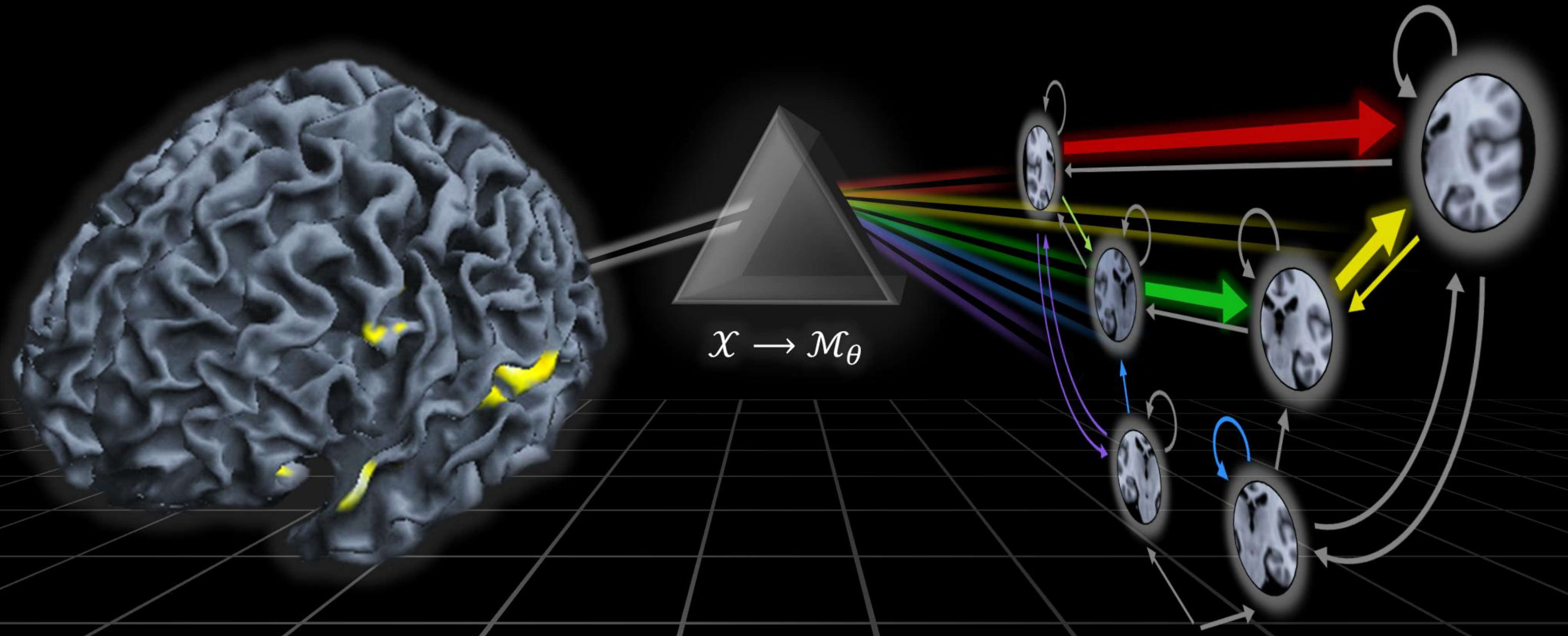
Generative embedding for fMRI

Kay H. Brodersen[1,2]

[1] Department of Computer Science, ETH Zurich, Switzerland
[2] Department of Economics, University of Zurich, Switzerland



$\mathcal{X} \longrightarrow \mathcal{M}_\theta$

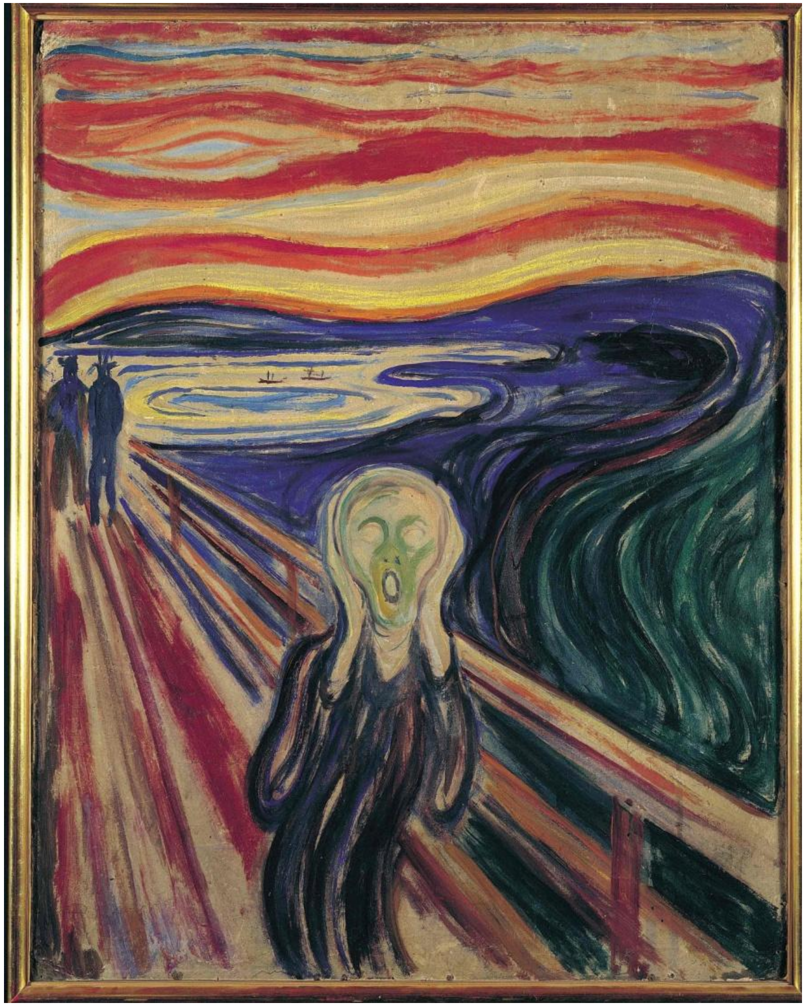
ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

University of Zurich

Psychiatric spectrum diseases

Schizophrenia, depression, mania, etc.

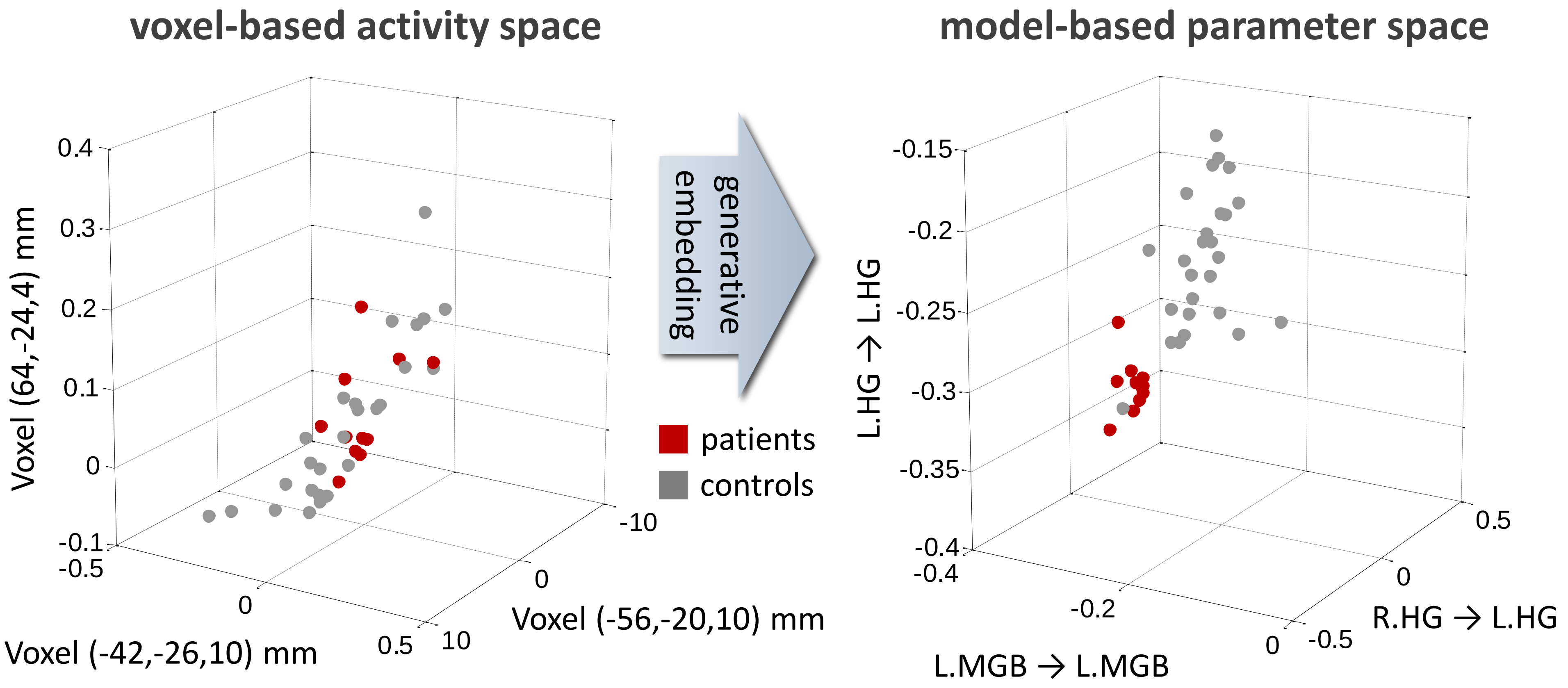
- diverse genetic basis, strong gene-environment interactions
⇨ genetically based diagnoses impossible
- multiple pathophysiological mechanisms
⇨ even when symptoms are similar, causes can differ across patients
- variability in treatment response and outcome



Consequences?

- need to infer on pathophysiological mechanisms in individual patients!

Dissecting diseases into physiologically defined subgroups



voxel-based activity space

generative embedding

model-based parameter space

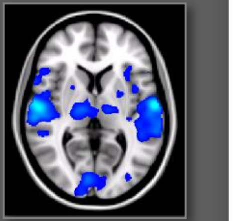
Brodersen et al. (2011) *PLoS Comput Biol*

Classification approaches by data representation

Structure-based analyses

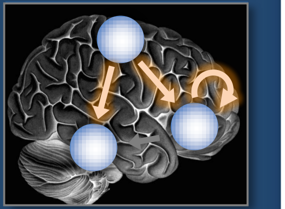
Which anatomical structures allow us to separate patients and healthy controls?

Activation-based analyses



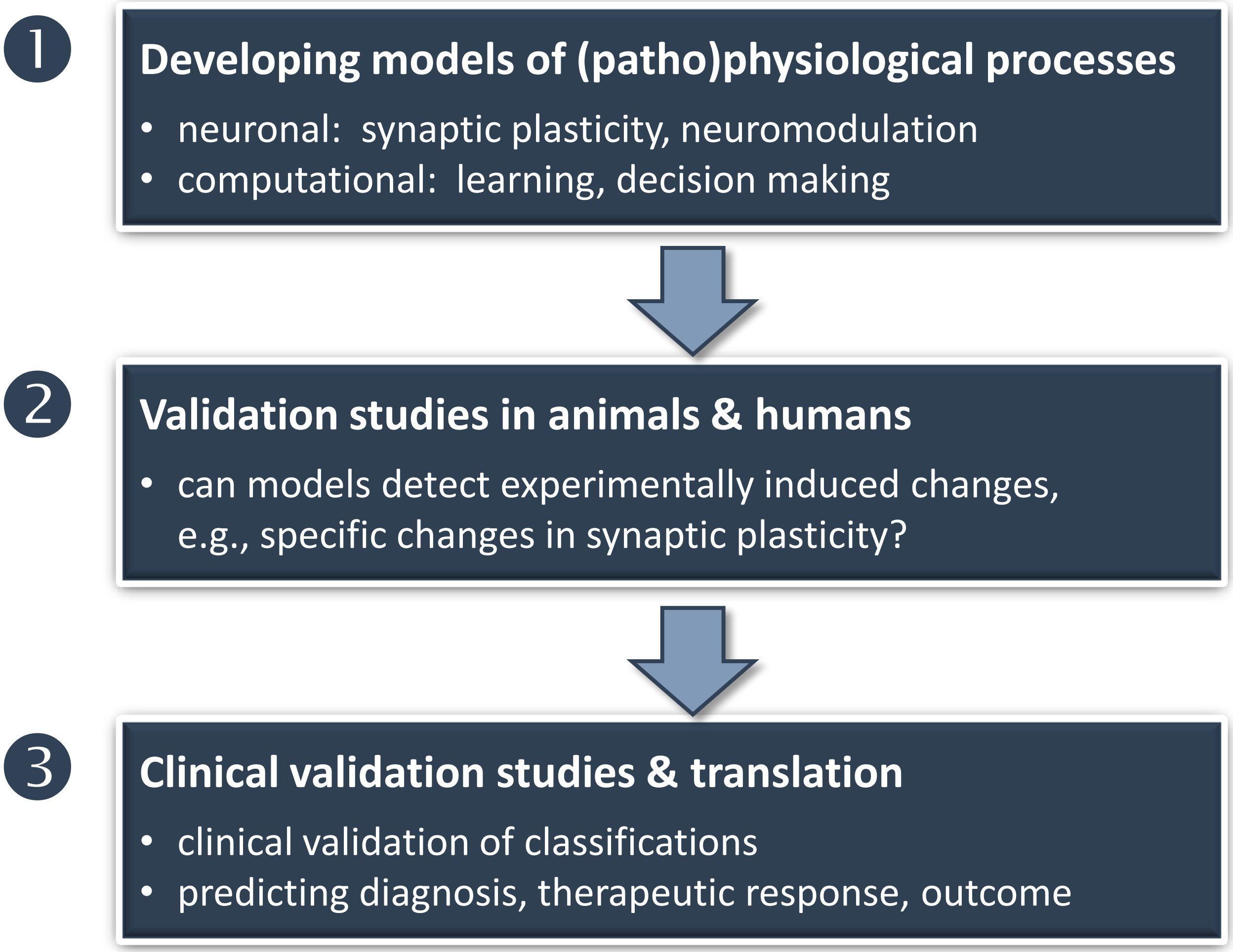
Which functional differences allow us to separate groups?

Model-based analyses



How do patterns of hidden quantities (e.g., connectivity among brain regions) differ between groups?

From models of pathophysiology to clinical applications



1

Developing models of (patho)physiological processes

- neuronal: synaptic plasticity, neuromodulation
- computational: learning, decision making

2

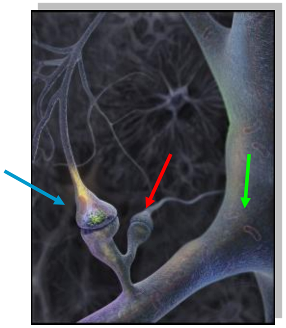
Validation studies in animals & humans

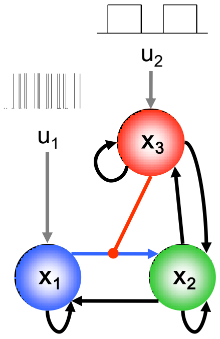
- can models detect experimentally induced changes, e.g., specific changes in synaptic plasticity?

3

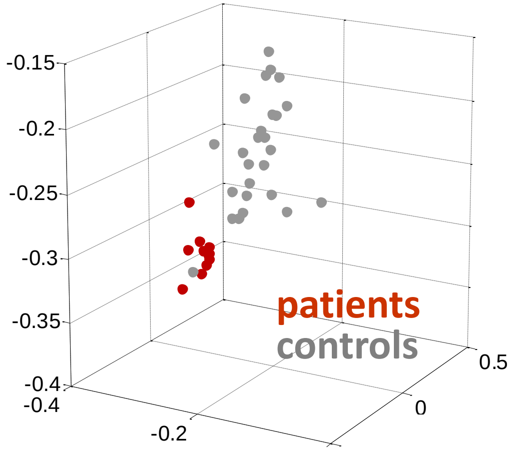
Clinical validation studies & translation

- clinical validation of classifications
- predicting diagnosis, therapeutic response, outcome





$$\frac{dx}{dt} = \left(A + \sum_{i=1}^{m} u_i B^{(i)} + \sum_{j=1}^{n} x_j D^{(j)}\right) x + Cu$$



From models of pathophysiology to clinical applications

1

Developing models of (patho)physiological processes

- neuronal: synaptic plasticity, neuromodulation
- computational: learning, decision making

2

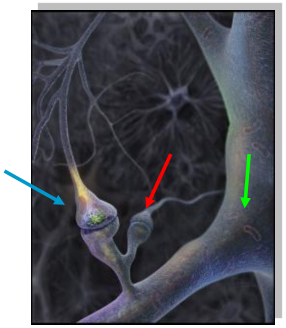
Validation studies in animals & humans

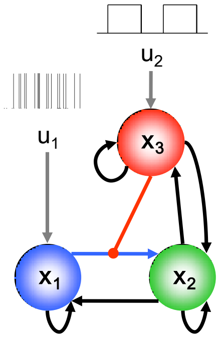
- can models detect experimentally induced changes, e.g., specific changes in synaptic plasticity?

3

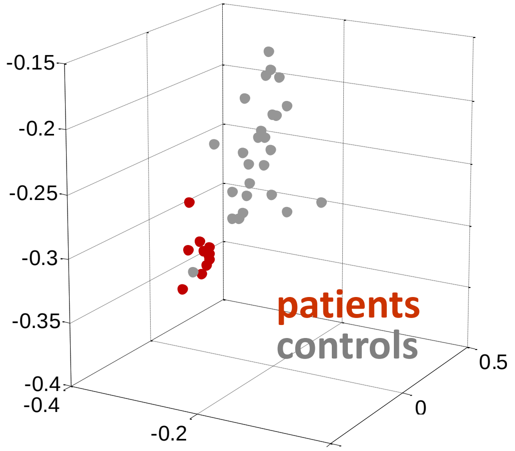
Clinical validation studies & translation

- clinical validation of classifications
- predicting diagnosis, therapeutic response, outcome

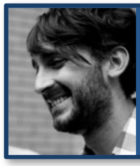




$$\frac{dx}{dt} = \left(A + \sum_{i=1}^{m} u_i B^{(i)} + \sum_{j=1}^{n} x_j D^{(j)} \right) x + Cu$$



Colleagues & collaborators



Thomas Schofield
University College London

Cheng Soon Ong
ETH Zurich

Kate Lomakina
University of Zurich · ETH Zurich

Alexander Leff
University College London

Christoph Mathys
University of Zurich · ETH Zurich

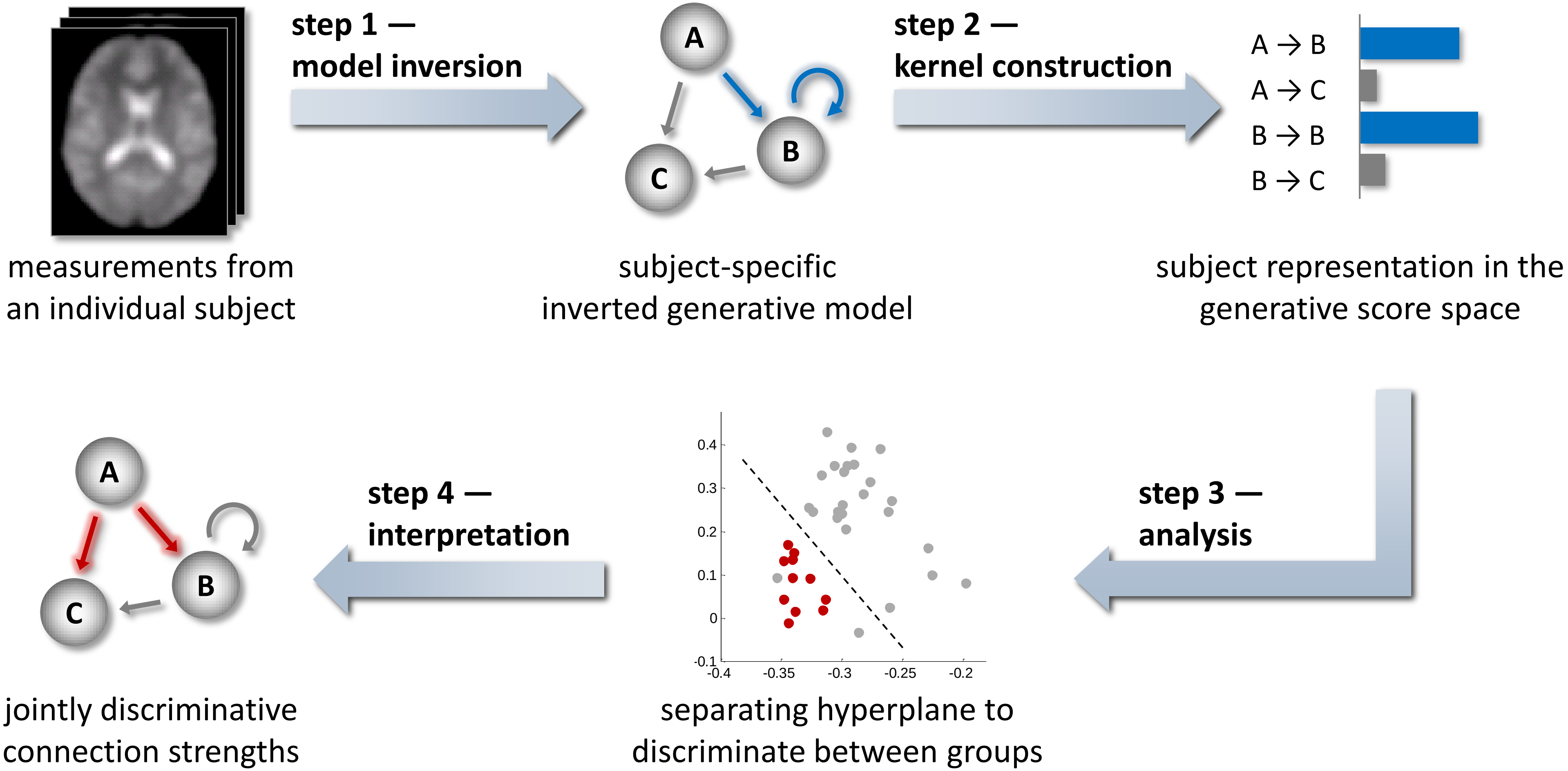
Justin R Chumbley
University of Zurich

Jean Daunizeau
University of Zurich · University College London

Joachim M Buhmann
ETH Zurich

Klaas Enno Stephan
University of Zurich · University College London

Model-based analysis by generative embedding



step 1 — model inversion

step 2 — kernel construction

step 3 — analysis

step 4 — interpretation

measurements from an individual subject

subject-specific inverted generative model

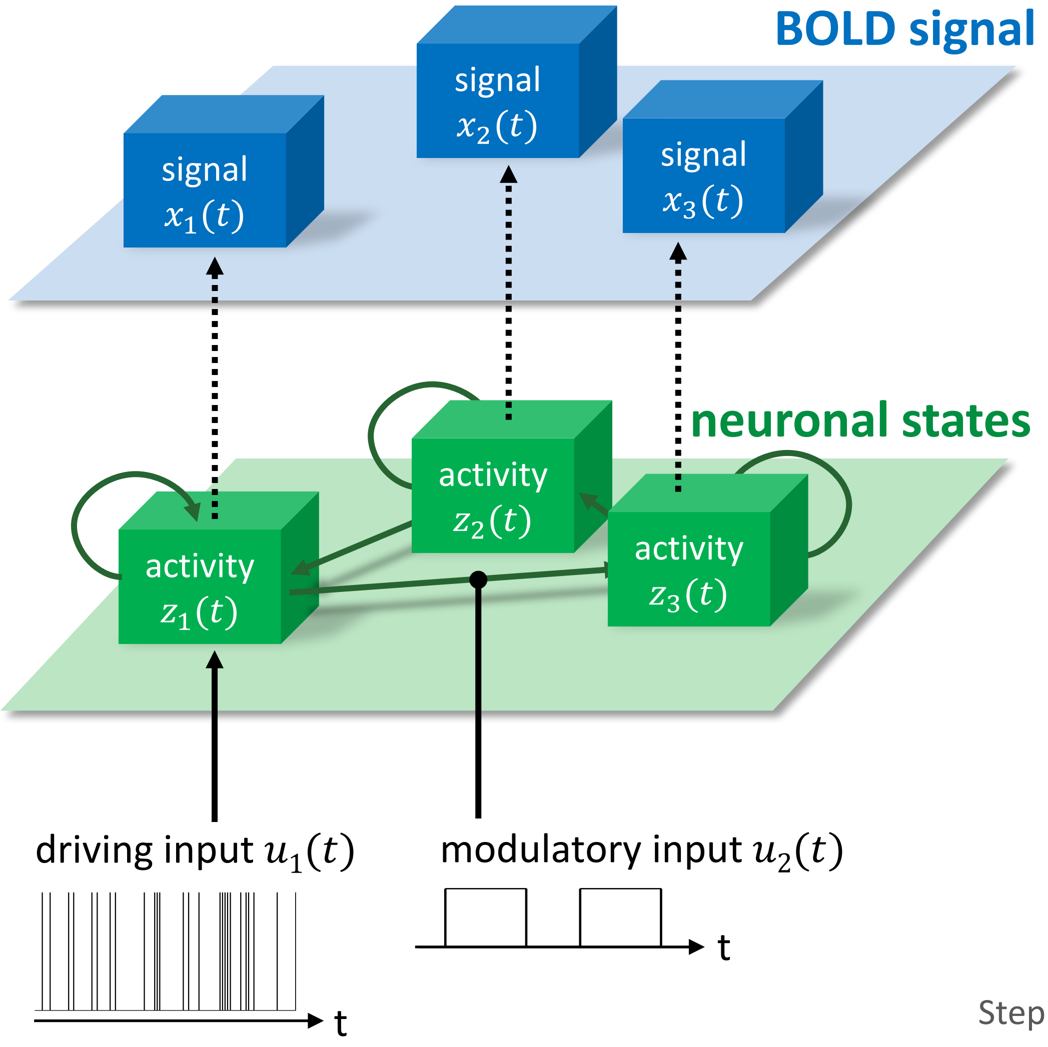
subject representation in the generative score space

separating hyperplane to discriminate between groups

jointly discriminative connection strengths

Brodersen et al. (2011) *NeuroImage;* Brodersen et al. (2011) *PLoS Comput Biol*

Choosing a generative model: DCM for fMRI



haemodynamic forward model

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

neural state equation

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

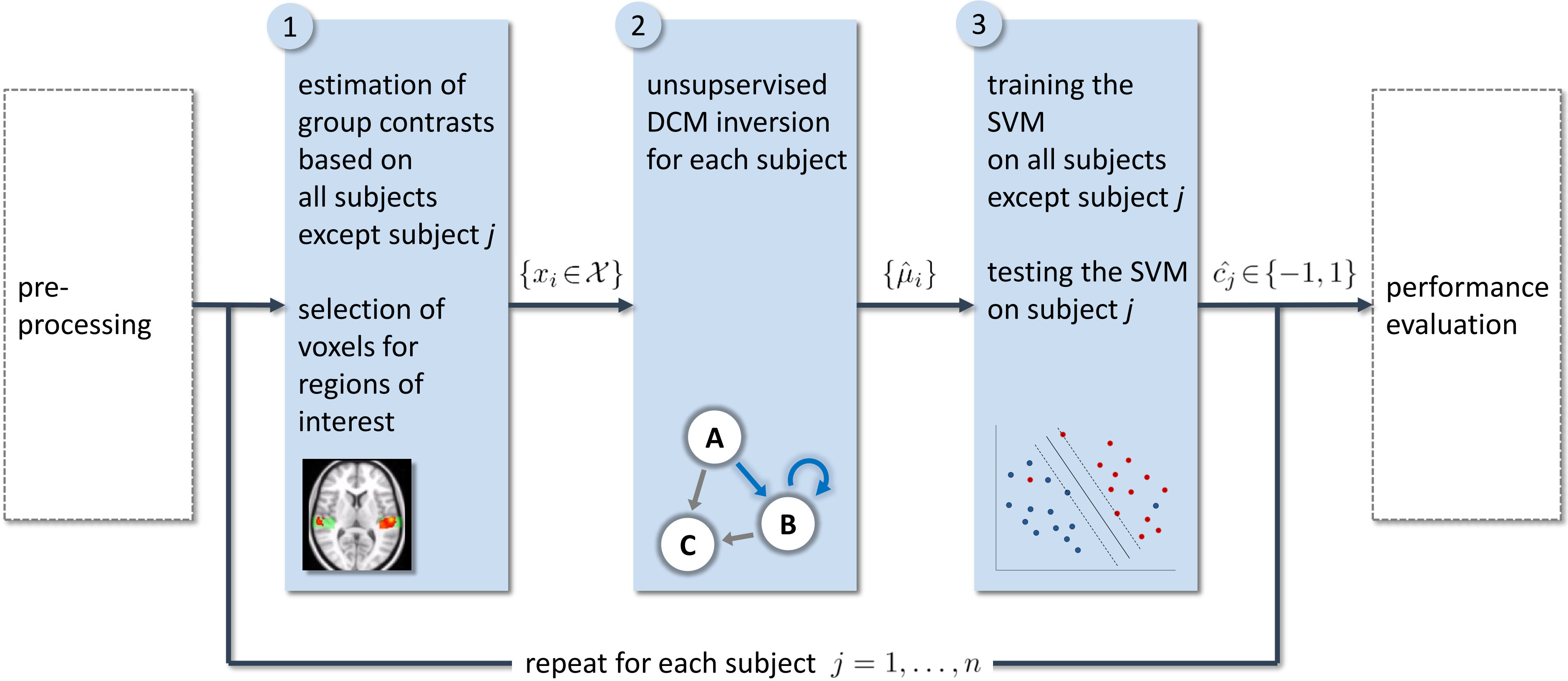
intrinsic connectivity

modulation of connectivity

direct inputs

Friston, Harrison & Penny (2003) *NeuroImage*
Stephan & Friston (2007) *Handbook of Brain Connectivity*

Summary of the analysis



pre-processing

1

estimation of group contrasts based on all subjects except subject *j*

selection of voxels for regions of interest

$\{x_i \in \mathcal{X}\}$

2

unsupservised DCM inversion for each subject

$\{\hat{\mu}_i\}$

3

training the SVM on all subjects except subject *j*

testing the SVM on subject *j*

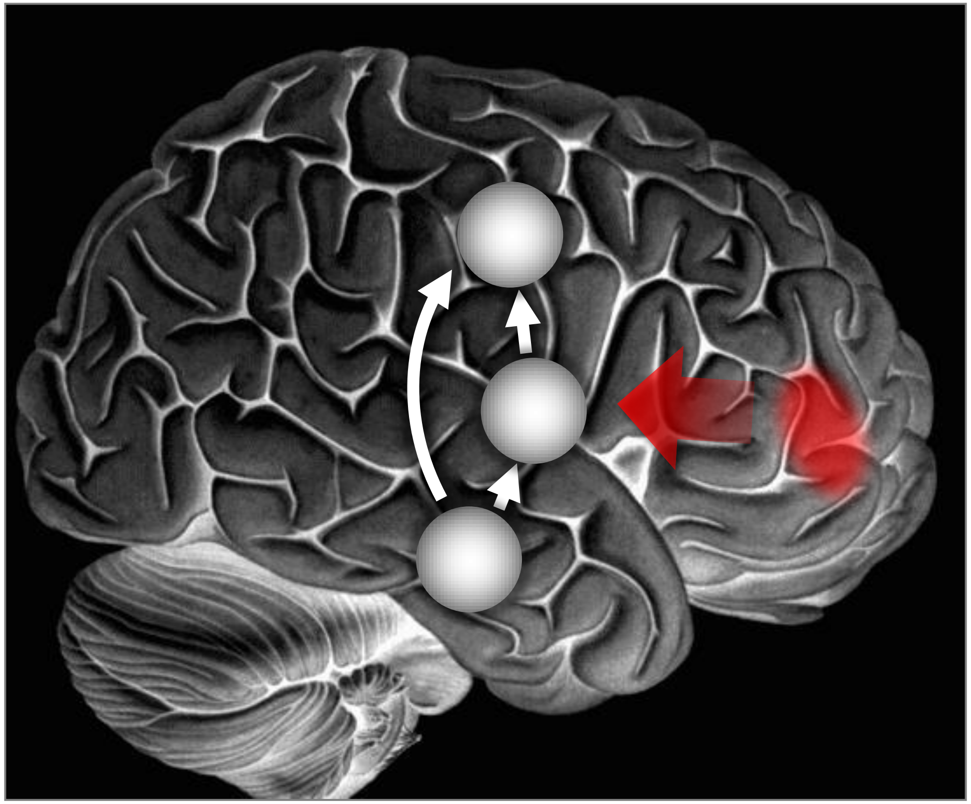
$\hat{c}_j \in \{-1, 1\}$

performance evaluation

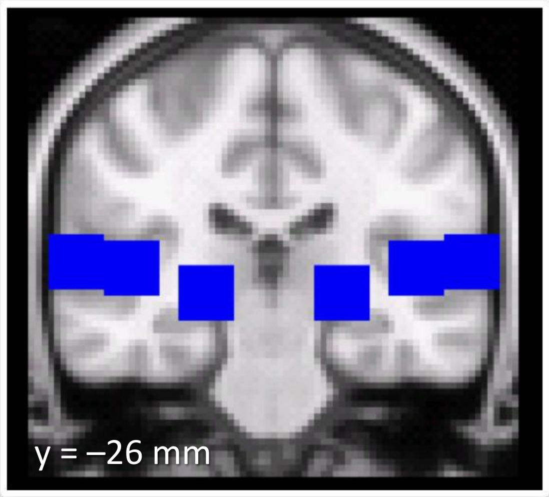
repeat for each subject $j = 1, \ldots, n$

Example: diagnosing stroke patients

To illustrate our approach, we aimed to distinguish between stroke patients and healthy controls, based on non-lesioned regions involved in speech processing.

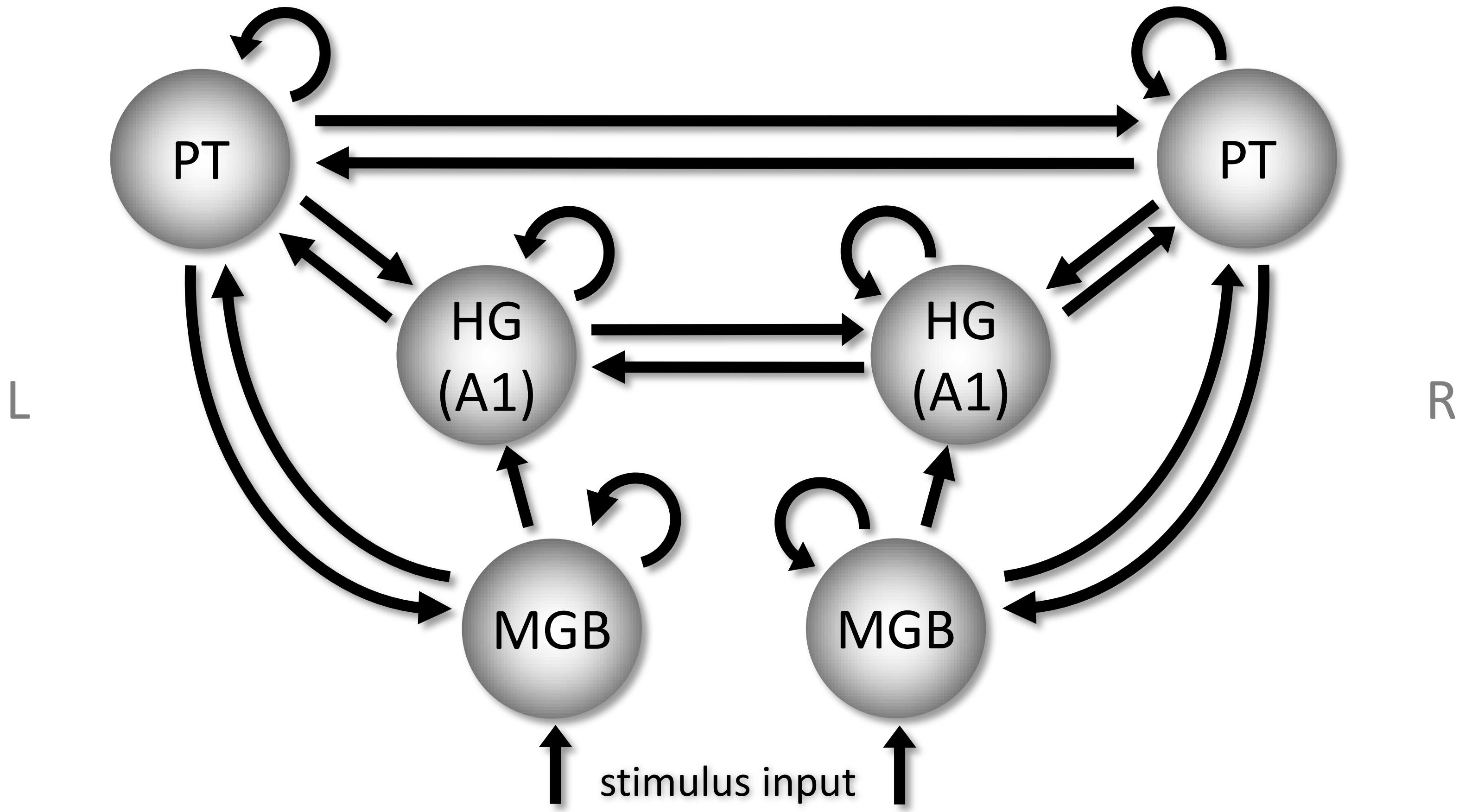


Example: diagnosing stroke patients

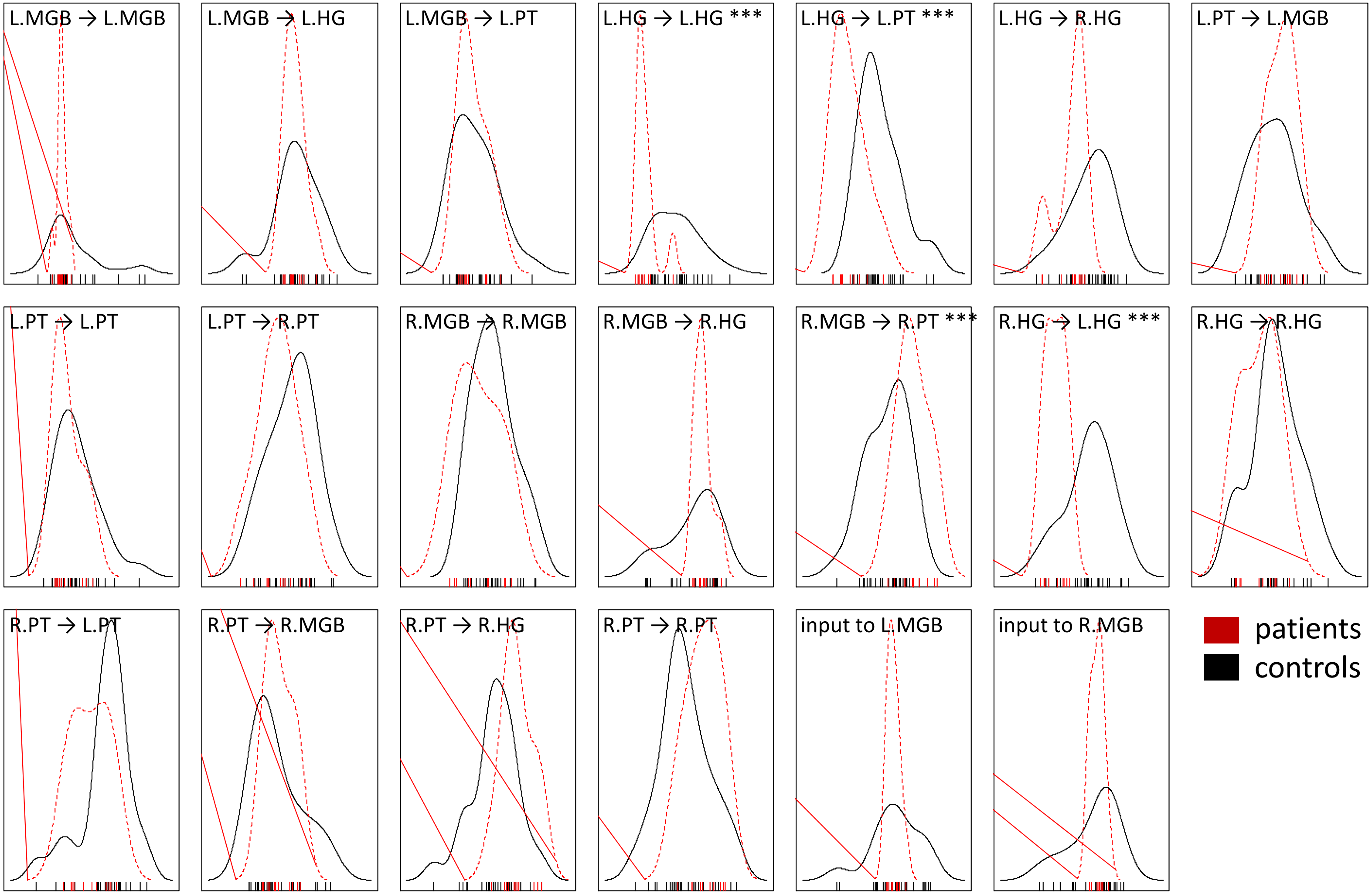


🟦 anatomical regions of interest

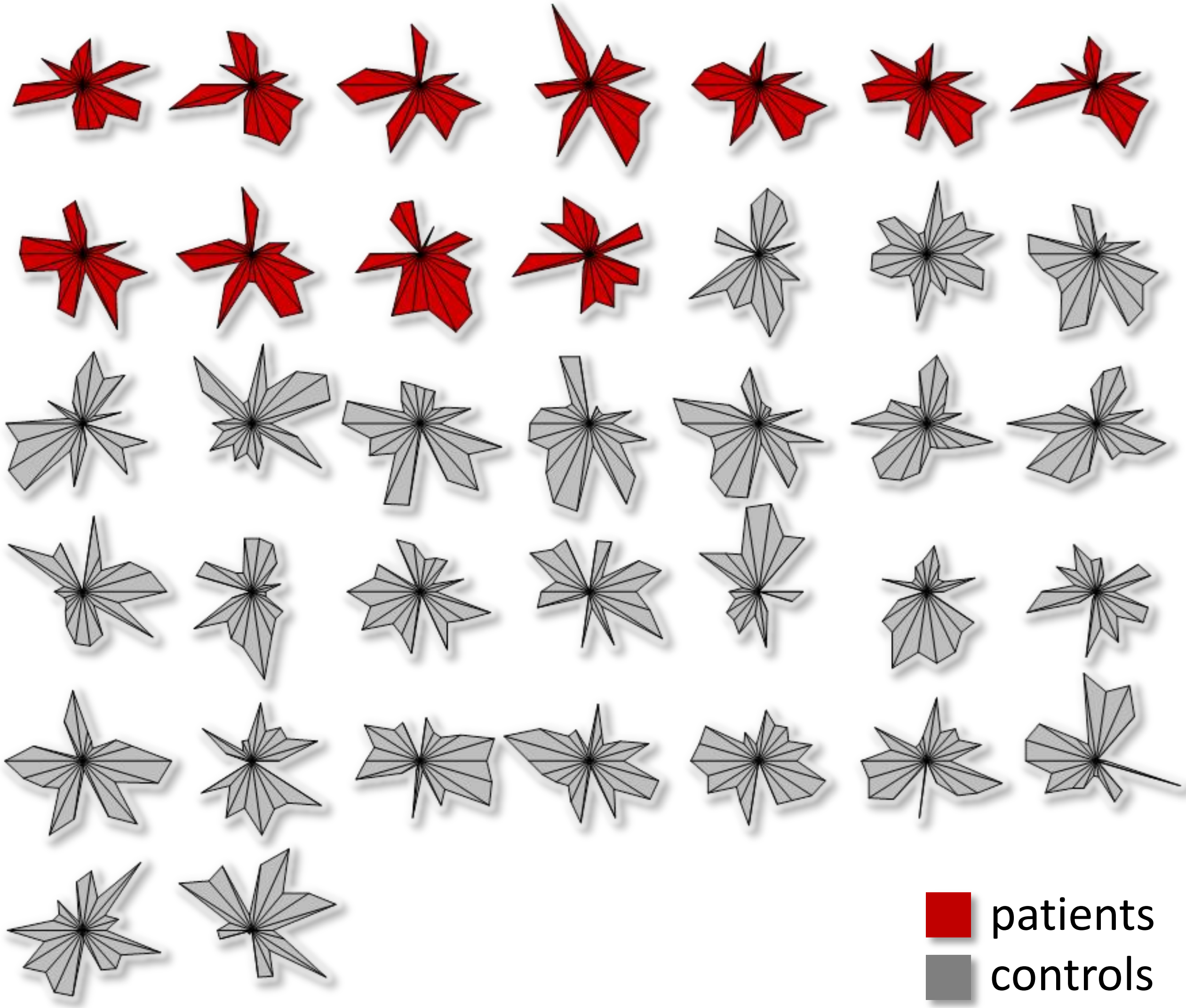
Example: diagnosing stroke patients



Univariate analysis: parameter densities

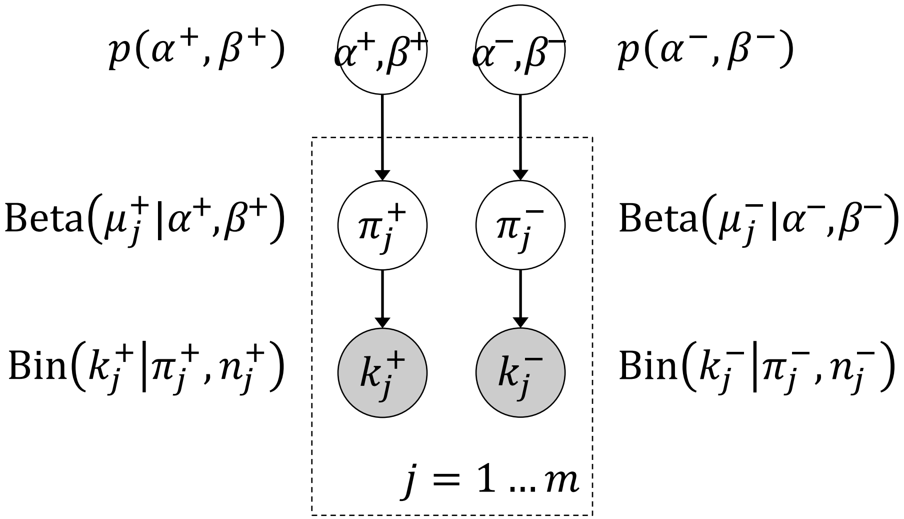


Multivariate analysis: connectional fingerprints



Full Bayesian approach to performance evaluation

Full Bayesian
mixed-effects inference
(beta-binomial model)



$p(\alpha^+,\beta^+)$

$p(\alpha^-,\beta^-)$

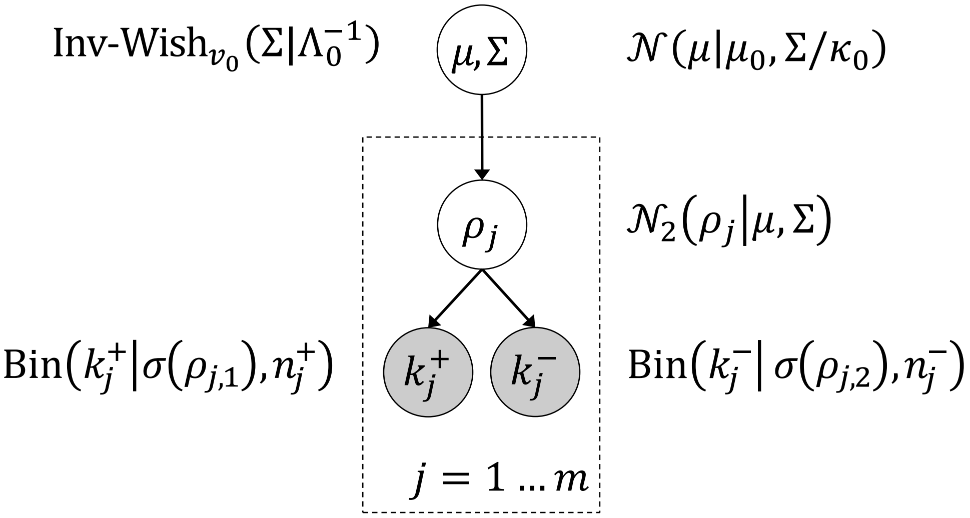
$\mathrm{Beta}\left(\mu_j^+|\alpha^+,\beta^+\right)$

$\mathrm{Beta}\left(\mu_j^-|\alpha^-,\beta^-\right)$

$\mathrm{Bin}\left(k_j^+|\pi_j^+,n_j^+\right)$

$\mathrm{Bin}\left(k_j^-|\pi_j^-,n_j^-\right)$

Full Bayesian
mixed-effects inference
(normal-binomial model)



$\text{Inv-Wish}_{\nu_0}(\Sigma|\Lambda_0^{-1})$

$\mathcal{N}(\mu|\mu_0,\Sigma/\kappa_0)$

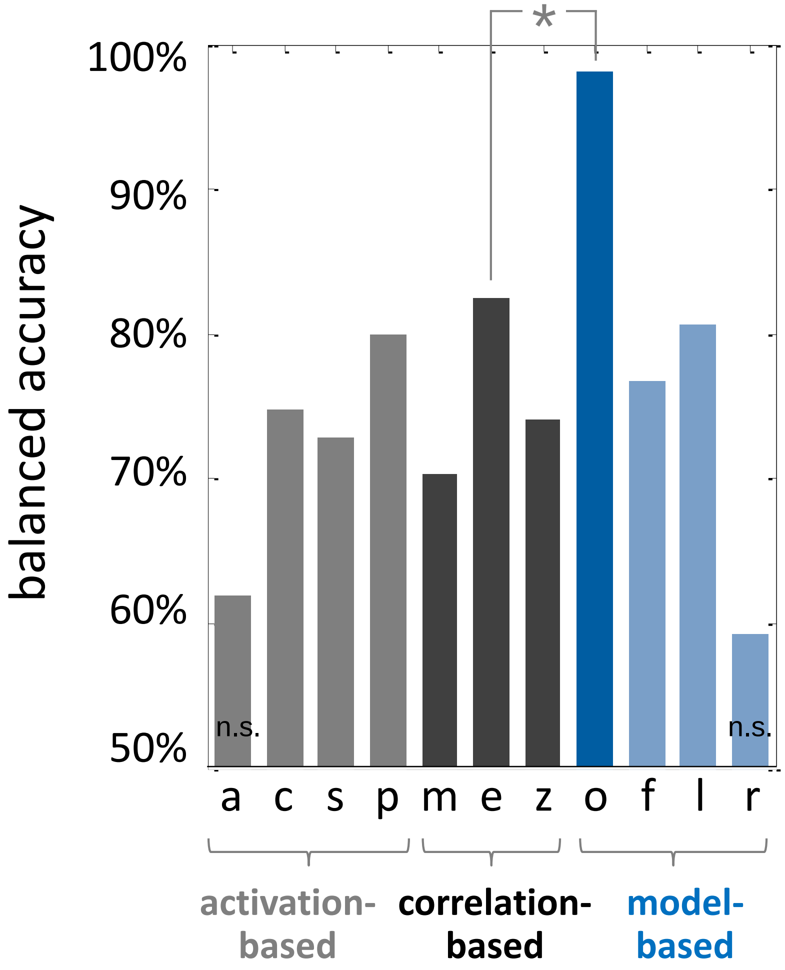
$\mathcal{N}_2\left(\rho_j|\mu,\Sigma\right)$

$\mathrm{Bin}\left(k_j^+|\sigma\left(\rho_{j,1}\right),n_j^+\right)$

$\mathrm{Bin}\left(k_j^-|\,\sigma\left(\rho_{j,2}\right),n_j^-\right)$

Brodersen et al. *(in preparation)*

Classification performance



Activation-based analyses

a anatomical feature selection

c mass-univariate contrast feature selection

s locally univariate searchlight feature selection

p PCA-based dimensionality reduction

Correlation-based analyses

m correlations of regional means

e correlations of regional eigenvariates

z Fisher-transformed eigenvariates correlations

Model-based analyses

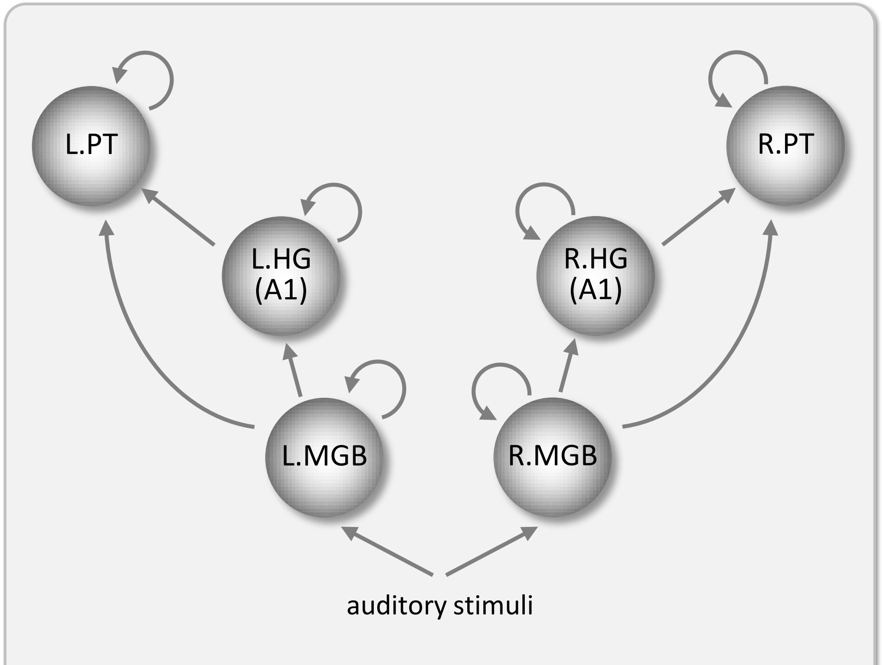
o gen.embed., original full model

f gen.embed., less plausible feedforward model

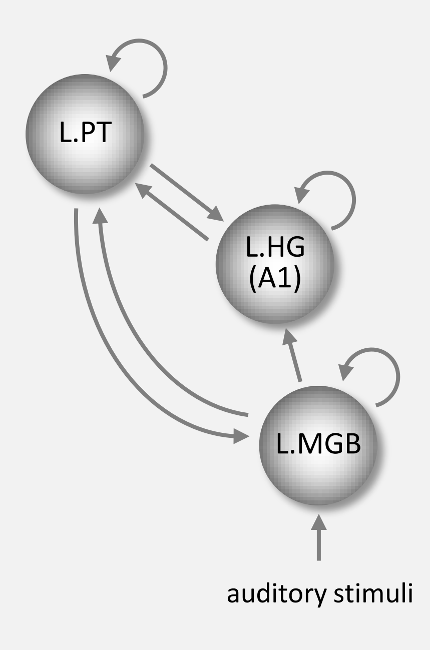
l gen.embed., left hemisphere only

r gen.embed., right hemisphere only

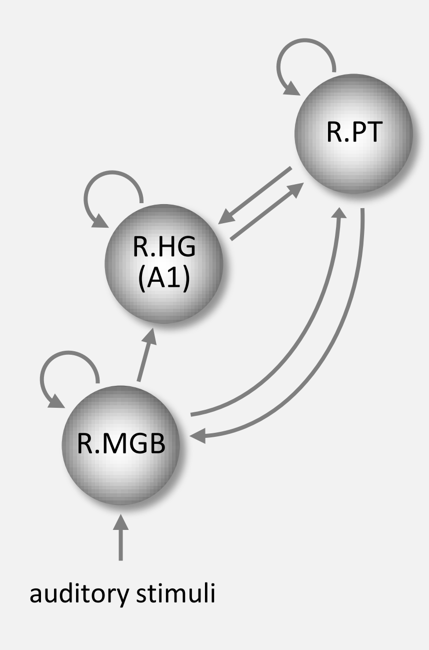
Biologically less plausible models perform poorly



feedforward connections only
accuracy: 77%



left hemisphere only
accuracy: 81%



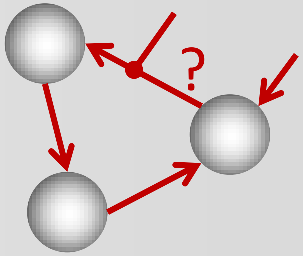
right hemisphere only
accuracy: 59%

Generative embedding and DCM

Question 1 – What do the data tell us about hidden processes in the brain?

⇨ compute the posterior

$$p(\theta|y,m) = \frac{p(y|\theta,m)p(\theta|m)}{p(y|m)}$$

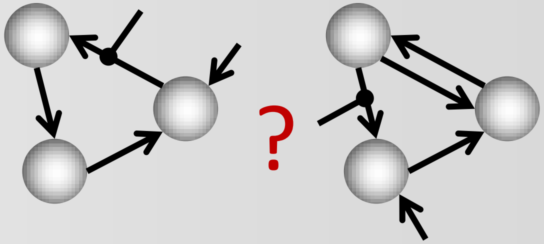


Question 2 – Which model is best w.r.t. the observed fMRI data?

⇨ compute the model evidence

$$p(m|y) \propto p(y|m)p(m)$$

$$= \int p(y|\theta,m)p(\theta|m)d\theta$$

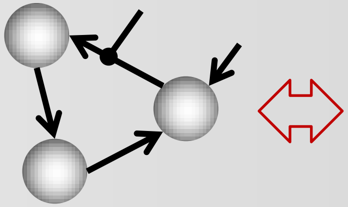


Question 3 – Which model is best w.r.t. an external criterion?

⇨ compute the classification accuracy

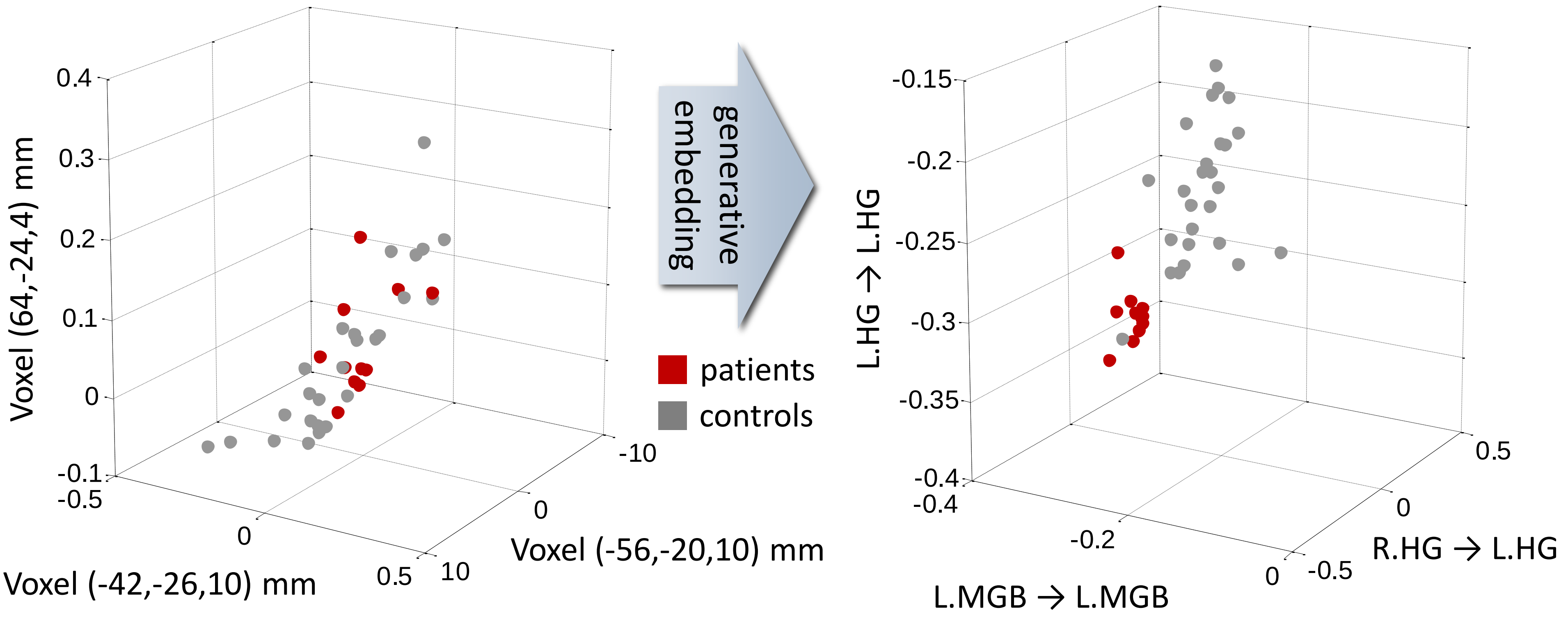
$$p(h(y) = x|y)$$

$$= \iiint p(h(y) = x|y, y_{\text{train}}, x_{\text{train}})\, p(y)\, p(y_{\text{train}})\, p(x_{\text{train}})\; dy\; dy_{\text{train}}\; x_{\text{train}}$$



**{ patient,
control }**

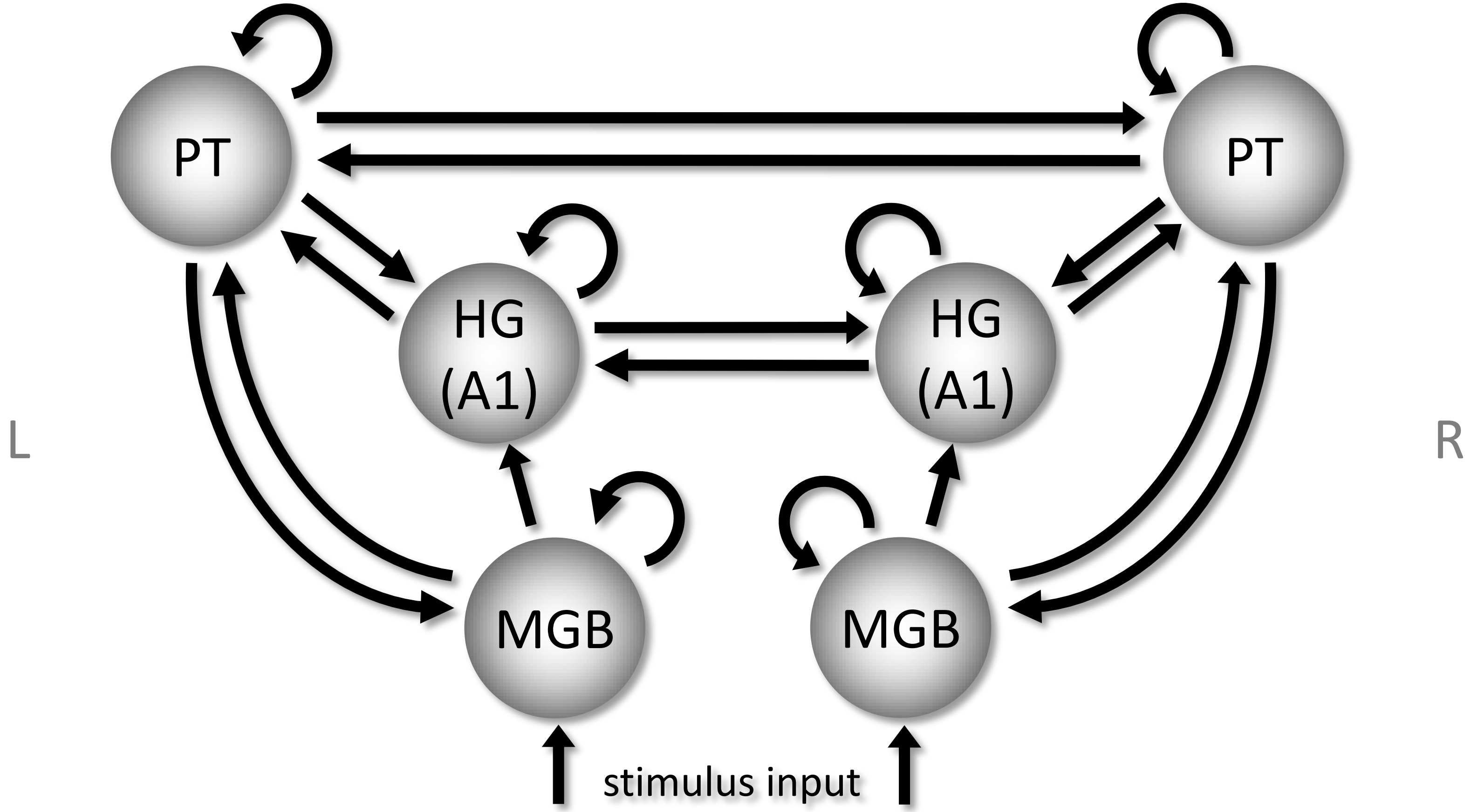
The generative projection



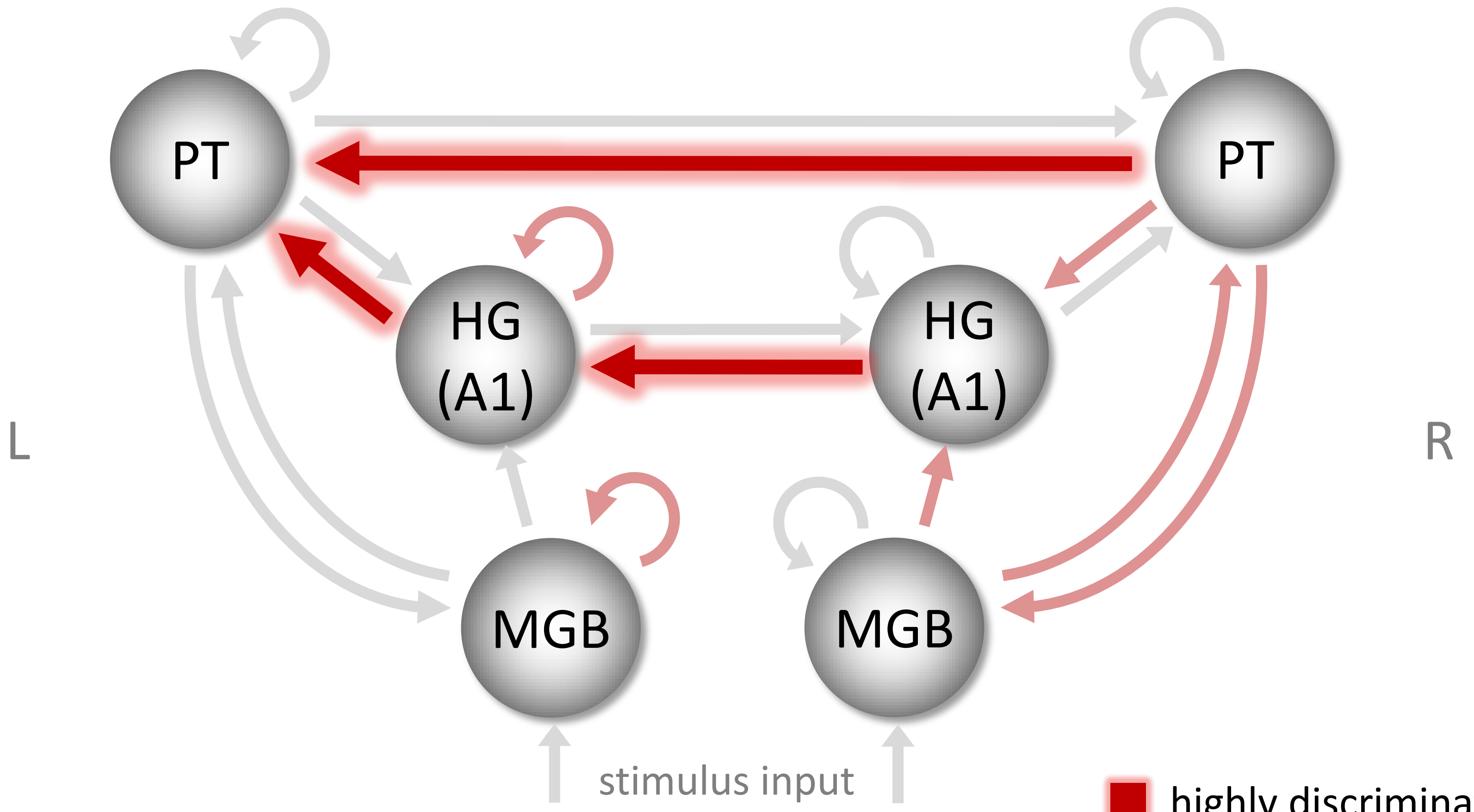
voxel-based activity space

model-based parameter space

Discriminative features in model space



Discriminative features in model space



PT

PT

HG (A1)

HG (A1)

MGB

MGB

L

R

stimulus input

🟥 highly discriminative

🟧 somewhat discriminative

⬜ not discriminative

Summary: generative embedding for fMRI

❶ **Strong classification performance.** Generative embedding exploits the rich discriminative information encoded in ‘hidden’ quantities, such as coupling parameters.

❷ **Creation of an interpretable feature space.** High-dimensional fMRI data are replaced by low-dimensional subject-specific fingerprints with biologically interpretable axes.

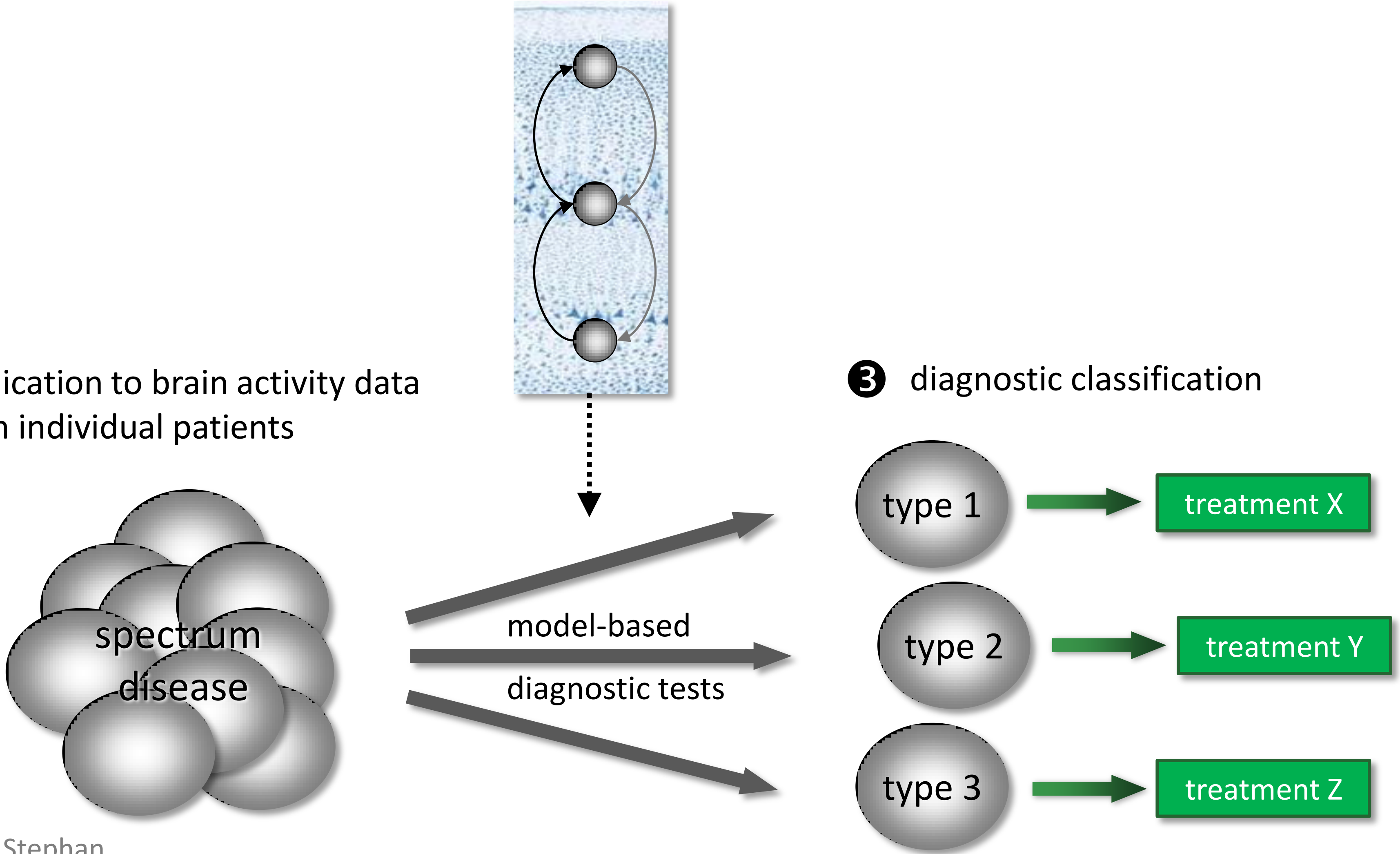
❸ **Future applications.** Generative embedding could help dissect spectrum disorders into physiologically defined subgroups.

Outlook: model-based inference on *individual* pathophysiology

❶ model of neuronal (patho)physiology

❷ application to brain activity data from individual patients

❸ diagnostic classification



spectrum disease

model-based

diagnostic tests

type 1

treatment X

type 2

treatment Y

type 3

treatment Z