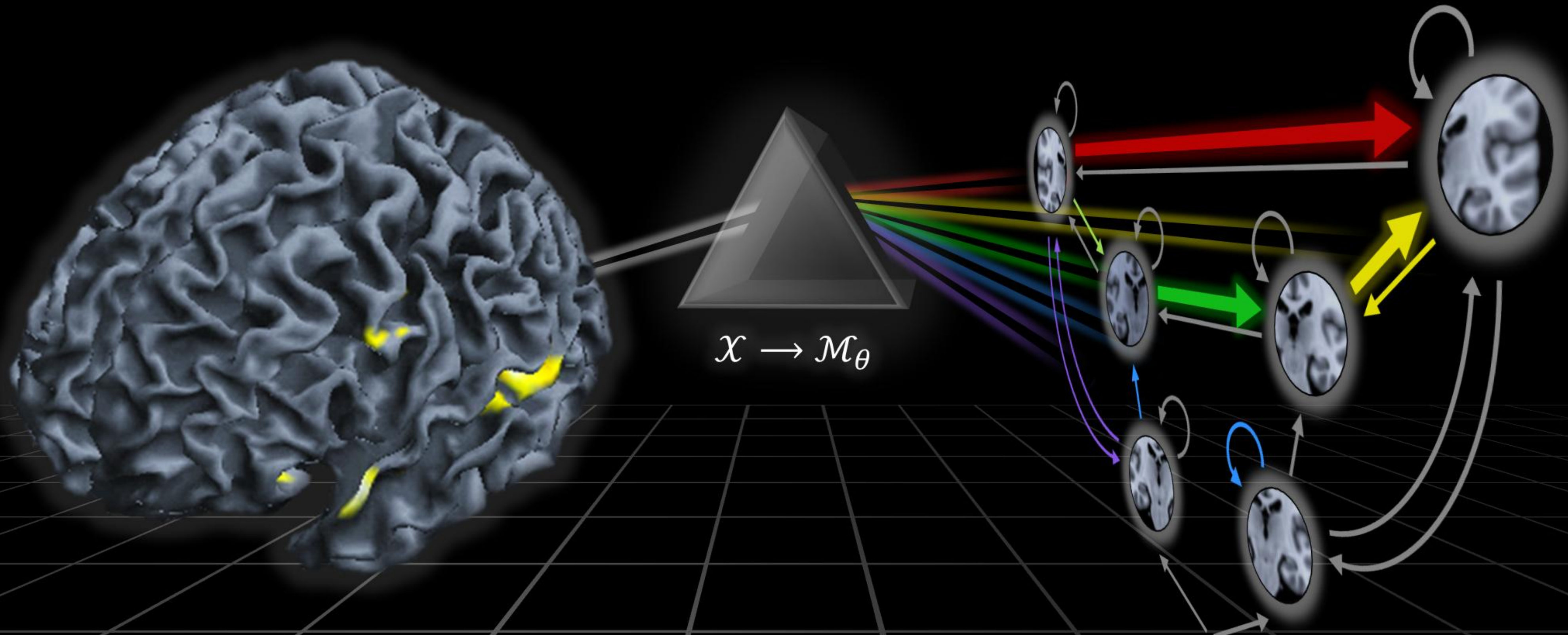
Generative embedding for model-based classification

Kay H. Brodersen[1,2], Thomas M. Schofield[3], Alexander P. Leff[3], Cheng Soon Ong[1], Ekaterina I. Lomakina[1,2], Joachim M. Buhmann[1], Klaas E. Stephan[2,3]

[1] Machine Learning Laboratory, Department of Computer Science, ETH Zurich, Switzerland
[2] Laboratory for Social and Neural Systems Research, Department of Economics, University of Zurich, Switzerland
[3] Wellcome Trust Centre for Neuroimaging, University College London, UK



ETH Eidgenössische Technische Hochschule Zürich Swiss Federal Institute of Technology Zurich
University of Zurich
UCL

Classification approaches by data representation

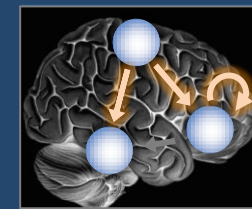
Structure-based classification

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

Activation-based classification

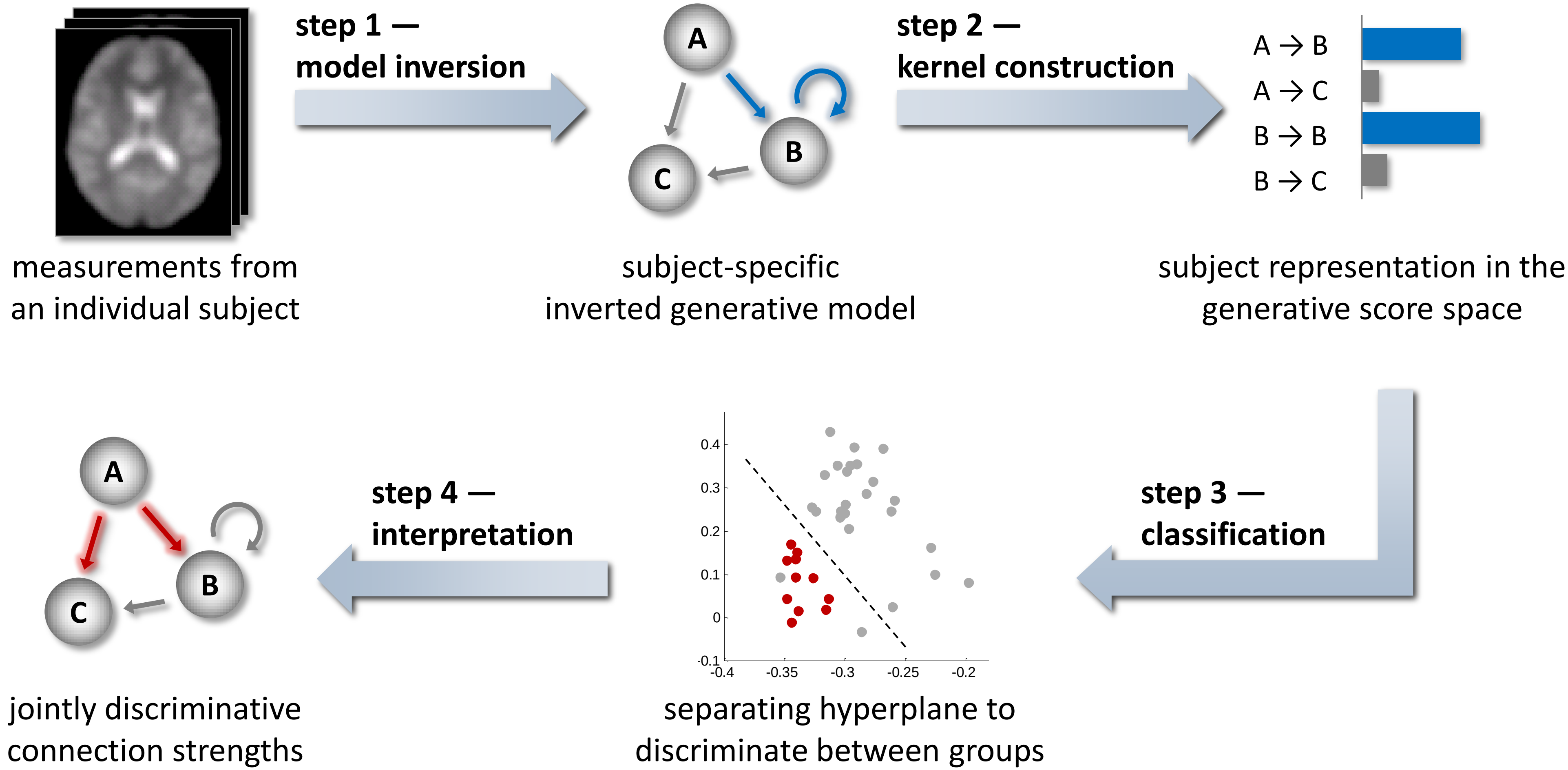
Which functional differences allow us to separate groups?

Model-based classification



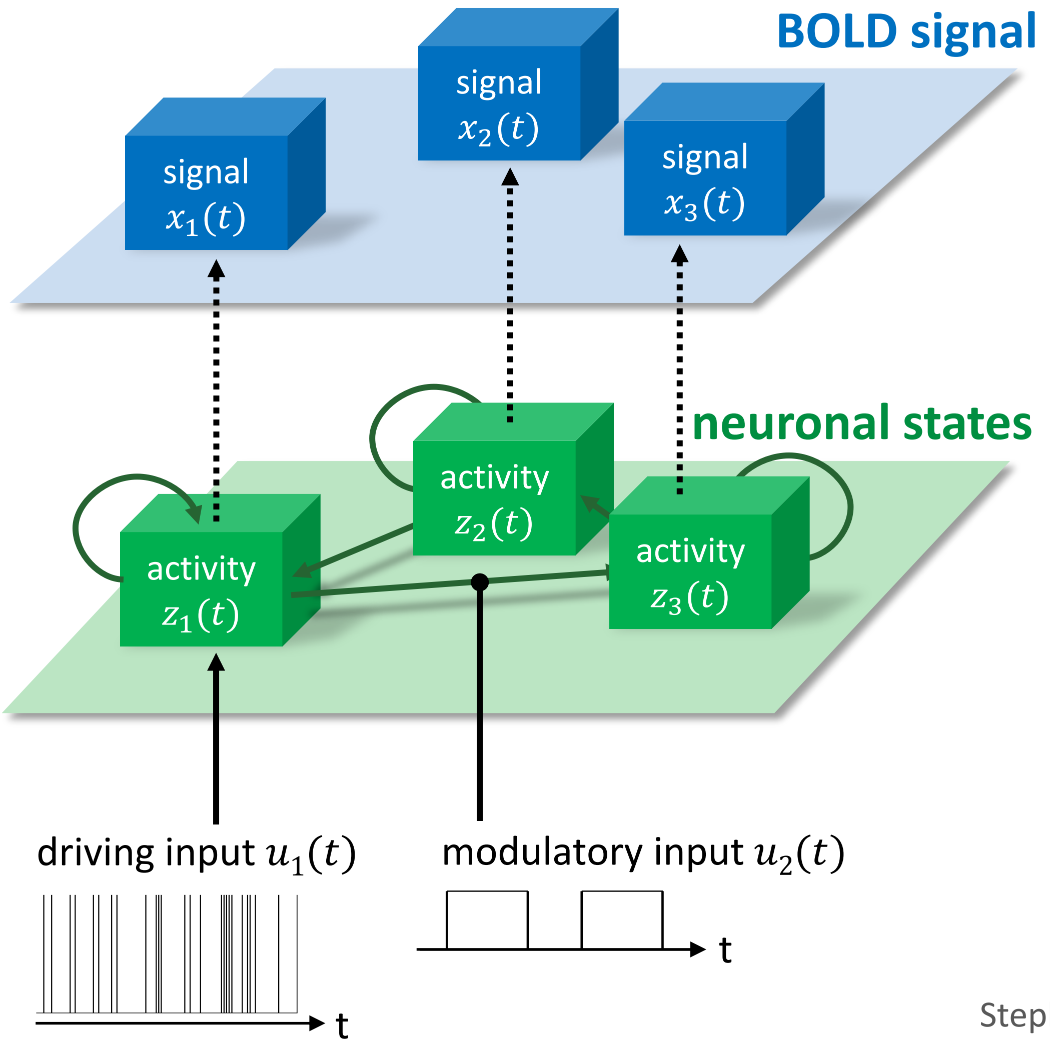
How do patterns of hidden quantities (e.g., connectivity among brain regions) differ between groups?

Model-based classification through generative embedding



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

Choosing a generative model: DCM for fMRI



hemodynamic 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

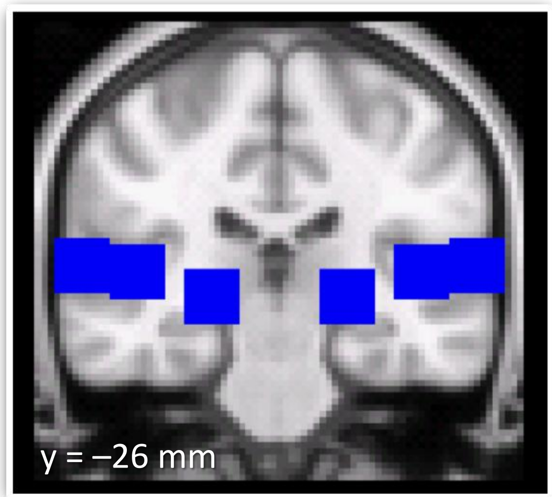
direct inputs

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

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.

L

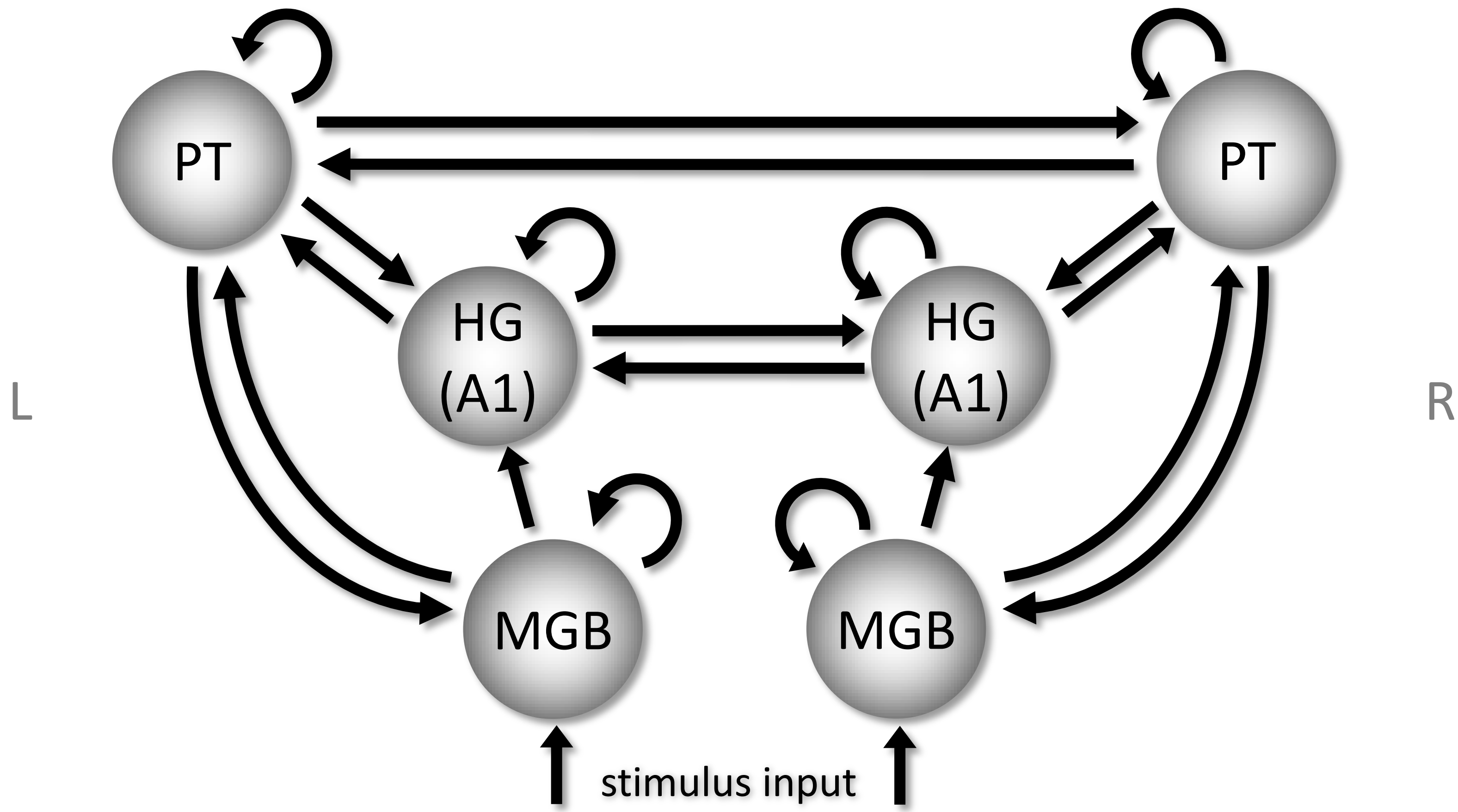


R

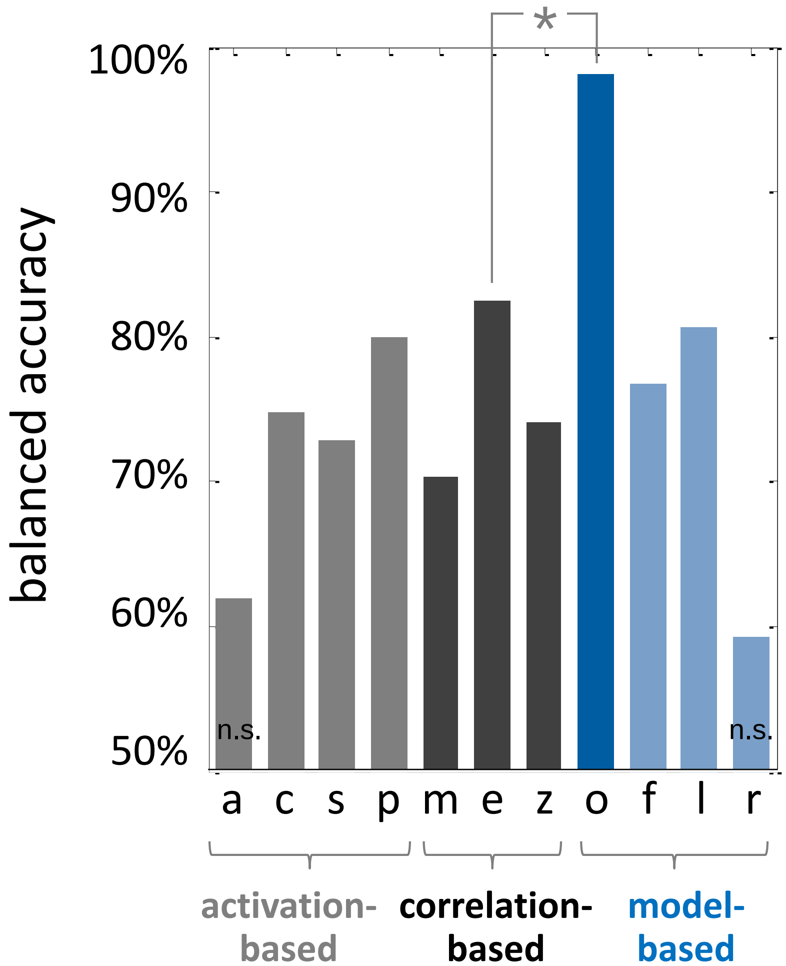
anatomical
regions of interest

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.



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

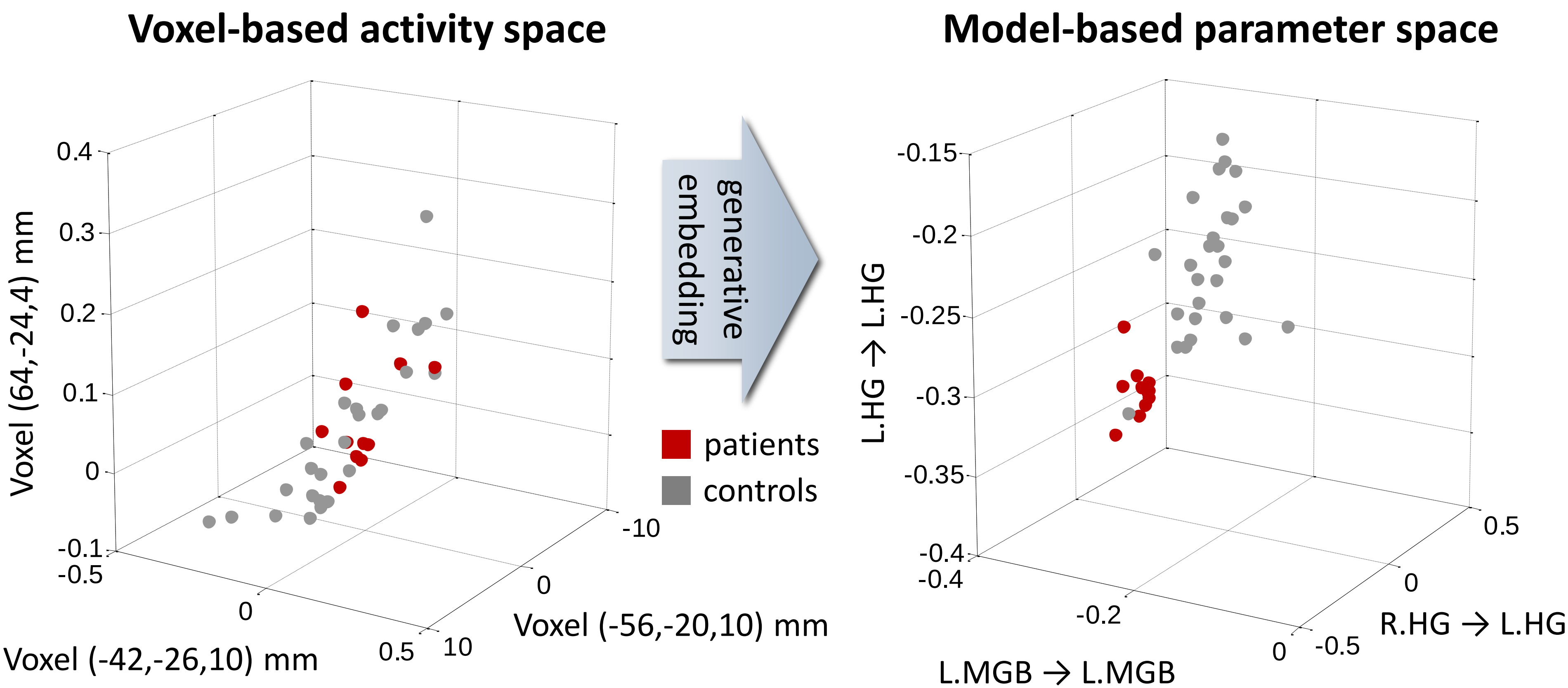
Correlation-based analyses

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

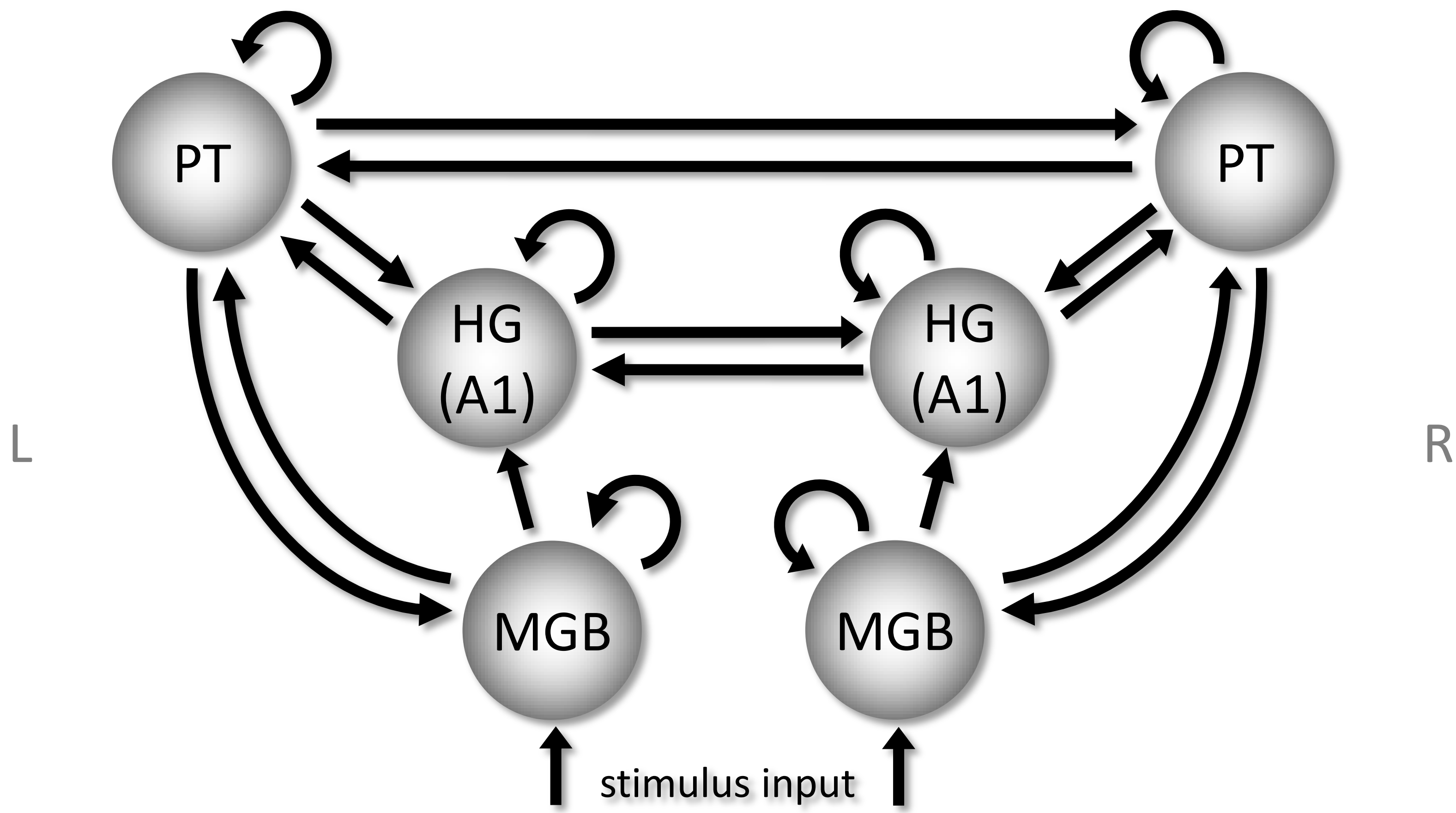
Model-based analyses

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

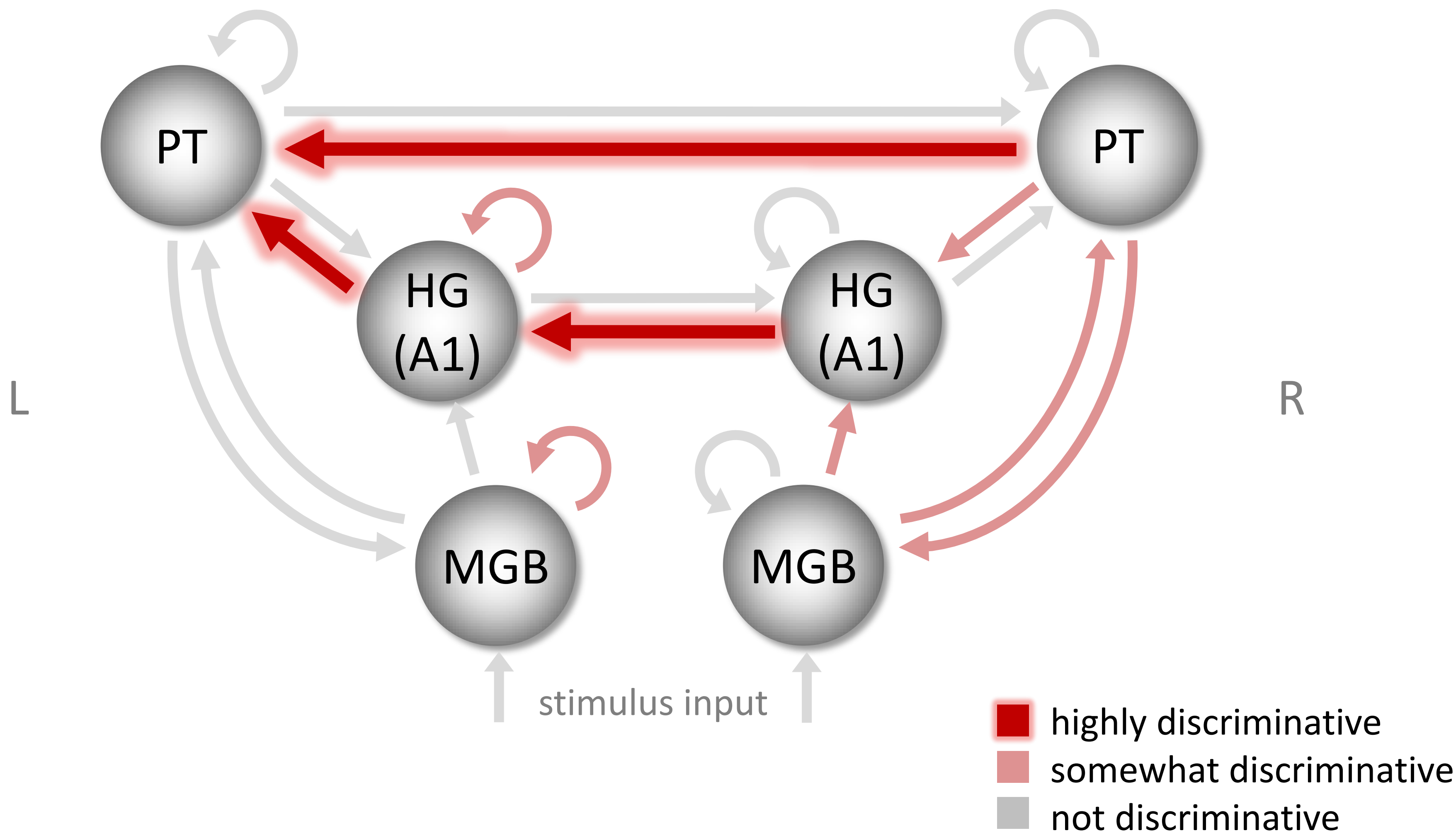
The generative projection



Discriminative features in model space



Discriminative features in model space



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.

For details, see: Brodersen et al. (2011) *PLoS Comp Biol*