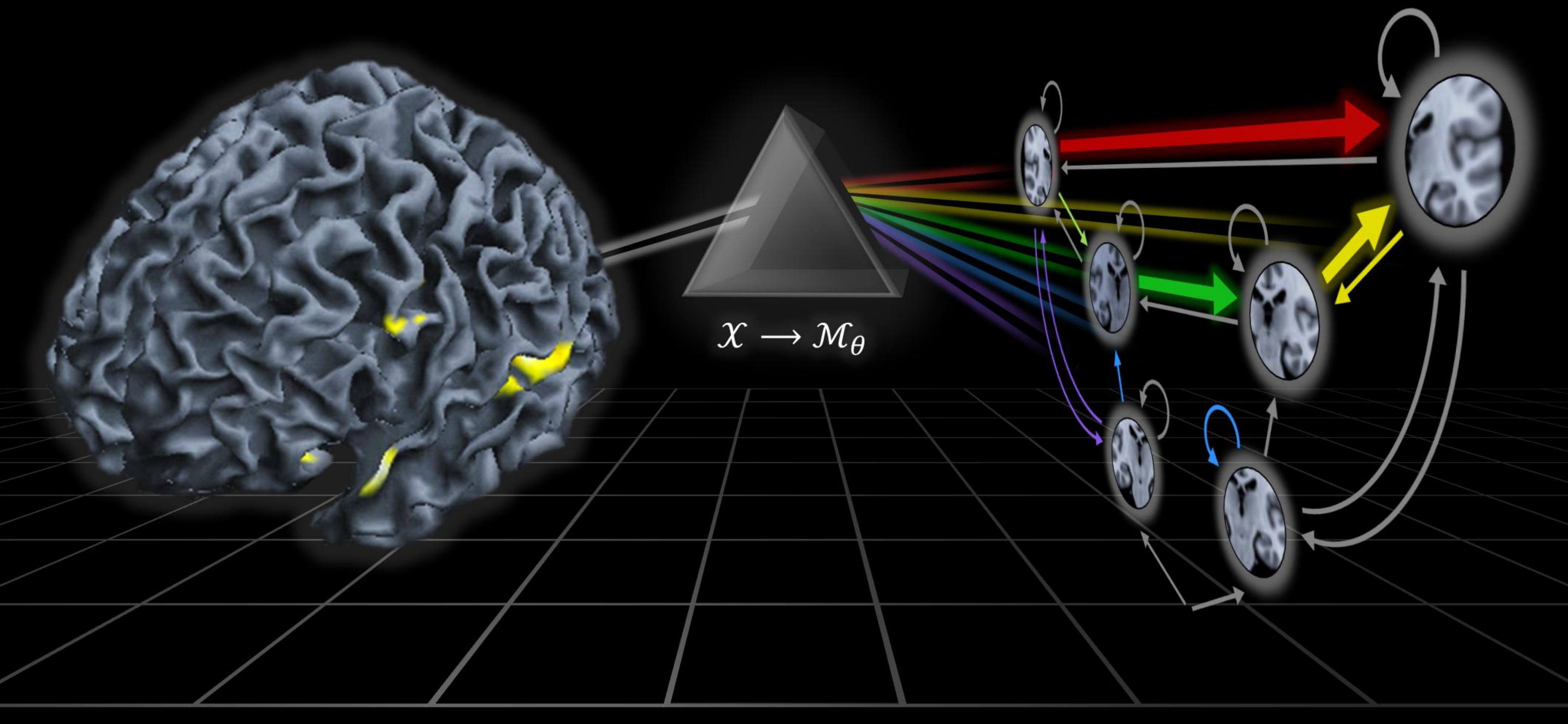
Model-based classification through generative embedding

Kay Henning Brodersen

Machine Learning and Pattern Recognition Group
Department of Computer Science, ETH Zurich

Computational Neuroeconomics Group
Department of Economics, University of Zurich

http://people.inf.ethz.ch/bkay/



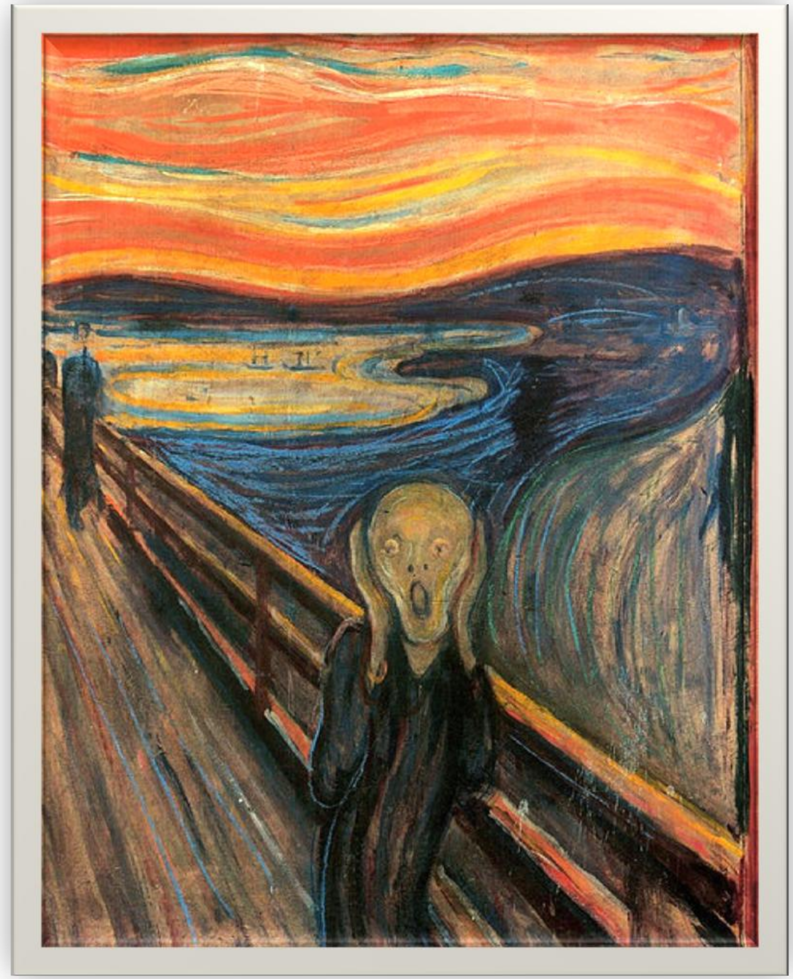
Computational science and psychiatry

Spectrum diseases

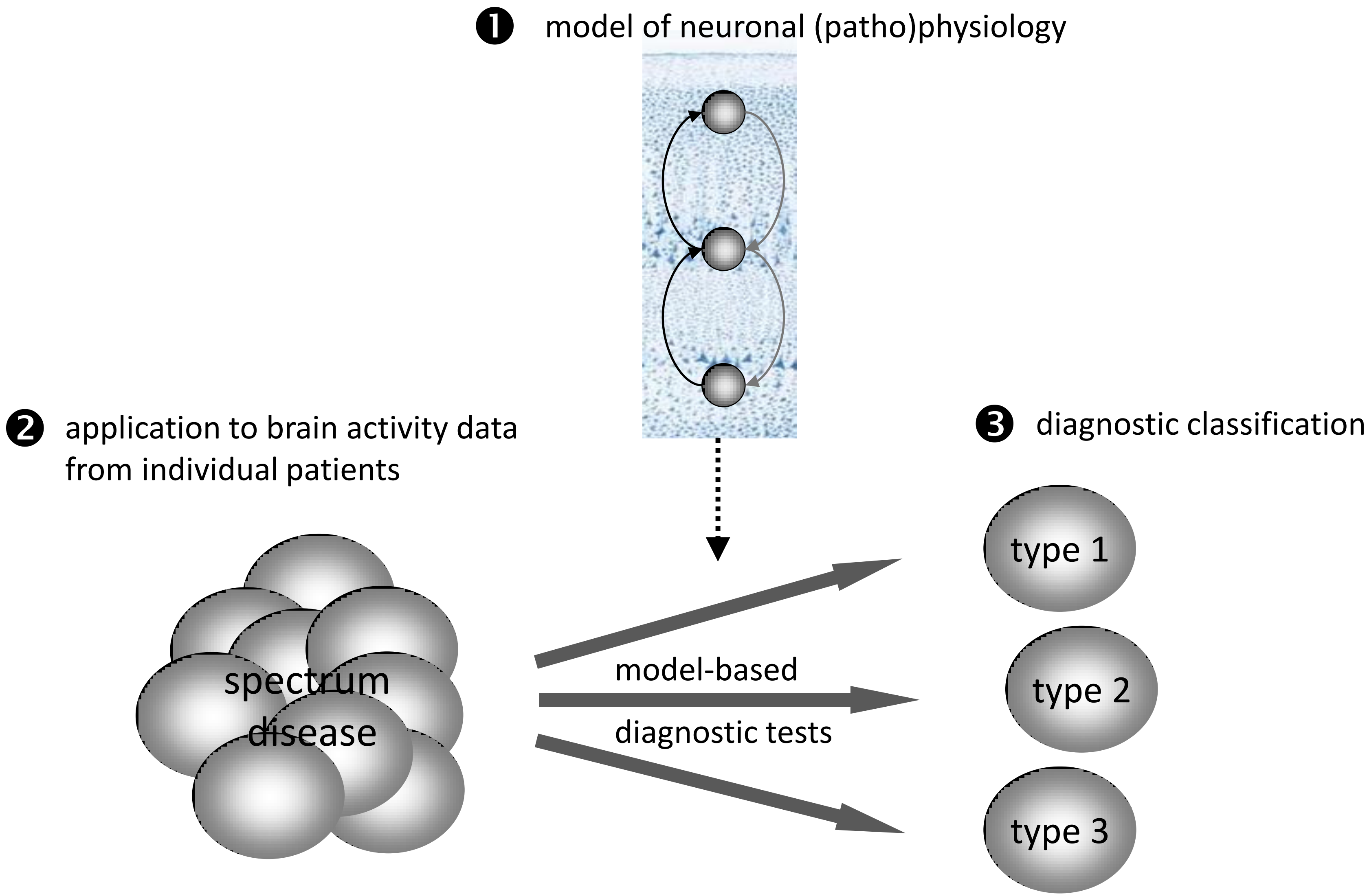
- diverse genetic basis
- strong gene-environment interactions
- variability in treatment response and outcome

Consequences

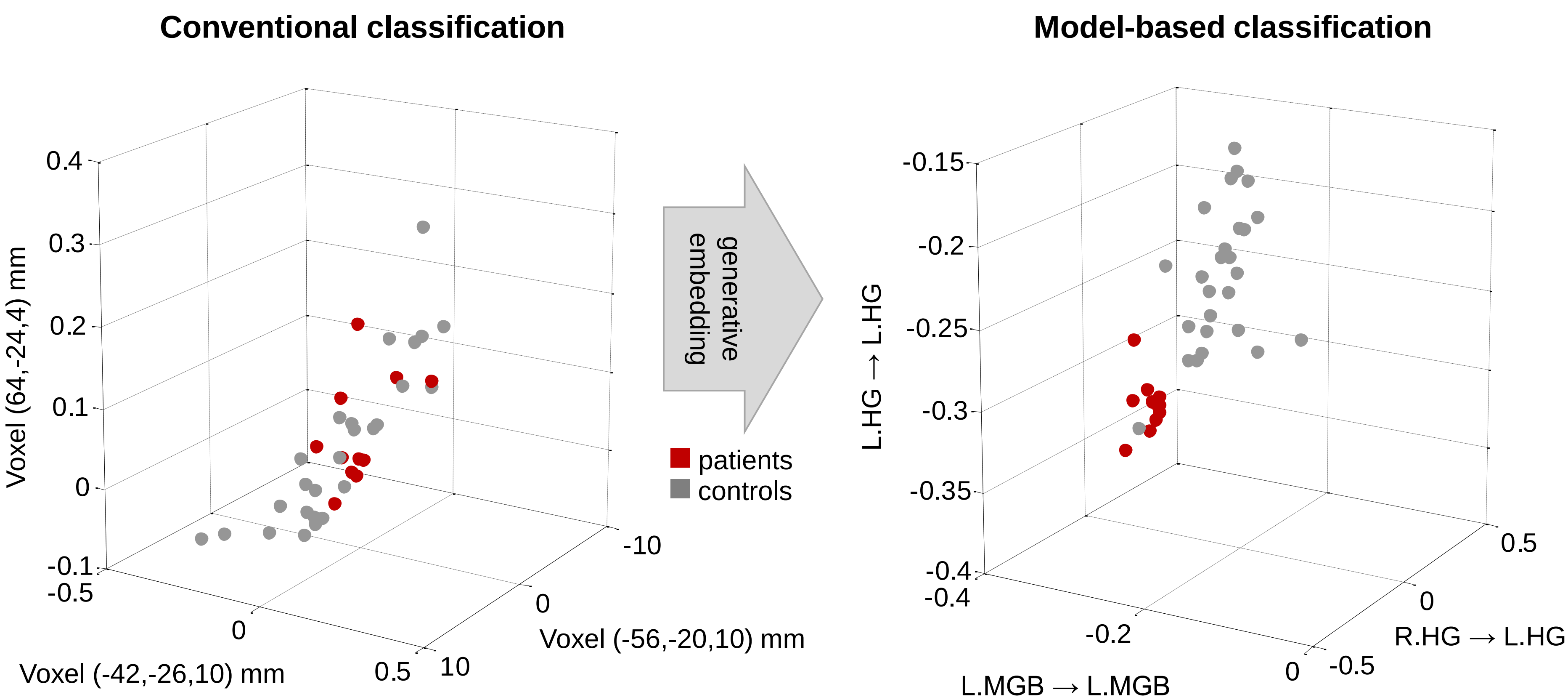
- multiple pathophysiological mechanisms
- even when symptoms are similar, causes may differ across patients
- need to infer pathophysiological mechanisms in individual patients!



Model-based inference on individual pathophysiology



Conventional vs. model-based classification



Colleagues & collaborators

Thomas Schofield
University College London

Cheng Soon Ong
ETH Zurich

Kate Lomakina
University of Zurich · ETH Zurich

Alexander Leff
University College London

Joachim M Buhmann
ETH Zurich

Klaas Enno Stephan
University of Zurich · University College London

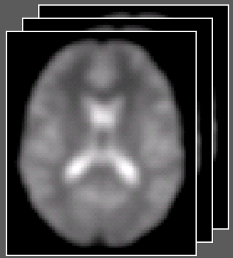
Data representations in classification analyses

Model-based classification

Can we exploit the rich discriminative information encoded in individual patterns of connection strengths?

Activation-based classification

- depression
- schizophrenia
- mild cognitive impairment

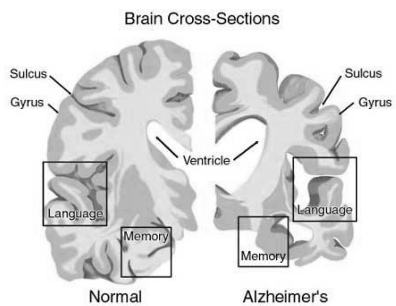


Structure-based classification

- mild traumatic brain injury
- Alzheimer's disease
- autistic spectrum disorder
- frontotemporal dementia
- mild cognitive impairment
- schizophrenia
- aphasia

Prediction versus inference

The goal of **prediction** is to find a highly accurate encoding or decoding function.

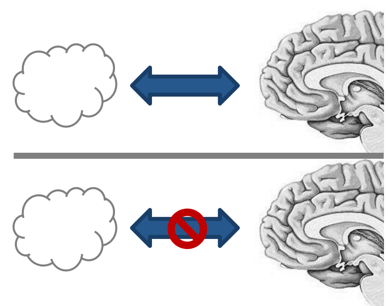


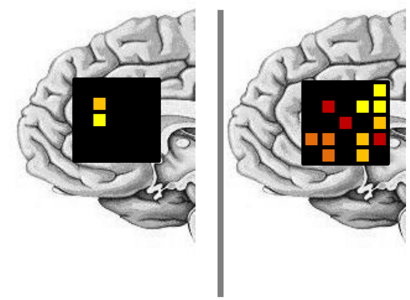
predicting a cognitive state using a brain-machine interface

predicting a subject-specific diagnostic status

⇨ powerful discriminative algorithms for classification

The goal of **inference** is to decide between competing hypotheses about mechanisms or representations in the brain.



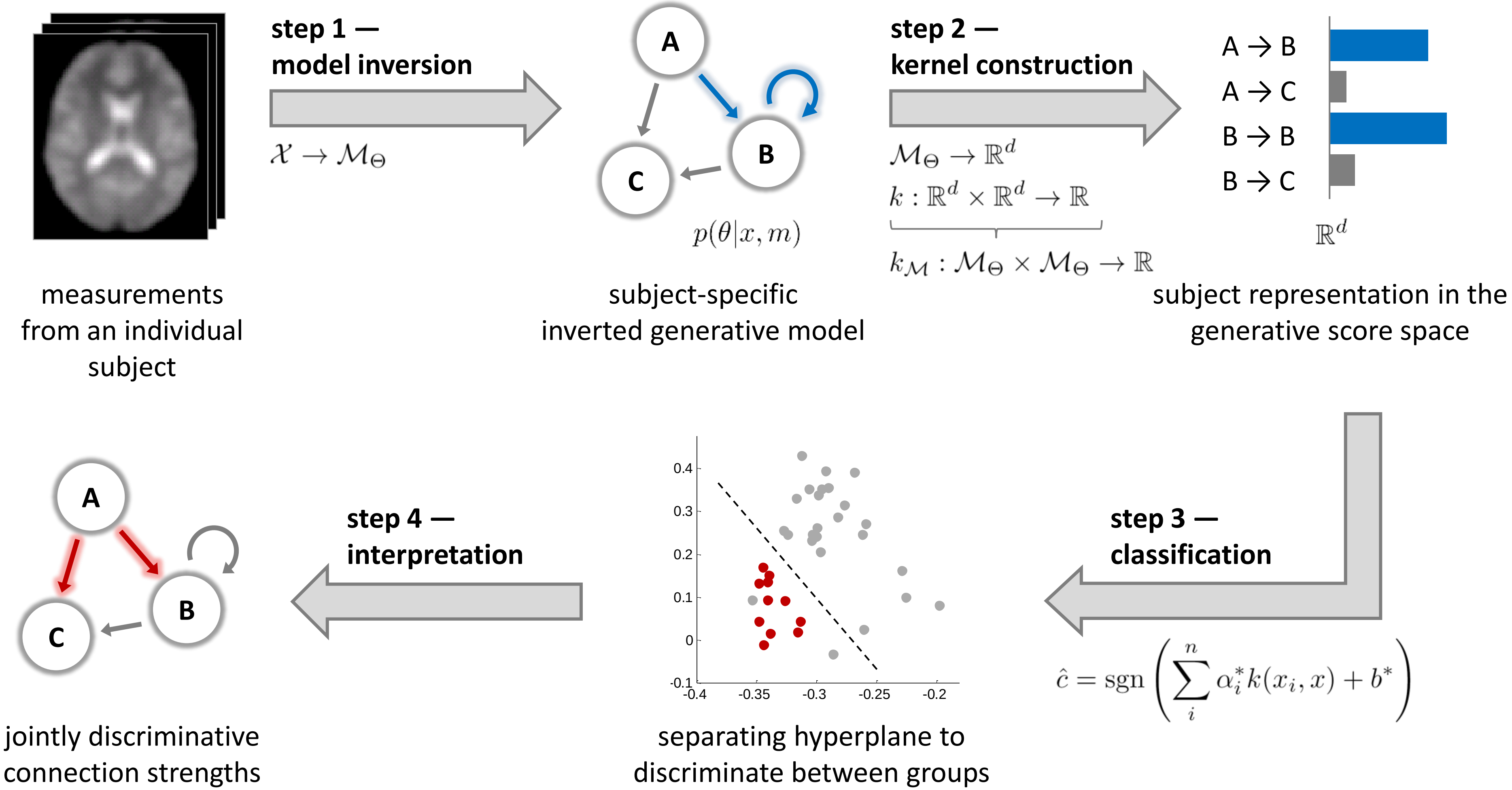


comparing a model that links distributed neuronal activity to a cognitive state with a model that does not

weighing the evidence for sparse coding vs. dense coding

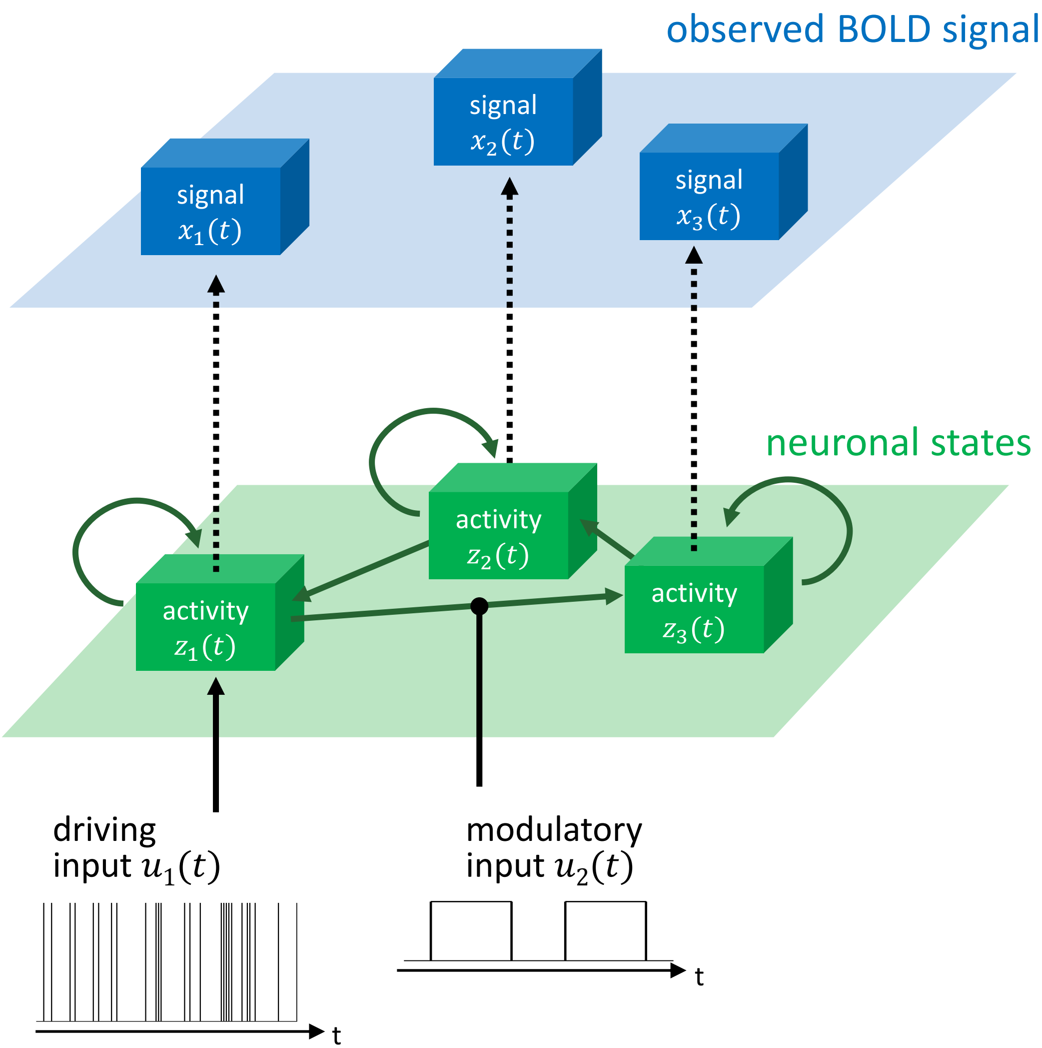
⇨ mechanistically interpretable generative models of brain function

Model-based classification through generative embedding



Brodersen, Haiss, Ong, Jung, Tittgemeyer, Buhmann, Weber, Stephan (2011) *NeuroImage*
Brodersen, Schofield, Leff, Ong, Lomakina, Buhmann, Stephan (2011) *PLoS Comp Biol*

Choosing a generative model: DCM for fMRI



haemodynamic forward model

$$x = g(z, \theta_h)$$

neural state equation

$$\dot{z} = (A + \sum u_j B^{(j)})z + Cu$$

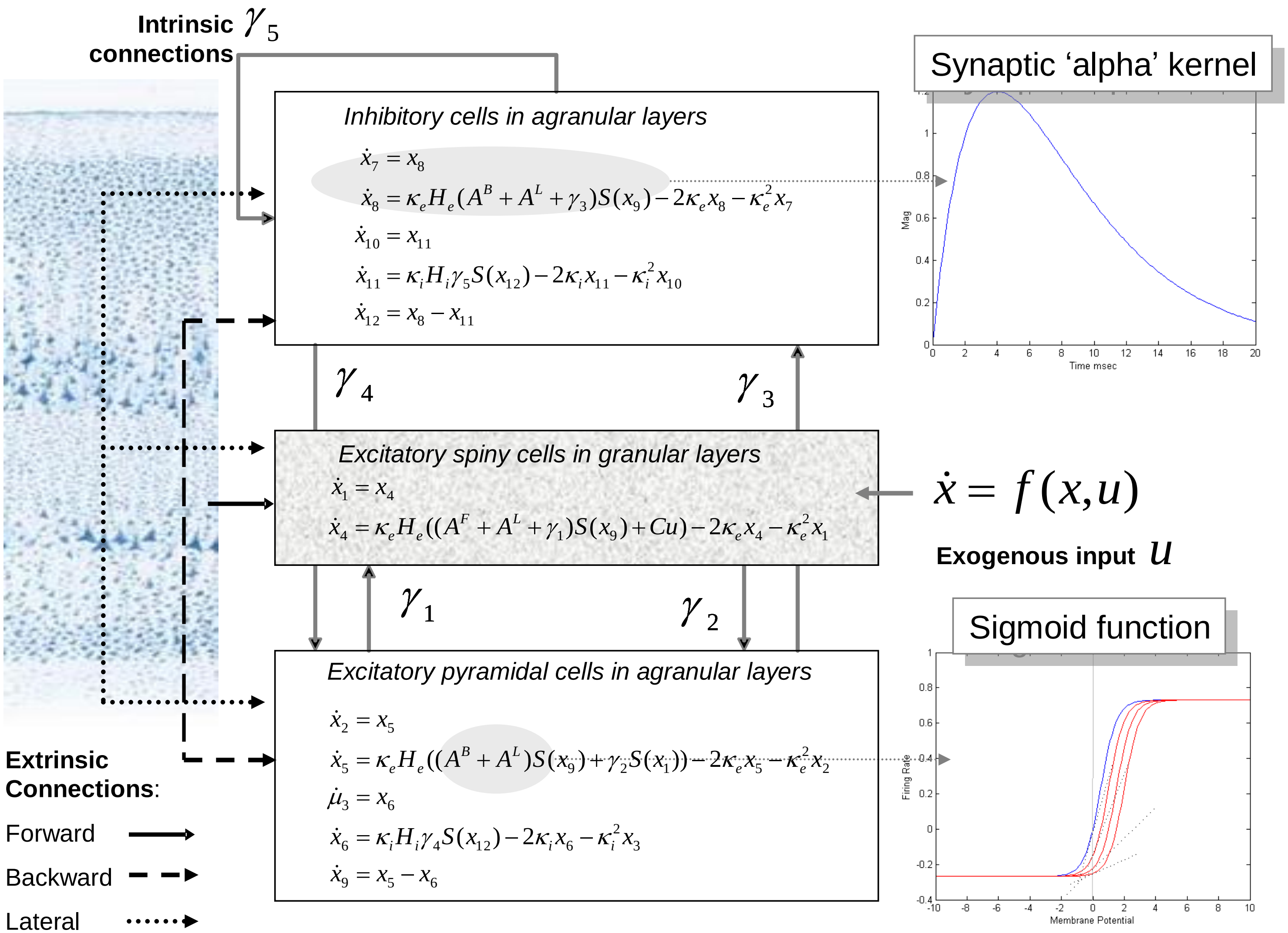
intrinsic connectivity → $A = \frac{\partial \dot{z}}{\partial z}$

modulation of connectivity → $B^{(j)} = \frac{\partial}{\partial u_j}\frac{\partial \dot{z}}{\partial z}$

direct inputs → $C = \frac{\partial \dot{z}}{\partial u}$

Jansen & Rit (1995) *Biological Cybernetics*
Friston, Harrison & Penny (2003) *NeuroImage*
Stephan & Friston (2007), *Handbook of Brain Connectivity*

Choosing a generative model: DCM for LFP/EEG



Moran et al. 2009 *NeuroImage*

Training and testing a model-based classifier

Training a kernel-based discriminant classifier:

$$\max_\alpha \mathcal{L}(\alpha) = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j c_i c_j \boxed{k(x_i, x_j)} + \sum_{i=1}^{n}\alpha_i$$

$$s.t. \ \sum_{i=1}^{n} c_i\alpha_i = 0$$

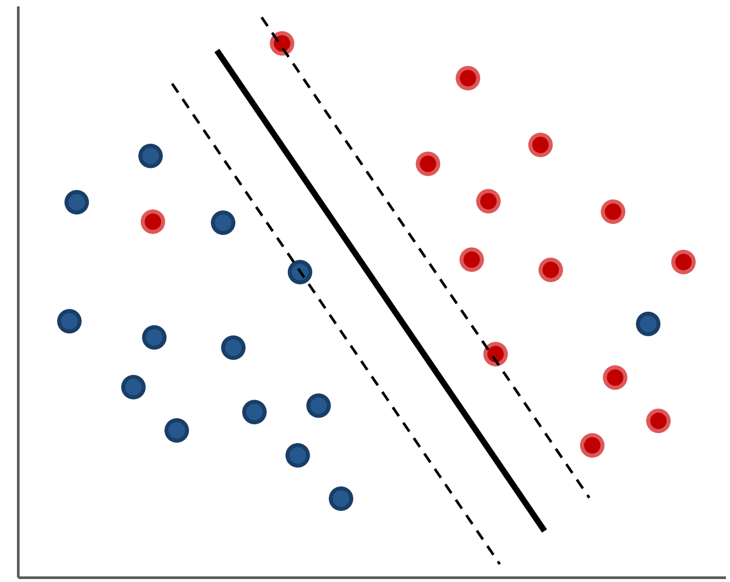
$$0 \le \alpha_i \le C \quad \forall i = 1,\ldots,n$$

Using the model to make predictions:

$$f(x_{n+1}) = \sum_{i=1}^{n}\alpha_i^* \boxed{k(x_i, x_{n+1})} + b^*$$

$$\hat{c}_{n+1} := \mathrm{sgn}(f(x_{n+1}))$$

Linear SVM



In the case of generative embedding:
$k(x_i, x_j) = x_i^T x_j$

Specifying and inverting the model – how?

Definition of ROIs
Are regions of interest defined
anatomically or functionally?

anatomically

A **1 ROI definition**
and *n* model inversions
☑ unbiased estimate

functionally

Functional contrasts
Are the functional contrasts defined
across all subjects or between groups?

across
subjects

B **1 ROI definition and *n* model inversions**
☒ slightly optimistic estimate:
voxel selection for training set and test set
based on test data

C **Repeat *n* times:**
1 ROI definition and *n* model inversions
☑ unbiased estimate

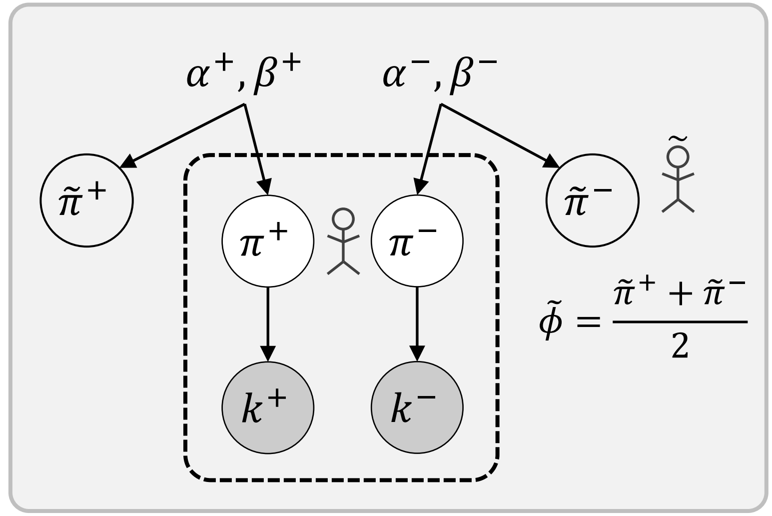
between
groups

D **1 ROI definition and *n* model inversions**
☒ highly optimistic estimate:
voxel selection for training set and test set
based on test data and test labels

E **Repeat *n* times:**
1 ROI definition and 1 model inversion
☒ slightly optimistic estimate:
voxel selection for training set based on
test data and test labels

F **Repeat *n* times:**
1 ROI definition and *n* model inversions
☑ unbiased estimate

Full Bayesian approach to performance evaluation



Model

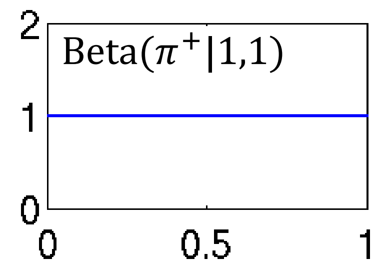
We model the likelihood functions for k^+ positive and k^- negative correct predictions as:

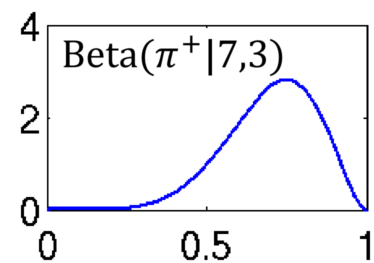
$$p(k^+|\pi^+, n^+) = \mathrm{Bin}(k^+|\pi^+, n^+)$$
$$p(k^-|\pi^-, n^-) = \mathrm{Bin}(k^-|\pi^-, n^-)$$

The class-specific accuracies π^+ and π^- can be modelled as latent random variables with conjugate Beta priors:

$$p(\pi^+|\alpha^+, \beta^+) = \mathrm{Beta}(\pi^+|\alpha^+, \beta^+)$$
$$p(\pi^-|\alpha^-, \beta^-) = \mathrm{Beta}(\pi^-|\alpha^-, \beta^-)$$

This prior is uninformative when using the hyperparameters $\alpha^+ = \beta^+ = \alpha^- = \beta^- = 1$. The balanced accuracy is given by $\phi := \frac{1}{2}(\pi^+ + \pi^-)$.





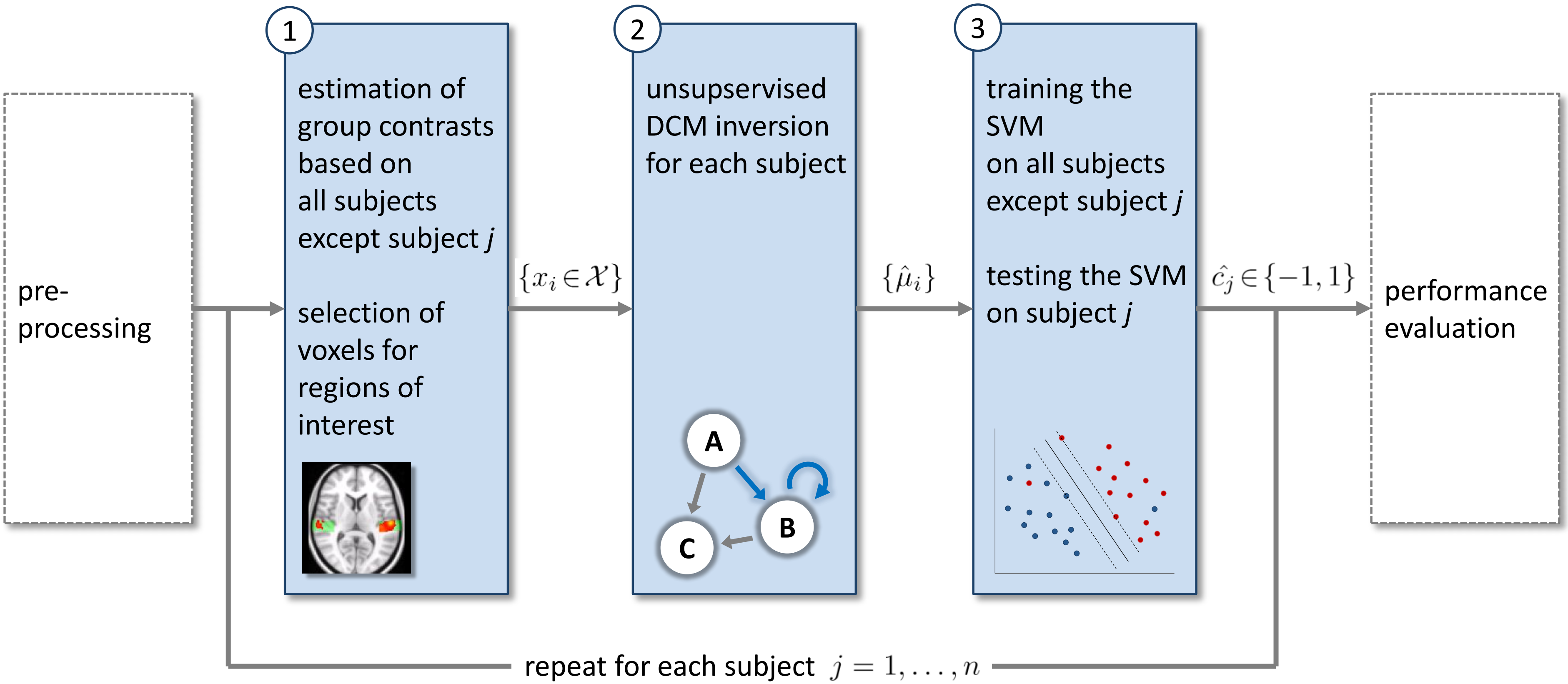
Inference

Inverting the model yields the posterior balanced classification accuracy,

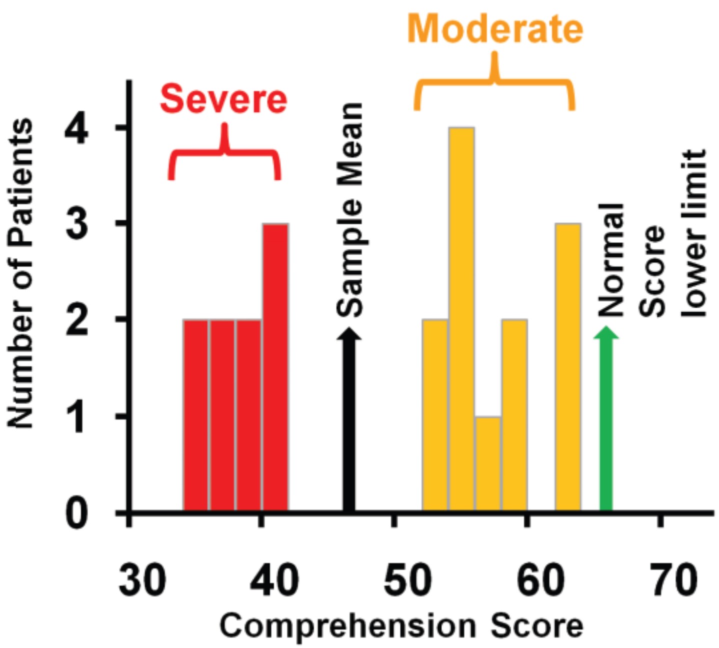
$$p(\phi|k^+, k^-, n^+, n^-, \alpha^+, \beta^+, \alpha^-, \beta^-)$$
$$= \int_0^1 \mathrm{Beta}(2(\phi - z)|\alpha_n^+, \beta_n^+)\ \mathrm{Beta}(2z|\alpha_n^-, \beta_n^-)\ dz\,.$$

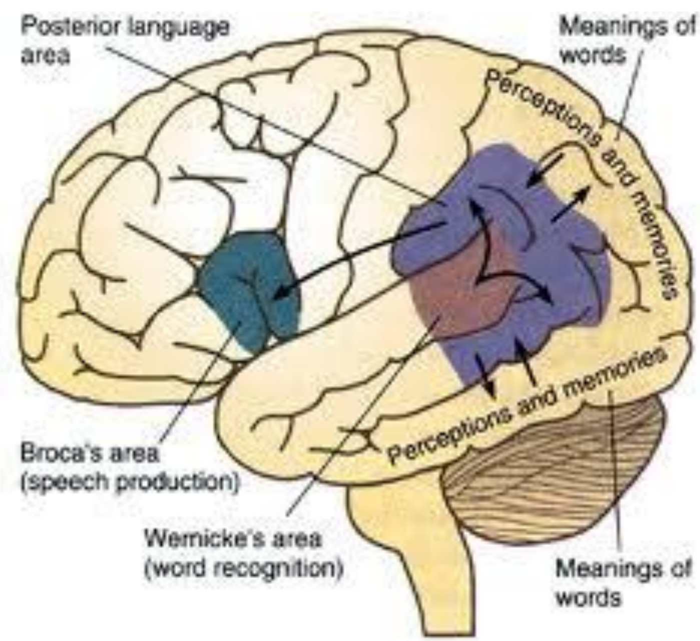
Brodersen, Chumbley, Mathys, Daunizeau, Ong, Buhmann & Stephan *(in preparation)*

Summary of the analysis

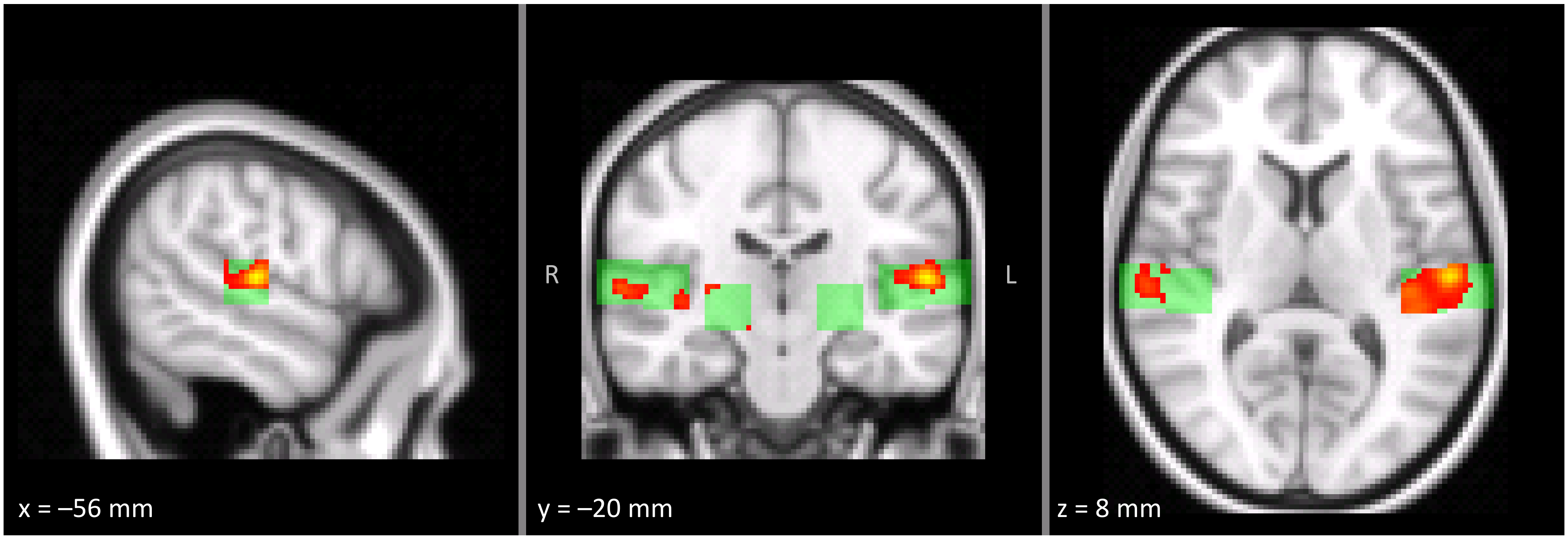


Example: diagnosis of moderate aphasia

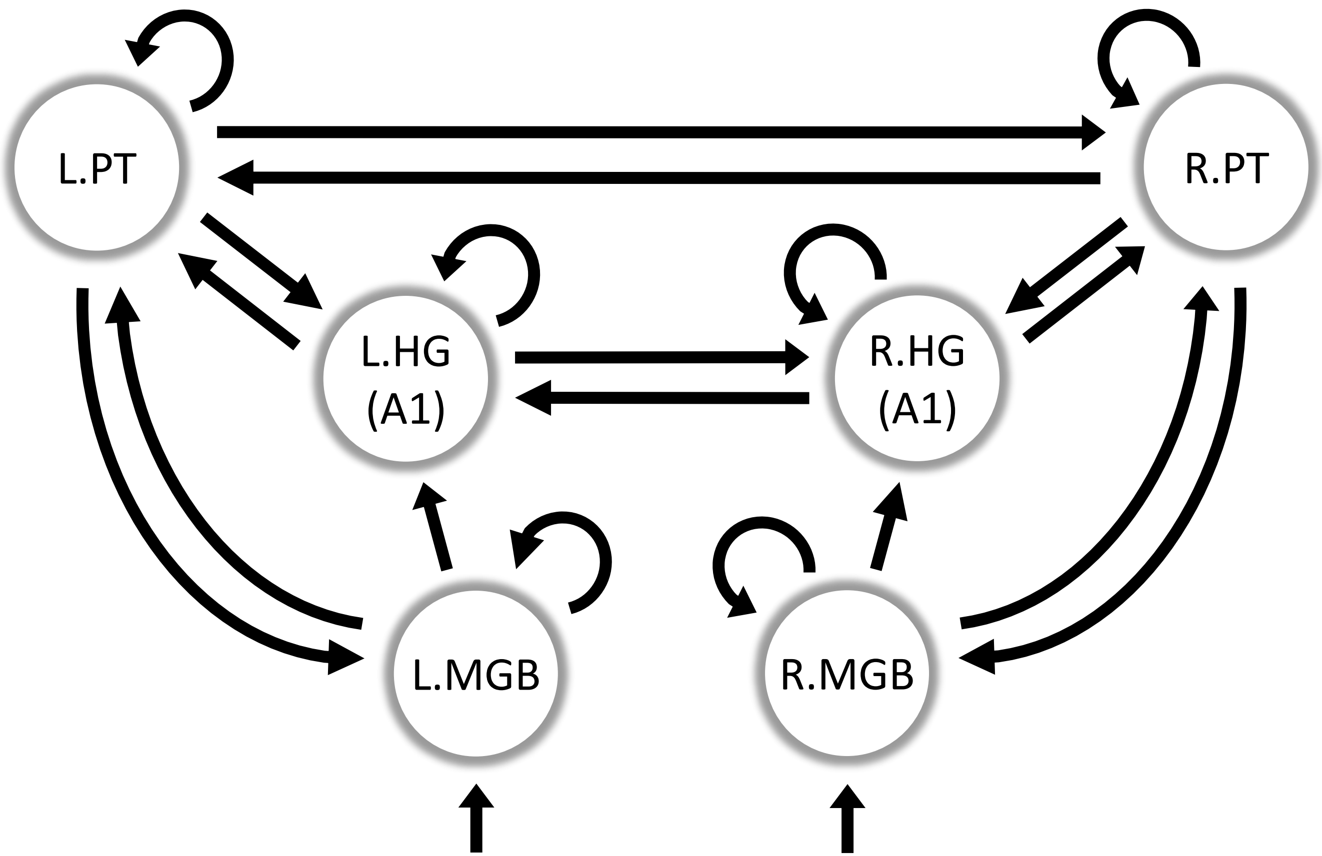




Regions of interest

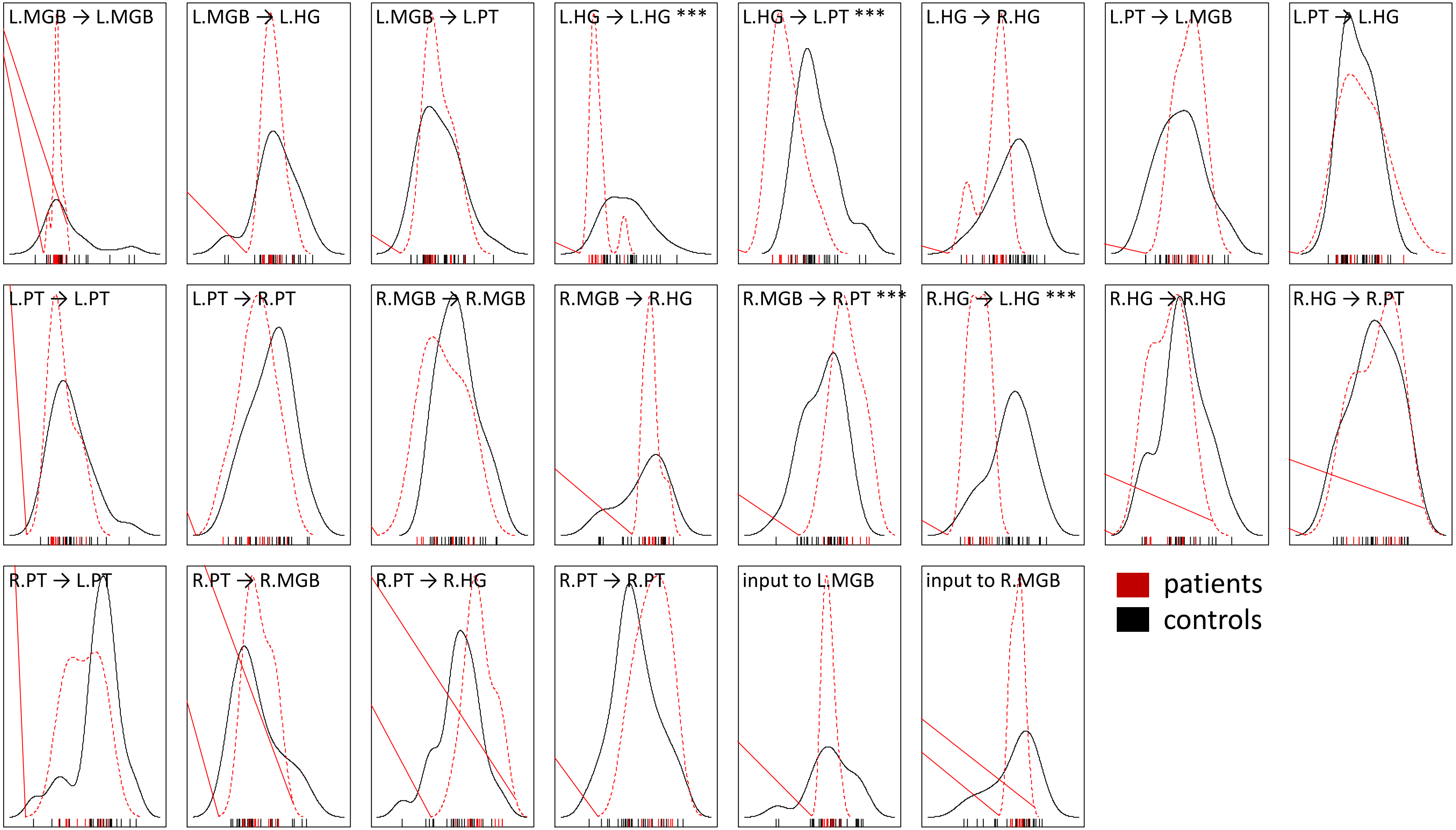


Neuronal model

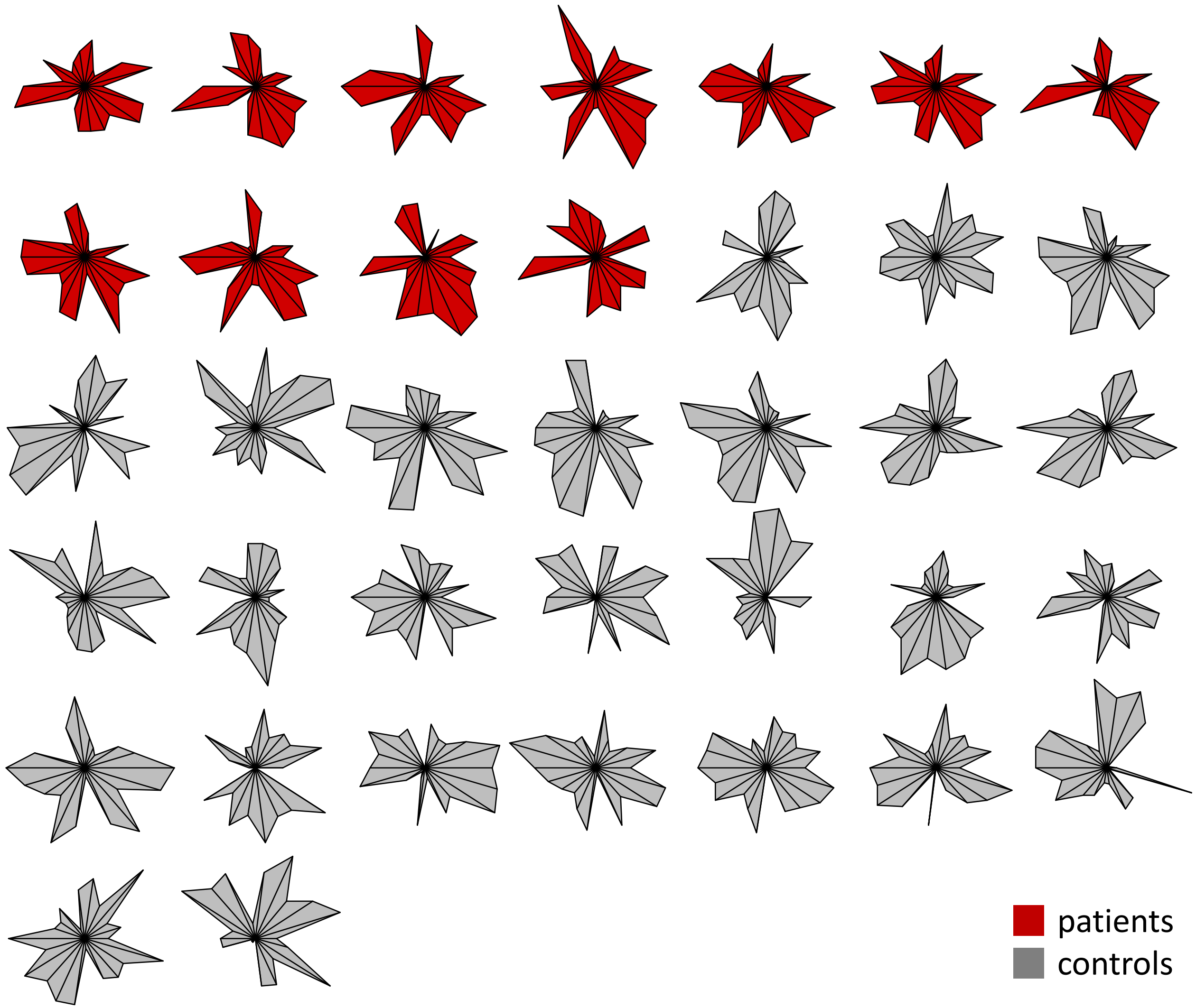


Schofield, Penny, Stephan, Crinion, Thompson, Price & Leff *(under review)*
Brodersen, Schofield, Leff, Ong, Lomakina, Buhmann & Stephan (2011) *PLoS Comp Biol*

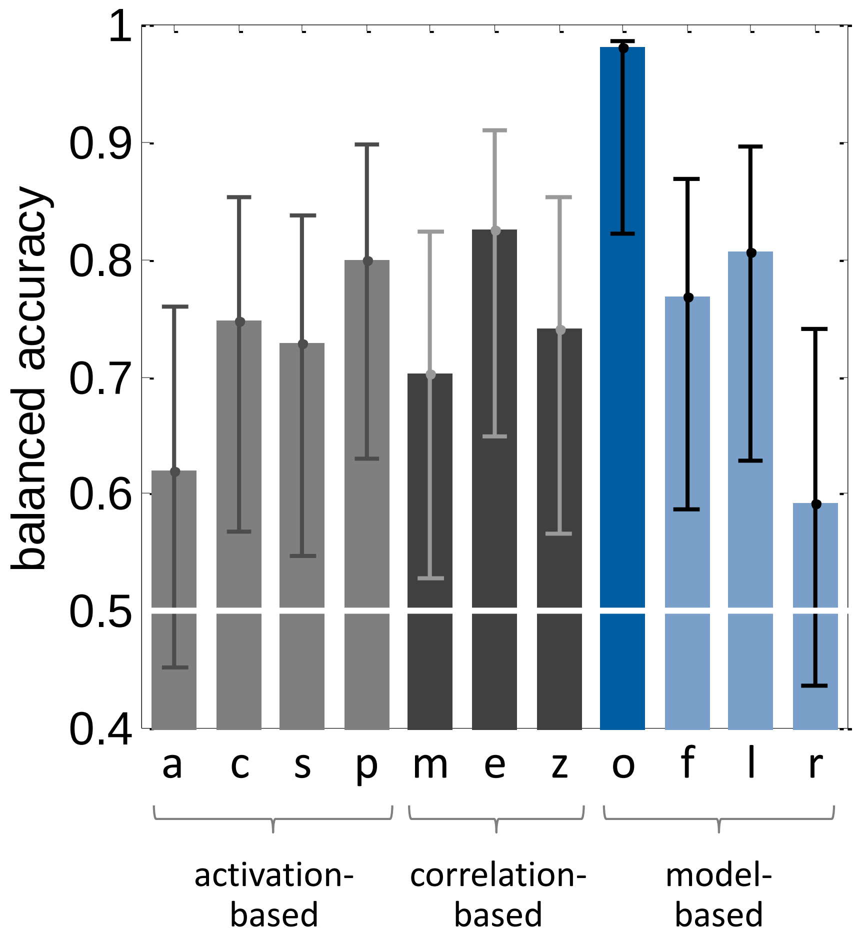
Univariate analysis

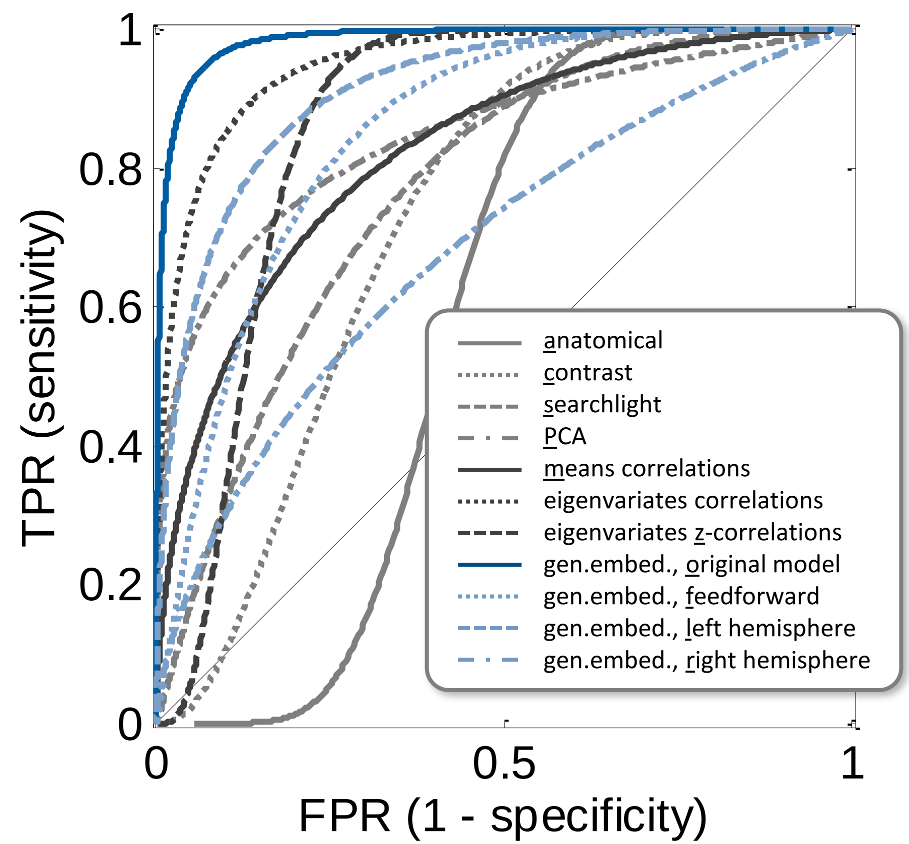


Connectional fingerprints

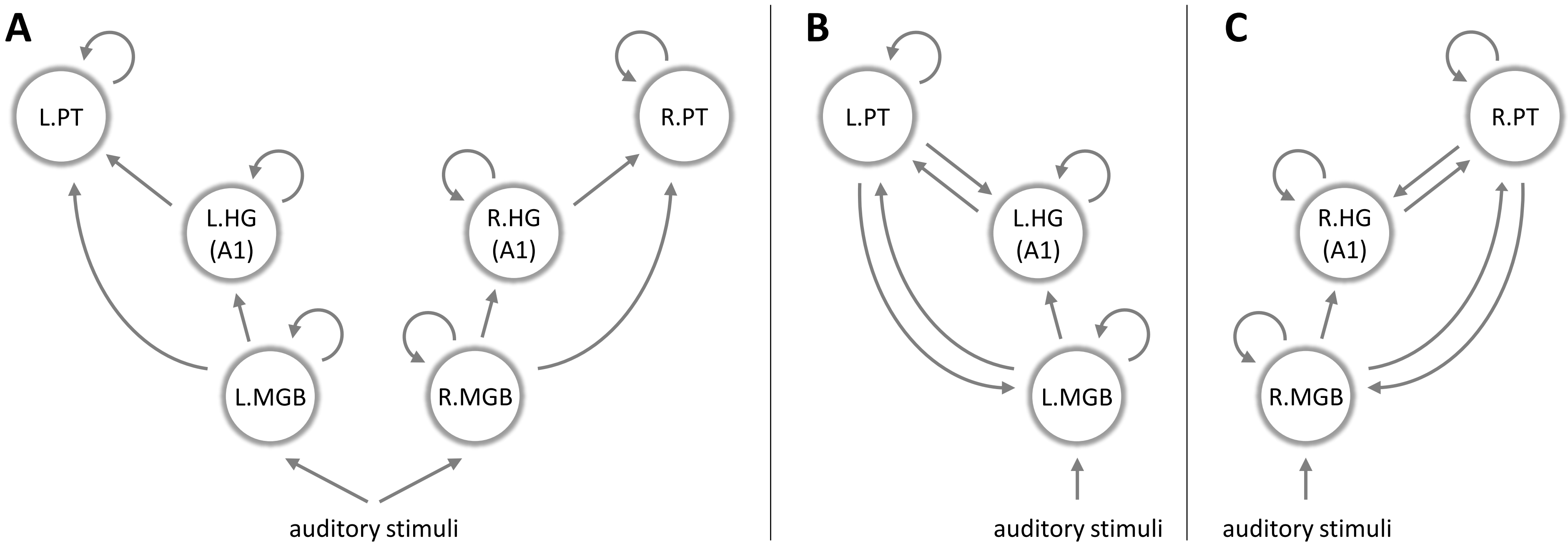


Classification performance

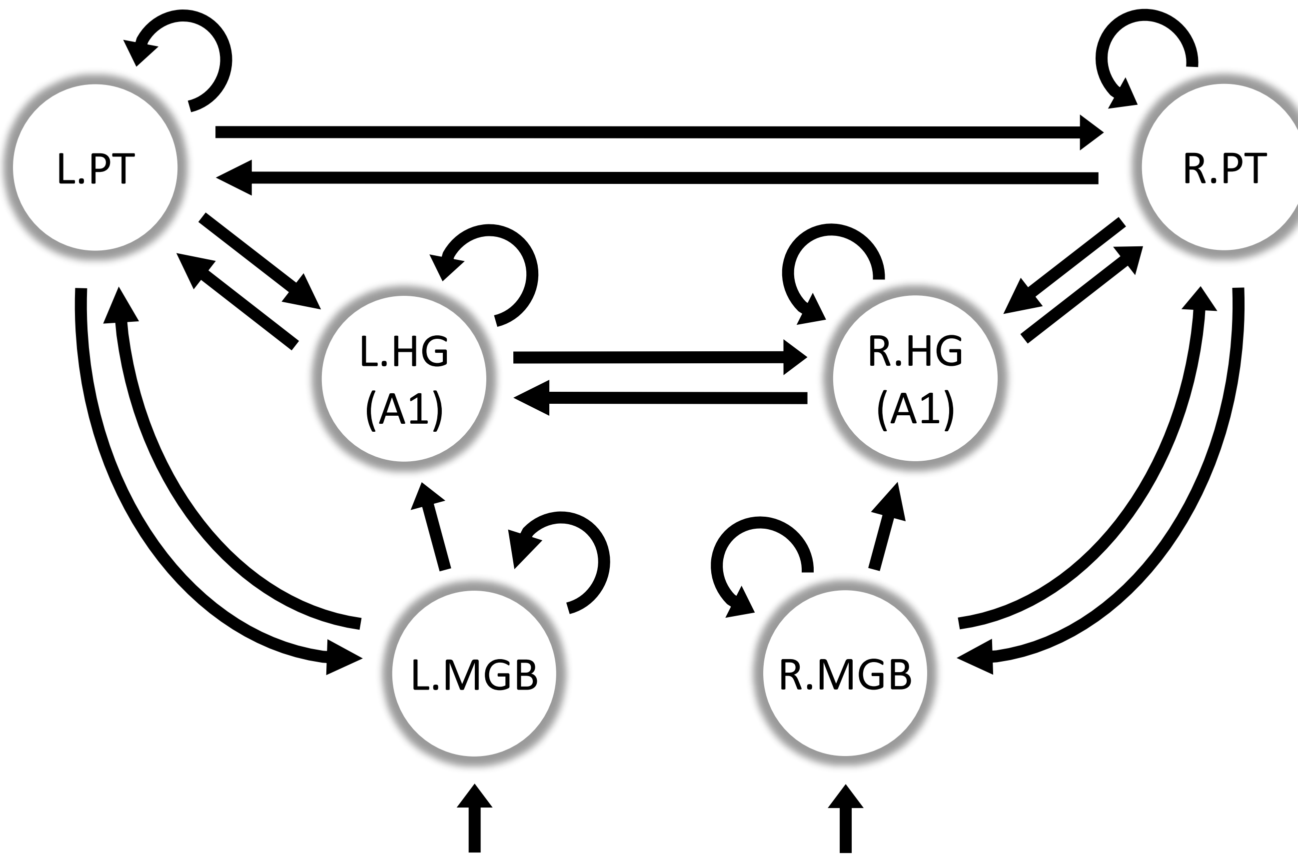




Biologically implausible models perform poorly



Discriminative features in model space



Discriminative features in model space

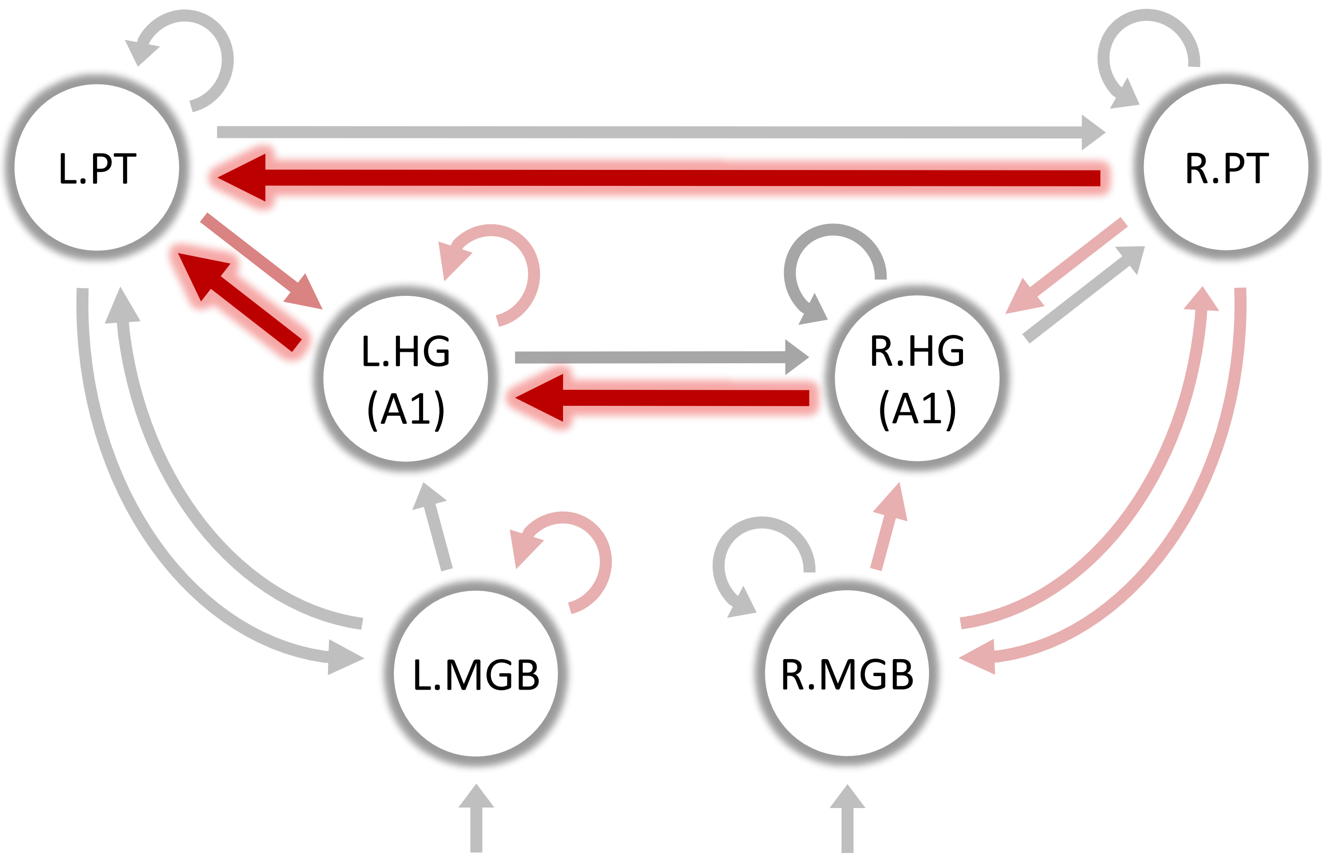
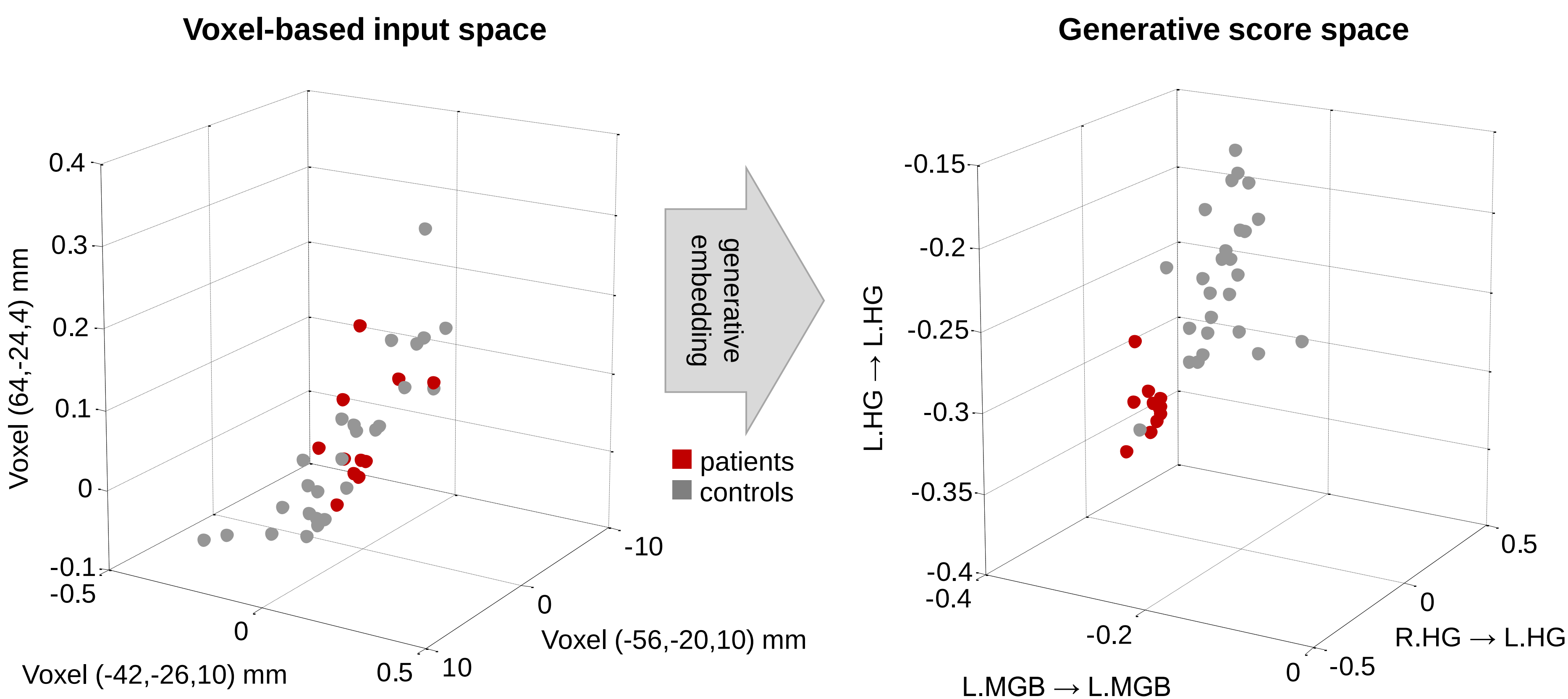


Illustration of the generative score space



Summary

1. **Strong classification performance**

   Generative embedding exploits the rich discriminative information encoded in ‘hidden’ quantities, such as coupling parameters. It may therefore outperform conventional schemes.

2. **Creation of a low-dimensional, interpretable feature space**

   The approach replaces high-dimensional fMRI data by a low-dimensional subject-specific fingerprint, where each dimension has a specific biological interpretation.

3. **Domains of application**

   Generative embedding can be used both for trial-by-trial decoding (EEG, MEG, or LFP data) and for subject-by-subject classification analyses (fMRI data).