Fine-grain atlases of functional modes for fMRI analysis
Kamalaker Dadi, Gaël Varoquaux, Antonia Machlouzarides-Shalit, Krzysztof Jacek Gorgolewski, Demian Wassermann, Bertrand Thirion, Arthur Mensch

To cite this version:
Kamalaker Dadi, Gaël Varoquaux, Antonia Machlouzarides-Shalit, Krzysztof Jacek Gorgolewski, Demian Wassermann, et al.. Fine-grain atlases of functional modes for fMRI analysis. 2020. hal-02496156

HAL Id: hal-02496156
https://hal.science/hal-02496156
Preprint submitted on 3 Mar 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Fine-grain atlases of functional modes for fMRI analysis

Kamalaker Dadi\textsuperscript{a}, Gaël Varoquaux\textsuperscript{a}, Antonia Machlouzarides-Shalit\textsuperscript{a}, Krzysztof J. Gorgolewski\textsuperscript{c}, Demian Wassermann\textsuperscript{a}, Bertrand Thirion\textsuperscript{a}, Arthur Mensch\textsuperscript{a,b}

\textsuperscript{a}Inria, CEA, Université Paris-Saclay, Palaiseau, 91120, France
\textsuperscript{b}ENS, DMA, 45 rue d’Ulm, 75005 Paris
\textsuperscript{c}Department of Psychology, Stanford University, California, USA

Abstract

Population imaging markedly increased the size of functional-imaging datasets, shedding new light on the neural basis of inter-individual differences. Analyzing these large data entails new scalability challenges, computational and statistical. For this reason, brain images are typically summarized in a few signals, for instance reducing voxel-level measures with brain atlases or functional modes. A good choice of the corresponding brain networks is important, as most data analyses start from these reduced signals. We contribute finely-resolved atlases of functional modes, comprising from 64 to 1024 networks. These dictionaries of functional modes (DiFuMo) are trained on millions of fMRI functional brain volumes of total size 2.4TB, spanned over 27 studies and many research groups. We demonstrate the benefits of extracting reduced signals on our fine-grain atlases for many classic functional data analysis pipelines: stimuli decoding from 12,334 brain responses, standard GLM analysis of fMRI across sessions and individuals, extraction of resting-state functional-connectomes biomarkers for 2,500 individuals, data compression and meta-analysis over more than 15,000 statistical maps. In each of these analysis scenarios, we compare the performance of our functional atlases with that of other popular references, and to a simple voxel-level analysis. Results highlight the importance of using high-dimensional “soft” functional atlases, to represent and analyse brain activity while capturing its functional gradients. Analyses on high-dimensional modes achieve similar statistical performance as at the voxel level, but with much reduced computational cost and higher interpretability. In addition to making them available, we provide meaningful names for these modes, based on their anatomical location. It will facilitate reporting of results.

Keywords: Brain imaging atlases; Functional networks; Functional parcellations; Multi-resolution;

1. Introduction

Population imaging has been bringing in terabytes of high-resolution functional brain images, uncovering the neural basis of individual differences (Elliott et al., 2008). While these great volumes of data enable fitting richer statistical models, they also entail massive data storage (Poldrack et al., 2013; Gorgolewski et al., 2017) and challenging high-dimensional data analysis. A popular approach to facilitate data handling is to work with image-derived phenotypes (IDPs), i.e. low-dimensional signals that summarize the information in the images while keeping meaningful representations of the brain (Miller et al., 2016).

While brain atlases originated in characterizing the brain’s microstructure (Brodmann, 1909), today they are widely used to study functional connectomes (Sporns et al., 2005; Varoquaux and Craddock, 2013) and for data reduction in functional imaging (Thirion et al., 2006; Craddock et al., 2012). For these applications, the choice of brain regions conditions the signal captured in the data analysis. To define regions well suited to brain-imaging endeavors, there is great progress in building atlases from the neuroimaging data itself (Eickhoff et al., 2018). Yet, most functional atlases describe the brain as parcellations, locally-uniform functional units, and thus do not represent well functional gradients (Huntenburg et al., 2018).

For functional imaging, brain structures delineated by an atlas should capture the main features of the functional signal, e.g. the functional networks (Smith et al., 2011). In a nutshell, there are two approaches to define well-suited structures. These can strive to select homogeneous neural populations, typically via clustering approaches (Goutte et al., 1999; Bellc et al., 2010; Craddock et al., 2012; Thirion et al., 2014; Schaefer et al., 2017). They can also be defined via continuous modes that map intrinsic brain functional networks (Damoiseaux et al., 2006; Varoquaux et al., 2011; Harrison et al., 2015). These functional modes have been shown to capture well functional connectivity, with techniques such as Independent Component Analysis (Kiviniemi et al., 2009; Pervaiz et al., 2019) or sparse dictionary learning (Mensch et al., 2016b; Dadi et al., 2019).

High-resolution atlases can give a fine-grained division of the brain and capture more functionally-specific regions and rich descriptions of brain activity (Schaefer et al., 2017). Yet, there is to date no highly-resolved set of “soft” functional modes available, presumably because increasing the dimensionality raises significant computa-
tional and statistical challenges (Mensch et al., 2016a; Pervaiz et al., 2019). In this paper, we address this need with high-order dictionaries of functional modes (DiFuMo) extracted at a large scale both in terms of data size (3 million volumes of total data size 2.4TB) and resolution (up to 1024 modes). For this, we leverage the wealth of openly-available functional images (Poldrack et al., 2013) and efficient dictionary-learning algorithms to fit on large data. This is unlike ICA which is hard to use for a high number of modes (Pervaiz et al., 2019).

**Contributions.** We provide Dictionaries of Functional Modesootnote{https://parietal-inria.github.io/DiFuMo} “DiFuMo” that can serve as atlases to extract functional signals, e.g. provide IDPs, with different dimensionalities (64, 128, 256, 512, and 1024). These modes are optimized to represent BOLD data well, over a wide range of experimental conditions. They are more finely-resolved than existing brain decompositions with continuous networks. By providing validated fine functional atlases, our goal is to streamline fMRI analysis with reduced representations, to facilitate large-cohort and inter-studies work. Through thorough benchmarking over classic data analysis tasks, we show that these modes gives IDPs that ground better analysis of functional images. Finally, we provide a meaningful label to each mode, summarizing its anatomical location, to facilitate reporting of results.

### 2. Methods: data-driven fine-grain functional modes

We describe in this section the models and methods underlying our definition of brain structures to extract IDPs.

#### 2.1. Context: Image Derived Phenotypes

While analysis of brain images has been pioneered at the voxel level (Friston et al., 1995), image-derived phenotypes (IDP) are increasingly used in the context of population imaging. Trading voxel-level signals for IDPs has several motivations. First and foremost, it greatly facilitates the analysis on large cohorts: the data are smaller, easier to share, requiring less disk storage, computer memory, and computing power to analyze. It can also come with statistical benefits. For instance, in standard analysis of task responses, e.g. in mass-univariate brain mapping, the statistical power of hypothesis test at the voxel level is limited by multiple comparisons (Friston et al., 1995), while working at the level of IDPs mitigates this problem (Thirion et al., 2006). For predictive modeling, e.g. in multi-variate decoding (Mourão-Miranda et al., 2005), the high-dimensionality of the signals is a challenge to learning models that generalize well—a phenomenon known as the curse of dimensionality in machine learning (Hastie et al., 2009). Finally, for functional connectomes, working at voxel-level is computationally and statistically intractable as it entails modeling billions of connections. The standard approach is therefore to average signals on regions or networks (Varoquaux and Craddock, 2013).

Functional neuroimaging is currently largely dependent on neuroanatomy for mapping function to structure (Destrieux et al., 2010; Devlin and Poldrack, 2007). Some anatomical structures support well a direct mapping to specific functions (Brett et al., 2002; Rademacher et al., 1993), e.g. the primary visual areas. Yet other functional units are not simply defined from anatomical features, for instance in high-level regions such as the default mode, which is defined from functional data (Leech et al., 2011; Greicius et al., 2003).

#### 2.2. Methods for data-driven functional atlases

Compared to anatomical atlases, defining regions from the functional signal can lead to a better explanation of behavioral outcomes (Dadi et al., 2019), as they capture the functional structure of the brain. Clustering of fMRI timeseries has been heavily used to define brain parcellations (Goutte et al., 1999), or for data reduction in predictive models (Michel et al., 2012). Reference functional brain parcellations have been defined with various clustering algorithms on resting-state fMRI (Bellec et al., 2010; Yeo et al., 2011; Craddock et al., 2012). Another class of approaches seeks modes of brain activity, decomposing the signal as a product of spatial maps and corresponding time-series (Figure 1). The most popular model in neuroimaging is independent component analysis (ICA, Hyvärinen and Oja, 2000), which optimizes spatial independence between extracted maps. It has been extensively used to define resting-state networks (Kiviniemi et al., 2003; Beckmann et al., 2005; Calhoun et al., 2001) and implicitly outlines soft parcellations of the brain at high order (Kiviniemi et al., 2009; Varoquaux et al., 2010b). ICA-defined networks are used to extract the official IDPs of UK BioBank, the largest brain-imaging cohort to date; these have been shown to relate to behavior (Miller et al., 2016).

We rely on another decomposition model, dictionary learning (Olshausen and Field, 1997), that enforces sparsity and non-negativity instead of independence on the

![Figure 1: Linear decomposition model of fMRI time-series for estimating brain networks](image)

The fMRI time series $X$ are factorized into a product of two matrices, $D$ which contain spatial modes and $A$ temporal loadings of each mode. $p$ - number of features, $n$ - number of volumes in fMRI image, $k$ - number of dictionaries.
spatial maps. While less popular than ICA in neuroimaging, sparsity brings the benefit of segmenting well functional regions on a zeroed-out background (Lee et al., 2010; Varoquaux et al., 2011). For our purposes, an important aspect of sparse models is that they have computationally-scalable formulations even with high model order and on large datasets (Mensch et al., 2016a, 2018). Functional modes defined from sparse dictionary learning have been used to predict Autism Spectrum Disorder (Abraham et al., 2017), or mental processes (Mensch et al., 2017).

Rest and task fMRI. Most functional brain atlases have been extracted from rest fMRI (Bellec et al., 2010; Power et al., 2011; Craddock et al., 2012; Ye et al., 2011; Miller et al., 2016; Schaefer et al., 2017). Brain networks can also be extracted from task fMRI data (Calhoun et al., 2008; Lee et al., 2010), and segment a similar intrinsic large-scale structure (Smith et al., 2009). In our work, we build functional modes from datasets with different experimental conditions, including task and rest. Our goal is to be as general as possible and capture information from different protocols. Indeed, defining networks on task fMRI can help representing these brain images and predicting the corresponding psychological conditions (Duff et al., 2012).

2.3. DiFuMo extraction: model and data

We consider BOLD time-series from fMRI volumes, resampled and registered to the MNI template. After temporal concatenation, those form a large matrix \( X \in \mathbb{R}^{p \times n} \), where \( p \) is the number of voxels of the images (around \( 2 \cdot 10^5 \)), and \( n \) is the number of brain images, of the order of \( 10^6 \) in the following. To extract DiFuMos, each brain volume is modeled as the linear combination of \( k \) spatial functional networks, assembled in a dictionary matrix \( D \in \mathbb{R}^{p \times k} \). We thus assume that \( X \) approximately factorizes as \( DA \), where the matrix \( A \in \mathbb{R}^{k \times n} \) holds in every column the loadings \( \alpha \) necessary to reconstruct the brain image \( x \) from the networks \( D \). The dictionary \( D \) is to be learned from data. For this, we rely on Stochastic Online Matrix Factorization\(^2\) (Mensch et al., 2018, SOMF), that is computationally tractable for matrices large in both directions, as with high-resolution large-scale fMRI data. SOMF solves the constrained \( \ell_2 \) reconstruction problem

\[
\min_{D \in \mathbb{R}^{p \times k}, A \in \mathbb{R}^{k \times n}} \| X - DA \|_F^2 + \lambda \| A \|_F^2,
\]

where \( \lambda \) is a regularization parameter that controls the sparsity of the dictionary \( D \), via the \( \ell_1 \) and positivity constraints. Encouraging sparsity in spatial maps is key to obtaining well-localized maps that outline few brain regions. The parameter \( \lambda \) is chosen so that the union of all maps approximately covers the whole brain.

Multi-dimensional DiFuMo atlases. We estimate dictionaries of dimensionality \( k \in \{64, 128, 256, 512, 1024\} \). This is useful as the optimal dimensionality for extracting IDPs often depends on the downstream data analysis task. The obtained functional modes segment well-localized regions, as illustrated in Figure 2.

2.4. Extracting signal on functional modes

The functional modes take continuous values (we refer to them as soft) and can have some overlap—though in practice this overlap is small. As a consequence, signal extraction calls for more than averaging on regions. The natural formulation is that the extracted signals (the IDPs) should best approximate the brain image \( x \in \mathbb{R}^p \) as a linear combination \( \alpha \in \mathbb{R}^k \) of the set of modes in the dictionary \( D \in \mathbb{R}^{p \times k} \). This is solved by linear regression:

\[
\alpha = \arg\min_{\alpha \in \mathbb{R}^k} \| x - D\alpha \|_2^2, \quad \text{i.e.} \quad \alpha = D^\dagger x, \quad (1)
\]

where \( D^\dagger = (D^T D)^{-1} D^T \in \mathbb{R}^{k \times p} \) is the pseudo-inverse of \( D \). For atlases composed of non-overlapping regions, such as classic brain parcellations—e.g. BASC (Bellec et al., 2010) or normalized cuts (Craddock et al., 2012)—linear regression simply amounts to averaging the images values in every cluster of \( D \). For overlapping modes as the ones of DiFuMo or the ICA maps used in UKBB (Miller et al., 2016), the linear regression formulation caters for the overlap and softness of the regions.

2.5. Region names: relation to anatomical structures

Relating IDPs to known brain structures facilitates interpretation and discussion of results. Though the DiFuMo atlases are defined from functional signal, we choose to reference their regions by their anatomical location, as it is a common and well-accepted terminology in neuroscience. For each resolution, we match the modes with regions in references of brain structure: the Harvard-Oxford atlas (Desikan et al., 2006), Destrieux atlas (Destrieux et al., 2010), the MIST atlas (Uurchs et al., 2019), Johns Hopkins University (JHU) atlas (Hua et al., 2008), and the Dierdrichsen cerebellum atlas (Dierdrichsen et al., 2009). We name each mode from the anatomical structure that

---

\(^2\)Available at: https://arthurmensch.github.io/modl/

---

\*Input fMRI data. We build the input data matrix \( X \) with BOLD time-series from 25 different task-based fMRI studies and 2 resting state studies, adding up to 2 192 functional MRI recording sessions. We gather data from OpenNeuro (Gorgolewski et al., 2017)—Table A4 lists the corresponding studies while Table A5 gives their data-acquisition parameters.

We use fMRIPrep (Esteban et al., 2019) for minimal preprocessing: brain extraction giving as a reference to correct for head-motion (Jenkinson et al., 2002), and co-registration to anatomy (Greve and Fischl, 2009). All the fMRI images are transformed to MNI template space. We then use MRIQC (Esteban et al., 2017) for quality control.
it most overlaps with. When the overlap was weak, a trained neuroanatomist (AMS) looked up the structure in standard classic anatomy references (Henri, 1999; Schmahmann et al., 1999; Rademacher et al., 1992; Ono et al., 1990; Catani and de Schotten, 2012). AppendixF gives more details on the naming of the brain areas.

3. Brain-image analysis on functional modes

We use the reduced representations (IDPs) introduced above for various functional-imaging analytic tasks: standard mass-univariate analysis of brain responses ($\S$3.2); decoding of mental processes from brain activity ($\S$3.3); prediction of phenotypes from functional connectomes ($\S$3.4); finally, we measure the quality of signal reconstruction after the dimension reduction, with an illustration on meta-analyses ($\S$3.5).

3.1. Benchmarking several functional atlases

To gauge the usefulness of the extracted IDPs, we compare each analysis pipeline across several functional atlases: DiFuMo and reference atlases are used to compute functional IDPs. We use the same signal-extraction function (1), but vary the spatial components $D$. As a baseline, we also perform the voxel-level analyses, though it entail significantly larger computational costs.

We consider other functional atlases that are multi-resolutions, accessible to download, and volumetric (Table 1): ICA maps with $k \in \{21, 55\}$ components, extracted on large-scale rs-fMRI from UKBB (Miller et al., 2016); bootstrap analysis of stable clusters (BASC) built with hierarchical clustering on rs-fMRI, with various number of clusters (Belloc et al., 2010); spatially-constrained clustering on rs-fMRI, with $k \in \{200, 400\}$ clusters (Craddock et al., 2012); $k = 333$ cortical areas derived from rs-fMRI using a local gradient approach (Gordon et al., 2014); $k \in \{90, 499\}$ functional regions covering cortical and subcortical gray matter with ICA and Ward clustering (Shirer et al. (2012), Altmann et al. (2015)); and brain parcellations derived with gradient-weighted Markov Random Field, with resolutions similar to ours (Schaefer et al., 2017, $k$ up to 1000).
### 3.2. Mapping brain response: standard task-fMRI analysis

Standard analysis in task fMRI relates psychological manipulations to brain activity separately for each voxel or region. It models the BOLD signal as a linear combination of experimental conditions—the General Linear Model (GLM, Friston et al., 1995). The BOLD signal forms a matrix $Y \in \mathbb{R}^{n \times p}$, where $p$ is the number of voxels. With data reduction, we use as input the reduced signal $Y_{\text{red}} = (Y_{\text{voxel}}(D))^\top \in \mathbb{R}^{q \times k}$ (Equation 1). The GLM models $Y$ or $Y_{\text{red}}$ as $Y = X\beta + \epsilon$ where $X \in \mathbb{R}^{n \times q}$ is the design matrix formed by $q$ temporal regressors of interest or nuisance and $\epsilon$ is noise (Friston et al., 1998). In our experiments, we use the Nistats library\(^3\).

With reduced input $Y_{\text{red}}$, we obtain one signal per region, as $\beta \in \mathbb{R}^{q \times k}$. The full $\beta$-maps can then be reconstructed by setting $\beta_{\text{rec}} = \beta D^\top \in \mathbb{R}^{q \times p}$. We transform the reconstructed $\beta$-maps into $z$-maps $z \in \mathbb{R}^{q \times p}$ using base contrasts, before thresholding them with Benjamini and Hochberg (1995) FDR correction for multiple comparisons. We then compare the $z$-maps obtained using voxels as input, and $z$-maps using reduced input and reconstructed $\beta$-maps, using the Dice (1945) similarity coefficient. We also perform an intra-subject analysis detailed in AppendixD.

**Data.** We consider the Rapid-Serial-Visual-Presentation (RSVP) language task of Individual Brain Charting (IBC) (see Pinho et al., 2018, for experimental protocol and pre-processing). We model six experimental conditions: complex meaningful sentences, simple meaningful sentences, jabberwocky, list of words, lists of pseudowords, consonant strings. $\beta$-maps are estimated for each subject using a fixed-effect model over 3 out of the 6 subject’s sessions. We randomly select 3 sessions 10 times to estimate the variance of the Dice index across sessions. As a baseline, we evaluate the mean and variance of the Dice index across $z$-maps when varying the sessions used in voxel-level GLM.

#### 3.3. Decoding experimental stimuli from brain responses

Decoding predicts psychological conditions from task-related $z$-maps (Haynes and Rees, 2006). The validity of a decoding model is evaluated on left-out data (following Varoquaux et al., 2017), e.g. left-out subjects for inter-subject decoding (Poldrack et al., 2009). We use linear decoding models: ridge regression for continuous target and Support Vector Machine (SVC, Hastie et al., 2009) for classification. For each study, we separate sessions (for intra-subject decoding) or subjects (for inter-subject decoding) into randomly-chosen train and test folds (20 folds with 30% test size), and measure the test accuracy. We compare the performance of predictive models using the voxel-level $z$-maps or using the data reduced with functional atlases.

**Data.** We use 6 open-access task fMRI studies. We perform inter-subject decoding in the emotional and sensitivity to pain experiences from Chang et al. (2015), and in three studies from HCP900 (Van Essen et al., 2012): working memory, gambling (Delgado et al., 2000), and relational processing (Smith et al., 2007). We perform intra-subject decoding using the several sessions of left and right button press responses in IBC (ARCHI protocol, Pined et al., 2007). The unthresholded $z$-maps used in the decoding pipeline are either obtained from Neurovault (Gorgolewski et al., 2015), or computed with the GLM following §3.2. Details are reported in AppendixB.1.

#### 3.4. Predicting phenotype from functional connectomes

Resting-state fMRI can be used to predict phenotypic traits (Richiardi et al., 2010). For this, each subject is represented by a functional connectivity matrix that captures the correlation between brain signals at various locations. Our functional-connectome prediction pipeline comprises three steps: 1) we extract a reduced representation of the BOLD signal, projecting voxel-level data onto a functional atlas as in §3.2; 2) we compute a functional connectome from the reduced BOLD signals; 3) we use it as input to a linear model. We compute a connectome from activations with the Ledoit and Wolf (2004) covariance estimator as Varoquaux and Craddock (2013); Brier et al. (2015). We then derive single-subject features from covariance matrices using their tangent space parametrization (Varoquaux et al., 2010a; Barachant et al., 2013; Pervaiz et al., 2019). Those are used to fit an $\ell_2$-penalized logistic regression for classification and a ridge regression for continuous targets.

---

\(^3\)https://nistats.github.io/

---

### Table 1: Functional atlases that we benchmark; they define IDPs for brain-images. analysis

<table>
<thead>
<tr>
<th>Name</th>
<th>Dimensionality</th>
<th># subj.</th>
<th>Soft</th>
<th>Extraction method</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>BASC</td>
<td>64, 122, 197, 325, 444</td>
<td>43</td>
<td>No</td>
<td>Hierarchical clustering</td>
<td>Belloc et al. (2010)</td>
</tr>
<tr>
<td>Craddock</td>
<td>200, 400</td>
<td>41</td>
<td>No</td>
<td>Spectral clustering</td>
<td>Craddock et al. (2012)</td>
</tr>
<tr>
<td>FIND(^a)</td>
<td>90, 499</td>
<td>15</td>
<td>Yes</td>
<td>ICA; Ward clustering</td>
<td>Shier et al. (2012); Altmann et al. (2015)</td>
</tr>
<tr>
<td>Gordon</td>
<td>333</td>
<td>120</td>
<td>No</td>
<td>Local-gradient approach</td>
<td>Gordon et al. (2014)</td>
</tr>
<tr>
<td>UKBB ICA</td>
<td>21, 55</td>
<td>4100</td>
<td>Yes</td>
<td>Sparse dictionary learning</td>
<td>Miller et al. (2016)</td>
</tr>
<tr>
<td>Schaefer</td>
<td>100, 200, 300, 400, 500, 1489</td>
<td>600, 800, 1000</td>
<td>No</td>
<td>Random-gradient weighted Markov</td>
<td>Schaefer et al. (2017)</td>
</tr>
<tr>
<td>DiFuMo</td>
<td>64, 128, 256, 512, 1024</td>
<td>2192</td>
<td>Yes</td>
<td>SMSparse dictionary learning</td>
<td>This paper</td>
</tr>
</tbody>
</table>

\(^a\)https://findlab.stanford.edu/functional_ROIs.html \(^b\)https://www.fmrib.ox.ac.uk/ukbiobank/ \(^c\)https://parietal-inria.github.io/DiFuMo
We assess predictive performance with 20 folds, random splits of subjects in train and test sets, with 25% test size.

**Data.** We use 7 openly-accessible datasets with diverse phenotypic targets, as summarized in Table A3. We predict diagnostic status for Alzheimer’s disease on ADNI (Mueller et al., 2005), PTSD on ADNIDOD; Autism Spektrum Disorder on ABIDE (Di Martino et al., 2014) and schizophrenia on COBRE (Calhoun et al., 2012); drug consumption on ACPI; IQ measures on HCP (Van Essen et al., 2013); and age (with a regression model) in normal aging with CamCan (Taylor et al., 2017).

### 3.5. Quality of image reconstruction

The signals extracted on a brain atlas can be seen as a compression, or simplification, of the original signal. Indeed, a full image can be reconstructed from these signals. We quantify the signal loss incurred by this reduction. For this, we project a brain map \( x \) onto an atlas (solving Eq. (1)), and compute the best reconstruction of \( x \) from the loadings \( \alpha \), namely \( \hat{x} = D \alpha \in \mathbb{R}^p \). We compare original and reconstructed images through the \( R^2 \) coefficient,

\[
R^2(x, \hat{x}) = 1 - \frac{\|x - \hat{x}\|_2^2}{\|x - \bar{x}\|_2^2}, \tag{2}
\]

where \( \bar{x} \in \mathbb{R} \) is the spatial mean of map \( x \). The \( R^2 \) coefficient is averaged across all images. Higher \( R^2 \) coefficients means that the reduced signals (IDPs) explain more variance of the original images, where \( R^2 = 1 \) corresponds to no signal loss. The larger the number of signals used, the easier it is to explain variance; it is therefore interesting to compare this measure across atlases with similar number of components.

**Data.** We use NeuroVault (Gorgolewski et al., 2015), the largest public database of statistical maps. To avoid circularity, we exclude maps derived from the studies used to extract the DiFuMo atlases, along with maps that fail semi-automated quality inspection (filtering out thresholded or non-statistical maps), resulting in 15,542 maps.

**Meta-analysis of contrasts maps.** Ideally, the extracted IDPs should allow to compute meta-analytical summaries of brain activity maps. In this setting, a single map, corresponding to a certain cognitive concept, is computed from many z-maps across different studies, associated to conditions that involve this cognitive concept. We compare the summaries obtained at voxel-level, i.e. averaging the maps \( \{x\} \), with the ones obtained using reconstructed images, i.e. averaging the maps \( \{\hat{x}\} \) used in Eq. (2). We use maps from our curated subset of NeuroVault annotated with terms motor, language and face recognition.

### 4. Results: comparing atlases for analyses

We report benchmarking results on the analytic tasks listed in the previous section.

![Figure 3: Overlap between GLM maps obtained with functional atlases and voxel-level analysis.](image)

**Figure 3:** Overlap between GLM maps obtained with functional atlases and voxel-level analysis. **Top:** The overlap is measured with the Dice similarity coefficient. The black line gives a baseline the mean overlap between voxel-level contrast maps over several random selections of sessions per subject. The figure gives Dice similarity scores between the GLM maps computed with signals extracted on functional atlases and at the voxel-level, after reconstruction of full z-maps and voxel-level thresholding with FDR control. The best similarity is achieved for highest dimensionality, though 1024-dimensional DiFuMo atlas largely dominates 1000-dimensional Schaefer parcellation. Each point is the mean and the error bar denotes the standard deviation over contrast maps. **Bottom:** The activity maps encoded on 1024-dimensional space capture the same information and much smoother to voxel-level.

### 4.1. Brain mapping: standard task-fMRI analysis

Figure 3 reports the results of standard analysis of task fMRI (GLM), comparing analysis at the voxel-level with analyses on signals extracted from functional atlases. Best correspondence is obtained at highest dimensionality, as the regions are finer. Notably, analysis with DiFuMo of dimensionality 1024 is markedly closer to voxel-level analysis than using the largest alternative, the 1000-dimensional Schaefer parcellation. In addition, the Dice index relative to the voxel-level gold standard is comparable to the Dice index between runs of voxel-level GLM estimated across folds. We note that using soft functional modes from only 55 ICA components shows excellent results, comparable to those obtained using the 1000 components Schaefer atlas. This stresses the benefit of continuous functional modes for the analysis of task responses. Overall, standard task-fMRI analysis on signals derived from 512 or 1024-dimensional DiFuMo gives results close to the voxel-level.
4.2 Decoding mental state from brain responses

Figure 4 shows the impact on decoding performance of reducing signals with various functional atlases. It reports the performance relative to the median across methods for each of the 6 tasks. These results clearly show the importance of high-dimensional functional modes for decoding. Indeed, the higher the atlas resolution, the better the predictions. Using DiFuMo \( k = 1024 \) or Schaefer \( k = 1000 \) gives the best performance. In addition, as these functional atlases segment sufficiently-fine regions, prediction from the corresponding signals tends to outperform voxel-level prediction. Indeed, applying multivariate models to a larger number of signals with a limited amount of data is more prone to overfitting—data reduction acts here as a welcome regularization. Qualitatively, brain maps containing decoding weights can be reconstructed. With high-dimensional atlases, they are interpretable and capture information similar to voxel-level analysis (Figure 5).

4.3 Predicting traits from functional connectomes

Figure 6 shows the impact of the choice of functional atlas when predicting phenotypes from functional connectomes. It reports the relative prediction accuracy for 7 different prediction problems (each composed of a dataset and a target phenotype); the lines give the median across the datasets. Here, we do not report a voxel-level baseline, as it requires to compute covariance matrices of dimensions around 300,000 × 100,000 and is therefore computationally and statistically intractable. Unlike with the previous results, high-resolution atlases do not provide the best performance, likely because the complexity of the statistical models increases with the square of the number of nodes. The best prediction overall is achieved using DiFuMo \( k = 256 \), followed by Craddock \( k = 400 \) and BASC \( k = 444 \). Figure A6 report absolute results for each prediction problem.
4.4. Fraction of the original signal captured

Figure 7 (left) displays the $R^2$ scores summarizing the loss of information when data are reduced on an atlas and reconstructed back to full images. Unsurprising, reducing the images with lower-order dimensions (atlases with fewer regions) yields a high loss of information across all methods. DiFuMo $k = 1024$ captures 70% of the original voxel-level signal. Qualitatively, the benefits of functional modes can be seen by comparing the meta-analytic maps related to motor tasks (Figure 7 right)—Figure A7 shows additional meta-analysis on other topics. The DiFuMo have clear visual benefits over brain discrete parcellations, such as BASC, as they better capture gradients.

5. Discussion

This paper introduces brain-wide soft functional modes, named DiFuMos and made of a few hundreds to a thousand of brain sub-divisions. They are derived from BOLD time-series across many studies to capture well functional images with a small number of signals. In the context of population imaging, these signals are known as image-derived phenotypes (IDP, Miller et al., 2016) and are crucial to easily scale statistical analysis, building a science of inter-individual differences by relating brain signals to behavioral traits (Dubois and Adolphs, 2016). Reducing the dimensionality of the signals not only come with a 1000× gain in storage, but also with 100× computational speed-up for the analysis (Table A1). Even small-scale studies may need functional nodes, e.g. for computing functional connectomes (Zalesky et al., 2010; Varoquaux and Craddock, 2013). There already exist many functional brain atlases; yet DiFuMos have the unique advantage of being both soft and highly resolved. These features are important to capture gradients of functional information.

Grounding better image-derived phenotypes. Signals extracted from a functional atlas should enable good statistical analysis of brain function. We considered quantitative measures for typical neuroimaging analytic scenarios and compared the fitness of extracting signal on DiFuMo with using existing functional brain atlases. The biggest gains in analysis come from increasing the dimensionality of brain sub-divisions, aside for functional connectome studies where an optimal is found around 200 nodes. Choosing the number of nodes then becomes a tradeoff between complexity of the representation and analytic performance. Importantly, the gains in analytic performance continue way beyond the dimensionality typically used for IDPs (e.g. 55 components from Miller et al., 2016). These results extend prior literature emphasizing the importance of high-dimensional parcellations for fMRI (Abou Elsayed et al., 2011; Thirion et al., 2014; Arslan et al., 2017; Sala-Llonch et al., 2019). To foster good analysis, the second most important aspect of a parcellation appears that it be soft, i.e. continuously-valued. For a given dimensionality, soft modes tend to outperform hard parcellations, whether they are derived with ICA or dictionary learning.

Modes well-adapted to the EPI signal. The functional modes are optimized to fit well a large number of EPI images: 2,192 sessions across 27 studies. As a result, they form a division of the brain well adapted to the signal. For instance, they define regions larger in the white matter and in the CSF than in the grey matter (Figure A1). A large dataset is needed to capture such implicit regularities of the signal with high-dimensional spatial decompositions. Indeed, running the same model on less data extracts modes with less spatial regularity (Figure A2). The combination of high dimensionality and large dataset leads to significant computational demands. The extraction of DiFuMos was possible thanks to fast algorithms for huge matrix factorization (Mensch et al., 2018), and gathering data representative of a wide variety of scanning protocols via openfMRI (Poldrack et al., 2013).

We did not limit the DiFuMo modes to gray matter, as measures outside gray matter can be useful in subsequent analysis, for instance to remove the global signal (Murphy and Fox, 2017). In addition, distributed modes extracted from full-brain EPI can separate out noise—such as movement artifacts—and help rejecting it in a later analysis (Perlbarg et al., 2007; Griffanti et al., 2014; Pruim et al., 2015). Some DiFuMo modes indeed segment ventricles or...
interfaces. Depending on the application, practitioners can choose to restrict signal extraction to a grey-matter mask.

The functional modes are sharp and anatomically relevant. To extract structures defined by brain anatomy or microstructure, atlasing efforts have used anatomical or multimodal imaging (Mori et al., 2005; Desikan et al., 2006; Eickhoff et al., 2007; Glasser et al., 2016). The DiFuMo atlases capture a different signal: brain activity. Yet, thanks to the sparsity and non-negativity constraint, they are made of localized modes which often have a natural anatomical interpretation. Consequently, we have labeled the modes with a unique name based on the most relevant anatomical structure, following Urchs et al. (2019) who also give anatomical labels to functional regions. Indeed, using a common vocabulary of brain structures is important for communication across the neuroimaging community. As visible on Figure 8, the modes are well anchored on anatomical structures such as the putamen. They are however not constrained to contain only one connected region. Smaller dimension DiFuMos indeed capture distributed networks, often comprising bilateral regions. As the dimensionality increases, the networks progressively separate in smaller networks which eventually form single regions. For instance, the left and right putamen appear in the same mode at dimension 64, and are first sub-divided along the antero-posterior direction, and later the left and right putamen are separated (Figure 8). Dimension choice is data driven: it should best explain the functional signal.

6. Conclusion

We provide multidimensional atlases of functional modes that can be used to extract functional signals: parietal-inria.github.io/DiFuMo. They give excellent performance for a wide variety of analytic tasks: GLM-based analysis, mental-process decoding or functional-connectivity analysis. Their availability reduces computational burdens: practitioners can readily perform analyses on a reduced signal, without a costly ROI-definition step. In addition, working on common functional modes across studies facilitates comparison and interpretations of results. To help communication, we have labeled every functional mode to reflect the neuroanatomical structures that it contains. To date, these are the only high-dimensional soft functional modes available. As they have been extracted from a variety of data (more than 2,000 sessions across 27 studies, 2.4TB in size) and improve many analytic tasks, the rich descriptions of neural activity that they capture is well suited for a broad set of fMRI studies.

7. Acknowledgments

This project has received funding from the European Union’s Horizon 2020 Research and Innovation Programme under Grant Agreement No. 785907 (HB P GSA2) and No 826421 (VirtualBrainCloud). The work of Arthur Mensch has been supported by the European Research Council (ERC Grant Noria). This work acknowledges the support of ANR NeuroRef and ERC-StG NeuroLang.

We also thank Pierre Bellec and Vincent Frouin for their helpful discussions on the experimental work, the neuroimaging community for giving access to fMRI datasets, and open-source contributors on the packages we build upon (including nilearn, fMRIPrep, and MRICQ).

Data collection and sharing for this project was funded by the Alzheimer’s Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbellVie, Alzheimer’s Association; Alzheimer’s Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpír, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research and Development, LLC.; Johnson and Johnson Pharmaceutical Research and Development LLC.; Lumosity; Lundbeck; Merck and Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer’s Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.


Appendix A. Performance of DiFuMos

As discussed in §3, we report how DiFuMoS components are well adapted to the MR EPI signal in Figure A1. Figure A2 qualitatively compare components obtained training on the whole data corpus and training on a fraction of it. Better component regularity is obtained with more data. Finally, Table A1 reports the computational speed-ups obtained using DiFuMoS IDPs instead of voxel in the decoding experiment. Similar speed-ups are observed in the other validation pipelines.

Appendix B. Details on stimulus decoding

We provide additional details for the decoding pipeline, to complete the description in §3.3.
Table A1: The comparison in computational times while predicting mental state on two set of brain features space: voxel-level \( \approx 200,000 \) and reduced voxels to DiFuMo 1024. We report the averaged time over 20 cross-validation folds for several task-fMRI conditions. Clearly, there are benefits trading for reduced representations in terms of computation time. On high-resolution brain images like HCP, these are decreased by a factor 200.

<table>
<thead>
<tr>
<th>Task</th>
<th># samples</th>
<th>Representation</th>
<th>Time (sec)</th>
<th>Speedup</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emotion</td>
<td>4924</td>
<td>Voxel-level</td>
<td>77.7</td>
<td>46×</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reduced</td>
<td>1.7</td>
<td></td>
</tr>
<tr>
<td>Pain</td>
<td>84</td>
<td>Voxel-level</td>
<td>1.5</td>
<td>250×</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reduced</td>
<td>0.006</td>
<td></td>
</tr>
<tr>
<td>Working mem.</td>
<td>3140</td>
<td>Voxel-level</td>
<td>874.7</td>
<td>240×</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reduced</td>
<td>3.7</td>
<td></td>
</tr>
<tr>
<td>Gambling</td>
<td>1574</td>
<td>Voxel-level</td>
<td>298.7</td>
<td>270×</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reduced</td>
<td>1.12</td>
<td></td>
</tr>
<tr>
<td>Relational</td>
<td>1572</td>
<td>Voxel-level</td>
<td>263.1</td>
<td>405×</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reduced</td>
<td>0.65</td>
<td></td>
</tr>
</tbody>
</table>

Table A2: Dataset, prediction tasks and dataset size for each of the 6 decoding tasks we consider in §3.3. z-maps from HCP and IBC were computed using the GLM, while NeuroVault directly provided the \( \beta \)-maps for Emotion and Pain. NV: NeuroVault.

Appendix B.1. Input data and pre-processing pipelines

The decoding pipeline classifies input unthresholded statistical maps. Table A2 summarizes the task-based studies used to obtain these statistical maps.

Pre-encoded maps downloaded from Neurovault.org. We download maps related to emotion and pain (Chang et al., 2015) using Neurovault, querying the collections 503 and 504. We use the “Rating” & “PainLevel” labels as predictive targets. We predict emotion using ridge regression, and pain-level over 3 classes using Linear SVC. The supervised learning pipeline, that includes cross-validation and linear models is implemented with Python based scikit-learn (Pedregosa et al., 2011). We use nilearn (Abraham et al., 2014) to download maps from Neurovault.org interface (Gorgolewski et al., 2015). The data acquisition parameters, preprocessing details and estimation of statistical maps are described in Chang et al. (2015).

Statistical maps encoded using the GLM. We compute z-maps from HCP900 (Van Essen et al., 2012) and IBC (Pinho et al., 2018) studies, that comprise high-quality task-fMRI experiments.

HCP. We download fMRI data from the HCP900 release; those are already preprocessed using HCP pipelines (Glasser et al., 2013). We use MNINonLinear-based registered data as input for the GLM, that outputs one z-map per condition per subject. We consider three task-based studies, namely: for Working Memory, we con-
sider z-maps based on condition: “0-back faces”, “2-back faces”, “0-back places”, “2-back places”. Similarly, for Gambling (Delgado et al., 2000), we consider z-maps for the conditions “loss” and “reward”; finally, on Relational processing, we consider z-maps for the conditions “relational processing” and “matching”. For each study, we use Linear SVC on encoded z-maps to predict psychological conditions. The predictive model therefore perform a 2-class or 4-class classification. The experimental protocol and data acquisition parameters are detailed in Van Essen et al. (2012).

IBC. We consider the Archi Standard (Pinel et al., 2007) motor task, where subjects are asked to press “left” and “right” button press based on audio and visual instructions. We perform within-subject classification between left and right button press, using z-maps corresponding to each repetition of the instruction. For each of the 13 available subjects, a linear model is trained on the z-maps from all but one session and prediction is performed on the left-out session. Each subject provides 80 encoded z-maps across 4 sessions. We use data preprocessed following the pipelines of Pinho et al. (2018).

GLM specification. For both datasets, the input z-maps are estimated from the raw fMRI data by fitting a GLM. We use Nistats⁴, a Python package for the statistical analysis of fMRI data. The temporal regressors of the model are specified according to the timing of stimulus presentations convolved with hemodynamic models (spm + derivative). We use polynomial model to capture the low-frequency drifts in the data.

Appendix B.2. Detailed results

To complete the summarizing Figure 4, we report the raw prediction scores separately for each decoding tasks in Figure A3. Prediction accuracy increases with the size of functional atlases. Using 1024 atlases allows to match or pass the performance of voxel-based prediction. In terms of interpretation, the weights are much smoother and blobs are clearly visible in the weight classification maps obtained using DiFuMo. This is illustrated on Figure A4 for face-versus-place decoding in the working-memory HCP study.

Appendix C. Details on biomarker prediction

We consider multiple datasets to account for the diversity of prediction targets in biomarker prediction problem. We report datasets, prediction groups and prediction targets in Table A3.

Appendix C.1. Input data and prediction settings

The connectivity features built from functional atlases predict various clinical outcomes (neuro-degenerative and neuro-psychiatric disorders, drug abuse impact) and psychological traits.

⁴https://nistats.github.io/

Appendix B.2. Detailed results

To complete the summarizing Figure 4, we report the raw prediction scores separately for each decoding tasks in Figure A3. Prediction accuracy increases with the size of functional atlases. Using 1024 atlases allows to match or pass the performance of voxel-based prediction. In terms of interpretation, the weights are much smoother and blobs are clearly visible in the weight classification maps obtained using DiFuMo. This is illustrated on Figure A4 for face-versus-place decoding in the working-memory HCP study.

Group classification. We use the Alzheimer’s Disease Neuroimaging Initiative⁵ (ADNI) and (ADNI-DOD) (Mueller et al., 2005) to predict neuro-degenerative diseases. We discriminate between Alzheimer’s Disease (AD) from Mild

⁵www.adni-info.org
Cognitive Impairment (MCI) group on ADNI. We discriminate between post-traumatic stress disorder (PTSD) and healthy individuals on ADNIDOD. We use data from the Center for Biomedical Research Excellence\textsuperscript{6} (COBRE Calhoun et al. (2012) to predict schizophrenia diagnosis of individuals. We classify autism and healthy individuals on Autism Brain Imaging Data Exchange database (ABIDE, Di Martino et al. (2014), Finally, we consider data from Addiction Connectome Preprocessed Initiative\textsuperscript{7} (ACPI), where we discriminate Marijuana consumers versus control subjects.

Psychological traits. We first stratify individuals from HCP900 release (Van Essen et al., 2013) into groups of high and low IQ, and perform binary classification on these. The details about the stratification into these groups are described in Dadi et al. (2019).

Age regression. We use Cambridge Center for Ageing and Neuroscience (CamCAN) dataset (Taylor et al., 2017) to study brain ageing. This study comprises wide range of age groups spanning from 24 – 86. We use ridge regression to predict age on this dataset.

AppendixC.2. Data acquisition parameters and pre-processing pipelines

The data acquisition details for ADNI, ADNIDOD, COBRE, ABIDE, ACPI and HCP are described in Dadi et al. (2019); those for CamCAN in Taylor et al. (2017). We pre-process individuals from CamCAN, ADNI, ADNIDOD and COBRE. All rs-fMRI acquisitions are pre-processed with standard steps, described in Dadi et al.

\textsuperscript{6}https://www.mrn.org/research/details/cobre
\textsuperscript{7}http://fcon_1000.projects.nitrc.org/indi/ACPI/html/

### Table A3: Resting-state fMRI datasets used in the pipeline described on §3.4 for predicting phenotypic labels from functional connectomes.

<table>
<thead>
<tr>
<th>Rest-fMRI</th>
<th>Prediction groups</th>
<th>Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>HCP900</td>
<td>High IQ vs Low IQ</td>
<td>443 subjects</td>
</tr>
<tr>
<td></td>
<td>213/230</td>
<td></td>
</tr>
<tr>
<td>ABIDE</td>
<td>Autism vs control</td>
<td>866 subjects</td>
</tr>
<tr>
<td></td>
<td>402/464</td>
<td></td>
</tr>
<tr>
<td>ACPI</td>
<td>Marijuana use vs control</td>
<td>126 subjects</td>
</tr>
<tr>
<td></td>
<td>62/64</td>
<td></td>
</tr>
<tr>
<td>ADNI</td>
<td>Alzheimers vs MCI</td>
<td>136 subjects</td>
</tr>
<tr>
<td></td>
<td>40/96</td>
<td></td>
</tr>
<tr>
<td>ADNIDOD</td>
<td>PTSD vs control</td>
<td>167 subjects</td>
</tr>
<tr>
<td></td>
<td>89/78</td>
<td></td>
</tr>
<tr>
<td>COBRE</td>
<td>Schizophrenia vs control</td>
<td>142 subjects</td>
</tr>
<tr>
<td></td>
<td>65/77</td>
<td></td>
</tr>
<tr>
<td>CamCAN</td>
<td>Age</td>
<td>626 subjects</td>
</tr>
<tr>
<td></td>
<td>24 – 86</td>
<td></td>
</tr>
</tbody>
</table>

Figure A5: Connectome prediction scores for each brain atlases and target: Each marker denotes the mean performance of using a certain brain atlas; error bars are the standard deviation of the prediction scores for this atlas. BASC and DiFuMo-based atlases give good prediction scores up to $k = 256$ ROIs.
(2019). The other considered datasets provide preprocessed data. We report the total number of subjects included in the analysis in Table A3, after excluding for severe scanning artifacts, head movements with amplitude larger than 2 mm and individuals who have more than one clinical diagnosis,

Confound removal and temporal signal pre-processing. The strategy we use for cleaning temporal signals is the same as in Dadi et al. (2019). We briefly outline these steps here. We regress out 10 CompCor (Behzadi et al., 2007) components on the whole brain and six motion related signals which are provided in the ADNI, ADNIDOD, COBRE, CamCAN datasets. We do not perform any additional preprocessing steps on ABIDE, ACPI and HCP. For all datasets, the signal of each region is normalized, detrended and bandpass-filtered between 0.01 and 0.1Hz. All these steps are done with nilearn (Abraham et al., 2014).

Appendix C.3. Detailed results

Figure 6 summarizes the impact of the brain atlases and ROIs in predicting diverse targets on rs-fMRI images. Figure A5 shows the absolute prediction scores for each target separately. DiFuMo-based predictions are on par with those using UKBB ICA components, Craddock et al. (2012) and BASC atlases.

Appendix D. Intra-subject encoding

In §3.2, we compare group-level z-maps computed at the voxel-level and on reduced data using the DICE similarity coefficient. We also performed an intra-subject, across sessions, standard analysis. We consider the Rapid-Serial-Visual-Presentation (RSVP) language task of Individual Brain Charting (IBC) (see Pinho et al. (2018) for details on experimental protocol and data pre-processing).

Encoding model. In this setting, we fit a GLM on the several acquisition sessions of each subject considered separately. That is, we compute a single β-map per session and condition, forming a set of maps \( \beta \in \mathbb{R}^{n \times p} \). \( \beta \) is either computed directly at the voxel-level or using functional atlases, in which case we set \( \beta = \beta_{sel}D^T \), with \( \beta \in \mathbb{R}^{n \times k} \).

We then use a leave-one-session-out cross-validation scheme to compare the observed, single-session, time series \( Y \in \mathbb{R}^{n \times p} \) to the reconstructed time-series \( \hat{Y} = X\hat{\beta} \), where \( \hat{\beta} \) are the average β-maps across the 5 training sessions. We obtain \( R^2 \)-maps, where each voxel holds the proportion of variance explained by the model

\[
r_i = 1 - \frac{\|y_i - \hat{y}_i\|^2}{\|y_i - \bar{y}_i\|^2},
\]

where \( y_i \) is the univariate time-series in \( \mathbb{R}^n \) associated to voxel \( i \) and \( \bar{y}_i \) is its temporal mean. We finally average

\[R^2\] scores across leave-one-session-out folds, and threshold non-positive values. The resulting \( R^2 \)-maps provides information on how much encoded β-maps are able to predict univariate voxel activation on new sessions. A value close to 1 means that the voxel activation is well predicted by the encoding model, while a 0 value means that the voxel activation cannot be predicted. We compare the \( R^2 \)-maps across the various data-reduction methods for estimating \( \beta \).
Validation. To measure the difference between $R^2$ maps $R$ computed from voxels and $R^2$ maps $\tilde{R}$ computed from DiFuMos, we report correlation coefficients $\rho$ between $R$ and $\tilde{R}$, and the slope $s$ predicting the activations $\tilde{R}$ from the activations $R$. This slope indicates a form of signal loss due to using functional atlases. We expect it to be smaller than 1, in part because projection on functional atlases have a noise reduction effect.

Results. Figure A6, using higher order DiFumo atlases leads to a loss of explained variance $R^2$ of only 6% compared to working directly with voxels, which may imputed to a denoising effect. Qualitatively, the $R^2$ maps are much comparable. DiFuMo ($k = 1024$) is therefore suitable for intra-subject encoding tasks; they make these much less costly. Using lower-order atlases yield stronger signal loss.

AppendixE. Extra meta-analysis maps

Figure A7 shows the meta-analysis summary images for two additional cognitive topics: language and face. We compare non-reduced images with reduced images using DiFuMo ($k = 1024$) and BASC ($k = 444$). The images reduced with DiFuMos are easier to interpret than the ones reduced with BASC for both topics. Quantitatively, we recall that Figure 7 shows the better performance of Di-FuMos for image compression.

AppendixF. DiFuMos naming details

A measure of overlap with a reference anatomical atlas allows to match each DiFuMo component with a specific anatomical region, e.g. “postcentral gyrus”. Where there are more than one component for each anatomical region, the functional atlas region are further characterized by an anatomical spatial descriptions, e.g. “postcentral gyrus inferior”. Finally, we append the localisation of the region in the left or right hemisphere, e.g. “postcentral gyrus inferior RH”. Some of the nodes from DiFuMo atlases overlaps a fraction of several regions in the anatomical atlas—those are named by a trained neuroanatomist.
Figure A7: **Meta-analysis on cognitive topics – language (a.) and face (b.) – from statistical images**: We compare images reconstructed with DiFuMo ($k = 1024$) and BASC ($k = 444$) with voxel-level averages (right). The topic-related activations are better visualized using DiFuMo (middle) than using BASC (left). DiFuMo results are closer to voxel-level averages, as the signal loss is minimal when projecting on this atlas.

Figure A8: **Interpretation of higher-dimensional modes of DiFuMo**: The putamen segmentation is refined as dimension of DiFuMos increases. A single mode contain the left and right putamen in lower dimension (a), when higher order atlases holds separate components for them. Larger atlases model the detailed organization within the sub-structures, which may be crucial in discriminative tasks.
<table>
<thead>
<tr>
<th>fMRI study</th>
<th>Version</th>
<th>Cognitive task</th>
<th>#Subjects</th>
<th>#Sessions</th>
<th>#Runs</th>
<th>Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Schonberg et al., 2012)</td>
<td>ds000001_R2.0.4</td>
<td>Balloon task</td>
<td>16</td>
<td>-</td>
<td>3</td>
<td>balloon analog risk</td>
</tr>
<tr>
<td>(Aron et al., 2006)</td>
<td>ds000002_R2.0.5</td>
<td>Classification learning</td>
<td>17</td>
<td>-</td>
<td>2</td>
<td>deterministic classification mixed event related prob</td>
</tr>
<tr>
<td>(Xue and Poldrack, 2007)</td>
<td>ds000003_R2.0.2</td>
<td>Rhyme judgment</td>
<td>13</td>
<td>-</td>
<td></td>
<td>probabilistic classification related prob</td>
</tr>
<tr>
<td>(Jimura et al., 2014)</td>
<td>ds000006_R2.0.1</td>
<td>ds000006</td>
<td>14</td>
<td>2</td>
<td>6</td>
<td>living nonlining decision</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>with plain or mirror reversed text</td>
</tr>
<tr>
<td>(Xue et al., 2008)</td>
<td>ds000007_R2.0.1</td>
<td>Stop-signal task with spoken &amp; manual responses</td>
<td>20</td>
<td>-</td>
<td>2</td>
<td>stop manual</td>
</tr>
<tr>
<td>(Aron et al., 2007)</td>
<td>ds000008_R2.0.0</td>
<td>Stop-signal task with unconditional and conditional stopping</td>
<td>14</td>
<td>-</td>
<td>3</td>
<td>conditional stop signal</td>
</tr>
<tr>
<td>(Foerde et al., 2006)</td>
<td>ds000011_R2.0.1</td>
<td>Classification learning and tone counting</td>
<td>14</td>
<td>-</td>
<td>2</td>
<td>Classification probe without feedback</td>
</tr>
<tr>
<td>(Rizk-Jackson et al., 2011)</td>
<td>ds000017_R2.0.1</td>
<td>Classification learning and stop-signal (1 year test-retest)</td>
<td>8</td>
<td>2</td>
<td>3</td>
<td>probabilistic classification progressive</td>
</tr>
<tr>
<td>(Alvarez and Poldrack, 2011)</td>
<td>ds000051_R2.0.2</td>
<td>Cross-language repetition priming</td>
<td>13</td>
<td>-</td>
<td>8</td>
<td>selective stop signal</td>
</tr>
<tr>
<td>(Poldrack et al., 2001)</td>
<td>ds000052_R2.0.0</td>
<td>Classification learning and reversal</td>
<td>14</td>
<td>-</td>
<td>2</td>
<td>abstract concrete judgment</td>
</tr>
<tr>
<td>(Mennes et al., 2013)</td>
<td>ds000101_R2.0.0</td>
<td>Simon task</td>
<td>21</td>
<td>-</td>
<td>2</td>
<td>weather prediction</td>
</tr>
<tr>
<td>(Kelly et al., 2008)</td>
<td>ds000102_R2.0.0</td>
<td>Flanker task (event-related)</td>
<td>26</td>
<td>-</td>
<td>2</td>
<td>reversal weather prediction</td>
</tr>
<tr>
<td>(Haxby et al., 2001)</td>
<td>ds000105_R2.0.2</td>
<td>Visual object recognition</td>
<td>6</td>
<td>-</td>
<td>12</td>
<td>object viewing</td>
</tr>
<tr>
<td>(O’Toole et al., 2005)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Hanson et al., 2004)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Duncan et al., 2009)</td>
<td>ds000107_R2.0.2</td>
<td>Word and object processing</td>
<td>49</td>
<td>-</td>
<td>2</td>
<td>1-back task</td>
</tr>
<tr>
<td>(Moran et al., 2012)</td>
<td>ds000109_R2.0.2</td>
<td>False belief task</td>
<td>36</td>
<td>-</td>
<td>2</td>
<td>theory of mind</td>
</tr>
<tr>
<td>(Uncapher et al., 2011)</td>
<td>ds000110_R2.0.1</td>
<td>Incidental encoding task (Posner Cueing Paradigm)</td>
<td>18</td>
<td>-</td>
<td>10</td>
<td>Incidental encoding task</td>
</tr>
<tr>
<td>(Gorgolewski et al., 2013)</td>
<td>ds000114_R2.0.1</td>
<td>A test-retest fMRI dataset for motor, language and spatial attention functions</td>
<td>10</td>
<td>2</td>
<td></td>
<td>covert verb generation</td>
</tr>
<tr>
<td>(Repovs and Barch, 2012)</td>
<td>ds000115_R2.0.0</td>
<td>Working memory in healthy and schizophrenic individuals</td>
<td>1</td>
<td>-</td>
<td></td>
<td>finger footprints</td>
</tr>
<tr>
<td>(Cera et al., 2014)</td>
<td>ds000133_R1.0.0</td>
<td>Modafinil alters intrinsic functional connectivity of the right posterior insula</td>
<td>26</td>
<td>2</td>
<td>3</td>
<td>overt verb generation</td>
</tr>
<tr>
<td>(Verstynen, 2014)</td>
<td>ds000164_R1.0.1</td>
<td>Stroop task</td>
<td>28</td>
<td>-</td>
<td></td>
<td>overt word generation</td>
</tr>
<tr>
<td>Gabitov et al., 2015</td>
<td>ds000170_R1.0.1</td>
<td>Learning and memory: motor skill consolidation and intermanual transfer</td>
<td>15</td>
<td>-</td>
<td>3</td>
<td>letter 0-back task</td>
</tr>
<tr>
<td>(Lepping et al., 2016a)</td>
<td>ds000171_R1.0.0</td>
<td>Neural Processing of Emotional Musical and Nonmusical Stimuli in Depression</td>
<td>39</td>
<td>-</td>
<td>5</td>
<td>letter 1-back task</td>
</tr>
<tr>
<td>(Lepping et al., 2016b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>letter 2-back task</td>
</tr>
<tr>
<td>(Iannilli et al., 2016)</td>
<td>ds000200_R1.0.0</td>
<td>Pre-adolescents Exposure to Manganese</td>
<td>1</td>
<td>-</td>
<td></td>
<td>rest</td>
</tr>
<tr>
<td>(Stephan-Otto et al., 2017)</td>
<td>ds000203_R1.0.2</td>
<td>Visual imagery and false memory for pictures</td>
<td>26</td>
<td>-</td>
<td>2</td>
<td>rest</td>
</tr>
<tr>
<td>(Kim et al., 2016)</td>
<td>ds000205_R1.0.0</td>
<td>Affective Videos</td>
<td>11</td>
<td>-</td>
<td>2</td>
<td>false memory</td>
</tr>
<tr>
<td>(Romaniuk et al., 2016)</td>
<td>ds000214_R1.0.0</td>
<td>EUPD Cyberball</td>
<td>40</td>
<td>-</td>
<td></td>
<td>rest</td>
</tr>
<tr>
<td>(Roy et al., 2017)</td>
<td>ds000220_R1.0.0</td>
<td>Cost Analysis TBI</td>
<td>26</td>
<td>2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table A4: Large-scale fMRI datasets downloaded from OpenNeuro to build our multi-scale functional atlases. Data are pre-processed using fMRIPrep. The data acquisition parameters of each study are listed on Table A5. The corpus is 2.4TB in total.
<table>
<thead>
<tr>
<th>MRI study</th>
<th>MR scanner</th>
<th>Slice orientation</th>
<th>FoV (mm)</th>
<th>Voxel size (mm)</th>
<th>Matrix size</th>
<th>TR (msec)</th>
<th>TE (msec)</th>
<th>Flip angle (°)</th>
<th>Number of volumes</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Schonberg et al., 2012)</td>
<td>3T Siemens AG Allegra (Erlangen, Germany)</td>
<td>axial</td>
<td>-</td>
<td>4 x 4 x 4</td>
<td>64 x 64</td>
<td>2000</td>
<td>30</td>
<td>90</td>
<td>300</td>
</tr>
<tr>
<td>(Aron et al., 2006)</td>
<td>3T Siemens Allegra</td>
<td>-</td>
<td>-</td>
<td>4 x 4 x 4</td>
<td>64 x 64</td>
<td>2000</td>
<td>30</td>
<td>90</td>
<td>180</td>
</tr>
<tr>
<td>(Xue and Poldrack, 2007)</td>
<td>3T Siemens Allegra (Iselin, NJ)</td>
<td>-</td>
<td>200</td>
<td>4 x 4 x 4</td>
<td>64 x 64</td>
<td>2000</td>
<td>30</td>
<td>90</td>
<td>160</td>
</tr>
<tr>
<td>(Jimura et al., 2014)</td>
<td>3T Siemens Allegra (Erlangen, Germany)</td>
<td>-</td>
<td>200</td>
<td>4 x 4 x 4</td>
<td>64 x 64</td>
<td>2000</td>
<td>30</td>
<td>90</td>
<td>205</td>
</tr>
<tr>
<td>(Xue et al., 2008)</td>
<td>3T Siemens Allegra</td>
<td>-</td>
<td>200</td>
<td>4 x 4 x 4</td>
<td>64 x 64</td>
<td>2000</td>
<td>30</td>
<td>90</td>
<td>182</td>
</tr>
<tr>
<td>(Aron et al., 2007)</td>
<td>3T Siemens Allegra</td>
<td>-</td>
<td>200</td>
<td>4 x 4 x 4</td>
<td>64 x 64</td>
<td>2000</td>
<td>30</td>
<td>90</td>
<td>176</td>
</tr>
<tr>
<td>(Foerde et al., 2006)</td>
<td>3T Siemens Allegra</td>
<td>-</td>
<td>200</td>
<td>4 x 4 x 4</td>
<td>64 x 64</td>
<td>2000</td>
<td>30</td>
<td>-</td>
<td>208</td>
</tr>
<tr>
<td>(Poldrack et al., 2001)</td>
<td>3T Siemens Allegra</td>
<td>axial</td>
<td>200</td>
<td>5 x 5 x 5</td>
<td>64 x 64</td>
<td>3000</td>
<td>30</td>
<td>-</td>
<td>225</td>
</tr>
<tr>
<td>(Mennes et al., 2013)</td>
<td>3T Siemens Allegra</td>
<td>-</td>
<td>192</td>
<td>3 x 3 x 3</td>
<td>64 x 64</td>
<td>2000</td>
<td>30</td>
<td>80</td>
<td>101</td>
</tr>
<tr>
<td>(Kelly et al., 2008)</td>
<td>3T Siemens Allegra</td>
<td>-</td>
<td>192</td>
<td>3 x 3 x 3</td>
<td>64 x 64</td>
<td>2000</td>
<td>30</td>
<td>80</td>
<td>146</td>
</tr>
<tr>
<td>(Haxby et al., 2001)</td>
<td>3T GE sagittal</td>
<td>240</td>
<td>3.5 x 3.5 x 3.5</td>
<td>-</td>
<td>2500</td>
<td>30</td>
<td>90</td>
<td>121</td>
<td></td>
</tr>
<tr>
<td>(Duncan et al., 2009)</td>
<td>1.5T Siemens Allegra</td>
<td>-</td>
<td>192</td>
<td>3 x 3 x 3</td>
<td>64 x 64</td>
<td>3000</td>
<td>50</td>
<td>-</td>
<td>165</td>
</tr>
<tr>
<td>(Moran et al., 2012)</td>
<td>3T Siemens Tim Trio</td>
<td>axial</td>
<td>-</td>
<td>3 x 3 x 3</td>
<td>64 x 64</td>
<td>2000</td>
<td>35</td>
<td>-</td>
<td>179</td>
</tr>
<tr>
<td>(Uncapher et al., 2011)</td>
<td>3T GE Signa</td>
<td>axial</td>
<td>-</td>
<td>3.44 x 3.44 x 3.44</td>
<td>64 x 64</td>
<td>2000</td>
<td>30</td>
<td>75</td>
<td>186</td>
</tr>
<tr>
<td>(Gorgolewski et al., 2013)</td>
<td>1.5T GE Signa</td>
<td>-</td>
<td>256</td>
<td>4 x 4 x 4</td>
<td>64 x 64</td>
<td>2500</td>
<td>50</td>
<td>90</td>
<td>varied</td>
</tr>
<tr>
<td>(Repovs and Barch, 2012)</td>
<td>3T Tim Trio</td>
<td>-</td>
<td>256</td>
<td>4 x 4 x 4</td>
<td>64 x 64</td>
<td>2500</td>
<td>27</td>
<td>90</td>
<td>137</td>
</tr>
<tr>
<td>(Cera et al., 2014)</td>
<td>3T Philips transaxial</td>
<td>256</td>
<td>4 x 4 x 4</td>
<td>64 x 64</td>
<td>1671</td>
<td>35</td>
<td>75</td>
<td>145</td>
<td></td>
</tr>
<tr>
<td>(Verstynen, 2014)</td>
<td>3T Siemens</td>
<td>-</td>
<td>-</td>
<td>3.2 x 3.2 x 4</td>
<td>-</td>
<td>1500</td>
<td>20</td>
<td>90</td>
<td>370</td>
</tr>
<tr>
<td>(Gabitov et al., 2015)</td>
<td>3T GE</td>
<td>220</td>
<td>3.4 x 3.4 x 3.4</td>
<td>64 x 64</td>
<td>3000</td>
<td>35</td>
<td>90</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>(Lepping et al., 2016a)</td>
<td>3T Siemens Skyra (Erlangen, Germany)</td>
<td>axial</td>
<td>220</td>
<td>2.9 x 2.9 x 3</td>
<td>64 x 64</td>
<td>3000</td>
<td>25</td>
<td>90</td>
<td>105</td>
</tr>
<tr>
<td>(Iannilli et al., 2016)</td>
<td>1.5T Siemens Aera (Erlangen, Germany)</td>
<td>axial</td>
<td>-</td>
<td>3.3 x 3.3 x 3.3</td>
<td>64 x 64</td>
<td>2500</td>
<td>50</td>
<td>-</td>
<td>120</td>
</tr>
<tr>
<td>(Stephan-Otto et al., 2017)</td>
<td>1.5T GE Signa</td>
<td>axial</td>
<td>240</td>
<td>4 x 4 x 4</td>
<td>64 x 64</td>
<td>2000</td>
<td>40</td>
<td>90</td>
<td>267</td>
</tr>
<tr>
<td>(Kim et al., 2016)</td>
<td>3T Siemens Tri (Erlangen)</td>
<td>axial</td>
<td>-</td>
<td>3 x 3 x 3</td>
<td>64 x 64</td>
<td>2200</td>
<td>35</td>
<td>90</td>
<td>365</td>
</tr>
<tr>
<td>(Romaniuk et al., 2016)</td>
<td>3T Siemens Magnetom Verio</td>
<td>axial</td>
<td>220</td>
<td>3.4 x 3.4 x 5</td>
<td>64 x 64</td>
<td>1560</td>
<td>26</td>
<td>66</td>
<td>341</td>
</tr>
<tr>
<td>(Roy et al., 2017)</td>
<td>3T Philips Achieva</td>
<td>axial</td>
<td>240</td>
<td>4 x 4 x 4</td>
<td>80 x 80</td>
<td>2000</td>
<td>30</td>
<td>90</td>
<td>144</td>
</tr>
</tbody>
</table>

Table A5: Data acquisition parameters for each fMRI study that we use for training DiFuMo atlases. Data are downloaded from OpenNeuro.