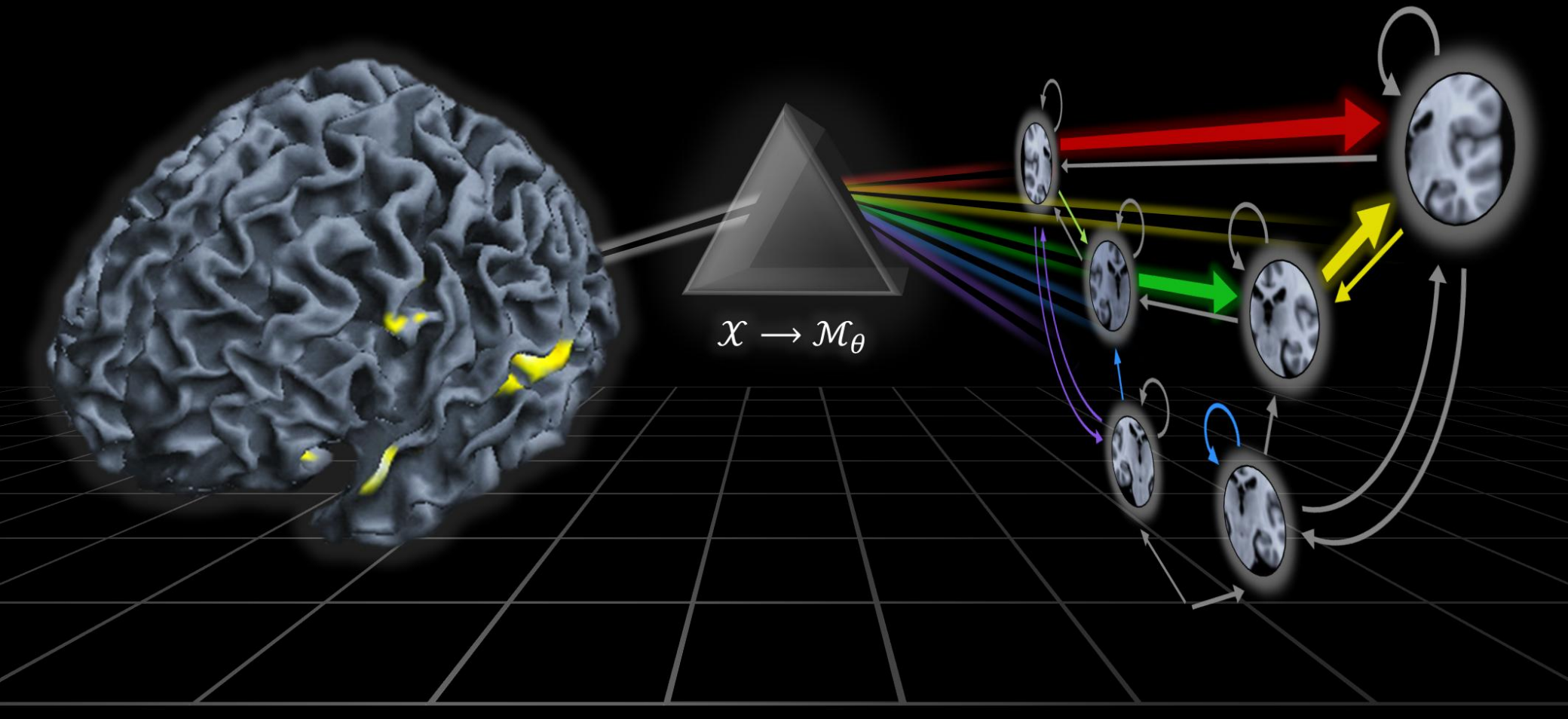
Model-based analysis of disease states of the brain using generative embedding

Kay H. Brodersen[1,2]

[1] Translational Neuromodeling Unit (TNU), Department of Biomedical Engineering, University of Zurich & ETH Zurich
[2] Machine Learning Laboratory, Department of Computer Science, ETH Zurich

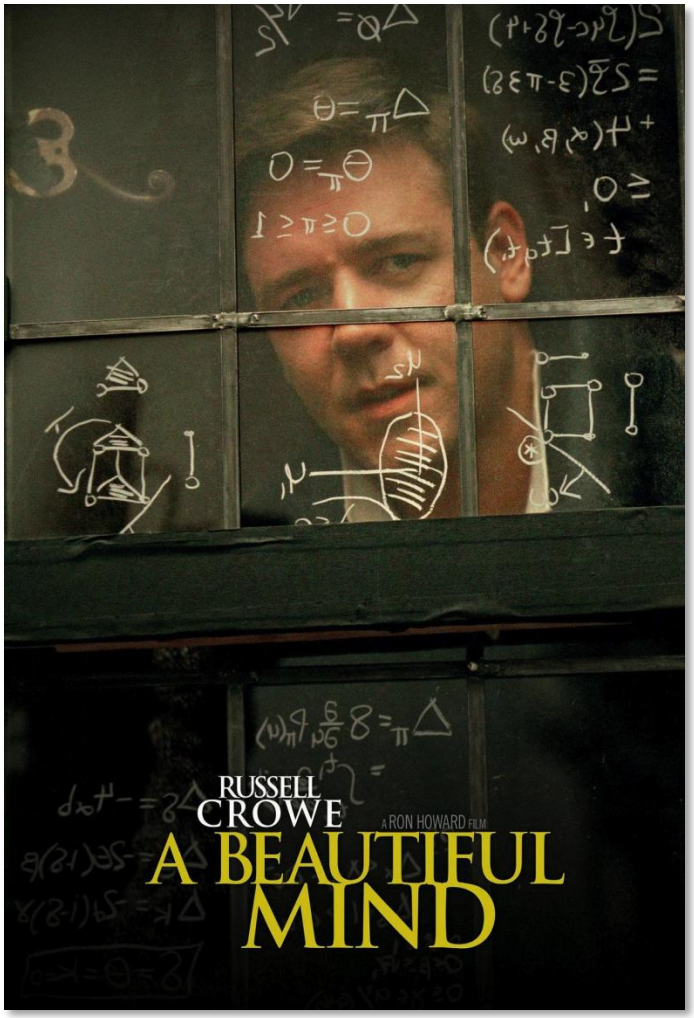


Psychiatric spectrum diseases

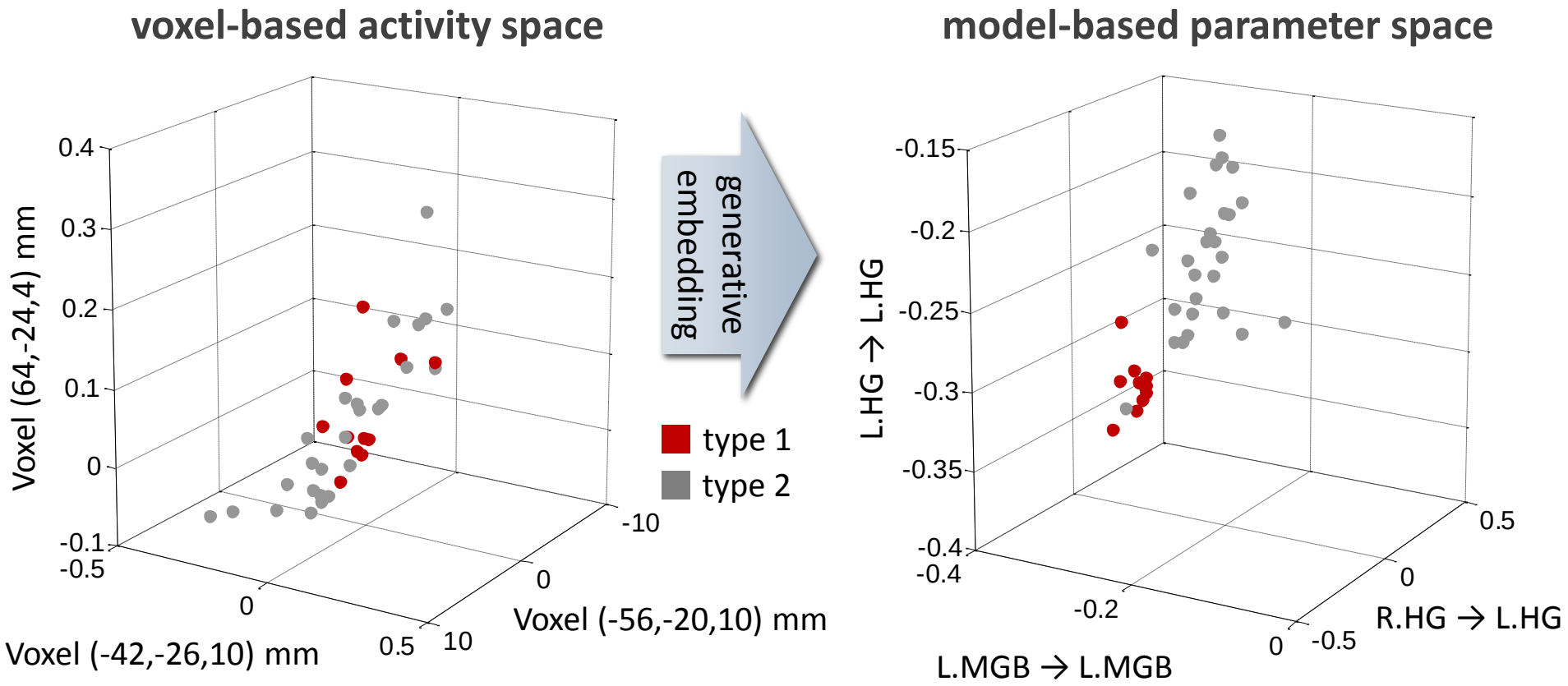
Schizophrenia, depression, mania, etc.

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

Consequences
need to infer on pathophysiological mechanisms in individual patients!



Dissecting diseases into physiologically defined subgroups



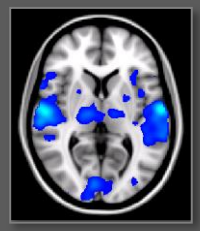
Brodersen et al. (2011) *PLoS Comput Biol*

Classification approaches by data representation

Model-based analyses

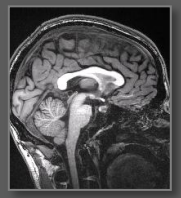
How do patterns of hidden quantities (e.g., connectivity among brain regions) differ between groups?

Activation-based analyses



Which functional differences allow us to separate groups?

Structure-based analyses



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

From models of pathophysiology to clinical applications

1

Developing models of (patho)physiological processes

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

2

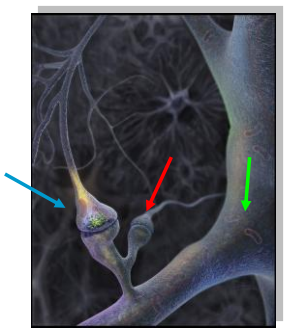
Validation studies in animals & humans

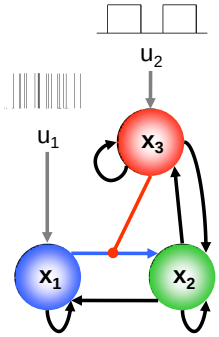
- can models detect experimentally induced changes, e.g., specific changes in synaptic plasticity?

3

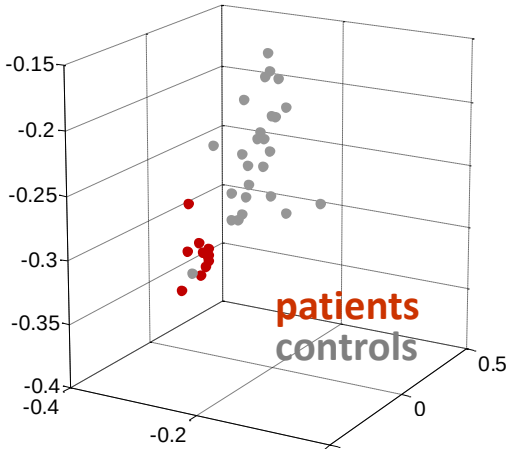
Clinical validation studies & translation

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





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



Klaas E. Stephan

Colleagues & collaborators

Thomas Schofield
University College London

Justin R Chumbley
University of Zurich

Cheng Soon Ong
National ICT Australia · University of Melbourne

Jean Daunizeau
ICM Paris · University College London

Kate Lomakina
University of Zurich · ETH Zurich

Joachim M Buhmann
ETH Zurich

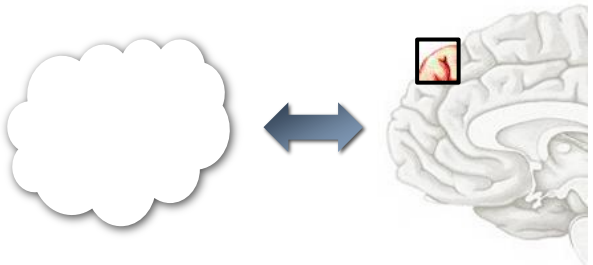
Alexander Leff
University College London (UCL)

Klaas Enno Stephan
University of Zurich · ETH Zurich · UCL

Christoph Mathys
University of Zurich · ETH Zurich

Univariate vs. multivariate models

A univariate model considers a single voxel at a time.

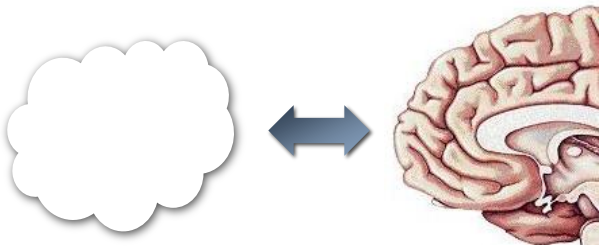


context
$X_t \in \mathbb{R}^d$

BOLD signal
$Y_t \in \mathbb{R}$

Spatial dependencies between voxels are only introduced afterwards, through random field theory.

A multivariate model considers many voxels at once.



context
$X_t \in \mathbb{R}^d$

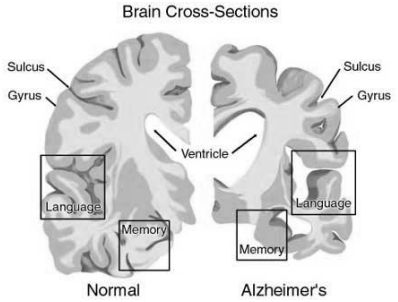
BOLD signal
$Y_t \in \mathbb{R}^v,\ \mathrm{v} \gg 1$

Multivariate models enable inferences on distributed responses without requiring focal activations.

Prediction vs. inference

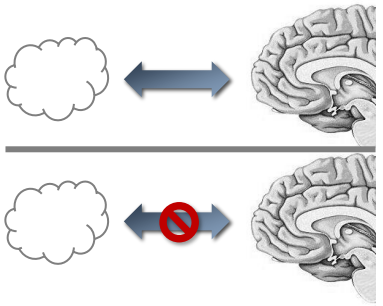
The goal of **prediction** is to find a highly accurate encoding or decoding function.

predicting a cognitive state using a brain-machine interface

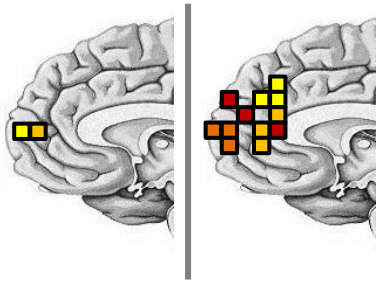


predicting a subject-specific diagnostic status

The goal of **inference** is to decide between competing hypotheses.



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



weighing the evidence for sparse vs. distributed coding

predictive density

$$p(X_{new}|Y_{new}, X, Y) = \int p(X_{new}|Y_{new}, \theta)p(\theta|X, Y)d\theta$$

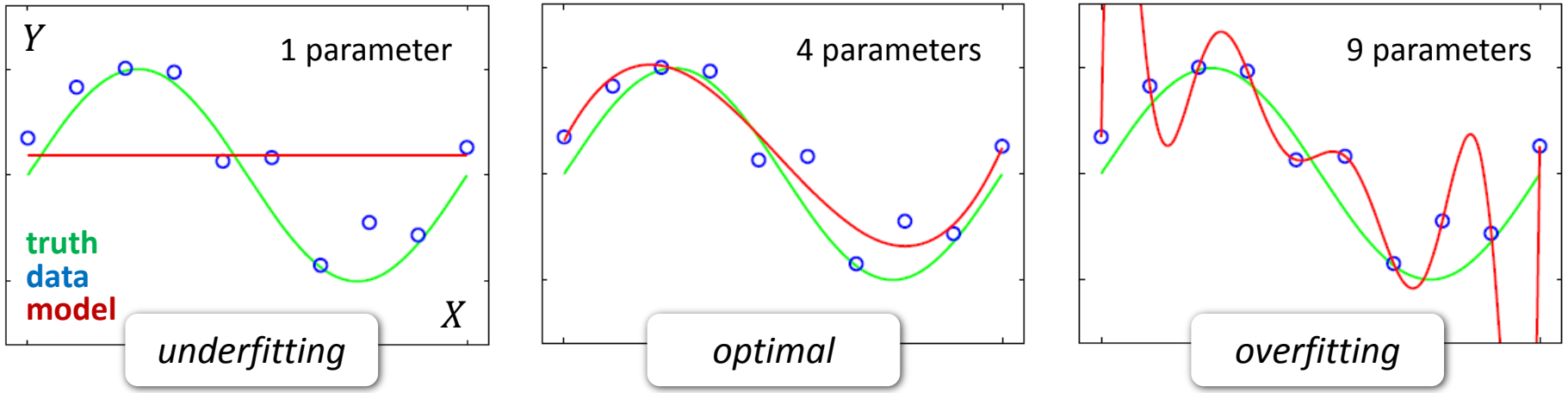
marginal likelihood (model evidence)

$$p(X|Y) = \int p(X|Y, \theta)p(\theta)d\theta$$

Goodness of fit vs. complexity

Goodness of fit is the degree to which a model explains observed data.

Complexity is the flexibility of a model (including, but not limited to, its number of parameters).



We wish to find the model that optimally trades off goodness of fit and complexity.

Bishop (2007) *PRML*

Constructing a classifier

A principled way of designing a classifier would be to adopt a probabilistic approach:

$Y_t \xrightarrow{\ f\ }$ that k which maximizes $p(X_t = k|Y_t, X, Y)$

In practice, classifiers differ in terms of how strictly they implement this principle.

Generative classifiers

use Bayes' rule to estimate $p(X_t|Y_t) \propto p(Y_t|X_t)p(X_t)$

- *Gaussian Naïve Bayes*
- *Linear Discriminant Analysis*

Discriminative classifiers

estimate $p(X_t|Y_t)$ directly without Bayes' theorem

- *Logistic regression*
- *Relevance Vector Machine*

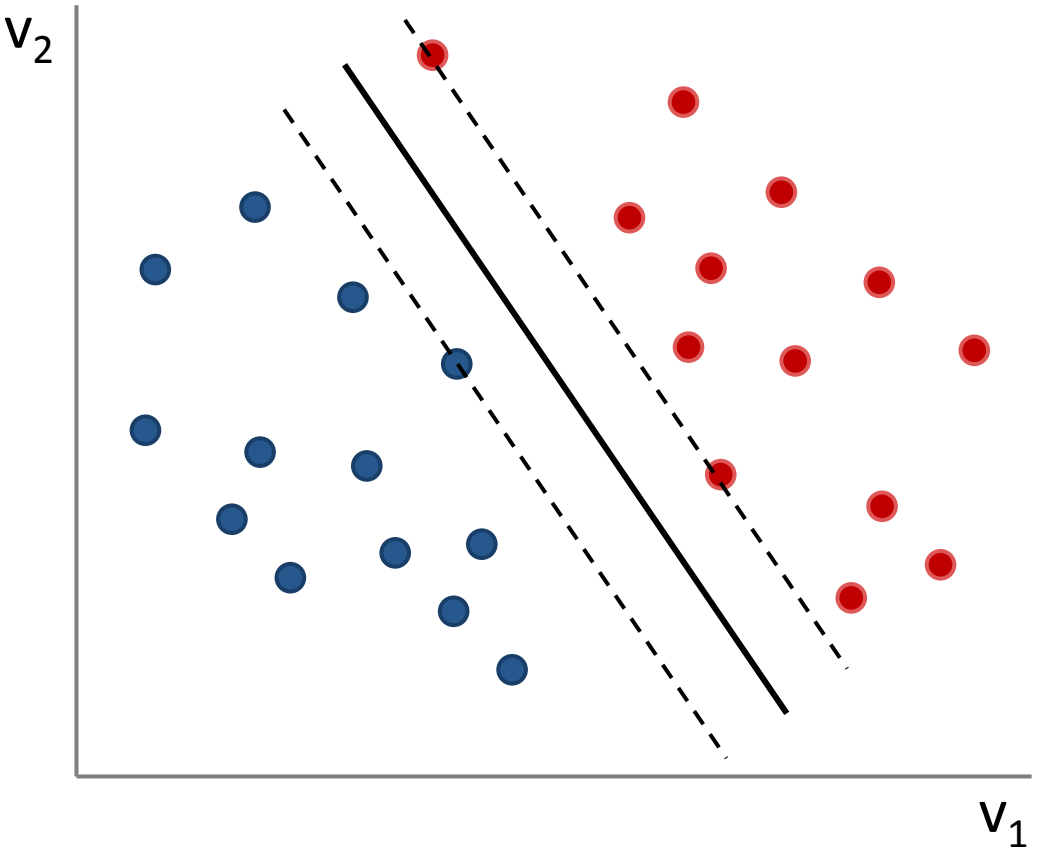
Discriminant classifiers

estimate $f(Y_t)$ directly

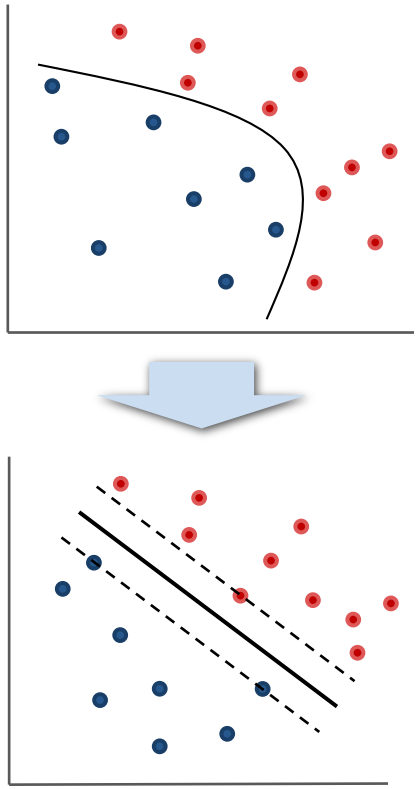
- *Fisher's Linear Discriminant*
- *Support Vector Machine*

Support vector machine (SVM)

Linear SVM

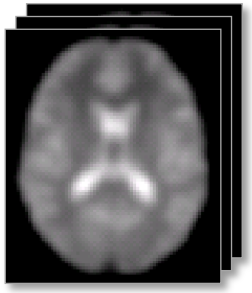


Nonlinear SVM



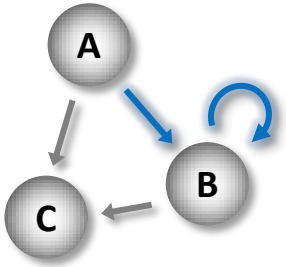
Vapnik (1999) Springer; Schölkopf et al. (2002) MIT Press

Model-based analysis by generative embedding



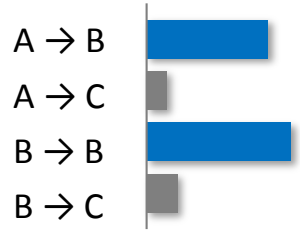
measurements from
an individual subject

**step 1 —
model inversion**



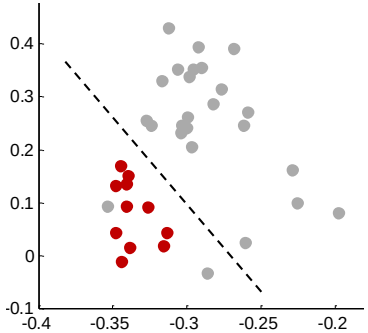
subject-specific
inverted generative model

**step 2 —
kernel construction**



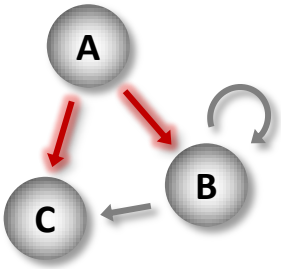
subject representation in the
generative score space

**step 3 —
analysis**



separating hyperplane to
discriminate between groups

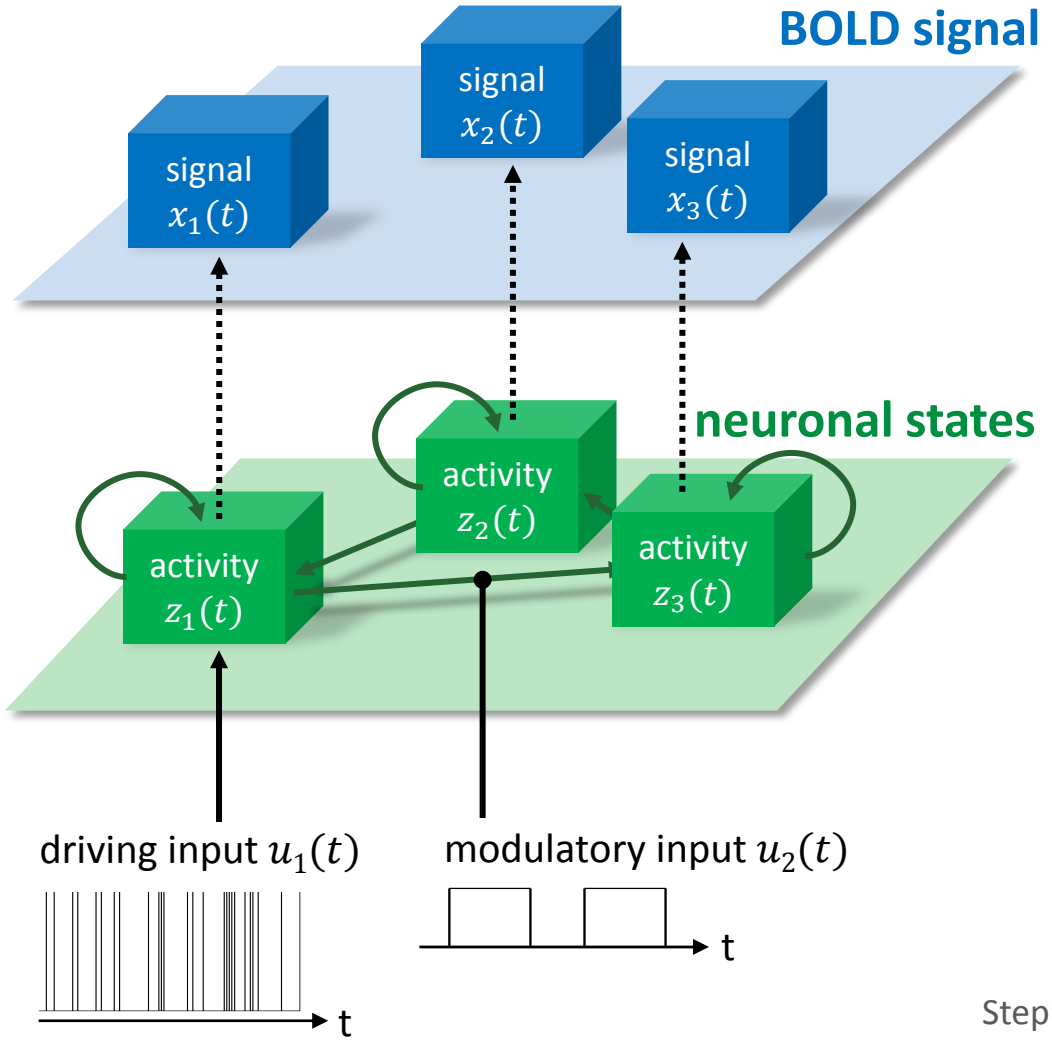
**step 4 —
interpretation**



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

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

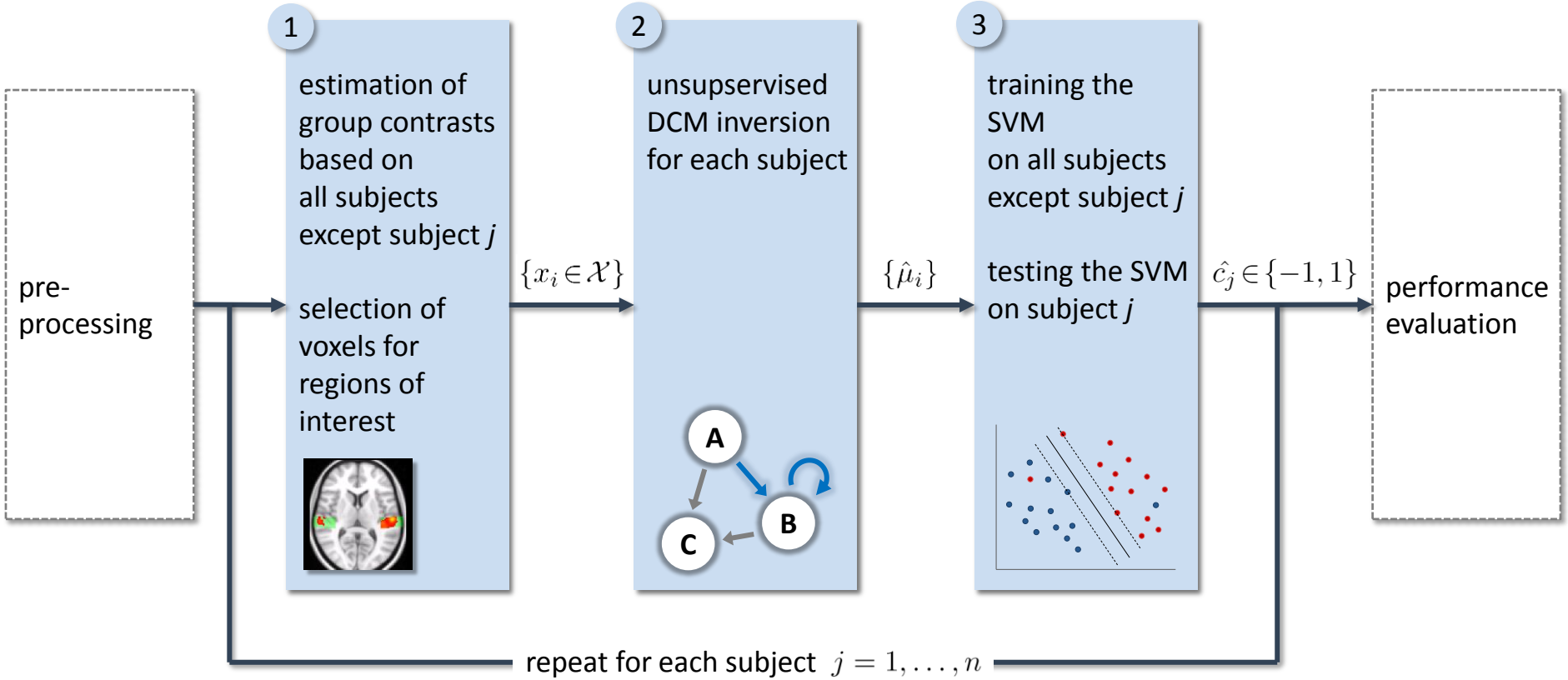
intrinsic connectivity

modulation of connectivity

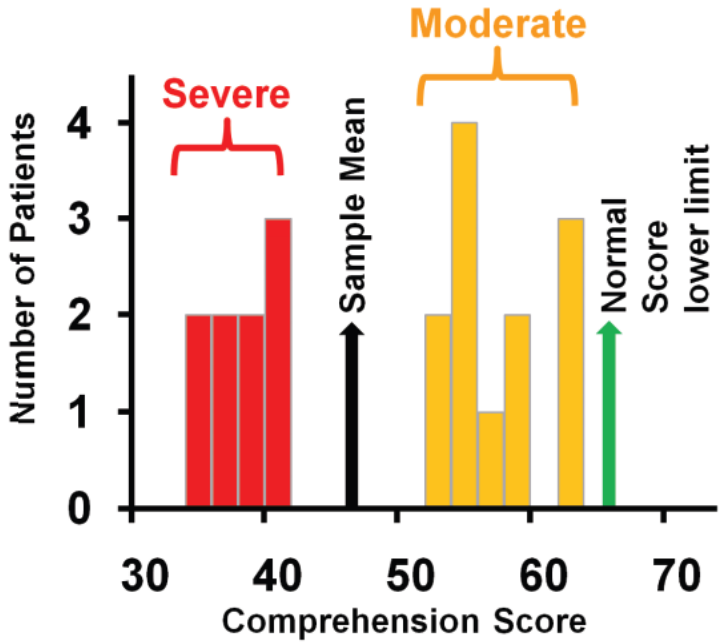
direct inputs

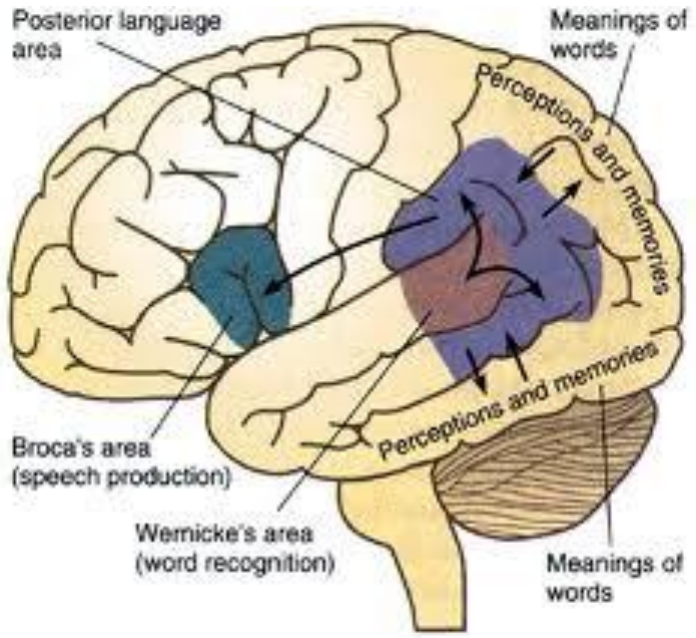
Friston, Harrison & Penny (2003) *NeuroImage*
Stephan & Friston (2007) *Handbook of Brain Connectivity*

Summary of the analysis



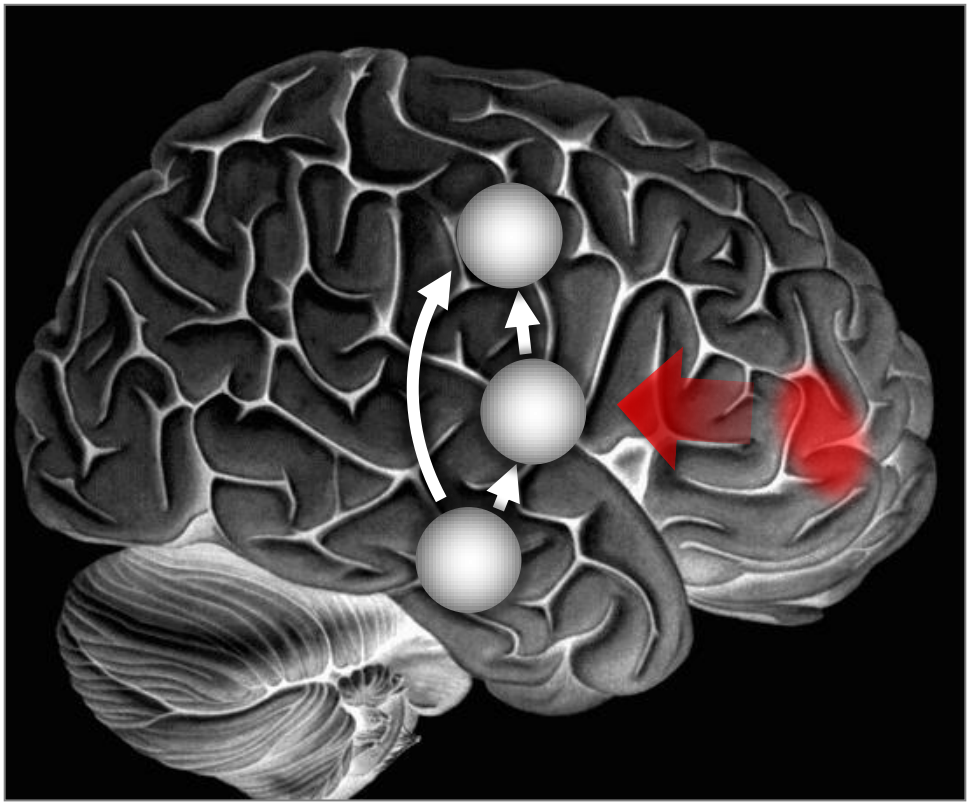
Example: diagnosis of moderate aphasia



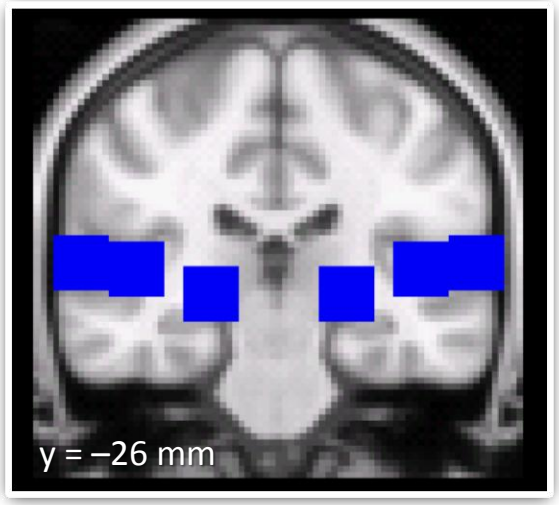


Example: diagnosing stroke patients

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

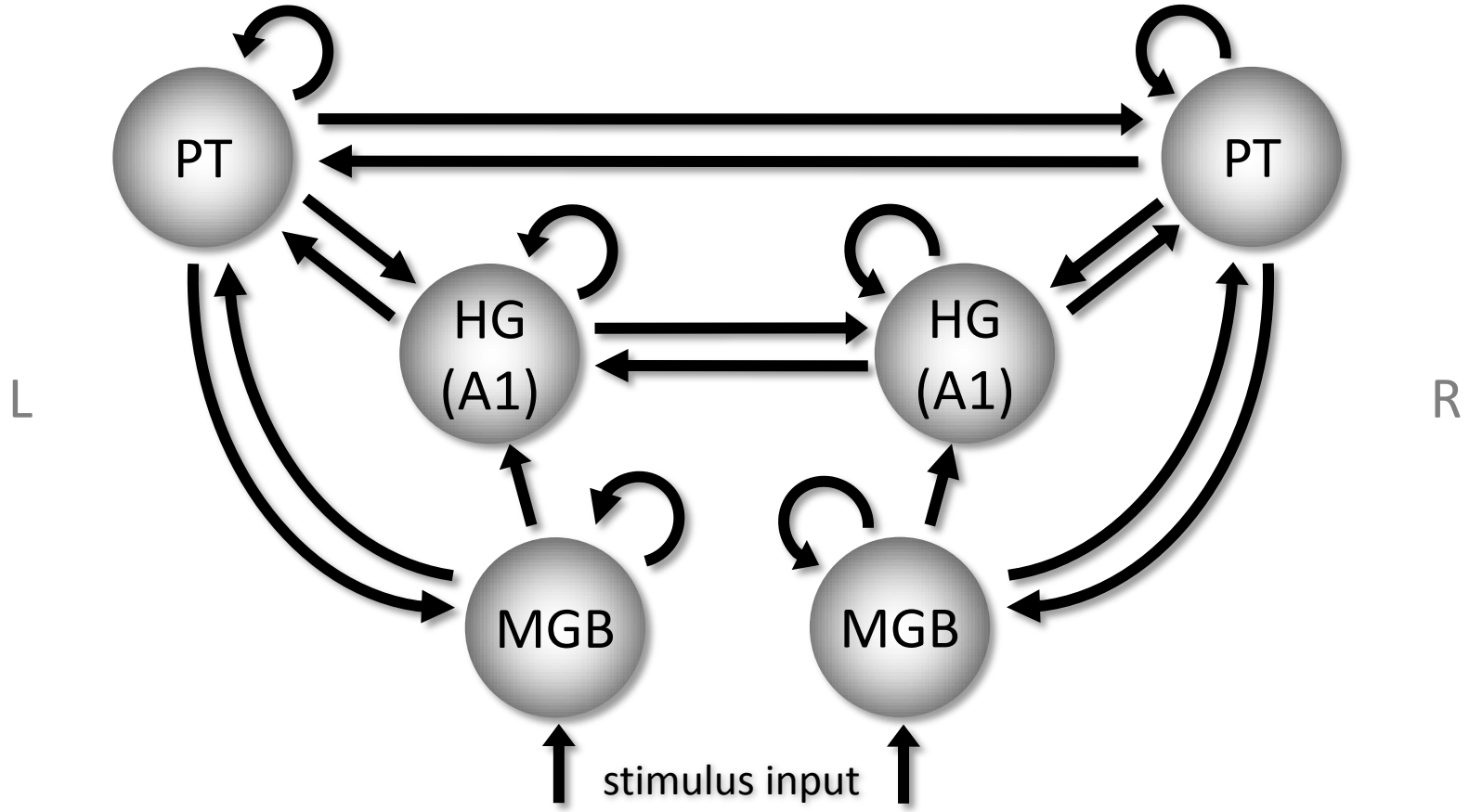


Example: diagnosing stroke patients

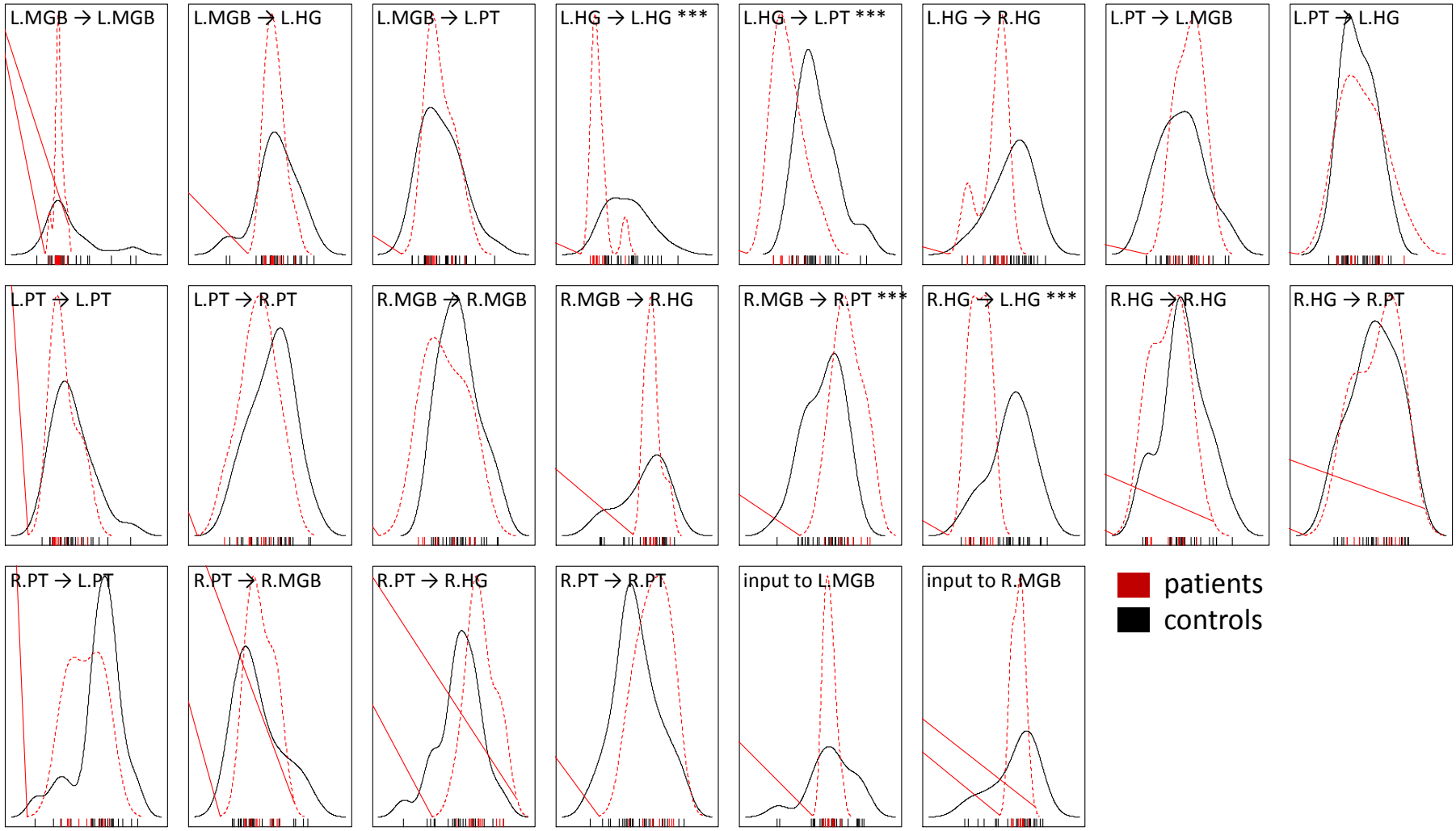


anatomical
regions of interest

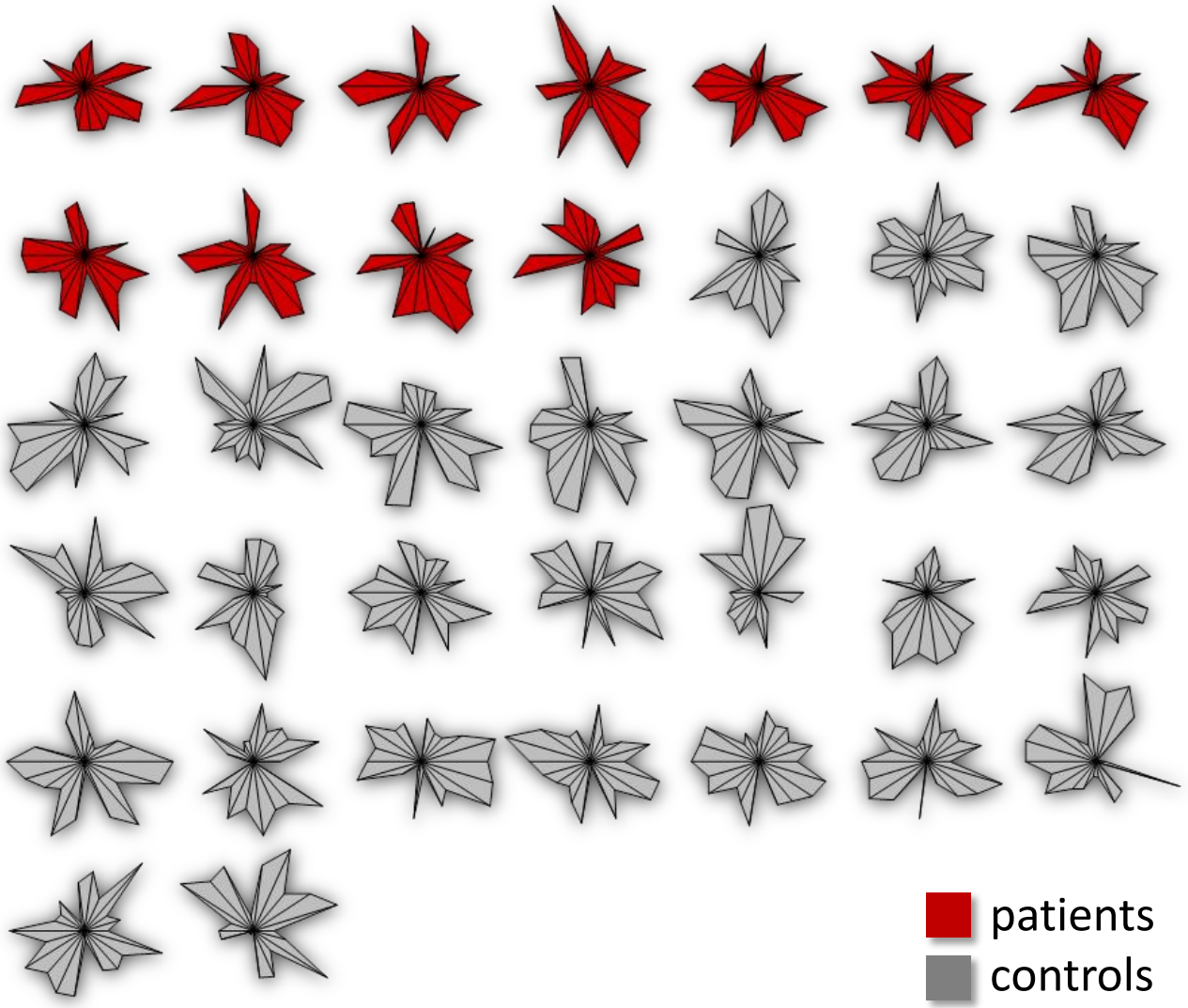
Example: diagnosing stroke patients



Univariate analysis: parameter densities

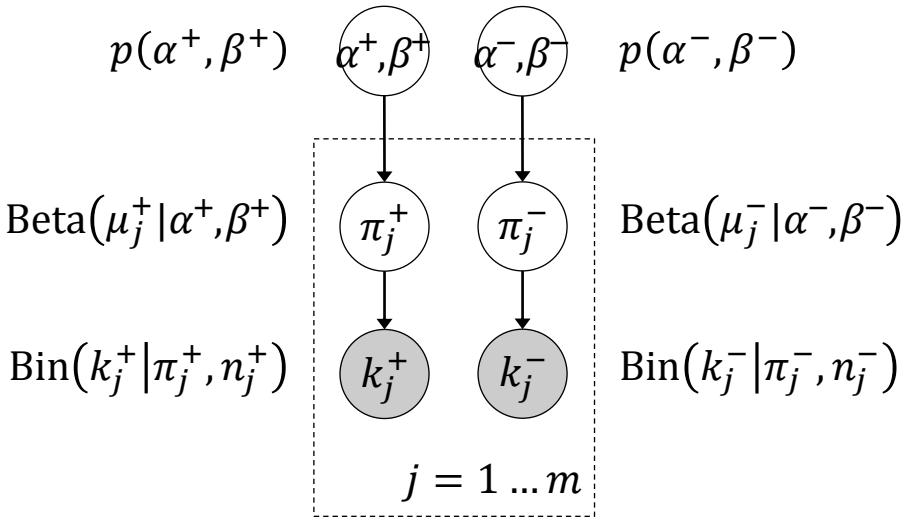


Multivariate analysis: connectional fingerprints

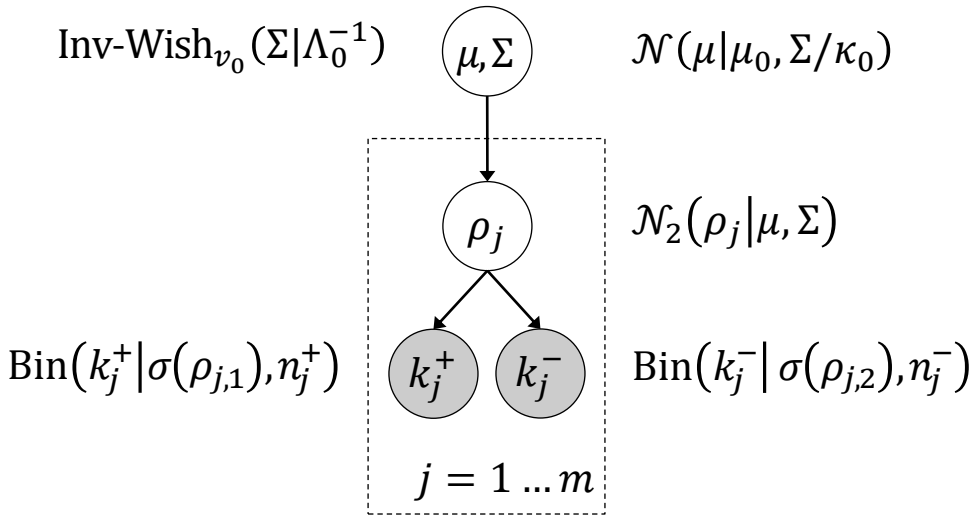


Full Bayesian approach to performance evaluation

Beta-binomial model

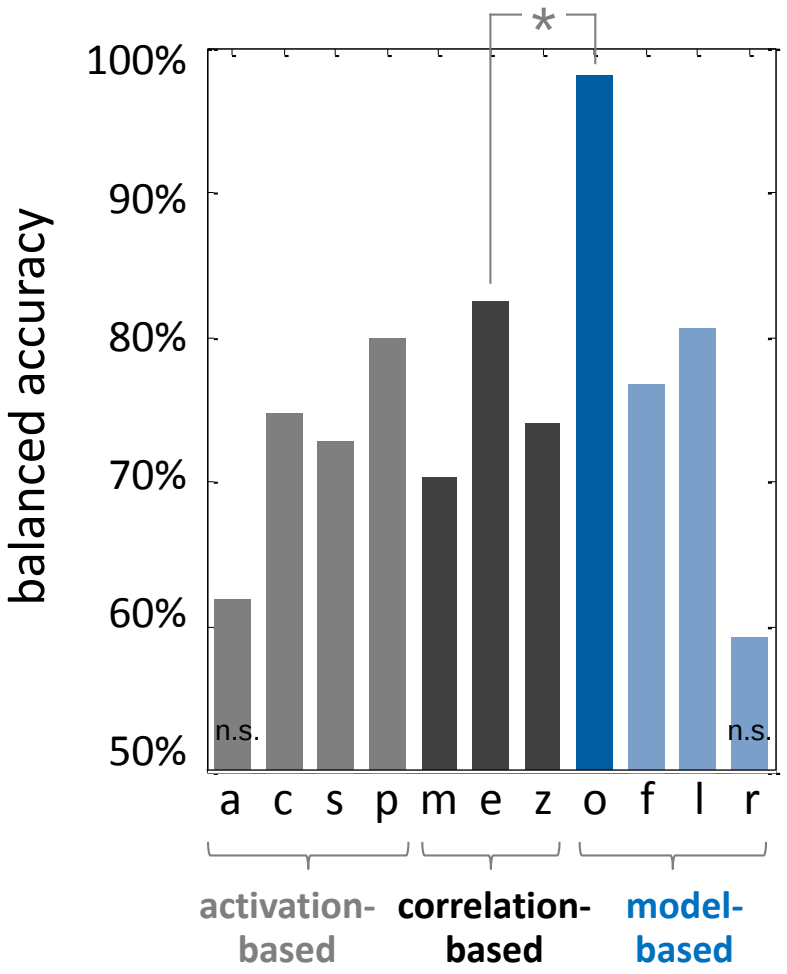


Normal-binomial model



Brodersen et al. *(under review)*

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

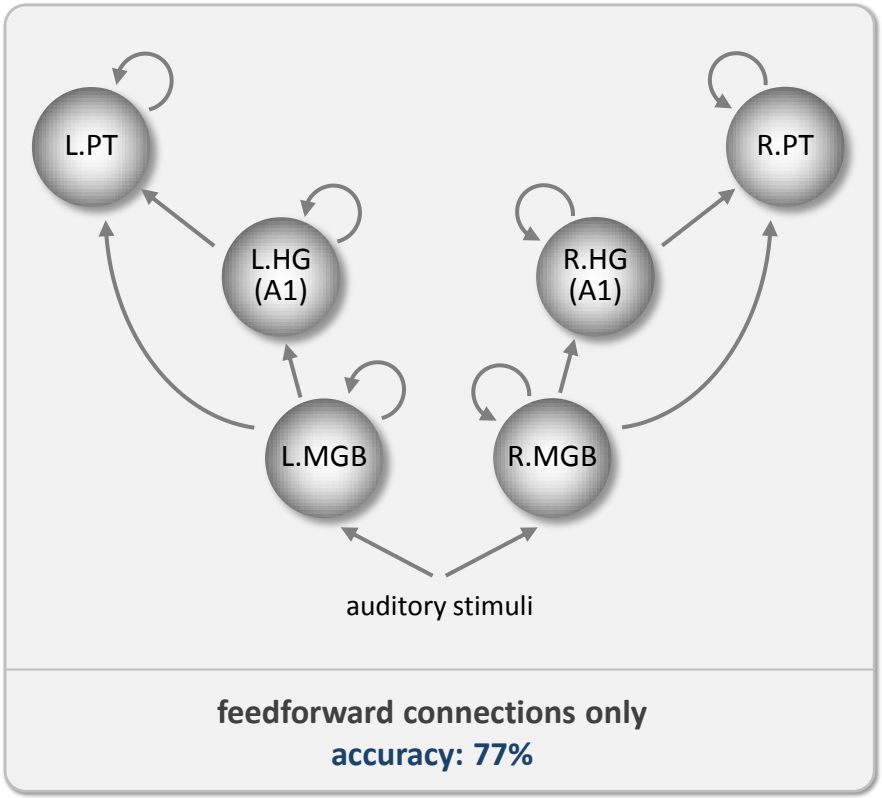
Correlation-based analyses

- **m** correlations of regional means
- **e** correlations of regional eigenvariates
- **z** Fisher-transformed eigenvariates correlations

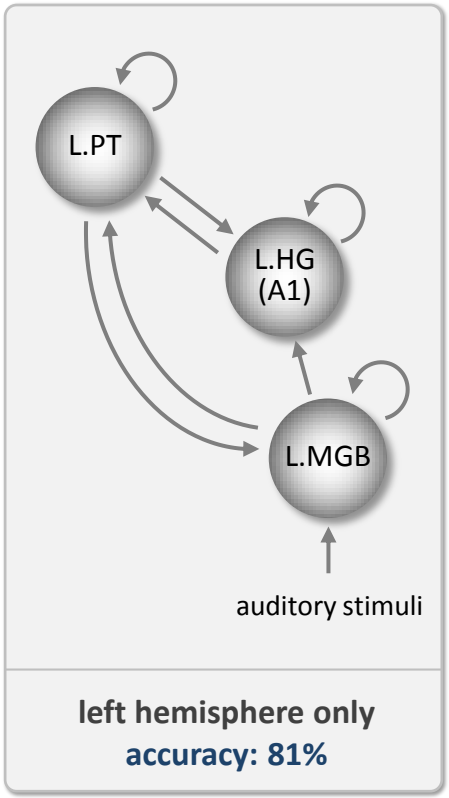
Model-based analyses

- **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

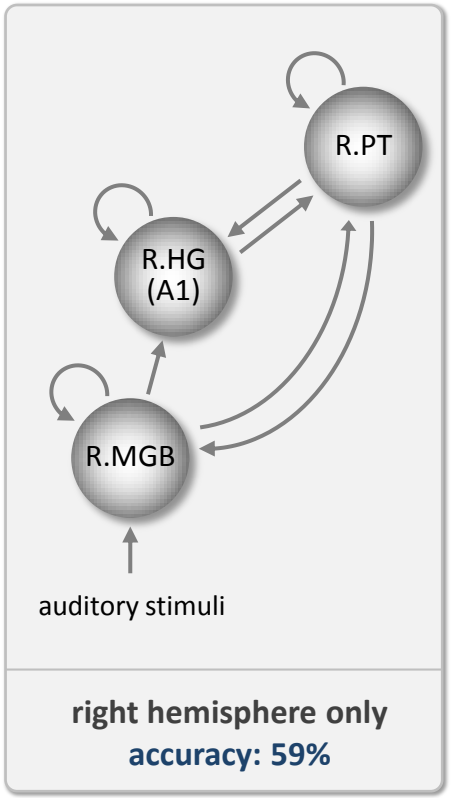
Biologically less plausible models perform poorly



feedforward connections only
accuracy: 77%

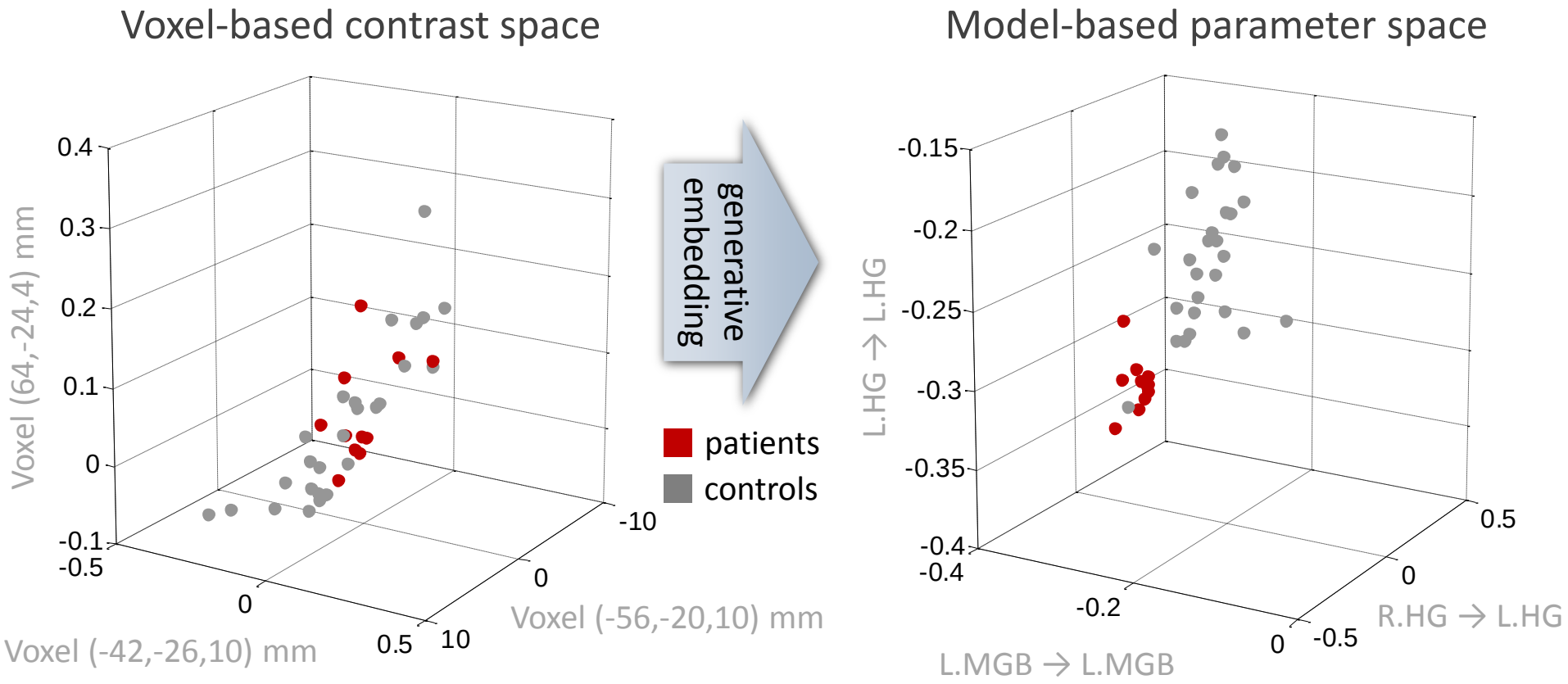


left hemisphere only
accuracy: 81%



right hemisphere only
accuracy: 59%

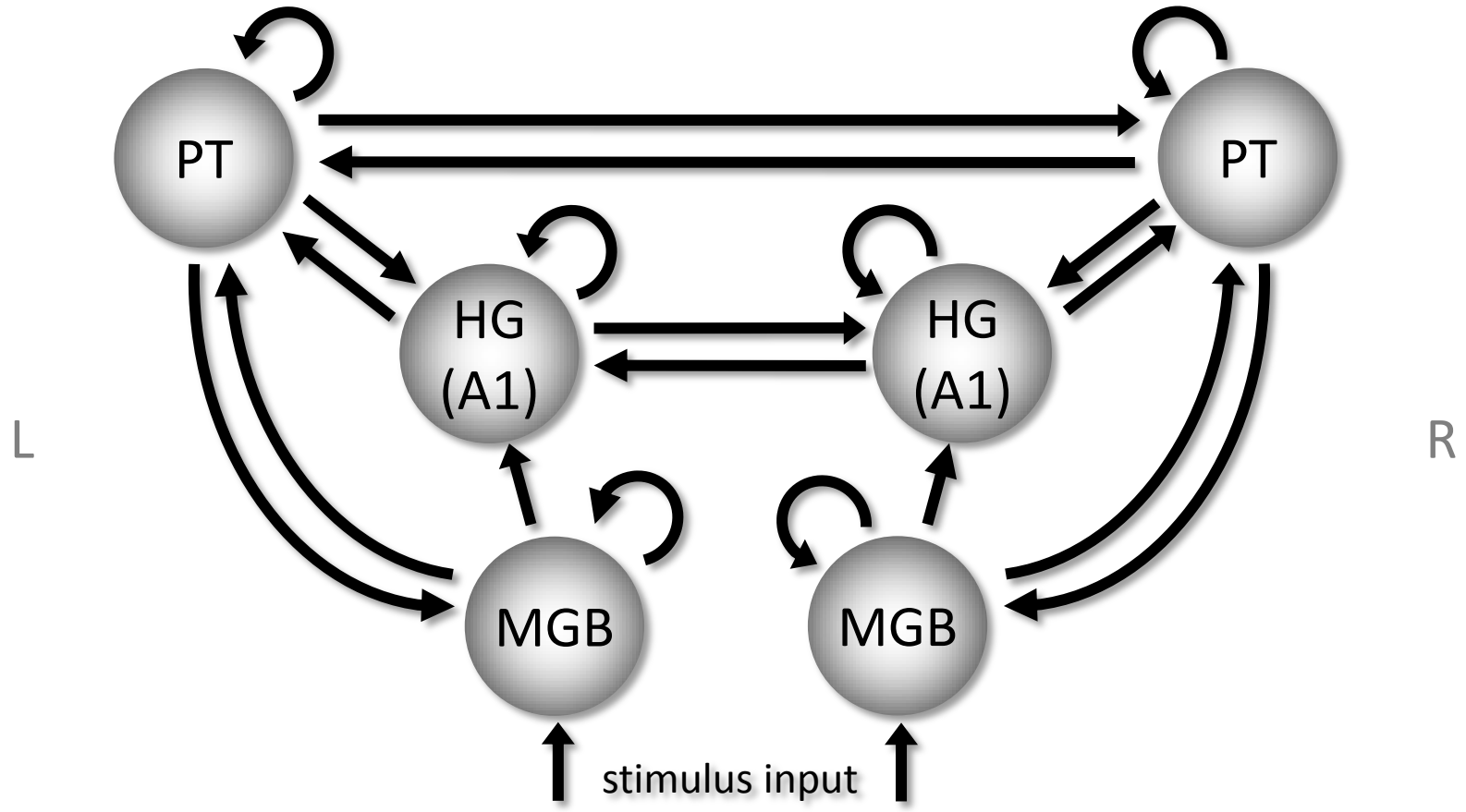
The generative projection



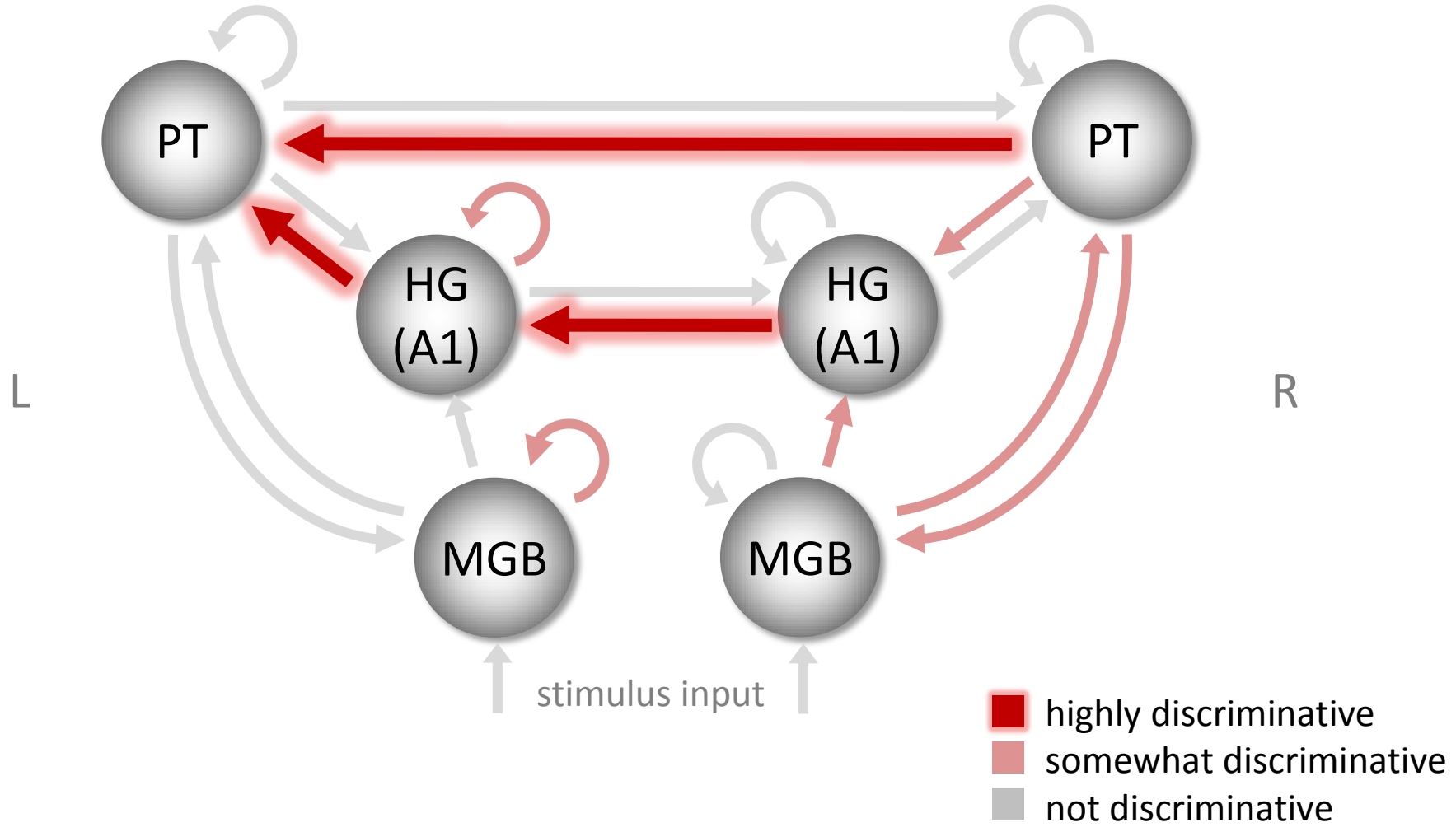
classification accuracy
(using all voxels in the regions of interest)
75%

classification accuracy
(using all 23 model parameters)
98%

Discriminative features in model space



Discriminative features in model space

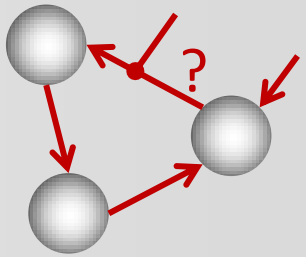


Generative embedding and DCM

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

⇨ compute the posterior

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

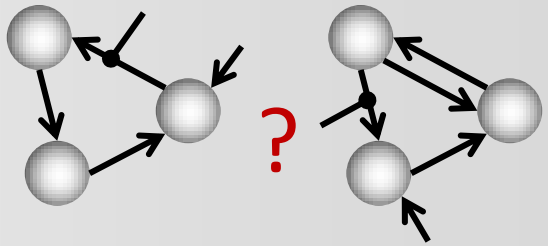


Question 2 – Which model is best w.r.t. the observed fMRI data?

⇨ compute the model evidence

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

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

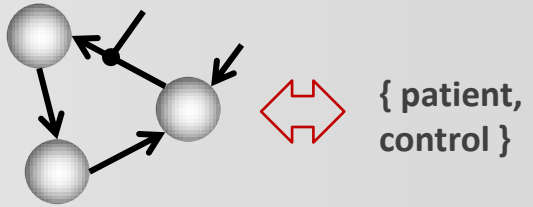


Question 3 – Which model is best w.r.t. an external criterion?

⇨ compute the classification accuracy

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

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

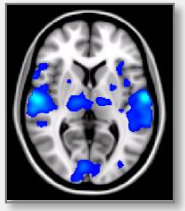


Model-based classification using DCM

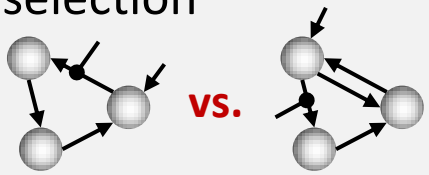
model-based classification

{ group 1, group 2 }

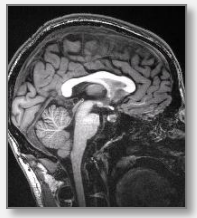
activation-based classification



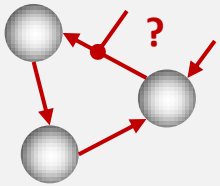
model selection



structure-based classification

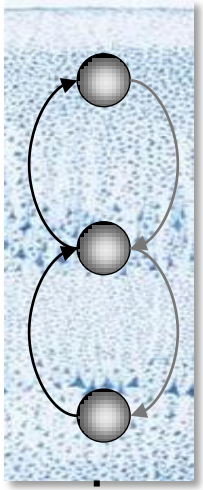


inference on model parameters

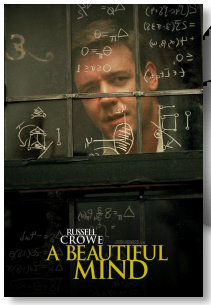


Model-based inference on *individual* pathophysiology

❶ model of neuronal (patho)physiology



❷ application to brain activity data from individual patients



spectrum disease

model-based

diagnostic tests

❸ diagnostic classification

type 1 → treatment X

type 2 → treatment Y

type 3 → treatment Z