a resolution effect: some systems, such as the visual cortex, require finer parcellations to obtain good results. Figure 6 depicts this situation, where doubling the number of parcels wrt Fig.5 improves accuracy in these ill-fitted regions. Overall, data-driven parcellations retain the advantage of adapting the resolution to better fit the signal.

However, the proposed approach has some limitations. We used a single metric, namely cosine distance [Manning et al., 2008], averaged over a selection of contrasts that display sufficient within-subject test-retest reliability. In particular, the metric is insensitive to signal magnitude, but in many context, such magnitude matters less than the topographic organization of the brain map. Following [Thirion et al., 2014], we could have also considered parcellation stability across individuals, but we believe that this type of analysis only leads to ambiguous results: what level of variability can indeed be considered to be tolerable? Does increased stability imply less sensitivity to the training data? In [Dadi et al., 2020], other criteria were taken into account, such as connectivity-based prediction, which is not addressed here. The relevance is less clear than the accuracy criterion we used, and more importantly, [Bijsterbosch et al., 2018] have shown that that the interpretation of differences in connectome-based prediction is quite complex, as it is hard to separate the connectivity estimates from the underlying atlas definition.

Finally, the work presented here relied quite extensively on a deep phenotyping approach, i.e., a collection

that yields abundant subject-specific data. In the present, we leveraged part of the IBC data, that provides a large variety of activation maps (160 contrasts were used here), highlighting functional characteristics of many brain systems. The price to pay is that data are obtained from a limited sample of participants, which may reduce the statistical validity of the results. We however observe that the results are very stable across 20 cross-validation folds, indicating that the conclusions are not too sensitive to sampling effects. This dataset also contains naturalistic data, that are extremely useful for parcellation purposes. This type of data can provide useful quantitative information on brain organization, and thus guide methodology for larger-scale studies, where many more individuals are sampled from a more limited set of experimental contexts. The present work highlights that parcellations learned from naturalistic stimuli (particularly, movie watching) are not well defined in somatosensory regions. We thus conjecture that learning them from dedicated task data, such as a motor localizer, would lead to a better result. Nevertheless, deep phenotyping allows us to address such questions, that we leave for future work. Another valuable direction of research is to compare parcellations learned from resting-state data, naturalistic data and functional contrasts. Finally, we expect that functional alignment could be used to enhance the similarity across participants, as demonstrated in [Thual et al., 2022], but this goes beyond the present work.

Conclusion This study provides evidence that individual data can advantageously be used to parcel brain imaging data, and that the resulting component definitions can be propagated to new individuals with some partial success: hotspots of brain functional organization variability will exhibit clear benefits, while other systems, prominently primary systems, do not. The nuances in our results imply that in practical situations, parcellations users will have to clearly assess the benefit of these representations. By default, relying on fixed scheme is a safe choice.

Acknowledgments. This project/research has received funding from the European Union's Horizon 2020 Framework Program for Research and Innovation under the Specific Grant Agreement No. 945539 (Human Brain Project SGA3). Bertrand Thirion is supported the KARAIB AI chair (ANR-20-CHIA-0025-01). Ana Luísa Pinho is the recipient of a BrainsCAN Postdoctoral Fellowship at Western University, funded by the Canada First Research Excellence Fund (CFREF).

References

[Abraham et al., 2014] Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., and Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. *Front Neuroinform*, 8:14.

[Amunts et al., 2014] Amunts, K., Hawrylycz, M. J., Van Essen, D. C., Van Horn, J. D., Harel, N., Poline, J.-B., De Martino, F., Bjaalie, J. G., Dehaene-Lambertz, G., Dehaene, S., Valdes-Sosa, P., Thirion, B., Zilles, K., Hill, S. L., Abrams, M. B., Tass, P. A., Vanduffel, W., Evans, A. C., and Eickhoff, S. B. (2014). Interoperable atlases of the human brain. *Neuroimage*, 99:525–532.

[Amunts et al., 2020] Amunts, K., Mohlberg, H., Bludau, S., and Zilles, K. (2020). Julich-Brain: A 3D probabilistic atlas of the human brain's cytoarchitecture. *Science*, 369(6506):988–992.

[Andersson et al., 2003] Andersson, J. L., Skare, S., and Ashburner, J. (2003). How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *Neuroimage*, 20(2):870 – 888.

[Badillo et al., 2013] Badillo, S., Vincent, T., and Ciuciu, P. (2013). Group-level impacts of within- and between-subject hemodynamic variability in fMRI. *Neuroimage*, 82:433–448.

[Behzadi et al., 2007] Behzadi, Y., Restom, K., Liau, J., and Liu, T. T. (2007). A component based noise correction method (compcor) for {BOLD} and perfusion based fMRI. *Neuroimage*, 37(1):90 – 101.

[Bijsterbosch et al., 2018] Bijsterbosch, J. D., Woolrich, M. W., Glasser, M. F., Robinson, E. C., Beckmann, C. F., Van Essen, D. C., Harrison, S. J., and Smith, S. M. (2018). The relationship between spatial configuration and functional connectivity of brain regions. *ELife*, 7:e32992.

[Bohland et al., 2009] Bohland, J. W., Bokil, H., Allen, C. B., and Mitra, P. P. (2009). The brain atlas concordance problem: quantitative comparison of anatomical parcellations. *PLoS One*, 4(9):e7200.

[Braga and Buckner, 2017] Braga, R. M. and Buckner, R. L. (2017). Parallel Interdigitated Distributed Networks within the Individual Estimated by Intrinsic Functional Connectivity. *Neuron*, 95(2):457–471.e5.

[Bryce et al., 2021] Bryce, N. V., Flournoy, J. C., Guassi Moreira, J. F., Rosen, M. L., Sambook, K. A., Mair, P., and McLaughlin, K. A. (2021). Brain parcellation selection: An overlooked decision point with meaningful effects on individual differences in resting-state functional connectivity. *Neuroimage*, 243(118487):118487.

[Campbell et al., 2015] Campbell, K. L., Shafto, M. A., Wright, P., Tsvetanov, K. A., Geerligs, L., Cusack, R., Tyler, L. K., Brayne, C., Bullmore, E., Calder, A., Cusack, R., Dalgleish, T., Duncan, J., Henson, R., Matthews, F., Marslen-Wilson, W., Rowe, J., Shafto, M., Campbell, K., Cheung, T., Davis, S., Geerligs, L., Kievit, R., McCarrey, A., Price, D., Taylor, J., Tsvetanov, K., Williams, N., Bates, L., Emery, T., Erzinçlioglu, S., Gadie, A., Gerbase, S., Georgieva, S., Hanley, C., Parkin, B., Troy, D., Allen, J., Amery, G., Amunts, L., Barcroft, A., Castle, A., Dias, C., Dowrick, J., Fair, M., Fisher, H., Goulding, A., Grewal, A., Hale, G., Hilton, A., Johnson, F., Johnston, P., Kavanagh-Williamson, T., Kwasniewska, M., McMinn, A., Norman, K., Penrose, J., Roby, F., Rowland, D., Sargeant, J., Squire, M., Stevens, B., Stoddart, A., Stone, C., Thompson, T., Yazlik, O., Dixon, M., Barnes, D., Hillman, J., Mitchell, J., Villis, L., and Tyler, L. K. (2015). Idiosyncratic responding during movie-watching predicted by age differences in attentional control. *Neurobiol Aging*, 36(11):3045 – 3055.

[Chen et al., 2015] Chen, P.-H. C., Chen, J., Yeshurun, Y., Hasson, U., Haxby, J., and Ramadge, P. J. (2015). A reduced-dimension fmri shared response model. In Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc.

[Dadi et al., 2019] Dadi, K., Rahim, M., Abraham, A., Chyzhyk, D., Milham, M., Thirion, B., Varoquaux, G., and Alzheimer's Disease Neuroimaging Initiative (2019). Benchmarking functional connectome-based predictive models for resting-state fMRI. *Neuroimage*, 192:115–134.

[Dadi et al., 2020] Dadi, K., Varoquaux, G., Machlouzarides-Shalit, A., Gorgolewski, K. J., Wassermann, D., Thirion, B., and Mensch, A. (2020). Fine-grain atlases of functional modes for fMRI analysis. *Neuroimage*, 221(117126):117126.

[Dohmatob et al., 2021] Dohmatob, E., Richard, H., Pinho, A. L., and Thirion, B. (2021). Brain topography beyond parcellations: Local gradients of functional maps. *Neuroimage*, 229:117706.

[Fischl et al., 1999] Fischl, B., Sereno, M. I., Tootell, R. B., and Dale, A. M. (1999). High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum Brain Mapp*, 8(4):272–284.

[Friston et al., 1995] Friston, K., Frith, C., Frackowiak, R., and Turner, R. (1995). Characterizing Dynamic Brain Responses with fMRI: a Multivariate Approach. *Neuroimage*, 2(2):166–172.

[Glasser et al., 2016] Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C. F., Jenkinson, M., Smith, S. M., and van Essen, D. C. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, 536(7615):171–178.

[Gordon et al., 2017] Gordon, E. M., Laumann, T. O., Gilmore, A. W., Newbold, D. J., Greene, D. J., Berg, J. J., Ortega, M., Hoyt-Drazen, C., Gratton, C., Sun, H., Hampton, J. M., Coalson, R. S., Nguyen, A. L., McDermott, K. B., Shimony, J. S., Snyder, A. Z., Schlaggar, B. L., Petersen, S. E., Nelson, S. M., and Dosenbach, N. U. (2017). Precision Functional Mapping of Individual Human Brains. *Neuron*, 95(4):791 – 807.e7.

[Gratton et al., 2018] Gratton, C., Laumann, T. O., Nielsen, A. N., Greene, D. J., Gordon, E. M., Gilmore, A. W., Nelson, S. M., Coalson, R. S., Snyder, A. Z., Schlaggar, B. L., Dosenbach, N. U., and Petersen, S. E. (2018). Functional Brain Networks Are Dominated by Stable Group and Individual Factors, Not Cognitive or Daily Variation. *Neuron*, 98(2):439 – 452.e5.

[Greene et al., 2020] Greene, D. J., Marek, S., Gordon, E. M., Siegel, J. S., Gratton, C., Laumann, T. O., Gilmore, A. W., Berg, J. J., Nguyen, A. L., Dierker, D., Van, A. N., Ortega, M., Newbold, D. J., Hampton, J. M., Nielsen, A. N., McDermott, K. B., Roland, J. L., Norris, S. A., Nelson, S. M., Snyder, A. Z., Schlaggar, B. L., Petersen, S. E., and Dosenbach, N. U. (2020). Integrative and Network-Specific Connectivity of the Basal Ganglia and Thalamus Defined in Individuals. *Neuron*, 105(4):742 – 758.e6.

[Hale et al., 2022] Hale, J. T., Campanelli, L., Li, J., Bhattasali, S., Pallier, C., and Brennan, J. R. (2022). Neurocomputational Models of Language Processing. *Annual Review of Linguistics*, 8(1):null.

[Harrison et al., 2020] Harrison, S. J., Bijsterbosch, J. D., Segerdahl, A. R., Fitzgibbon, S. P., Farahibozorg, S.-R., Duff, E. P., Smith, S. M., and Woolrich, M. W. (2020). Modelling subject variability in the spatial and temporal characteristics of functional modes. *Neuroimage*, 222(117226):117226.

[Haxby et al., 2011] Haxby, J. V., Guntupalli, J. S., Connolly, A. C., Halchenko, Y. O., Conroy, B. R., Gobbini, M. I., Hanke, M., and Ramadge, P. J. (2011). A Common, High-Dimensional Model of the Representational Space in Human Ventral Temporal Cortex. *Neuron*, 72(2):404 – 416.

[Kasties et al., 2021] Kasties, V., Karnath, H. O., and Sperber, C. (2021). Strategies for feature extraction from structural brain imaging in lesion-deficit modelling. *Hum Brain Mapp*, 42(16):5409–5422.

[Kong et al., 2019] Kong, R., Li, J., Orban, C., Sabuncu, M. R., Liu, H., Schaefer, A., Sun, N., Zuo, X. N., Holmes, A. J., Eickhoff, S. B., and Yeo, B. T. T. (2019). Spatial Topography of Individual-Specific Cortical Networks Predicts Human Cognition, Personality, and Emotion. *Cereb Cortex*, 29(6):2533–2551.

[Kong et al., 2023] Kong, R., Tan, Y. R., Wulan, N., Ooi, L. Q. R., Farahibozorg, S.-R., Harrison, S., Bijsterbosch, J. D., Bernhardt, B. C., Eickhoff, S., and Thomas Yeo, B. T. (2023). Comparison between gradients and parcellations for functional connectivity prediction of behavior. *Neuroimage*, 273(120044):120044.

[Lawrence et al., 2021] Lawrence, R. M., Bridgeford, E. W., Myers, P. E., Arvapalli, G. C., Ramachandran, S. C., Pisner, D. A., Frank, P. F., Lemmer, A. D., Nikolaidis, A., and Vogelstein, J. T. (2021). Standardizing human brain parcellations. *Sci. Data*, 8(1):78.

[Manning et al., 2008] Manning, C. D., Raghavan, P., and Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press, Cambridge, UK.

[Mantini et al., 2012] Mantini, D., Hasson, U., Betti, V., Perrucci, M. G., Romani, G. L., Corbetta, M., Orban, G. A., and Vanduffel, W. (2012). Interspecies activity correlations reveal functional correspondence between monkey and human brain areas. *Nat Methods*, 9(3):277–282.

[Mensch et al., 2021] Mensch, A., Mairal, J., Thirion, B., and Varoquaux, G. (2021). Extracting representations of cognition across neuroimaging studies improves brain decoding. *PLoS Comput. Biol.*, 17(5):e1008795.

[Menuet et al., 2022] Menuet, R., Meudec, R., Dockès, J., Varoquaux, G., and Thirion, B. (2022). Comprehensive decoding mental processes from web repositories of functional brain images. *Sci. Rep.*, 12(1):7050.

[Nastase et al., 2020] Nastase, S. A., Goldstein, A., and Hasson, U. (2020). Keep it real: rethinking the primacy of experimental control in cognitive neuroscience. *Neuroimage*, 222(117254):117254.

[Nickerson et al., 2017] Nickerson, L. D., Smith, S. M., Öngür, D., and Beckmann, C. F. (2017). Using dual regression to investigate network shape and amplitude in functional connectivity analyses. *Front. Neurosci.*, 11:115.

[Nishimoto et al., 2011] Nishimoto, S., Vu, A., Naselaris, T., Benjamini, Y., Yu, B., and Gallant, J. (2011). Reconstructing Visual Experiences from Brain Activity Evoked by Natural Movies. *Curr Biol*, 21:1641–6.

[Pedregosa et al., 2011] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-Learn: Machine Learning in Python. *J Mach Learn Res*, 12:2825–2830.

[Pervaiz et al., 2020] Pervaiz, U., Vidaurre, D., Woolrich, M. W., and Smith, S. M. (2020). Optimising network modelling methods for fMRI. *Neuroimage*, 211(116604):116604.

[Pinho et al., 2021] Pinho, A. L., Amadon, A., Fabre, M., Dohmatob, E., Denghien, I., Torre, J. J., Ginisty, C., Becuwe-Desmidt, S., Roger, S., Laurier, L., Joly-Testault, V., Médiouni-Cloarec, G., Doublé, C., Martins, B., Pinel, P., Eger, E., Varoquaux, G., Pallier, C., Dehaene, S., Hertz-Pannier, L., and Thirion, B. (2021). Subject-specific segregation of functional territories based on deep phenotyping. *Hum Brain Mapp*, 42(4):841–870.

[Pinho et al., 2020] Pinho, A. L., Amadon, A., Gauthier, B., Clairis, N., Knops, A., Genon, S., Dohmatob, E., Jesús Torre, J., Ginisty, C., Becuwe-Desmidt, S., Roger, S., Lecomte, Y., Berland, V., Laurier, L., Joly-Testault, V., Médiouni-Cloarec, G., Doublé, C., Martins, B., Salmon, E., Piazza, M., Melcher, D., Pessiglione, M., Van Wassenhove, V., Eger, E., Varoquaux, G., Dehaene, S., Hertz-Pannier, L., and Thirion, B. (2020). Individual Brain Charting dataset extension, second release of high-resolution fMRI data for cognitive mapping. *Sci Data*, 7(1).

[Pinho et al., 2018] Pinho, A. L., Amadon, A., Ruest, T., Fabre, M., Dohmatob, E., Denghien, I., Ginisty, C., Séverine-Becuwe, Roger, S., Laurier, L., Joly-Testault, V., Médiouni-Cloarec, G., Doublé, C., Martins, B., Pinel, P., Eger, E., Varoquaux, G., Pallier, C., Dehaene, S., Hertz-Pannier, L., and Thirion, B. (2018). Individual Brain Charting, a high-resolution fMRI dataset for cognitive mapping. *Sci Data*, 5:180105.

[Richard et al., 2021] Richard, H., Ablin, P., Thirion, B., Gramfort, A., and Hyvarinen, A. (2021). Shared independent component analysis for multi-subject neuroimaging. In Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W., editors, *Advances in Neural Information Processing Systems*, volume 34, pages 29962–29971. Curran Associates, Inc.

[Schaefer et al., 2017] Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X.-N., Holmes, A. J., Eickhoff, S. B., and Yeo, B. T. (2017). Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity mri. *Cereb Cortex*, 28(9):3095–3114.

[Tavor et al., 2016] Tavor, I., Jones, O. P., Mars, R. B., Smith, S. M., Behrens, T. E., and Jbabdi, S. (2016). Task-free MRI predicts individual differences in brain activity during task performance. *Science*, 352(6282):216–220.

[Thirion et al., 2006] Thirion, B., Flandin, G., Pinel, P., Roche, A., Ciuciu, P., and Poline, J.-B. (2006). Dealing with the shortcomings of spatial normalization: multi-subject parcellation of fMRI datasets. *Hum. Brain Mapp.*, 27(8):678–693.

[Thirion et al., 2021] Thirion, B., Thual, A., and Pinho, A. L. (2021). From deep brain phenotyping to functional atlasing. *Current Opinion in Behavioral Sciences*, 40:201–212.

[Thirion et al., 2014] Thirion, B., Varoquaux, G., Dohmatob, E., and Poline, J.-B. (2014). Which fMRI clustering gives good brain parcellations? *Front. Neurosci.*, 8:167.

[Thomas et al., 2022] Thomas, A., Ré, C., and Poldrack, R. (2022). Self-supervised learning of brain dynamics from broad neuroimaging da ta. In Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., and añd A. Oh, K. C., editors, *Advances in Neural Information Processing Systems*, volume 35, pages 21255–21269. Curran Associates, Inc.

[Thual et al., 2022] Thual, A., TRAN, Q. H., Zemskova, T., Courty, . N., Flamary, R., Dehaene, S., and Thirion, B. (2022). Aligning individual brains with fused unbalanced gromov wasserstein. In Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., and añd A. Oh, K. C., editors, *Advances in Neural Information Processing Systems*, volume 35, pages 21792–21804. Curran Associates, Inc.

[Tzourio-Mazoyer et al., 2002] Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., and Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, 15(1):273–289.

[Uddin et al., 2023] Uddin, L. Q., Betzel, R. F., Cohen, J. R., Damoiseaux, J. S., De Brigard, F., Eickhoff, S. B., Fornito, A., Grattoñ, C., Gordon, E. M., Laird, A. R., Larson-Prior, L., McIntosh, A. R., Nickerson, L. D., Pessoa, L., Pinho, A. L., Poldrack, R. u. A., Razi, A., Sadaghiani, S., Shine, J. M., Yendiki, A., Yeo, B. T. T., and Spreng, R. N. (2023). Controversies and progress on standardization of large-scale brain network nomenclature. *Netw Neurosci*, 7(3):864–905.

[van Essen et al., 2012] van Essen, D. C., Glasser, M. F., Dierker, D. L., Harwell, J., and Coalson, T. (2012). Parcellations and Hemispheric Asymmetries of Human Cerebral Cortex Analyzed on Surface-Based Atlases. *Cereb Cortex*, 22(10):2241.

[Varoquaux and Poldrack, 2019] Varoquaux, G. and Poldrack, R. (2019). Predictive models avoid excessive reductionism in cognitive neuroimaging. *Curr Opin Neurobiol*, 55.

[Varoquaux et al., 2013] Varoquaux, G., Schwartz, Y., Pinel, P., and Thirion, B. (2013). Cohort-Level Brain Mapping: Learning Cognitive Atoms to Single Out Specialized Regions. In Gee, J. C., Joshi, S., Pohl, K. M., Wells, W. M., and Zöllei, L., editors, *Inf Process Med Imaging*, volume 23, pages 438–449, Berlin, Heidelberg. Springer.

[Yarkoni et al., 2011] Yarkoni, T., Poldrack, R. A., Nichols, T. E., van Essen, D. C., and Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nat Methods*, 8(8):665–70.

[Yeo et al., 2011] Yeo, B. T. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., et al. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J Neurophysio*, 106:1125.

[Zhao et al., 2020] Zhao, J., Tang, C., and Nie, J. (2020). Functional parcellation of individual cerebral cortex based on functional MRI. *Neuroinformatics*, 18(2):295–306.

A List of the task contrasts used in the presented experiments

We list in the Table 2 the list of functional contrasts that were considered to evaluate parcellation accuracy [13] .

Table 2: List of functional tasks and contrasts of the IBC dataset considered in this study.

Task	Contrast	Description
archi standard	left-right button press	left vs. right hand button press
archi standard	horizontal-vertical	horizontal vs. vertical checkerboard
archi standard	computation-sentences	mental subtraction vs. sentence reading
archi standard	reading-listening	reading sentence vs. listening to sentence
archi standard	reading-checkerboard	read sentence vs. checkerboard
archi standard	motor-cognitive	button presses vs. narrative/computation
archi spatial	saccades	saccade vs. fixation
archi spatial	rotation side	hand palm or back vs. fixation

[13] see the documentation for details on these contrasts.

Table 2: List of functional tasks and contrasts of the IBC dataset considered in this study.

Task	Contrast	Description
archi spatial	hand-side	left or right hand vs. hand palm or back
archi spatial	object orientation	image orientation reporting
archi spatial	grasp-orientation	object grasping vs. orientation reporting
archi social	triangle random	randomly drifting triangle
archi social	triangle mental-random	mental motion vs. random motion
archi social	mechanistic video	reading a mechanistic story
archi social	mechanistic audio	listening to a mechanistic tale
archi social	false belief-mechanistic video	false-belief story vs. mechanistic story
archi social	false belief-mechanistic audio	false-belief tale vs. mechanistic tale
archi social	non speech sound	listen to natural sound
archi social	speech-non speech	listen to voice sound vs. natural sound
archi emotional	face gender-control	guess the gender from face image
archi emotional	face trusty-gender	assess face trustfulness vs. gender
archi emotional	expression gender-control	guess the gender from eyes image vs. view scrambled image
archi emotional	expression intention-gender	guess intention vs. gender from eyes image
hcp emotion	shape	shape comparison
hcp emotion	face-shape	emotional face comparison vs. shape comparison
hcp gambling	reward	gambling with positive outcome
hcp gambling	punishment-reward	negative vs. positive gambling outcome
hcp motor	tongue-avg	move tongue vs. hands and feet
hcp language	math	mental additions
hcp language	story-math	listening to tale vs. mental additions
hcp relational	match	visual feature matching vs. fixation
hcp relational	relational-match	relational comparison vs. matching
hcp social	random	random motion vs. fixation
hcp social	mental-random	mental motion vs. random motion
hcp wm	2back-0back	2-back vs. 0-back
rsvp language	consonant string	read and encode consonant strings vs. fixation
rsvp language	word-consonant string	read words vs. consonant strings
rsvp language	pseudo-consonant string	read pseudowords vs. consonant strings
rsvp language	word-pseudo	read words vs. pseudowords
rsvp language	complex-simple	read sentence with complex vs. simple syntax
rsvp language	sentence-word	read sentence vs. words
rsvp language	jabberwocky-pseudo	read jabberwocky vs. pseudowords

Table 2: List of functional tasks and contrasts of the IBC dataset considered in this study.

Task	Contrast	Description
rsvp language	sentence-jabberwocky	read sentence vs. jabberwocky
mtt we	we average reference	updating ones position in space and time in west-east island
mtt we	we all space cue	spatial cue of the next event in west-east island
mtt we	we all time cue	time cue of the next event in west-east island
mtt we	we all space-time cue	spatial vs. time cues in west-east island
mtt we	we all time-space cue	time vs. spatial cues in west-east island
mtt we	we average event	figuring out the space or time of an event in west-east island
mtt we	we space event	figuring out the position of an event in west-east island
mtt we	we time event	figuring out the time of an event in west-east island
mtt we	we space-time event	event in space vs. event in time in west-east island
mtt we	we time-space event	event in time vs. event in space in west-east island
mtt we	westside-eastside event	events occuring westside vs. eastside
mtt we	eastside-westside event	events occuring eastside vs. westside
mtt we	we before-after event	events occuring before vs. after in west-east island
mtt we	we after-before event	events occuring after vs. before in west-east island
mtt we	we all event response	motor responses performed after every event condition in the west-east island
mtt sn	sn average reference	updating ones position in space and time in south-north island
mtt sn	sn all space cue	spatial cue of the next event in south-north island
mtt sn	sn all time cue	time cue of the next event in south-north island
mtt sn	sn all space-time cue	spatial vs. time cues in south-north island
mtt sn	sn all time-space cue	time vs. spatial cues in south-north island
mtt sn	sn average event	figuring out the space or time of an event in south-north island
mtt sn	sn space event	figuring out the position of an event in south-north island
mtt sn	sn time event	figuring out the time of an event in south-north island
mtt sn	sn space-time event	event in space vs. event in time in south-north island
mtt sn	sn time-space event	event in time vs. event in space in south-north island
mtt sn	southside-northside event	events occuring southside vs. northside
mtt sn	northside-southside event	events occuring northsife vs. southside
mtt sn	sn before-after event	events occuring before vs. after in south-north island
mtt sn	sn after-before event	events occuring after vs. before in south-north island

Table 2: List of functional tasks and contrasts of the IBC dataset considered in this study.

Task	Contrast	Description
mtt sn	sn all event response	motor responses performed after all event condition in the south-north island
preference food	food constant	evaluation of food
preference food	food linear	linear effect of food preference
preference food	food quadratic	quadratic effect of food preference
preference paintings	painting constant	evaluation of paintings
preference paintings	painting linear	linear effect of paintings preference
preference paintings	painting quadratic	quadratic effect of paintings preference
preference faces	face constant	evaluation of faces
preference faces	face linear	linear effect of face preference
preference faces	face quadratic	quadratic effect of face preference
preference houses	house constant	evaluation of houses
preference houses	house linear	linear effect of houses preference
preference houses	house quadratic	quadratic effect of houses preference
theory of mind	photo	manipulation of fact judgments
theory of mind	belief-photo	belief vs. factual judgments
emotional pain	physical pain	reading physical pain story
emotional pain	emotional-physical pain	emotional vs. physical pain story
pain movie	movie pain	movie with physically painful events
pain movie	movie mental-pain	mental events vs. physically painful events
self	instructions	read instruction in form of a question
bang	talk-no talk	speech vs. non-speech sections in movie watching
bang	no talk	non-speech section in movie watching
lyon lec2	attend	response to attended text
lyon lec2	unattend	response to unattended text
lyon lec2	attend-unattend	response to attended vs. unattended text
lyon audi	silence	listen to silence
lyon audi	tear-silence	listen to tears
lyon audi	suomi-silence	listen to unknown language
lyon audi	yawn-silence	listen to yawning
lyon audi	human-silence	listen to human sounds
lyon audi	music-silence	listen to music
lyon audi	reverse-silence	listen to reversed speech
lyon audi	speech-silence	listen to speech
lyon audi	alphabet-silence	listen to letters
lyon audi	cough-silence	listen to coughing
lyon audi	environment-silence	listen to environment sounds
lyon audi	laugh-silence	listen to laugh

Table 2: List of functional tasks and contrasts of the IBC dataset considered in this study.

Task	Contrast	Description
lyon audi	animals-silence	listen to animals
lyon visu	scrambled	view a scrambled image
lyon visu	face-scrambled	view a face image
lyon visu	characters-scrambled	view a characters
lyon visu	scene-scrambled	view a scene
lyon visu	house-scrambled	view a house
lyon visu	animal-scrambled	view an animal
lyon visu	pseudoword-scrambled	view a pseudoword
lyon visu	tool-scrambled	view a tool
lyon lec1	random string	read a random string
lyon lec1	word-random string	read a word vs. a random string
lyon lec1	word-pseudoword	read a word vs. a pseudoword
lyon lec1	pseudoword-random string	read a pseudoword vs. a random string
lyon mveb	2 letters different-same	maintaining two letters vs. one
lyon mveb	4 letters different-same	maintaining four letters vs. one
lyon mveb	6 letters different-same	maintaining six letters vs. one
lyon mveb	6 letters different-2 letters different	maintaining six letters vs. two
lyon mvis	2 dots-2 dots control	maintain position of two dots vs. one
lyon mvis	4 dots-4 dots control	maintain position of four dots vs. one
lyon mvis	6 dots-6 dots control	maintain position of six dots vs. one
lyon mvis	6 dots-2 dots	maintain position of six dots vs. two
lyon moto	instructions	read instructions
lyon moto	finger right-fixation	right finger tapping vs. any movement
lyon moto	finger left-fixation	left finger tapping vs. any movement
lyon moto	foot left-fixation	move left foot vs. any movement
lyon moto	foot right-fixation	move right foot vs. any movement
lyon moto	hand left-fixation	move left hand vs. any movement
lyon moto	hand right-fixation	move right hand vs. any movement
lyon moto	saccade-fixation	saccade vs. any movement
lyon moto	tongue-fixation	move tongue vs. any movement
lyon mcse	salience left-right	looking for a symbol in left vs. right visual field
lyon mcse	low-high salience	looking for a low-salient symbol
audio	music-silence	listen to music vs. silence
audio	speech-silence	listen to speech vs. silence

B List of the brain systems considered

The list of brain systems considered in our experiments is given in Table 3.

Table 3: List of the systems used to parcel the cortical surface. These systems are those described in [Glasser et al., 2016], with the exception that V1 is not handled as an independent regions but associated with other primary visual areas.

System name	MMP1.0 region names	MMP1.0 regions indices	size (vertices) left/right
early visual	V1, V2, V3, V4	1,4,5,6	10714/10455
dorsal stream	V3A, V3B, V6, V6A, V7, IPS1	3,13,16,17,19,152	2692/2774
ventral stream	V8, VVC, PIT, FFC, VMV1, VMV2,VMV3	7, 18, 22, 153, 154, 160, 163	3117/3351
MT+ Complex	V3CD, LO1, LO2, LO3, V4t, FST, MT, MST, PH	2, 20, 21, 23, 138, 156, 157, 158, 159	3795/3243
somatosensory and motor	4, 3a, 3b, 1, 2	8, 9, 51, 52, 53	13388/13135
paracentral lobular and mid cingulate	24dd, 24dv, 6mp, 6ma, SCEF, 5m, 5L, 5mv	36, 37, 39, 40, 41, 43, 44, 55	7665/8240
premotor	55b, 6d, 6a, FEF, 6v, 6r, PEF	10, 11, 12, 54, 56, 78, 96	6184/6338
posterior opercular	43, FOP1 , OP4, OP1, OP2-3, PFcm	99, 100, 101, 102, 105, 113	3868/3913
early auditory	A1, MBelt, LBelt, Pbelt, RI	24, 104, 124, 173, 174	3245/2558
association auditory	A4, A5, STSdp, STSda, STSvp, STSva, STGa, TA2	107, 123, 125, 128, 129, 130, 175, 176	5640/5961
insular and frontal opercular	52, PI, Ig, PoI1, PoI2, FOP2, FOP3, MI, AVI, AAIC, Pir,FOP4, FOP5	103, 106, 108, 109, 110, 111, 112, 114, 115, 167, 168, 169, 178	8445/8305
medial temporal	H, PreS, EC, PeEc, PHA1, PHA2, PHA3	118, 119, 120, 122, 126, 127, 155	5660 /4398
lateral temporal	PHT, TE1p, TE1m, TE1a, TE2p, TE2a, TGv, TGd, TF	131, 132, 133, 134, 135, 136, 137, 172, 177	9284/8305
temporal-parietal-occipital junction	TPOJ1, TPOJ2, TPOJ3, STV, PSL	25, 28, 139, 140, 141	3406/4769
superior parietal and IPS	LIPv, LIPd, VIP, AIP, MIP, 7PC, 7AL, 7Am, 7PL, 7Pm	29, 42, 45, 46, 47, 48, 49, 50, 95, 117	8552/8125
inferior parietal	PGp, PGs, PGi, PFm, PF, PFt, PFop, IP0, IP1, IP2	116, 143, 144, 145, 146, 147, 148, 149, 150, 151	12814/12469
posterior cingulate	DVT, ProS, POS1, POS2, RSC, v23ab, d23ab, 31pv, 31pd, 31a, 23d, 23c, PCV, 7m	14, 15, 27, 30, 31, 32, 33, 34, 35, 38, 121, 142, 161, 162	10082/10304
anterior cingulate and medial prefrontal	33pr, p24pr, a24pr, p24, a24, p32pr, a32pr, d32, p32, s32, 8BM, 9m, 10v, 10r, 25	57, 58, 59, 60, 61, 62, 63, 64, 65, 69, 88, 164, 165, 179, 180	7874/8736
orbital and polar frontal	47s, 47m, a47r, 11l, 13l, a10p, p10p, 10pp, 10d., OFC, pOFC	66, 72, 77, 89, 90, 91, 92, 93, 94, 166, 170	7299/7772
inferior frontal	44, 45, IFJp, IFJa, IFSp, IFSa, 47l, p47r	74, 75, 76, 79, 80, 81, 82, 171	5163/4617
dorsolateral prefrontal	8C, 8Av, i6-8, s6-8, SFL, 8BL, 9p, 9a, 8Ad, p9-46v, a9-46v, 46, 9-46d	26, 67, 68, 70, 71, 73, 83, 84, 85, 86, 87, 97, 98	11087/11163

C Contrasts selected to test

In each brain system studied, the parcellation model was tested on a subset of the contrasts that achieve high test-retest reproducibility in average across subjects: the criterion is that the correlation of the maps restricted to a given brain system across runs has to be greater than 0.5. For completeness we list in Table 4 the contrasts used in each system in the left hemisphere. Accordingly, we list in Table 5 the contrasts used in each system in the right hemisphere.

Table 4: **Contrasts used in each system of the left hemisphere for performance assessment.** These contrasts have been selected so that the correlation of system-level activity maps is greater than .5 in average across runs.

early visual	reading-listening, reading-checkerboard, horizontal-vertical, rotation side, object orientation, triangle random, mechanistic video, mental-random, random, reward, shape, face-shape, match, face-avg, place-avg, consonant string, we average event, house constant, face constant, food constant, photo, physical pain, enumeration constant, encode other, recognition other hit, correct rejection, attend, scrambled, random string, word-pseudoword, word-random string, pseudoword-random string, letter occurrence response
dorsal stream	reading-checkerboard, saccades, rotation side, object orientation, triangle random, mental-random, random, reward, face-avg, pseudo-consonant string, consonant string, we average event, sn average event, house constant, food constant, photo, physical pain, movie pain, vstm constant, enumeration constant, encode other, recognition other hit, random string, word-pseudoword, letter occurrence response, low-high salience
ventral stream	reading-listening, rotation side, object orientation, mechanistic video, random, reward, face-shape, face-avg, place-avg, house constant, face constant, food constant, photo, physical pain, vstm constant, enumeration constant, attend, random string
MT+ Complex	reading-listening, saccades, rotation side, object orientation, triangle random, mechanistic video, story-math, mental-random, random, reward, face-shape, body-avg, face-avg, place-avg, pseudo-consonant string, consonant string, we average event, sn average event, house constant, face constant, food constant, photo, physical pain, vstm constant, enumeration constant, encode other, recognition other hit, unattend, attend, random string, letter occurrence response
somatosensory and motor	motor-cognitive, object orientation, random, left hand-avg, right hand-avg, right foot-avg, tongue-avg, we average event, sn average event, house constant, face constant, food constant, enumeration constant, tongue-fixation
paracentral lobular and mid cingulate	object orientation, right hand-avg, right foot-avg, 2back-0back, we average event, sn average event, food constant, photo, physical pain, enumeration constant

premotor	motor-cognitive, computation-sentences, saccades, rotation side, object orientation, story-math, random, reward, right hand-avg, tongue-avg, 2back-0back, we average event, sn average event, house constant, face constant, food constant, photo, physical pain, enumeration constant, attend, random string, pseudoword-random string, 6 letters different-same, 6 dots-2 dots
posterior opercular	motor-cognitive, object orientation, random, left hand-avg, tongue-avg, we average event, sn average event, house constant, face constant, food constant, enumeration constant, alphabet-silence
early auditory	reading-listening, mechanistic audio, non speech sound, story-math, we average event, sn average reference, sn average event, enumeration constant, tear-silence, suomi-silence, yawn-silence, human-silence, silence, reverse-silence, alphabet-silence, cough-silence, environment-silence, laugh-silence, animals-silence
association auditory	reading-listening, mechanistic audio, mechanistic video, non speech sound, expression intention-gender, story-math, sentence-jabberwocky, sentence-word, word-consonant string, we average reference, we average event, sn average reference, sn average event, photo, physical pain, emotional-physical pain, enumeration constant, tear-silence, suomi-silence, human-silence, reverse-silence, alphabet-silence
insular and frontal opercular	story-math, 2back-0back, we average event, sn average event, face constant, photo, physical pain, enumeration constant
medial temporal	story-math, face-shape, face-avg, place-avg, tools-avg, house constant, emotional-physical pain, enumeration constant
lateral temporal	story-math, 2back-0back, sentence-jabberwocky, we average event, sn average event, photo, physical pain, emotional-physical pain, enumeration constant
temporal-parietal-occipital junction	reading-listening, saccades, rotation side, object orientation, mechanistic audio, mechanistic video, non speech sound, expression intention-gender, story-math, mental-random, random, reward, sentence-jabberwocky, word-consonant string, we average event, sn average event, face constant, photo, physical pain, enumeration constant, attend, suomi-silence, alphabet-silence
superior parietal and IPS	reading-listening, motor-cognitive, computation-sentences, saccades, hand-side, rotation side, object orientation, story-math, random, reward, 2back-0back, word-consonant string, pseudo-consonant string, consonant string, we average event, sn average event, house constant, face constant, food constant, photo, physical pain, enumeration constant, pseudoword-random string, 6 letters different-same, 6 dots-2 dots
inferior parietal	motor-cognitive, computation-sentences, rotation side, object orientation, expression intention-gender, story-math, random, reward, 2back-0back, sentence-jabberwocky, we average event, sn average event, house constant, face constant, food constant, photo, physical pain, emotional-physical pain, movie pain, movie mental-pain, enumeration constant, encode other
posterior cingulate	story-math, random, reward, we average event, sn average event, photo, physical pain, enumeration constant

anterior cingulate and medial prefrontal	expression intention-gender, story-math, 2back-0back, we average event, sn average event, physical pain, enumeration constant, encode other
orbital and polar frontal	story-math, reward, we average event, photo, physical pain, enumeration constant, enumeration linear, attend
inferior frontal	computation-sentences, grasp-orientation, object orientation, mechanistic video, expression intention-gender, story-math, reward, 2back-0back, sentence-jabberwocky, word-consonant string, we average event, sn average event, photo, physical pain, enumeration constant, enumeration linear, attend, word-pseudoword, word-random string, pseudoword-random string
dorsolateral prefrontal	story-math, reward, 2back-0back, we average event, sn average event, photo, physical pain, enumeration constant

Table 5: **Contrasts used in each system of the right hemisphere for performance assessment.**These contrasts have been selected so that the correlation of system-level activity maps is greater than .5 in average across runs.

early visual	reading-listening, reading-checkerboard, horizontal-vertical, rotation side, object orientation, triangle random, mechanistic video, mental-random, random, reward, shape, face-shape, match, face-avg, place-avg, consonant string, we average event, house constant, face constant, food constant, photo, physical pain, vstm constant, enumeration constant, encode other, recognition other hit, correct rejection, attend, random string, word-pseudoword, word-random string, pseudoword-random string, letter occurrence response
dorsal stream	reading-listening, saccades, rotation side, object orientation, story-math, mental-random, random, reward, face-avg, place-avg, pseudo-consonant string, consonant string, we average event, sn average event, house constant, food constant, photo, physical pain, vstm constant, enumeration constant, random string, word-pseudoword
ventral stream	reading-listening, rotation side, object orientation, random, reward, face-shape, face-avg, place-avg, house constant, face constant, food constant, photo, physical pain, vstm constant, enumeration constant, letter occurrence response
MT+ Complex	reading-listening, reading-checkerboard, saccades, rotation side, object orientation, triangle random, mechanistic video, mental-random, random, reward, face-shape, body-avg, face-avg, place-avg, word-consonant string, pseudo-consonant string, consonant string, we average event, sn average event, house constant, face constant, food constant, photo, physical pain, movie pain, vstm constant, enumeration constant, animal-scrambled, random string, word-pseudoword
somatosensory and motor	left-right button press, motor-cognitive, left hand-avg, left foot-avg, tongue-avg, we average event, sn average event, enumeration constant

paracentral lobular and mid cingulate	object orientation, random, left foot-avg, we average event, sn average event, photo, physical pain, enumeration constant
premotor	[motor-cognitive, saccades, object orientation, story-math, random, reward, tongue-avg, 2back-0back, we average event, sn average event, house constant, face constant, food constant, photo, physical pain, vstm constant, enumeration constant
posterior opercular	motor-cognitive, tongue-avg, we average event, sn average event, enumeration constant, tear-silence, reverse-silence, alphabet-silence
early auditory	reading-listening, mechanistic audio, non speech sound, left foot-avg, we average event, sn average event, enumeration constant, tear-silence, suomi-silence, yawn-silence, human-silence, silence, reverse-silence, alphabet-silence, cough-silence, environment-silence, laugh-silence, animals-silence
association auditory	reading-listening, mechanistic audio, non speech sound, story-math, mental-random, we average reference, we average event, sn average reference, sn average event, physical pain, enumeration constant, tear-silence, suomi-silence, human-silence, reverse-silence, alphabet-silence, environment-silence, laugh-silence
insular and frontal opercular	story-math, 2back-0back, we average event, sn average event, photo, physical pain, enumeration constant, 6 letters different-same
medial temporal	story-math, face-shape, body-avg, face-avg, place-avg, house constant, photo, enumeration constant
lateral temporal	story-math, random, reward, face-shape, 2back-0back, we average event, sn average event, enumeration constant
temporal-parietal-occipital junction	reading-listening, motor-cognitive, saccades, object orientation, mechanistic audio, story-math, mental-random, random, we average reference, we average event, sn average reference, sn average event, physical pain, enumeration constant, alphabet-silence
superior parietal and IPS	reading-listening, saccades, hand-side, rotation side, object orientation, triangle random, story-math, random, reward, 2back-0back, word-consonant string, pseudo-consonant string, consonant string, we average event, sn average event, house constant, face constant, food constant, photo, physical pain, movie pain, enumeration constant, word-pseudoword, low-high salience
inferior parietal	motor-cognitive, object orientation, story-math, random, reward, match, 2back-0back, we average event, sn average event, house constant, face constant, food constant, photo, physical pain, emotional-physical pain, movie pain, movie mental-pain, enumeration constant, encode other
posterior cingulate	story-math, random, we average event, sn average event, photo, physical pain, movie mental-pain, enumeration constant
anterior cingulate and medial prefrontal	story-math, 2back-0back, we average event, sn average event, photo, physical pain, enumeration constant, encode other
orbital and polar frontal	story-math, reward, 2back-0back, we average event, sn average event, photo, physical pain, enumeration constant

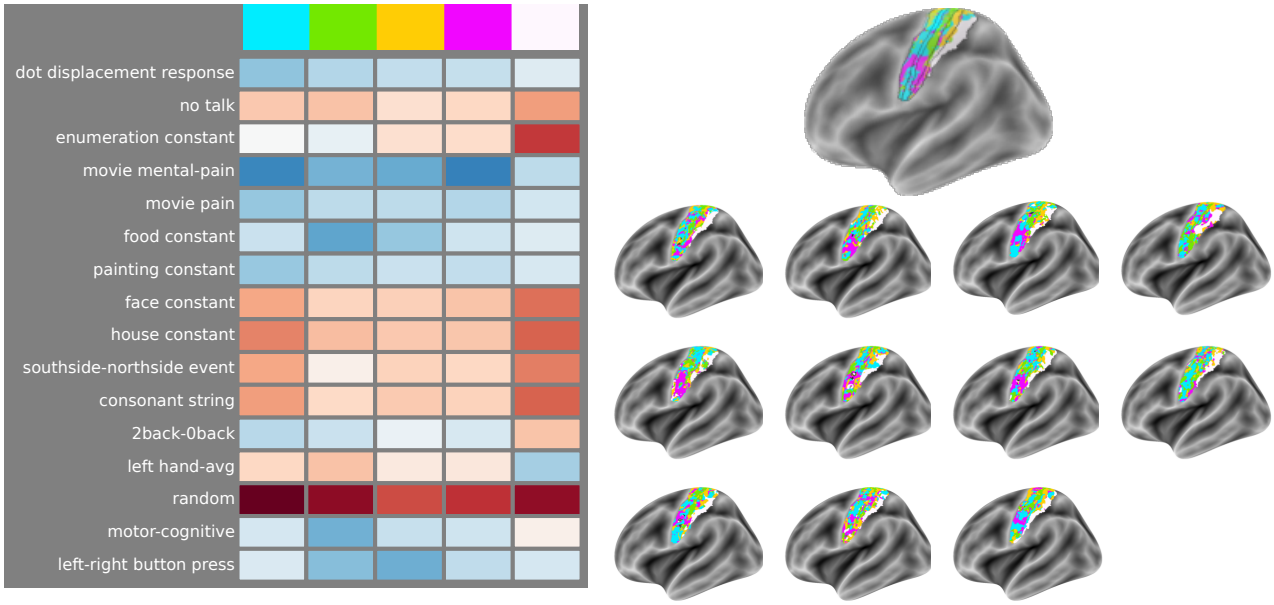

Figure 9: Example of an individualized parcellation scheme in the somato-sensory cortex. (left) For 5-components parcellation of the somato-sensory cortex, one can characterize their response to 160 contrasts; 16 are displayed for readability, and are related to various motor or non-motor tasks. The resulting group parcellation (top right) is superimposed with the original MMP1.0 parcellation. Individual parcellations are displayed on the bottom right.

inferior frontal	story-math, random, reward, 2back-0back, we average event, sn average event, photo, physical pain, enumeration constant, enumeration linear
dorsolateral prefrontal	story-math, reward, 2back-0back, we average event, sn average event, photo, enumeration constant, encode other

D Additional regional results

We provide in Fig.9 a view of the parcellation in the motor system of the MMP1.0 atlas, where data-driven parcellations achieve a poor fit. Indeed the parcellation does not seem to match somatotopic structures; this could be due to the fact that movie watching does not provide enough relevant information in the motor cortex.