Dynamic causal modeling for fMRI

Methods and Models for fMRI, HS 2016

Jakob Heinzle

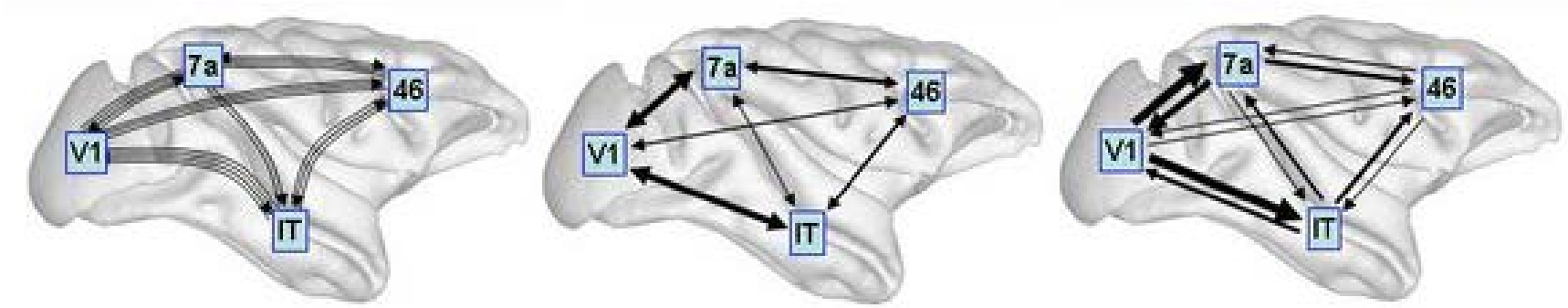
Translational Neuromodeling Unit

Universität Zürich[UZH]

ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Structural, functional & effective connectivity



Sporns 2007, *Scholarpedia*

anatomical/structural connectivity

- presence of physical connections
- DWI, tractography, tracer studies (monkeys)

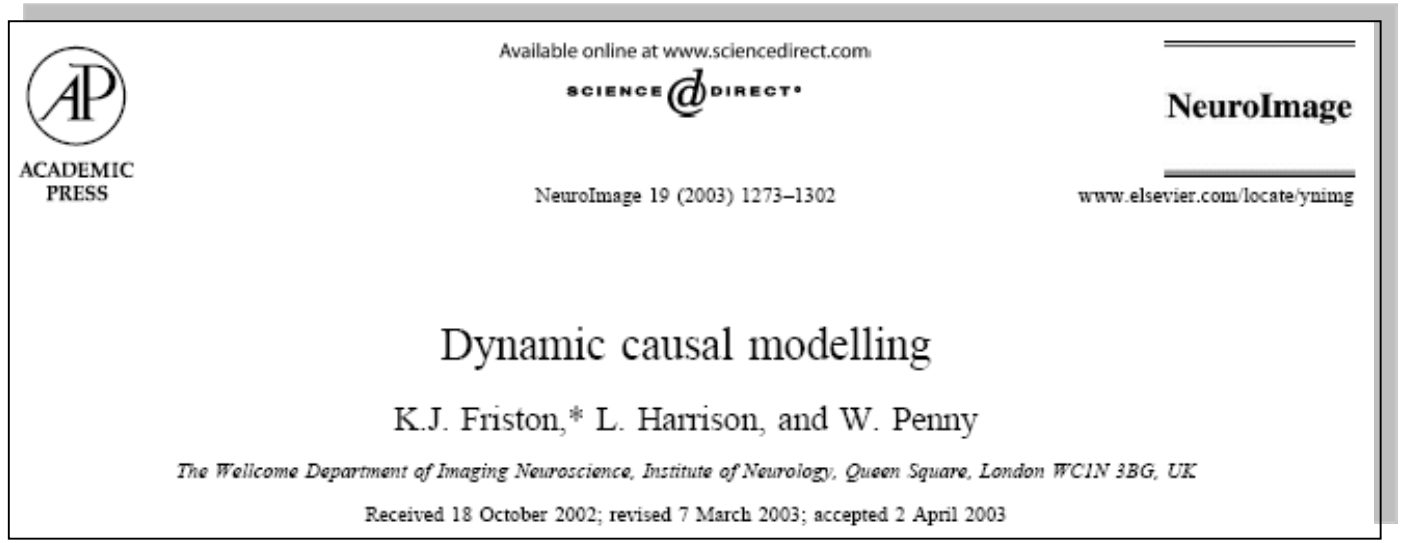
functional connectivity

- statistical dependency between regional time series
- correlations, ICA

effective connectivity

- causal (directed) influences between neuronal populations
- DCM

Dynamic causal modelling (DCM) for fMRI



ACADEMIC PRESS

Available online at www.sciencedirect.com
SCIENCE DIRECT®

NeuroImage

NeuroImage 19 (2003) 1273–1302

www.elsevier.com/locate/ynimg

Dynamic causal modelling

K.J. Friston,* L. Harrison, and W. Penny

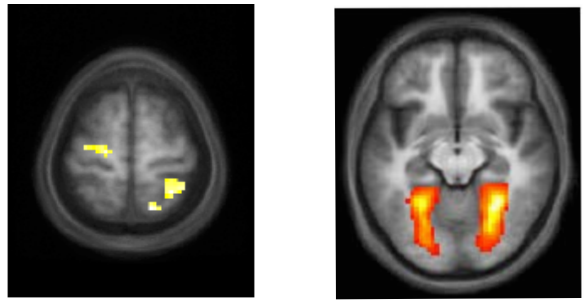
The Wellcome Department of Imaging Neuroscience, Institute of Neurology, Queen Square, London WC1N 3BG, UK

Received 18 October 2002; revised 7 March 2003; accepted 2 April 2003

- DCM framework was introduced in 2003 for fMRI by Karl Friston, Lee Harrison and Will Penny (NeuroImage 19:1273-1302)
- part of the SPM software package
- Allows to do an effective connectivity analysis

From functional segregation to functional integration

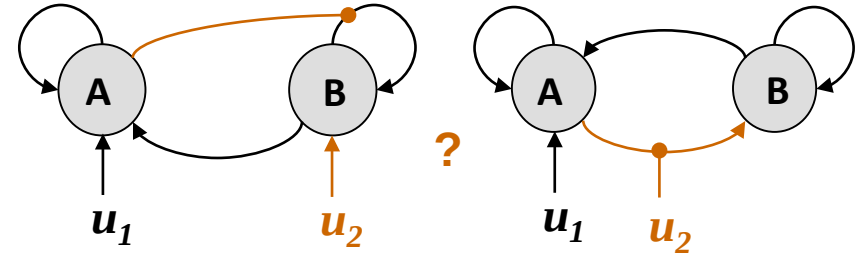
localizing brain activity:
functional segregation



u_1 $u_1 X u_2$

« Where, in the brain, did my experimental manipulation have an effect? »

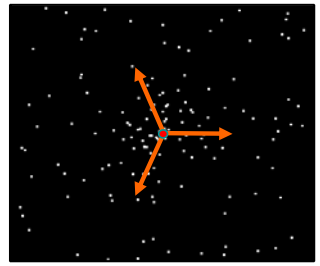
effective connectivity analysis:
functional integration

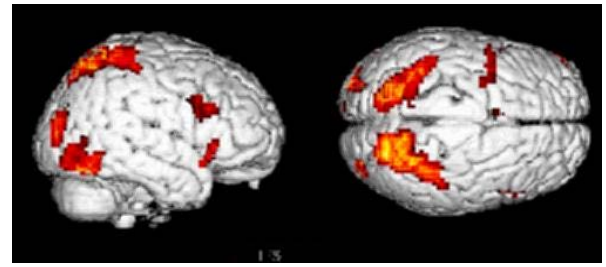


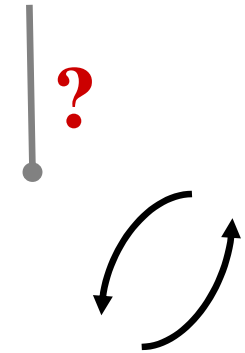
« How did my experimental manipulation propagate through the network? »

Introductory example: Attention to motion

Assess site of attention modulation during visual processing in fMRI paradigm reported by Büchel and Friston.





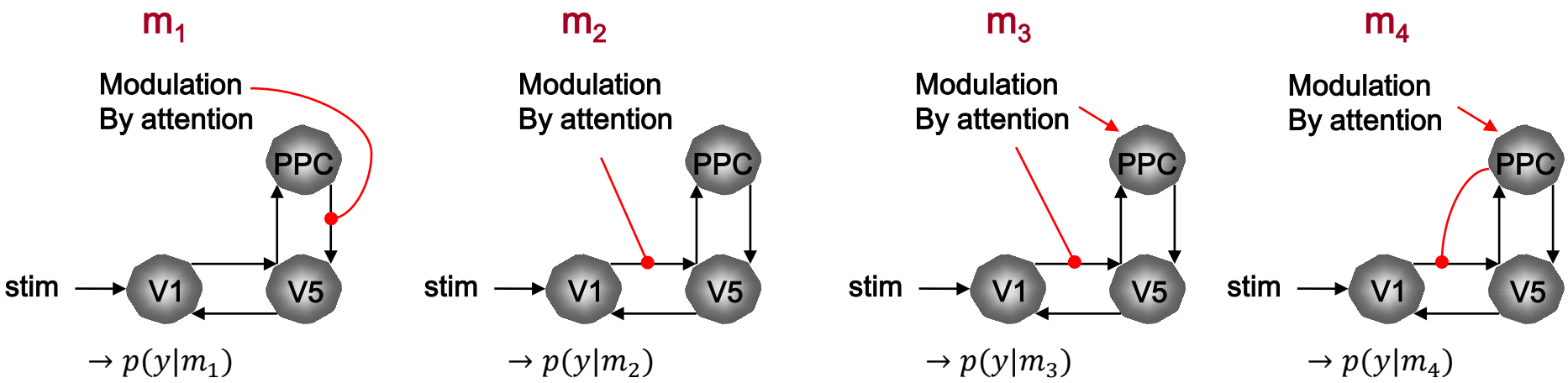


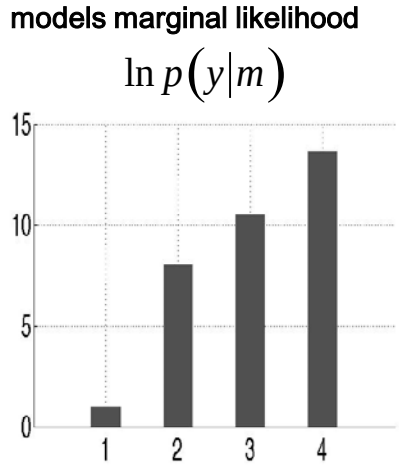
How can the classical results be explained mechanistically:

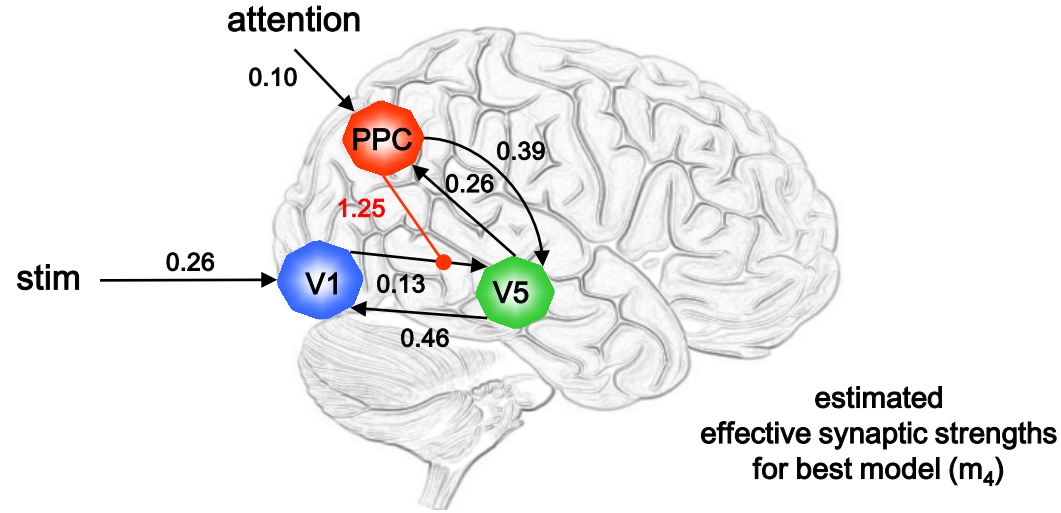
Photic → V1
+ Motion → V5
+ Attention → V5 + parietal cortex (SPC)

Friston et al. (2003) ***NeuroImage***

Bayesian model selection

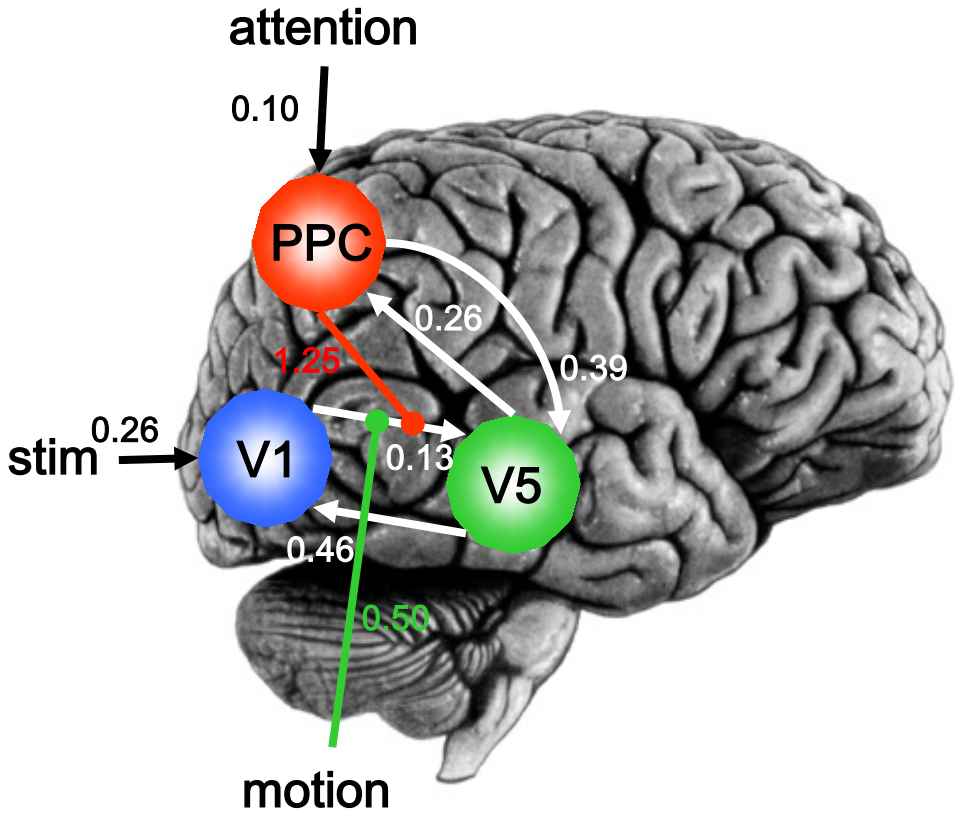


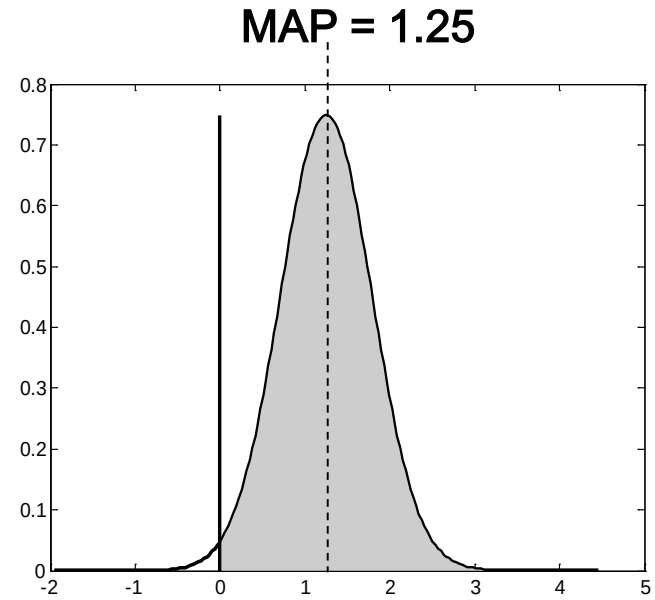




Stephan et al. 2008, *NeuroImage*

Parameter inference



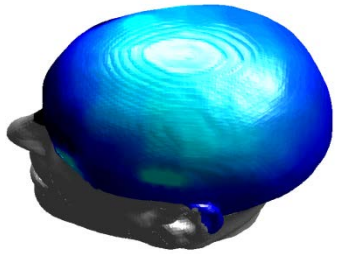


$$p(D^{PPC}_{V5,V1} > 0 \mid y) = 99.1\%$$

We will run this example in the tutorial

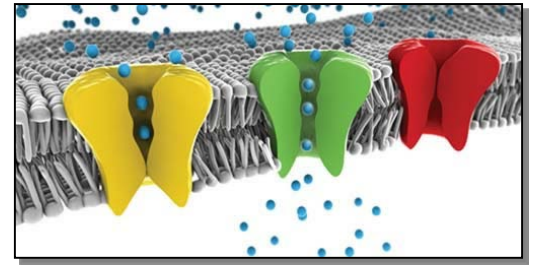
Stephan et al. 2008, *NeuroImage*

Generative model



$$p(y \mid \theta, m) \cdot p(\theta \mid m)$$

$$p(\theta \mid y, m)$$



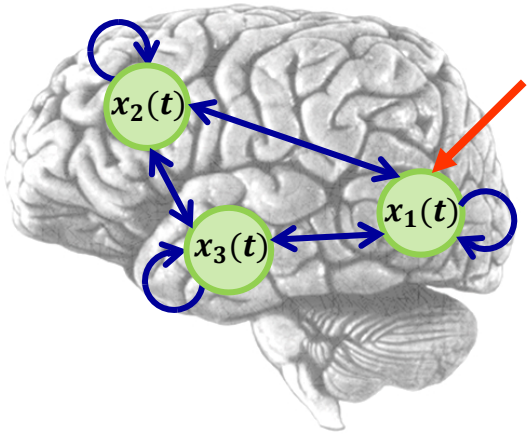
1. enforces mechanistic thinking: how could the data have been caused?
2. generate synthetic data (observations) by sampling from the prior – can model explain certain phenomena at all?
3. inference about model structure: formal approach to disambiguating mechanisms → p(m|y)
4. inference about parameters → p(θ|y)

DCM approach to effective connectivity

A simple model of a neural network …

… described as a dynamical system …

… causes the data (BOLD signal).



$x_i(t)$ Neural node

Input (u)

Connections (θ)

$$\dot{x} = f(x, u, \theta_x)$$

$$y = g(x, \theta_y)$$

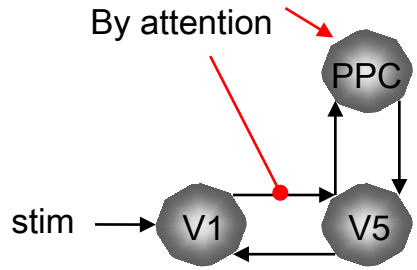
Let the system run with input (u) and parameters (θ_x, θ_y), and you will get a BOLD signal time course y that you can compare to the measured data.

Approximating $f(x, u, \theta_x)$

$$\frac{dx}{dt} = f(x,u) \approx f(x_0, 0) + \frac{\partial f}{\partial x}x + \frac{\partial f}{\partial u}u + \frac{\partial^2 f}{\partial x \partial u}xu + \frac{\partial^2 f}{\partial x^2}x^2 + \cdots$$

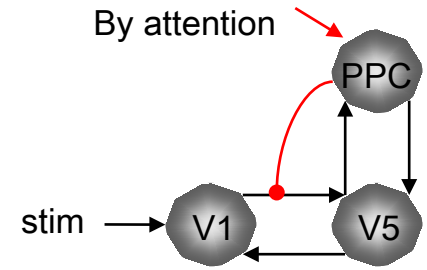
Bi-linear model

Modulation
By attention



Non-**linear model**

Modulation
By attention

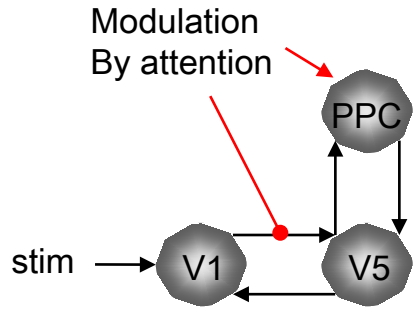


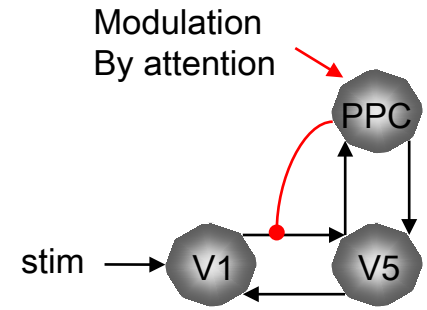
Approximating $f(x, u, \theta_x)$

$$\frac{dx}{dt} = f(x, u) \approx f(x_0, 0) + A\,x + C\,u + B\,xu + D\,x^2 + \cdots$$

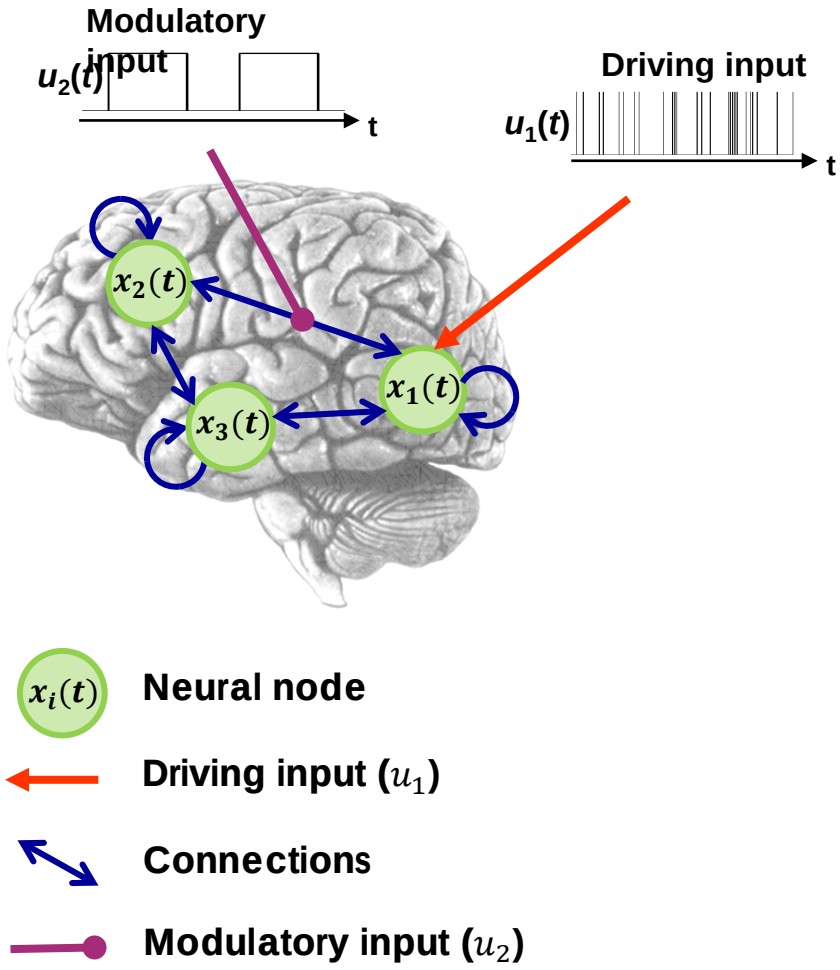
Bi-linear model

Non-linear model





The neural model – summary.



Parameter sets…

A - fixed connectivity

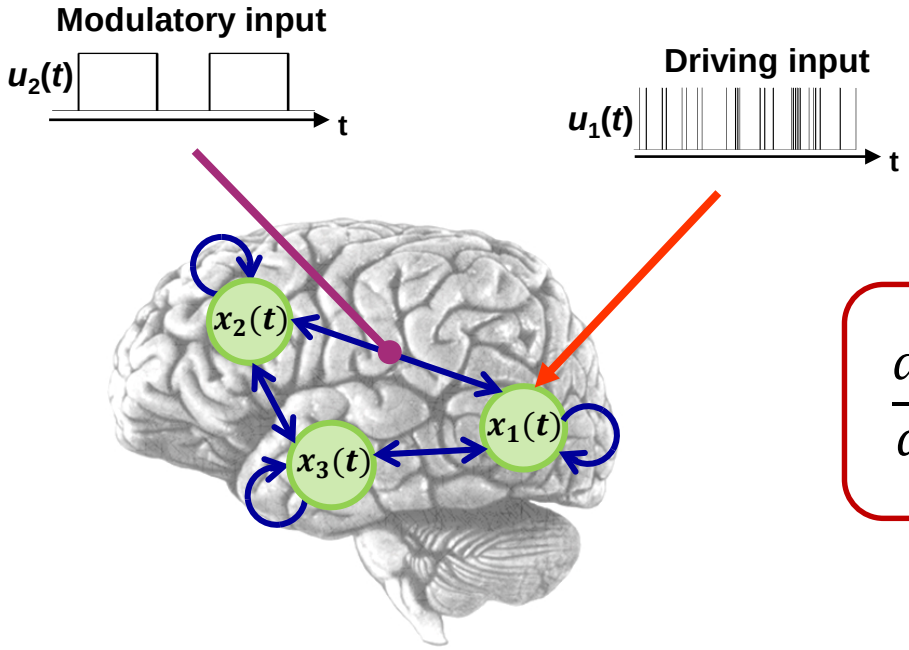
B - modulation of connectivity

C - weight of driving inputs

D – weight of non-linear terms

… determine dynamics!

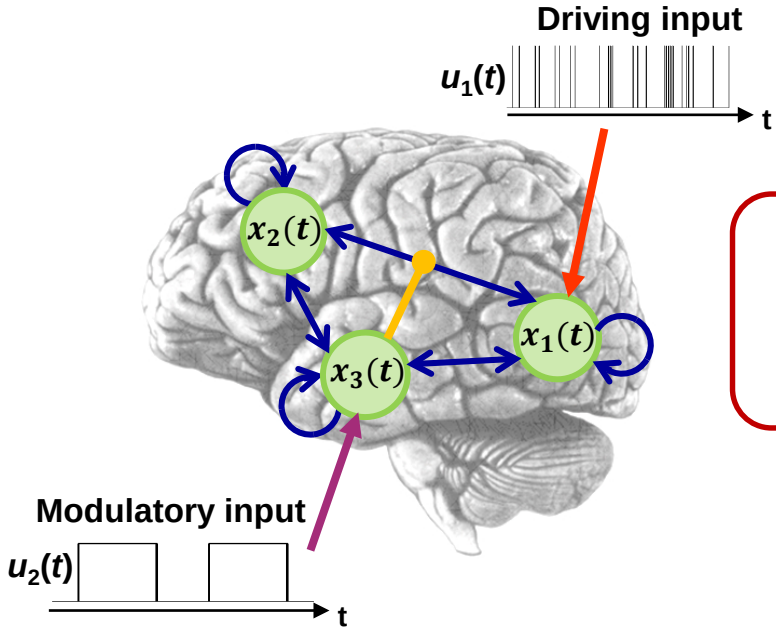
The neural equations – bilinear model



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

Parameters A, B and C define connectivity!

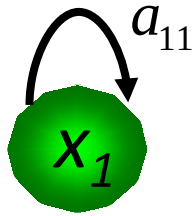
The neural equations – non-linear model



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

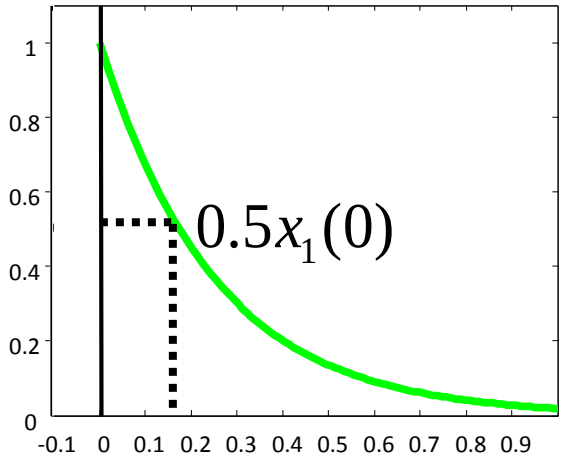
Parameters A, B, C and D define connectivity!

DCM parameters = rate constant

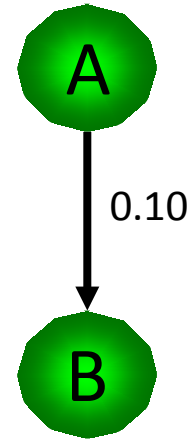


$$= \quad \frac{dx_1}{dt} = a_{11}x_1 \quad \rightarrow \quad x_1(t) = x_1(0)\exp(a_{11}t)$$

Decay function

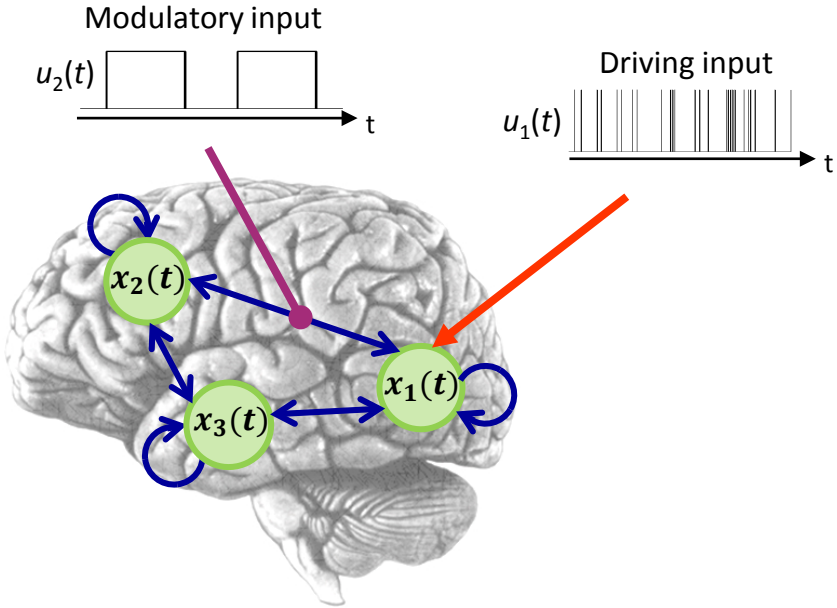


$$\tau = \ln 2 / a_{11}$$



If A→B is 0.10 s^{-1} this means that, per unit time, the increase in activity in B corresponds to 10% of the current activity in A

Bilinear neural state equation



State changes

External modulatory inputs

Current state

External driving inputs

$$\dot{\boldsymbol{x}} = \left(\boldsymbol{A} + \sum_{j=1}^{m} u_j \boldsymbol{B}^{(j)} \right) \boldsymbol{x} + \boldsymbol{C}\boldsymbol{u}$$

$$\boldsymbol{\theta} = \{\boldsymbol{A}, \boldsymbol{B}^{\boldsymbol{1}}, \ldots, \boldsymbol{B}^{\boldsymbol{m}}, \boldsymbol{C}\}$$

Fixed connectivity weights

Weights (strength) of connectivity modulation

Weights for direct inputs

The problem of the hemodynamic response

↑ neural activity

⬇

↑ blood flow

⬇

↑ oxyhemoglobin

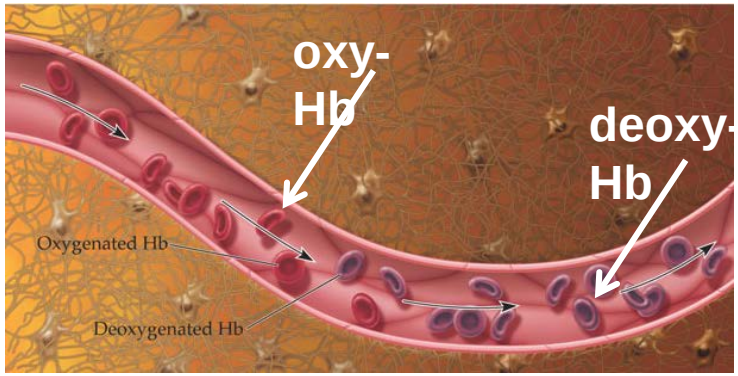
⬇

↑ T2*

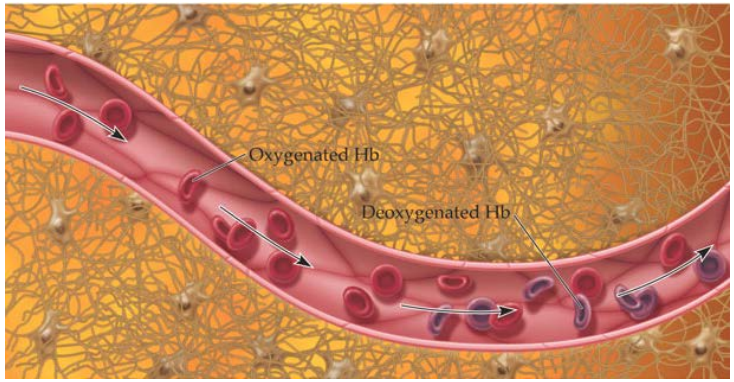
⬇

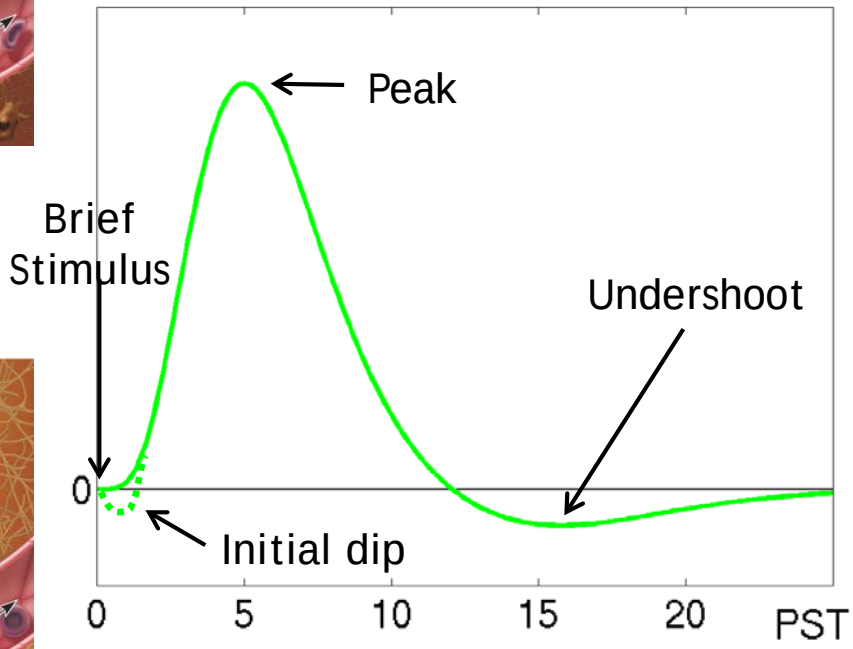
↑ MR signal

Rest



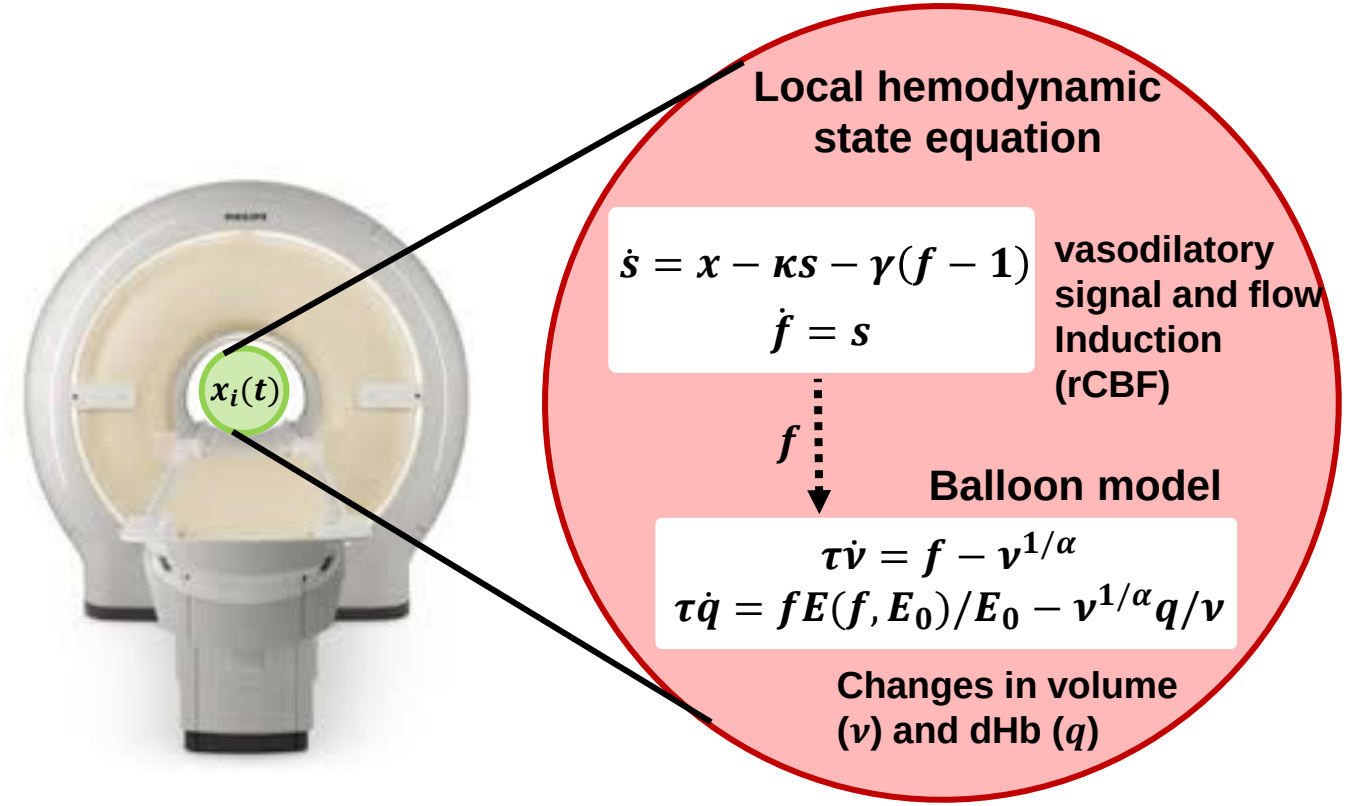
Activity





Source, Huettel et al, 2004, fMRI (Book)

From neural activity to the BOLD signal



BOLD signal change equation

$$y = \frac{\Delta S}{S_0} \approx V_0 \left[k_1(1 - q) + k_2\left(1 - \frac{q}{v}\right) + k_3(1 - v) \right]$$

cf. Simulations in Lecture 1

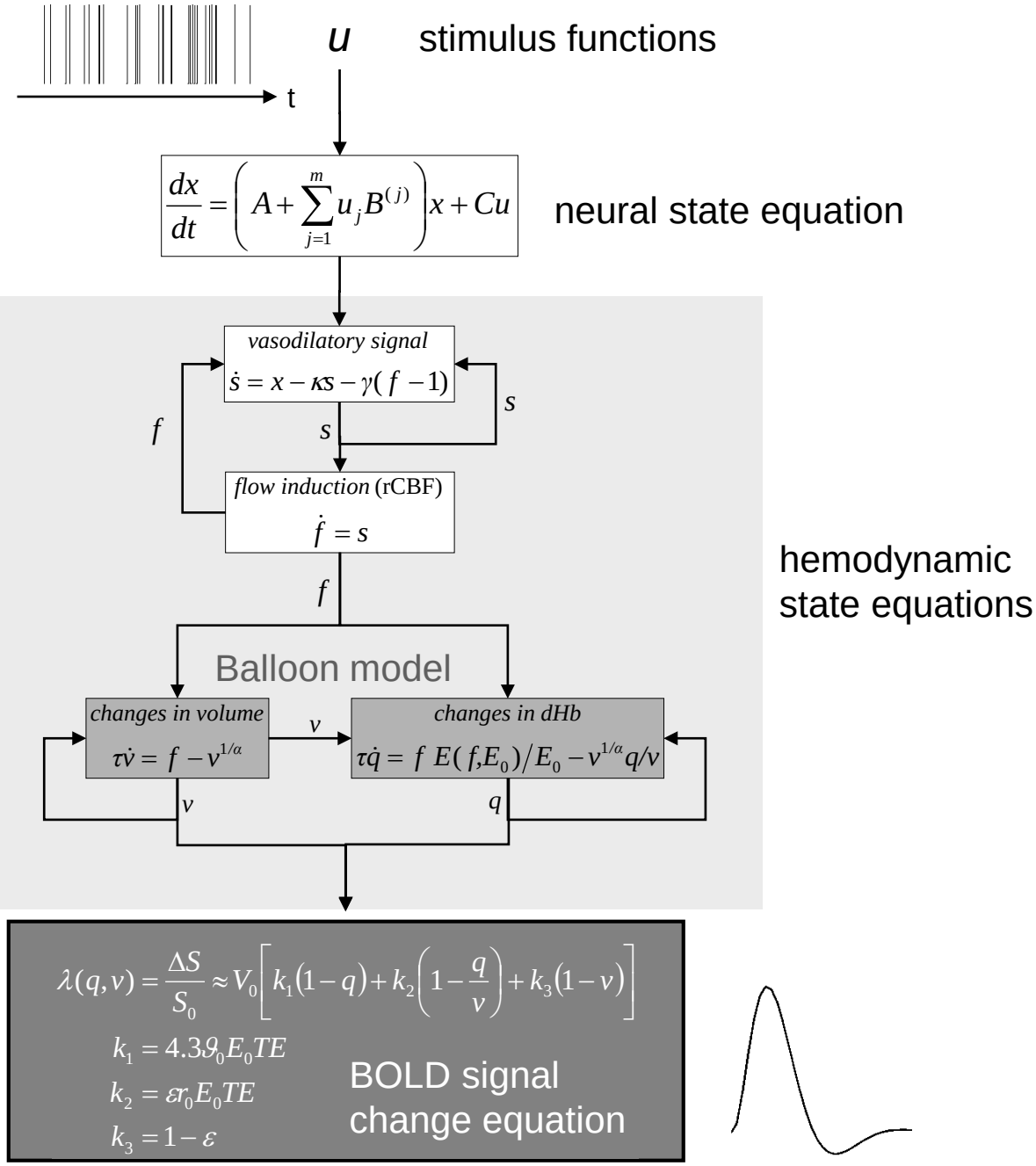
The hemodynamic model in DCM

- 6 hemodynamic parameters:

$$\theta^h = \{\kappa, \gamma, \tau, \alpha, \rho, \varepsilon\}$$

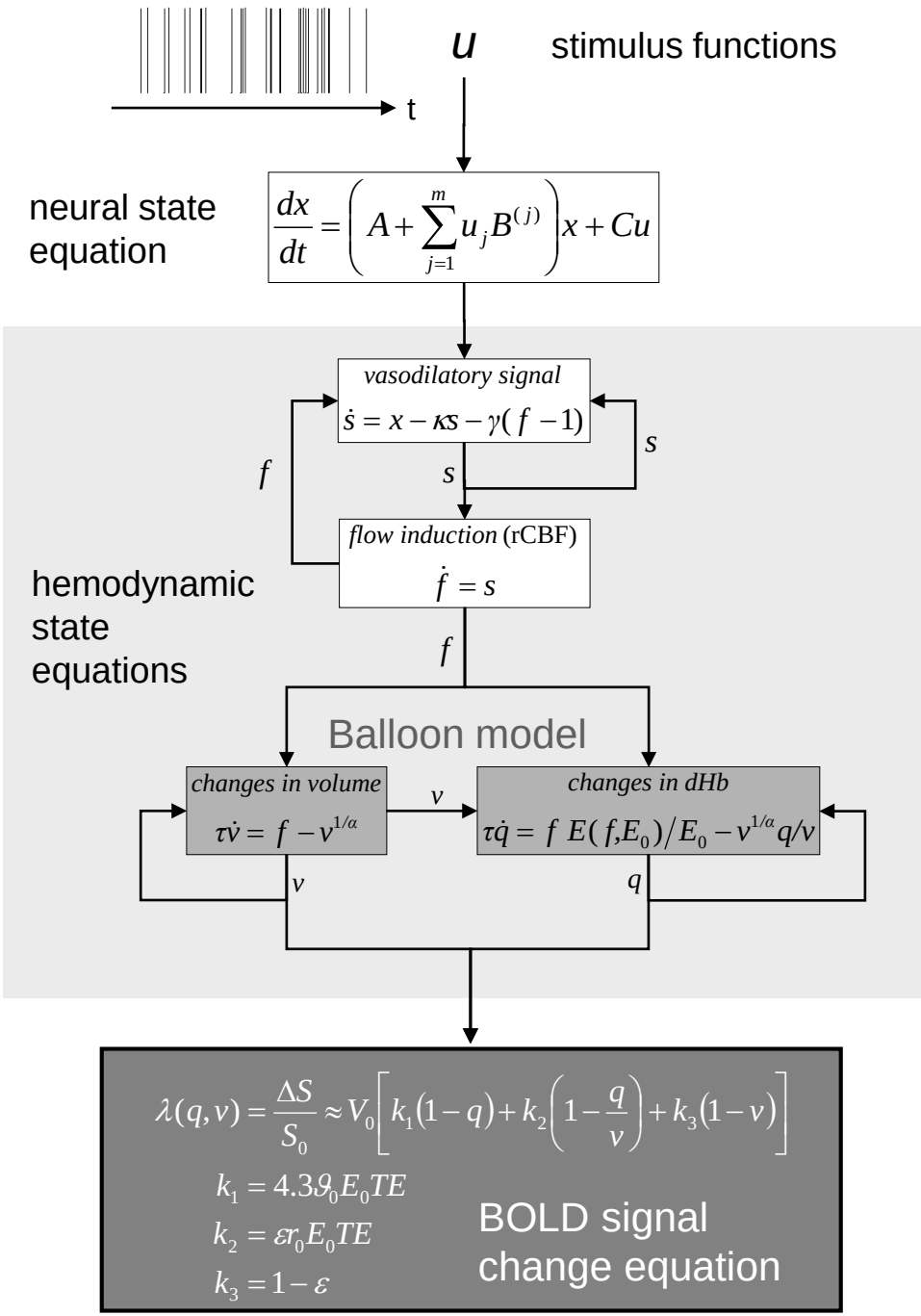
important for model fitting, but of no interest for statistical inference

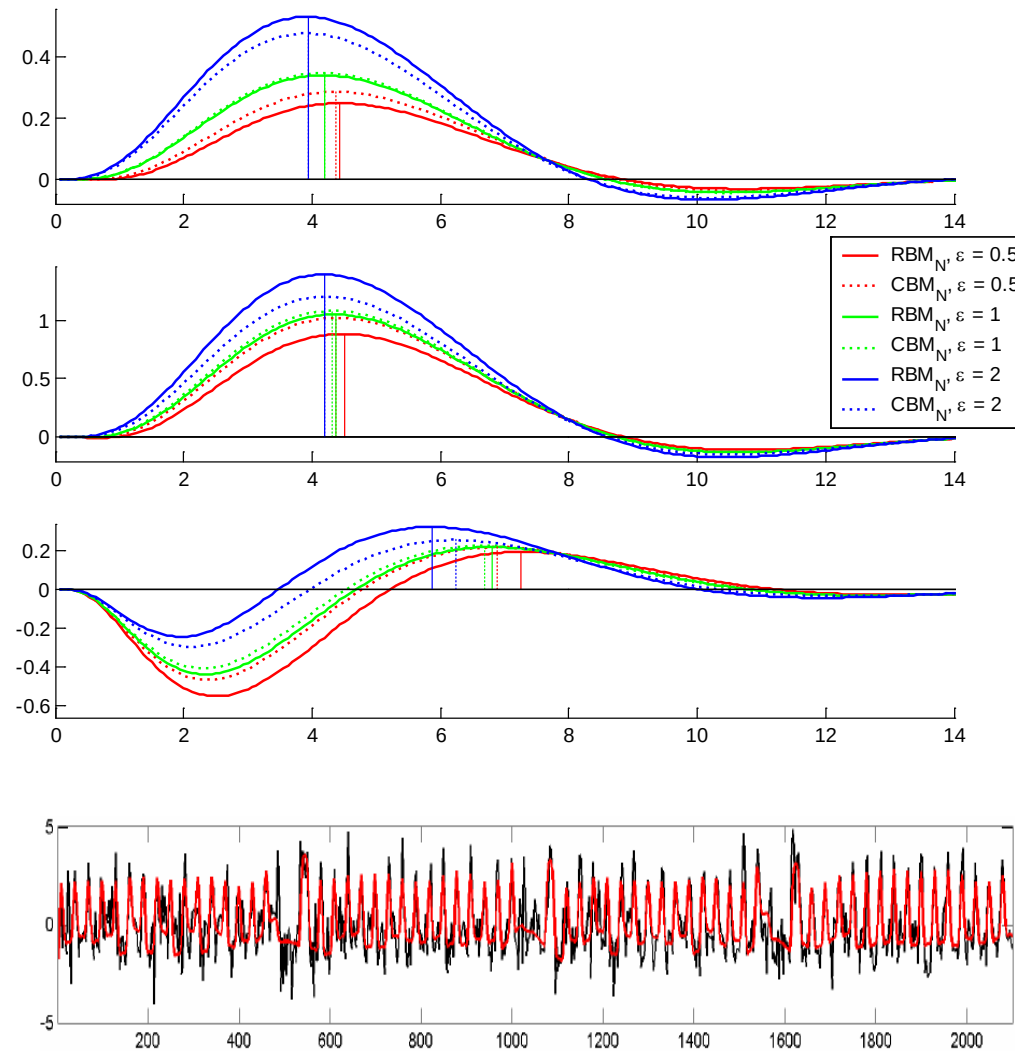
- Computed separately for each area (like the neural parameters)
  → region-specific HRFs!



Friston et al. 2000, *NeuroImage*
Stephan et al. 2007, *NeuroImage*

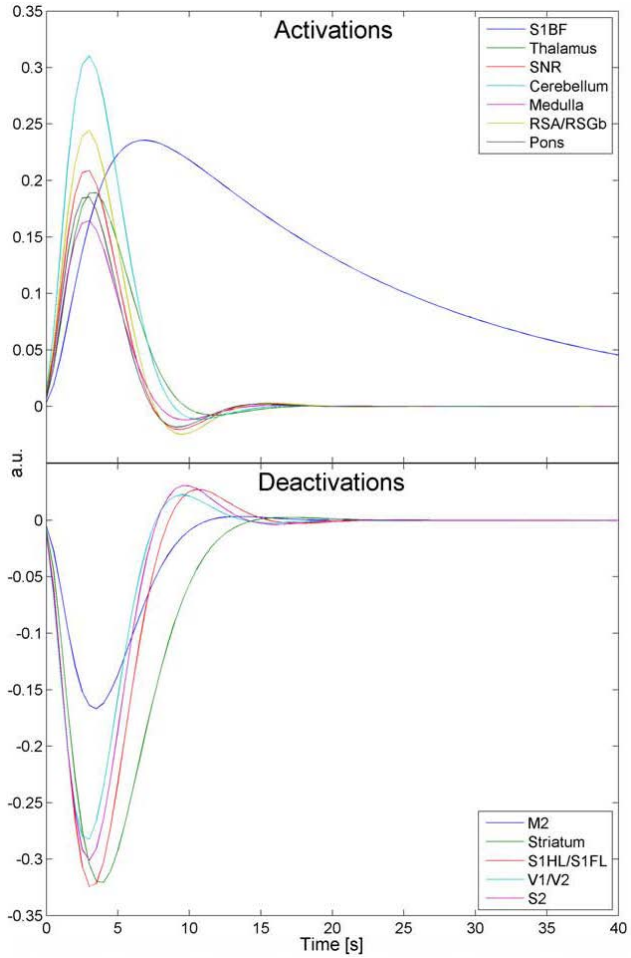
The hemodynamic model in DCM – role of ε



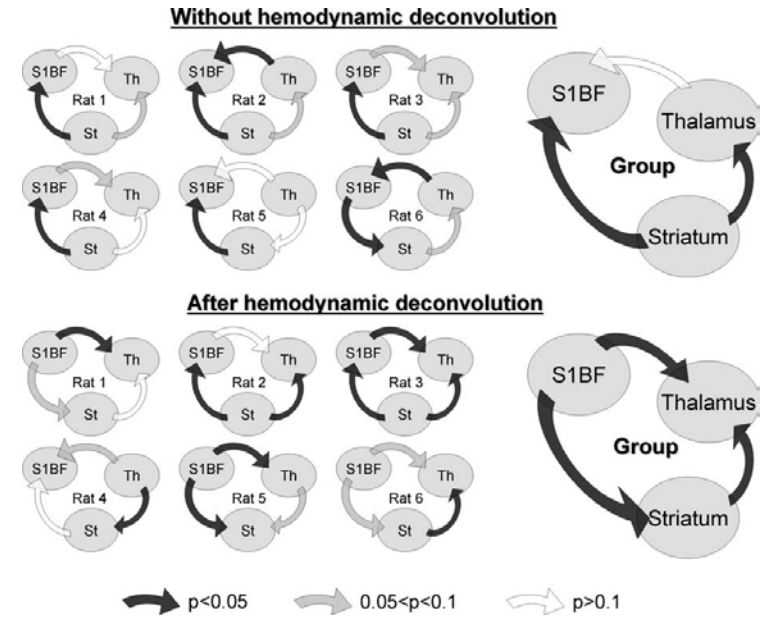


Stephan et al. 2007, *NeuroImage*

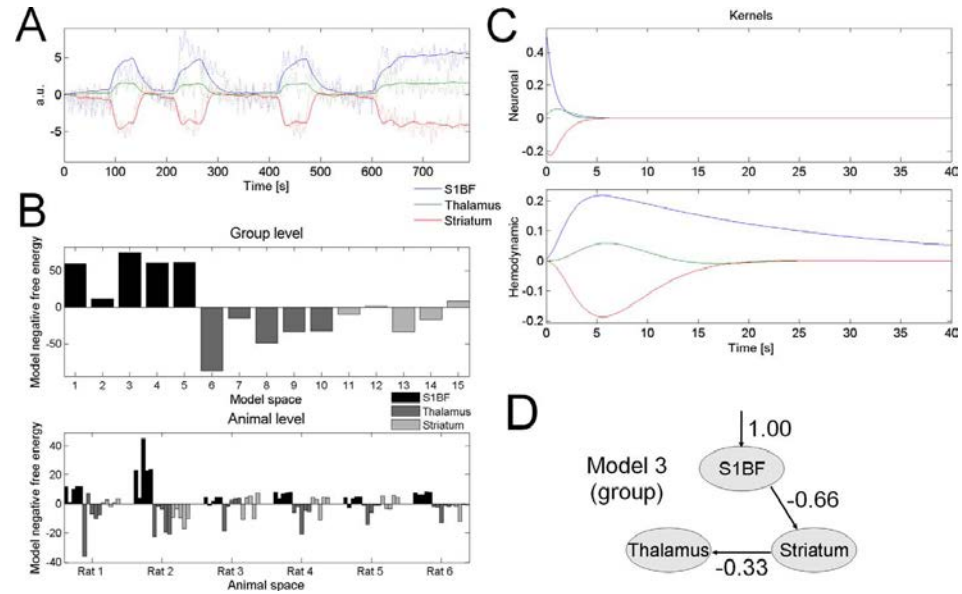
Hemodynamic forward models are important for connectivity analyses of fMRI data



Granger causality

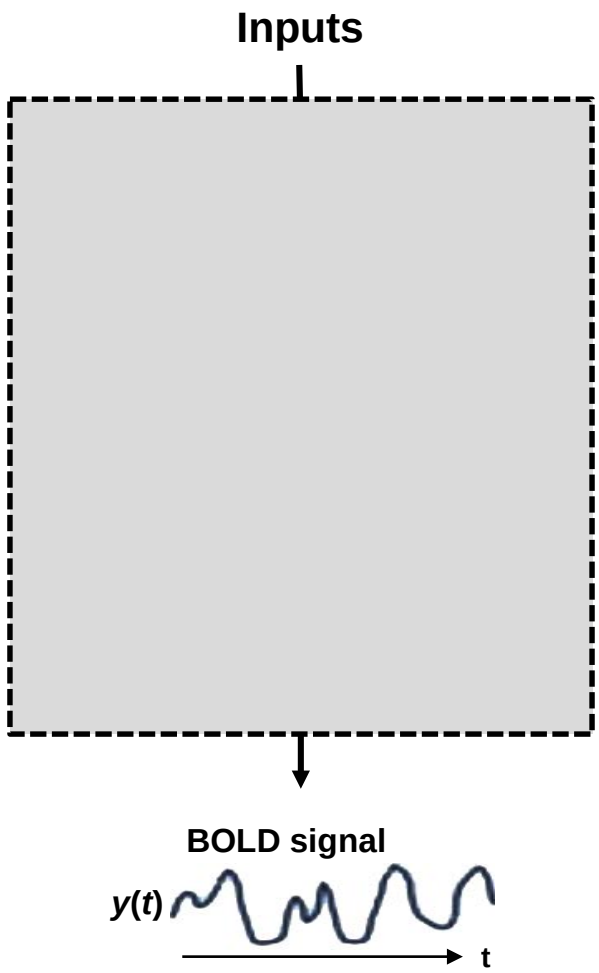


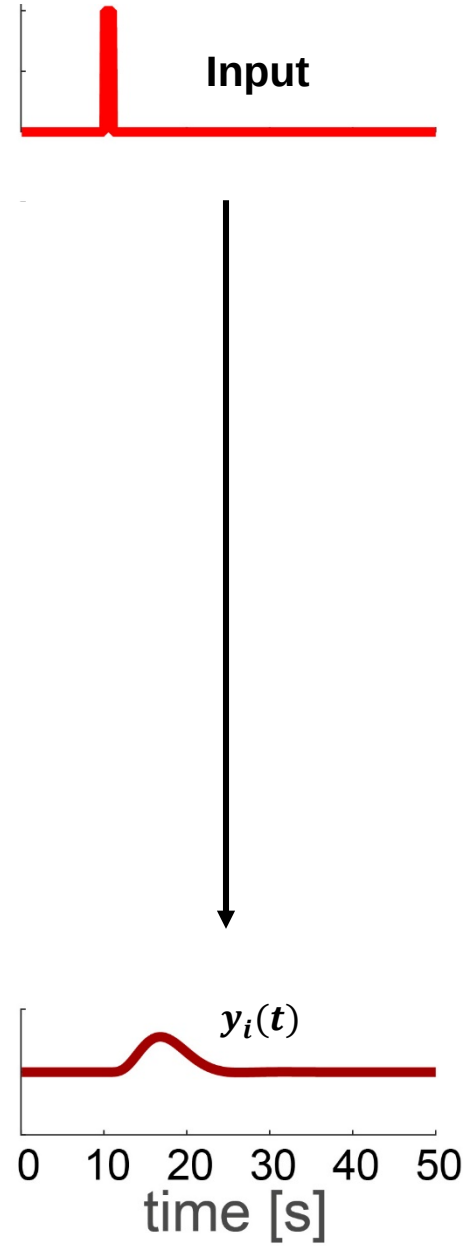
DCM



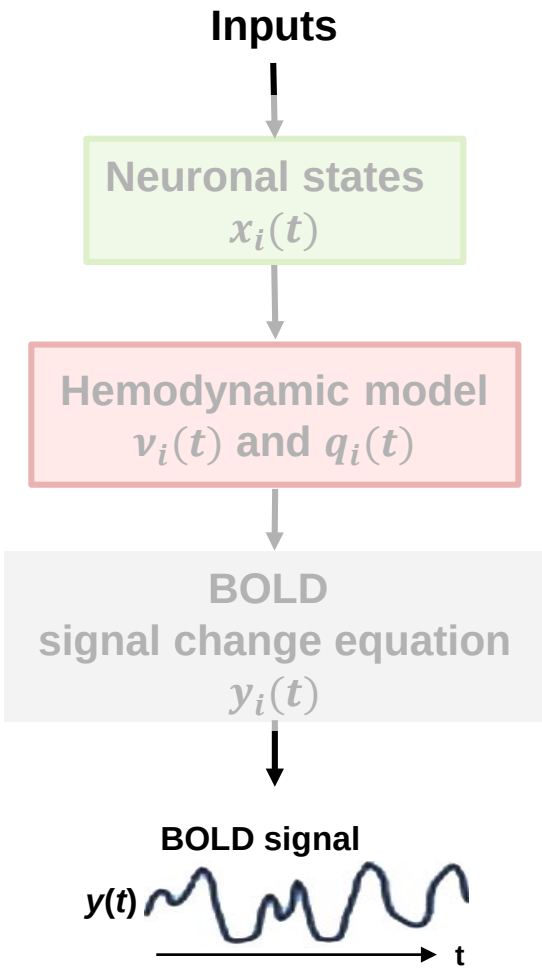
David et al. 2008, *PLoS Biol.*

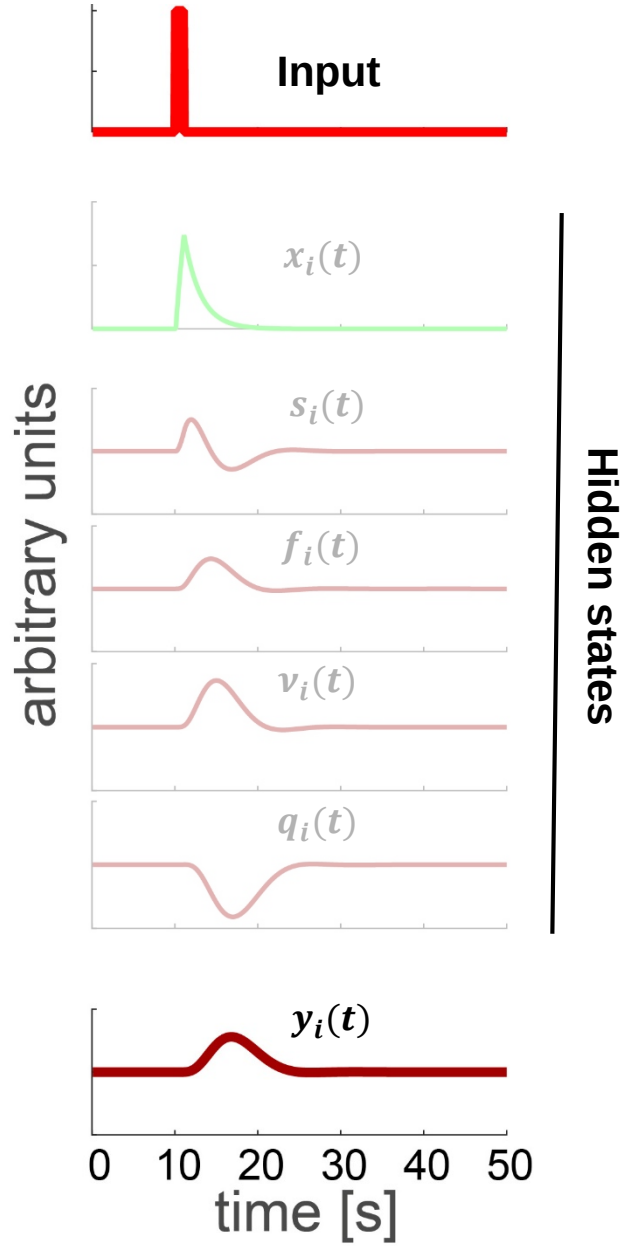
Summary – the full model



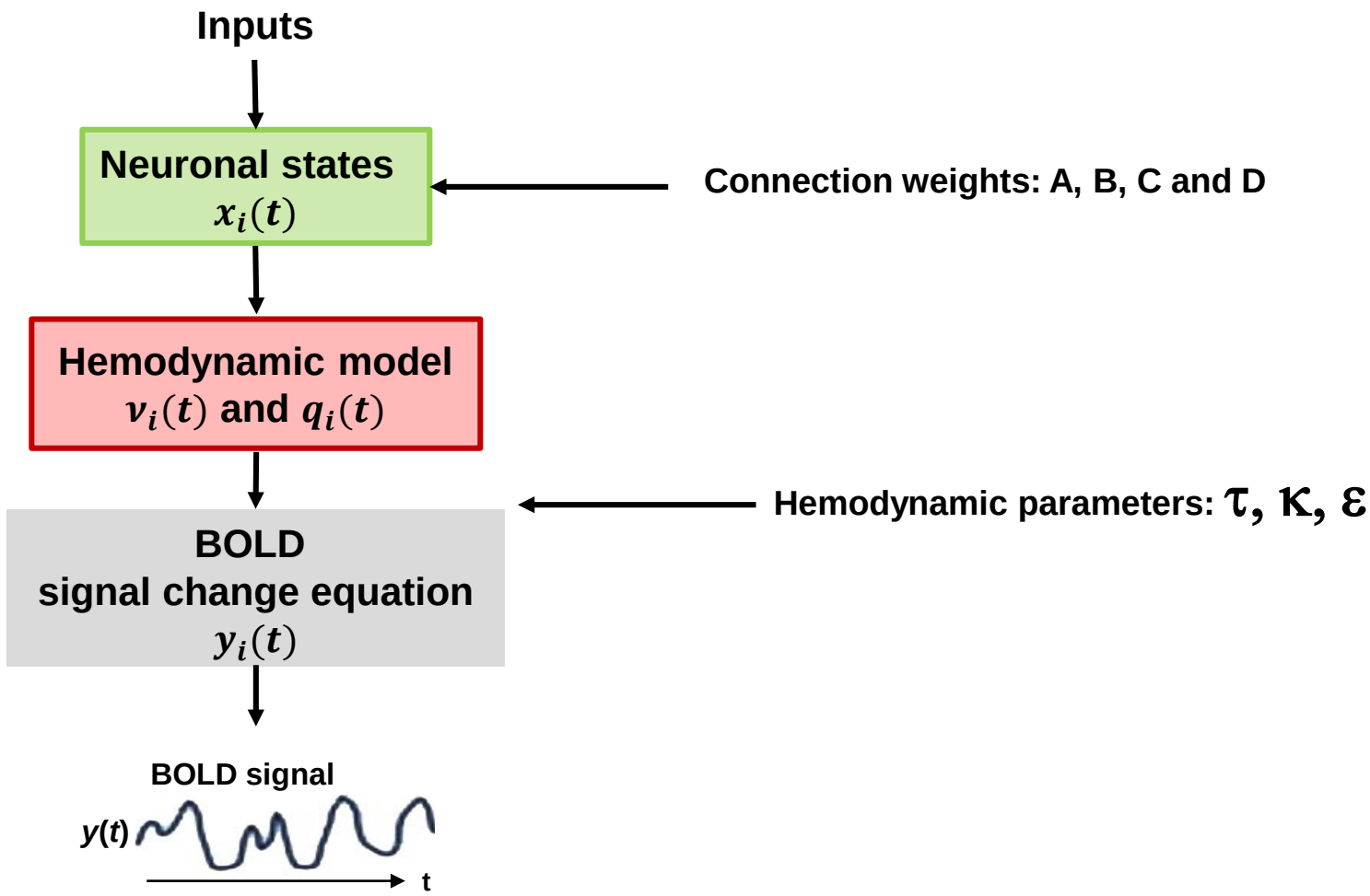


Summary – the full model

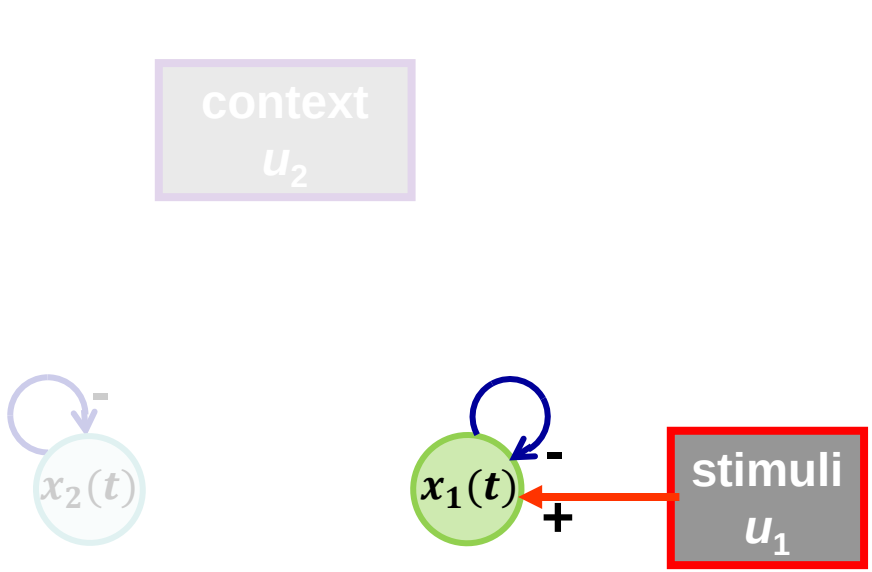


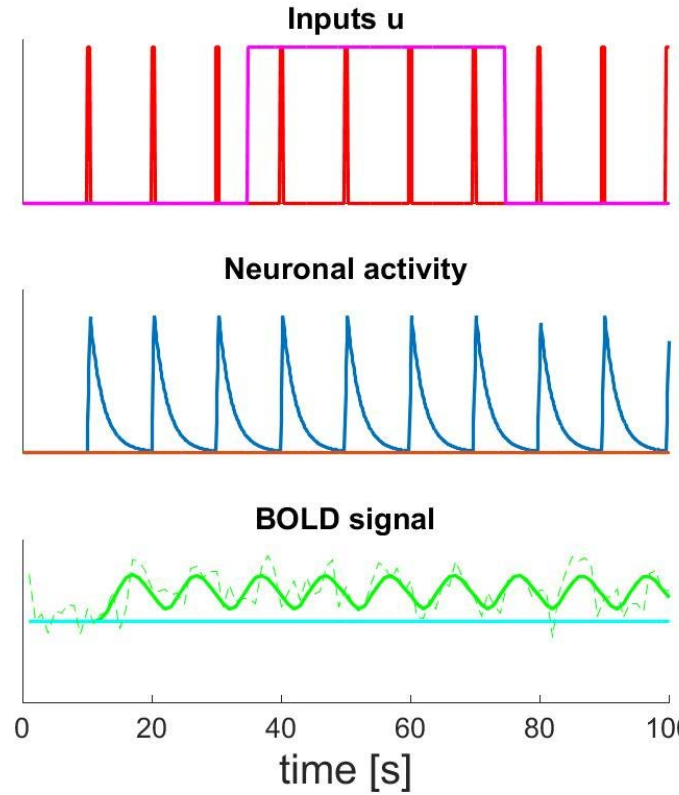


Summary – parameters of interest



Example traces 1: Single node

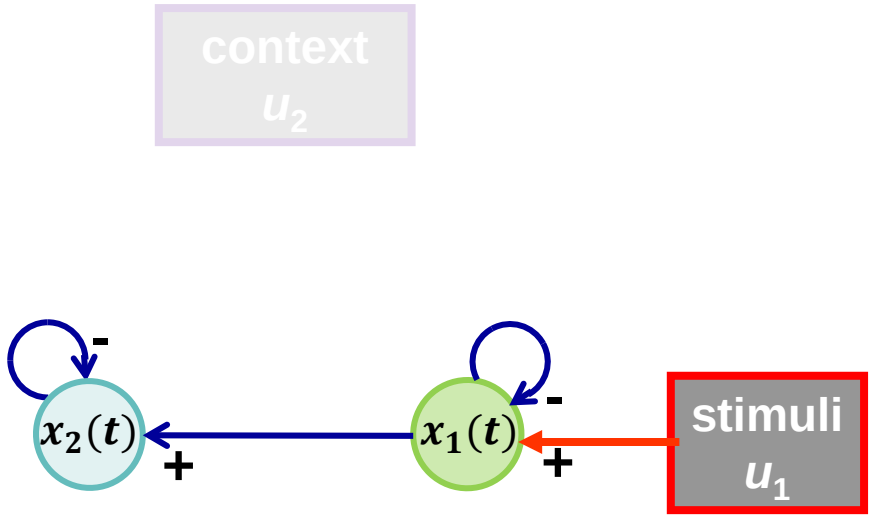


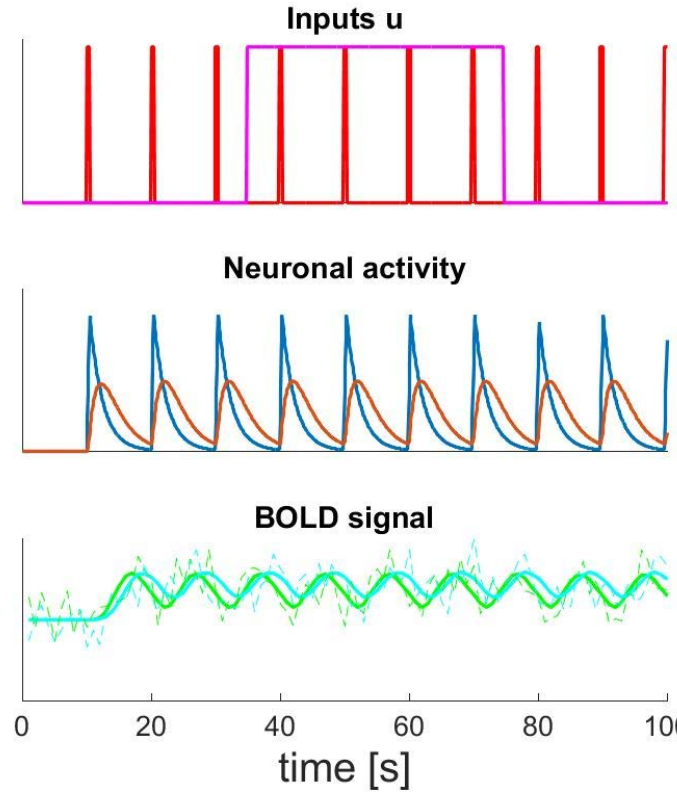


$$\dot{x} = Ax + u_2 B^{(2)} x + Cu_1$$

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} \sigma & 0 \\ 0 & \sigma \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + u_2 \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} c_{11} & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

Example traces 2: Connected nodes

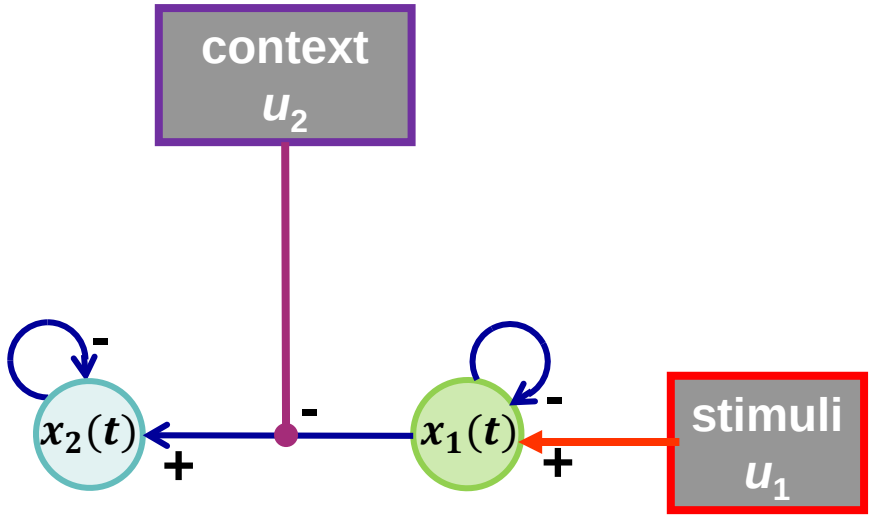


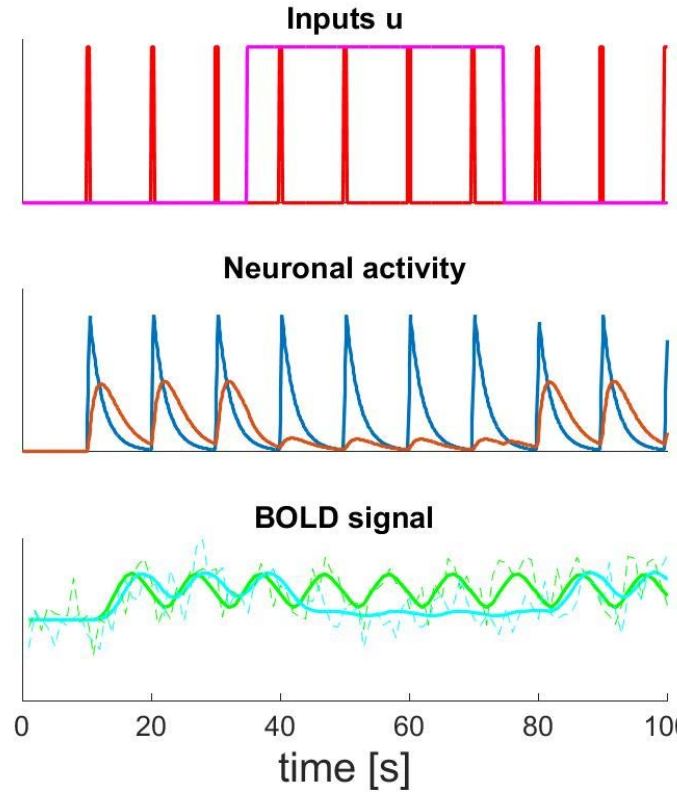


$$\dot{x} = Ax + u_2 B^{(2)} x + C u_1$$

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} \sigma & 0 \\ a_{12} & \sigma \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + u_2 \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} c_{11} & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

Example traces 3: Modulation of connection

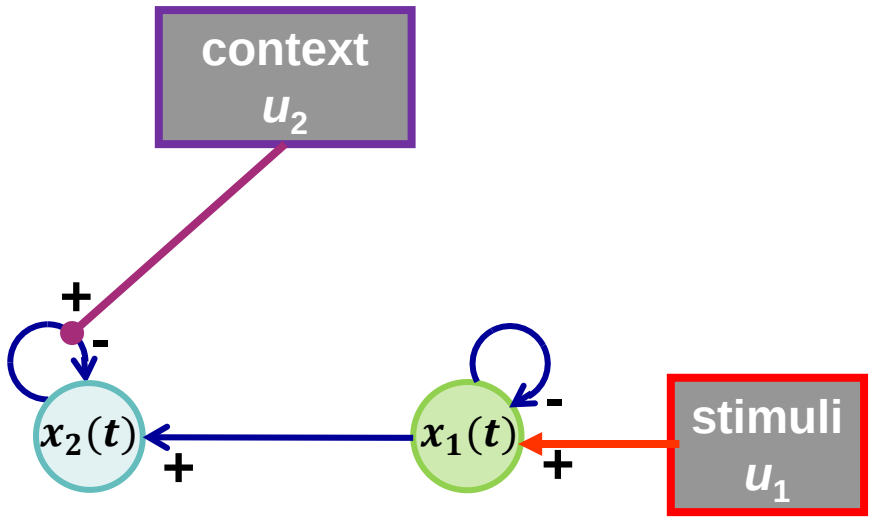


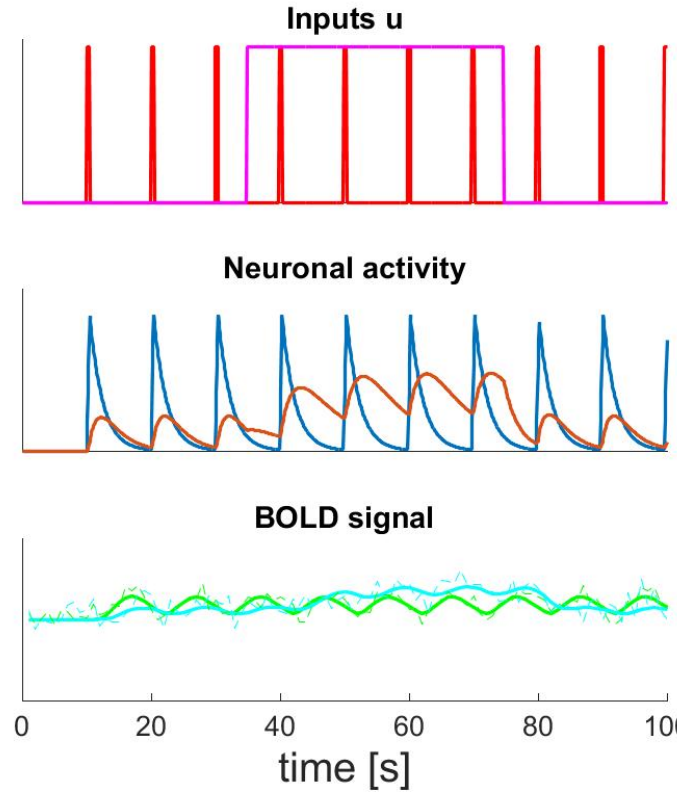


$$\dot{x} = Ax + u_2 B^{(2)} x + Cu_1$$

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} \sigma & 0 \\ a_{12} & \sigma \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + u_2 \begin{bmatrix} 0 & 0 \\ b_{12}^{(2)} & 0 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} c_{11} & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

Example traces 4: Modulation of self-connection

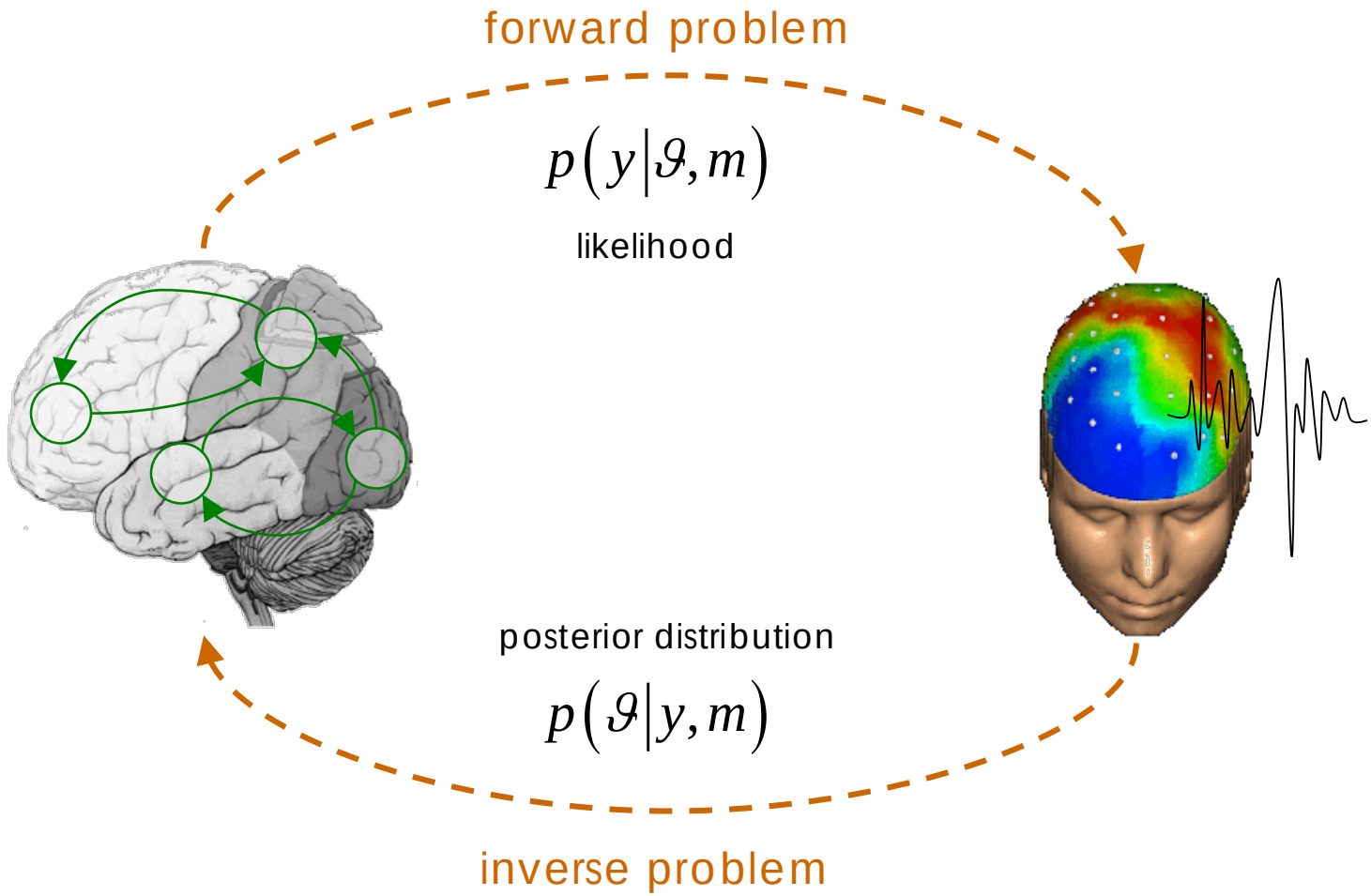




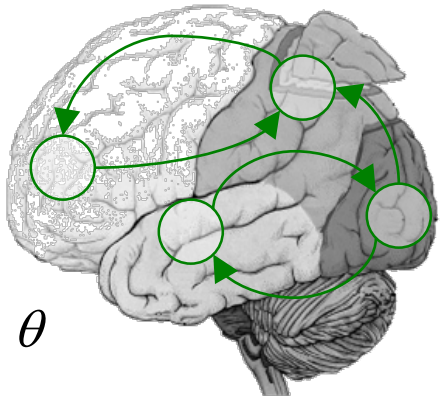
$$\dot{x} = Ax + u_2 B^{(2)} x + C u_1$$

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} \sigma & 0 \\ a_{12} & \sigma \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + u_2 \begin{bmatrix} 0 & 0 \\ 0 & b_{22}^{(2)} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} c_{11} & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

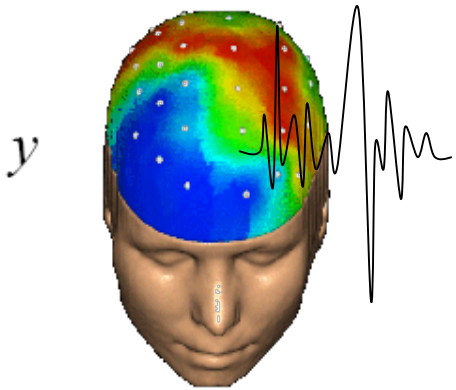
Bayesian inference – forward and inverse model



Bayesian inference



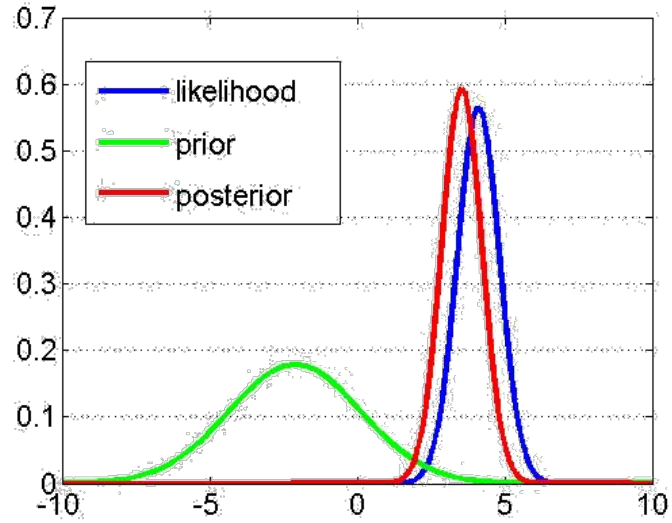
generative model *m*



Likelihood: $p(y|\theta,m)$

Prior: $p(\theta|m)$

Bayes rule: $p(\theta|y,m)=\dfrac{p(y|\theta,m)\,p(\theta|m)}{p(y|m)}$



Dynamical systems in Bayes

Assume data is normally distributed around the prediction from the dynamical model.v

$$\longrightarrow \quad p(y(t)|\theta, m) = \mathcal{N}(y(t), \theta_\sigma)$$

Dynamical model defines the likelihood!

Combining likelihood (data) and priors

Likelihood = Probability of data

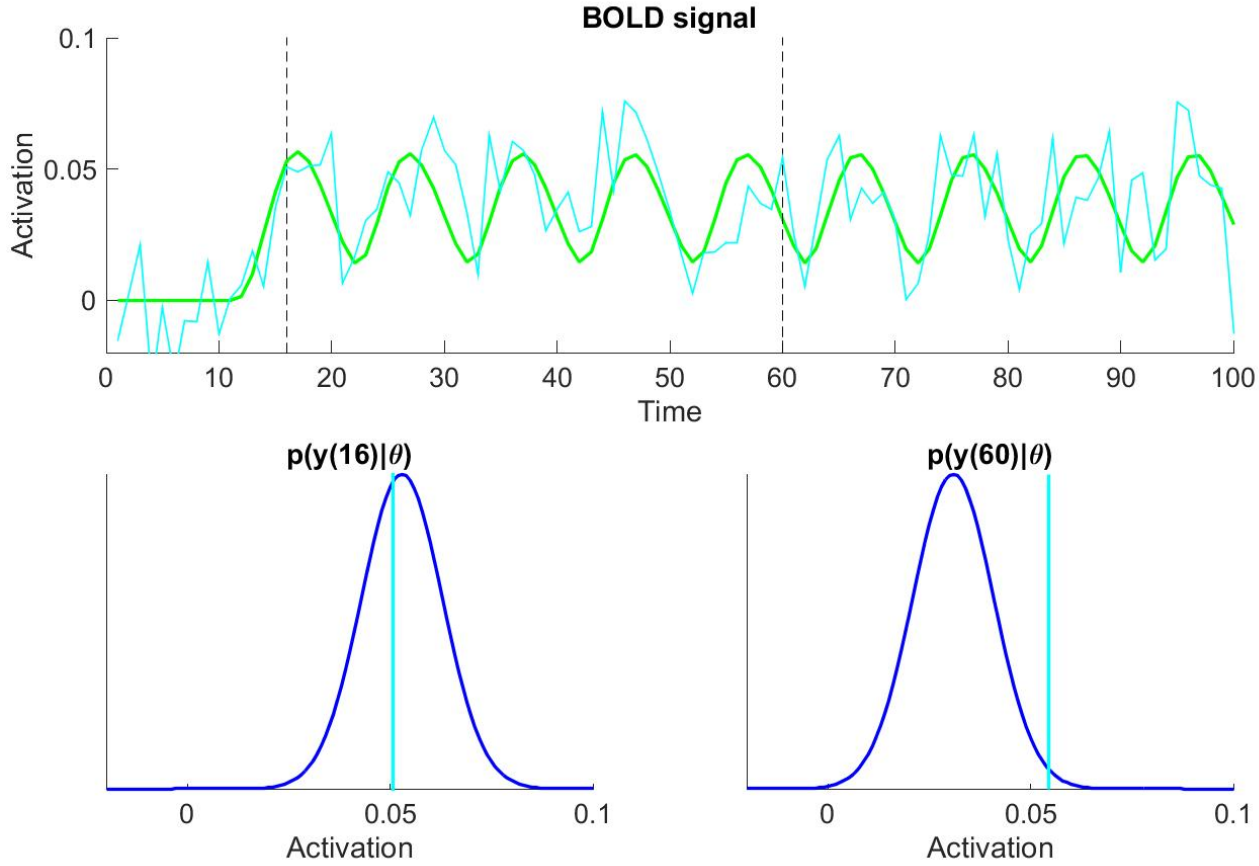
- Derived from dynamical system
  - Gaussian noise

Priors (constraints):

- Hemodynamic parameters
  - empirical
- Neural parameters
  - self connections: principled
  - other parameters (inputs, connections): shrinkage

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

The likelihood of the data



$$p(y|\theta, m) = \prod_t p(y(t)|\theta, m)$$

Type role and impact of priors

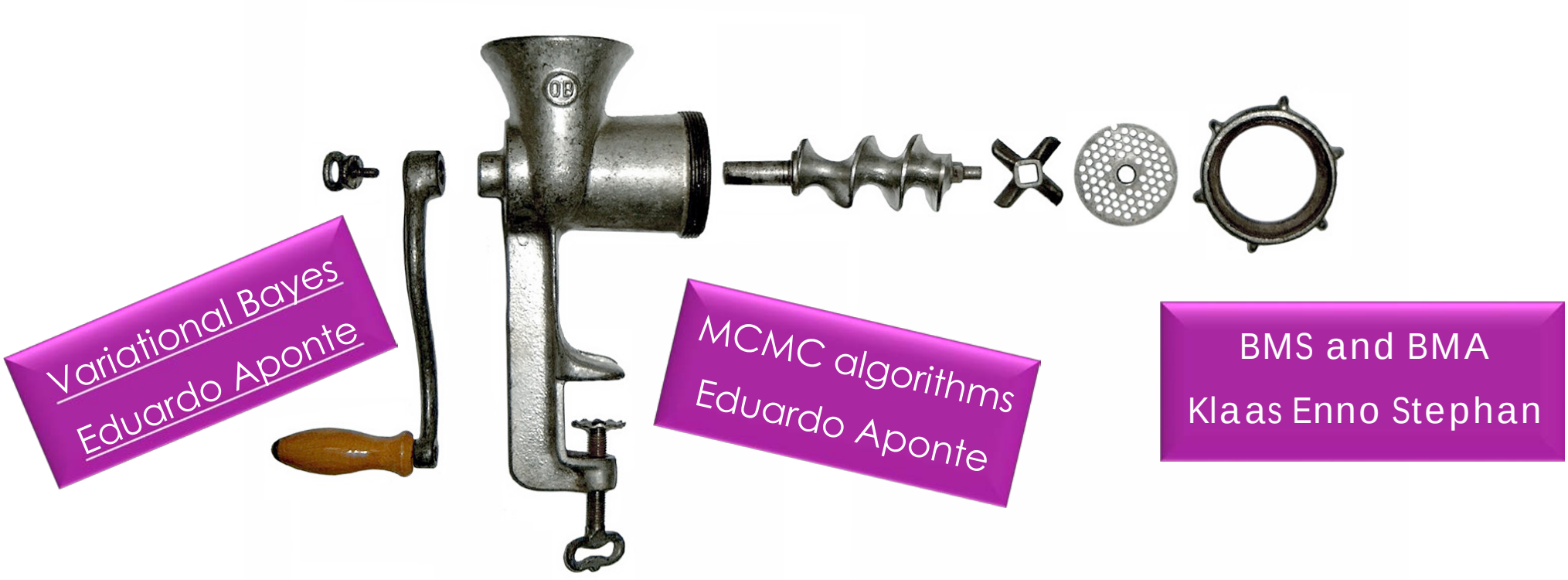
- **Types of priors:**
  - ✓ **Explicit priors on *model parameters* (e.g., connection strengths)**
  - ✓ **Implicit priors on *model functional form* (e.g., system dynamics)**
  - ✓ **Choice of "interesting" *data features* (e.g., regional timeseries vs ICA components)**

- **Role of priors (on model parameters):**
  - ✓ **Resolving the *ill-posedness* of the inverse problem**
  - ✓ **Avoiding *overfitting* (cf. generalization error)**

- **Impact of priors:**
  - ✓ **On parameter posterior distributions (cf. "shrinkage to the mean" effect)**
  - ✓ **On model evidence (cf. "Occam's razor")**
  - ✓ **On free-energy landscape (cf. Laplace approximation)**

Model estimation: running the machinery

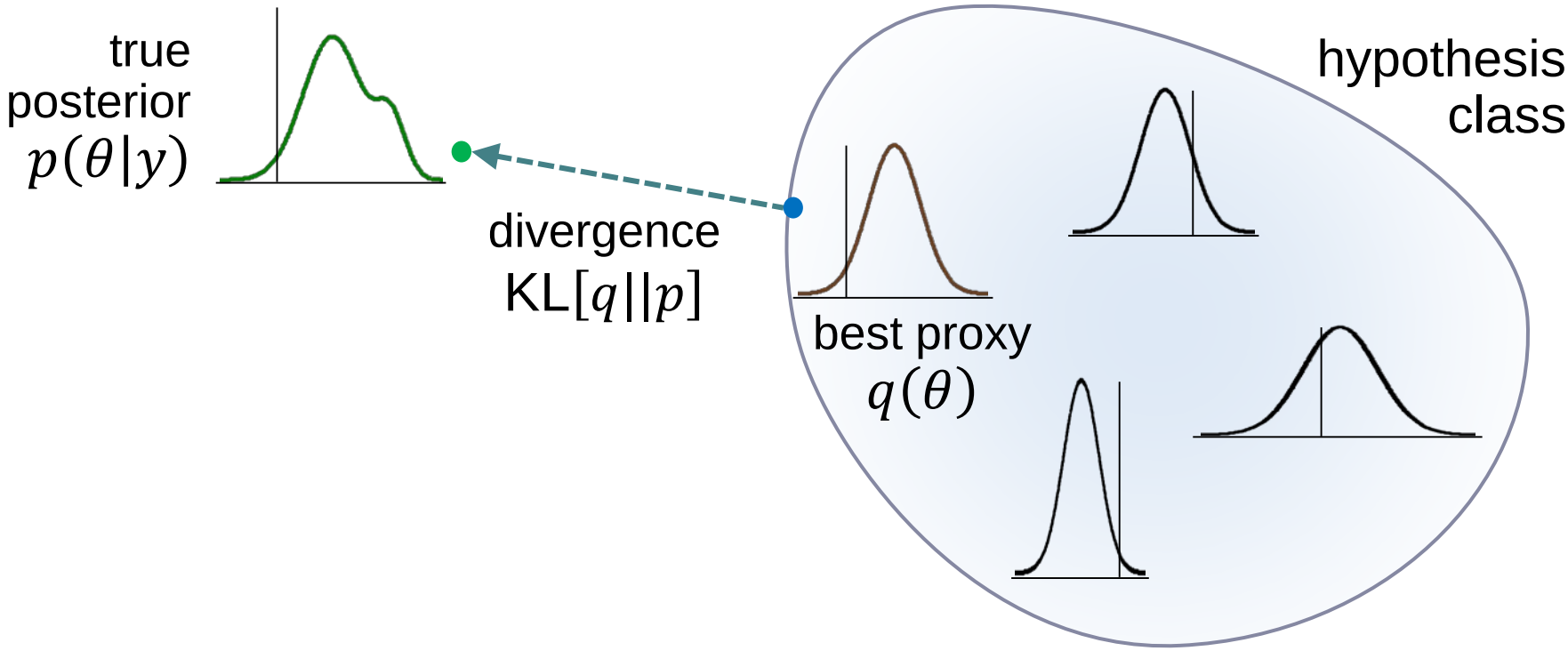
- Goal: Find posterior of parameters $p(\theta|y,m)$ that maximises model evidence $p(y|m)$ given data and priors
  - This is often not possible analytically → approximate methods are used



Variational Bayes (VB)

Idea: find an approximate density $q(\theta)$ that is maximally similar to the true posterior $p(\theta|y)$.

This is often done by assuming a particular form for q (fixed form VB) and then optimizing its sufficient statistics.



Variational Bayes

$$\ln p(y) = \underbrace{\mathrm{KL}[q||p]}_{\text{divergence}} + \underbrace{F(q,y)}_{\text{neg. free energy}}$$

divergence ≥ 0 (unknown)

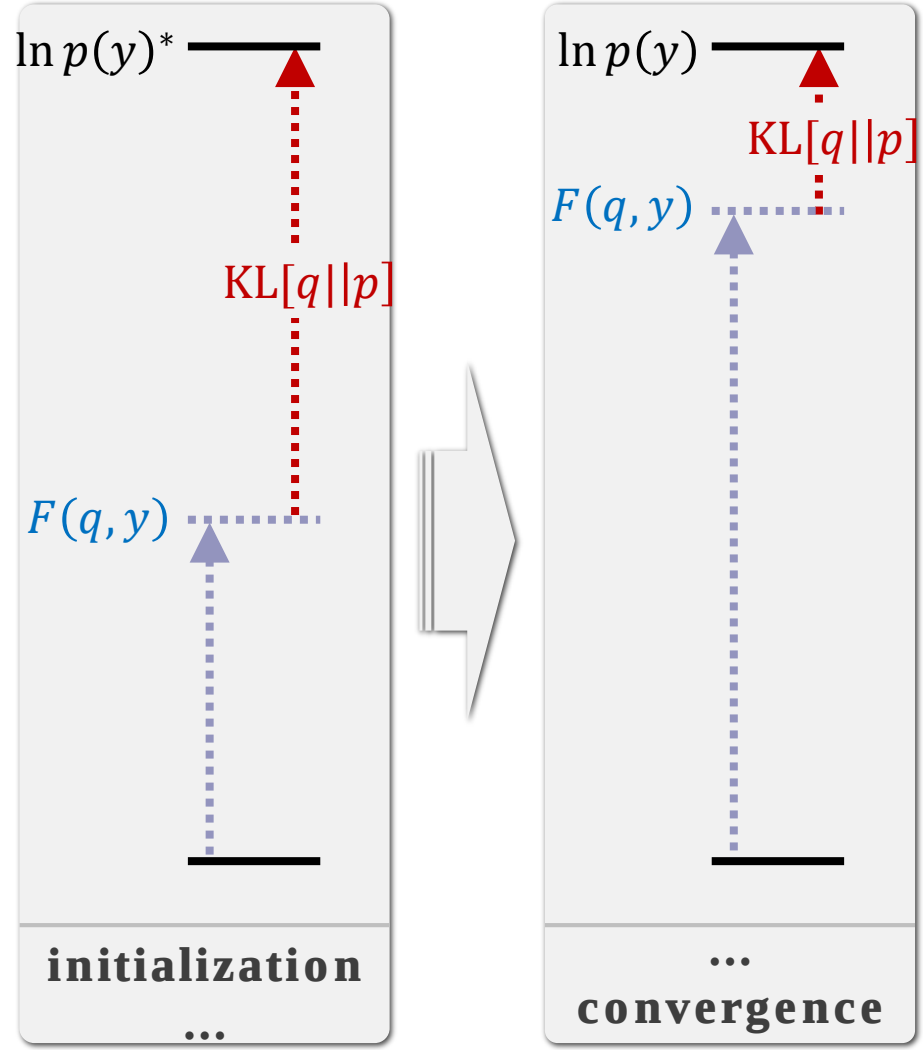
neg. free energy (easy to evaluate for a given q)

$F(q,y)$ is a functional wrt. the approximate posterior $q(\theta)$.

Maximizing $F(q,y)$ is equivalent to:

- minimizing $\mathrm{KL}[q||p]$
- tightening $F(q,y)$ as a lower bound to the log model evidence

When $F(q,y)$ is maximized, $q(\theta)$ is our best estimate of the posterior.



Bayesian system identification

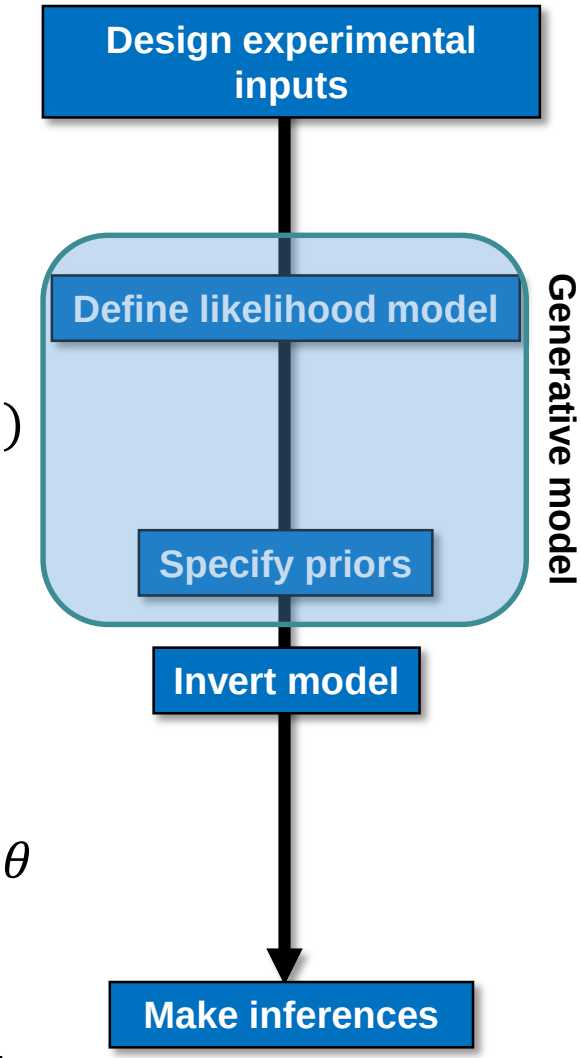
$u(t)$

Neural dynamics $\dot{x} = f(x, u, \theta_x)$

Observer function $y = g(x, \theta_y) + \varepsilon$

$$p(y|\theta, m) = \mathcal{N}(g(x, \theta_y), \Sigma(\theta_\sigma))$$

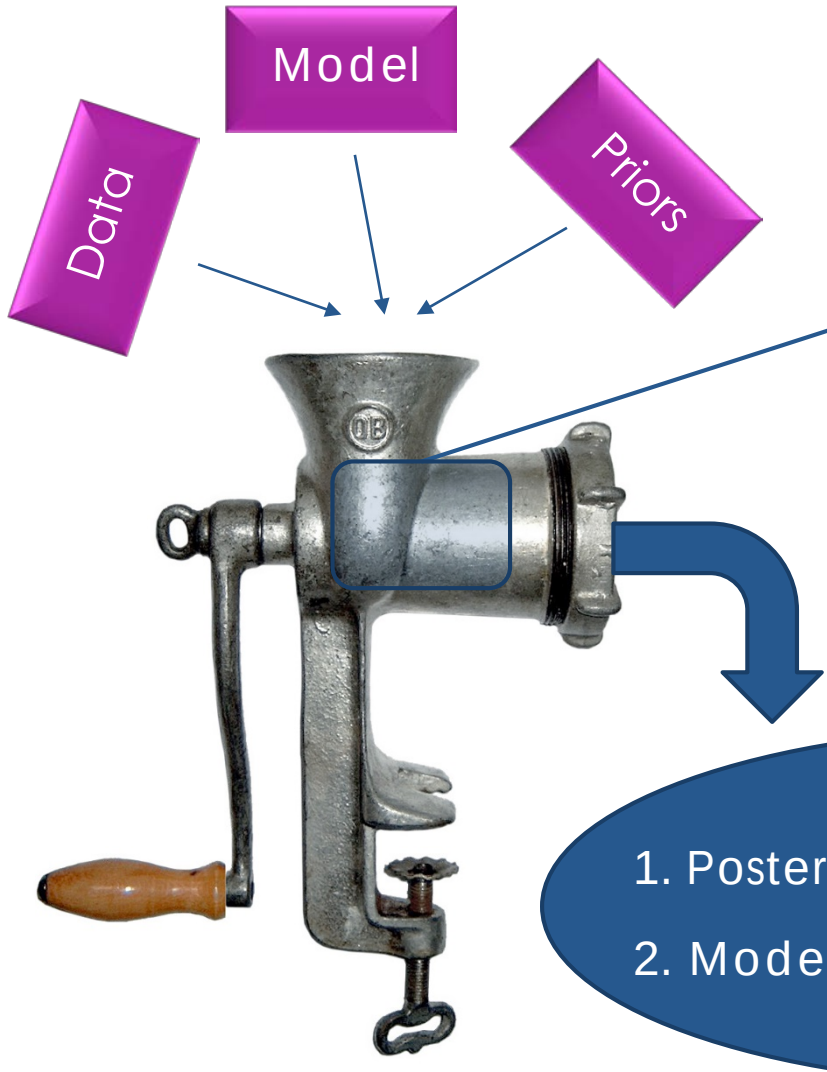
$$p(\theta|m) = \mathcal{N}(\mu_\theta, \Sigma_\theta)$$



Inference on model structure $p(y|m) = \int p(y|\theta, m)p(\theta|m)d\theta$

Inference on parameters $p(\theta|y, m) = \dfrac{p(y|\theta, m)p(\theta|m)}{p(y|m)}$

Model estimation machinery



Expectation-Maximization algorithm

Iterative procedure:

1. Compute model response using current set of parameters
2. Compare model response with data
3. Improve parameters, if possible

1. Posterior parameters - $\boldsymbol{p(\theta|y,m)}$
2. Model evidence - $\boldsymbol{p(y|m)}$

Generative models & model selection

- any DCM = a particular generative model of how the data (may) have been caused
- generative modelling: comparing competing hypotheses about the mechanisms underlying observed data

  $\rightarrow$ a priori definition of hypothesis set (model space) is crucial

  $\rightarrow$ determine the most plausible hypothesis (model), given the data

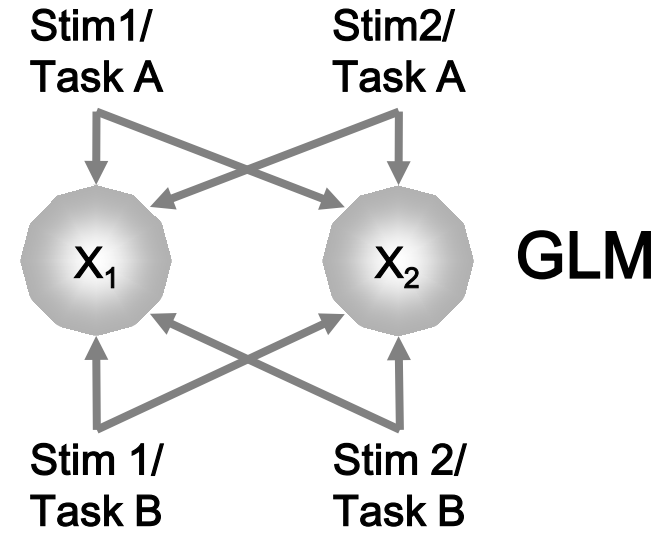
- model selection $\neq$ model validation!

  $\rightarrow$ model validation requires external criteria (external to the measured data)

Multifactorial design: explaining interactions with DCM

Stimulus factor	Task A	Task B
Stim 1	T_A/S_1	T_B/S_1
Stim 2	T_A/S_2	T_B/S_2

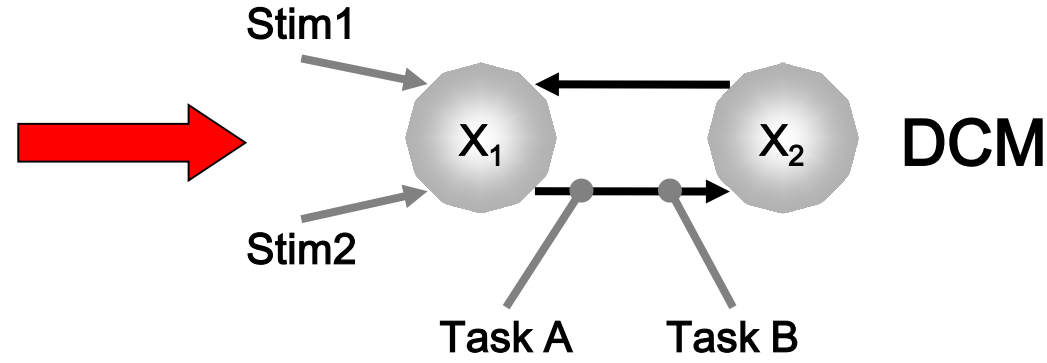
Task factor



GLM

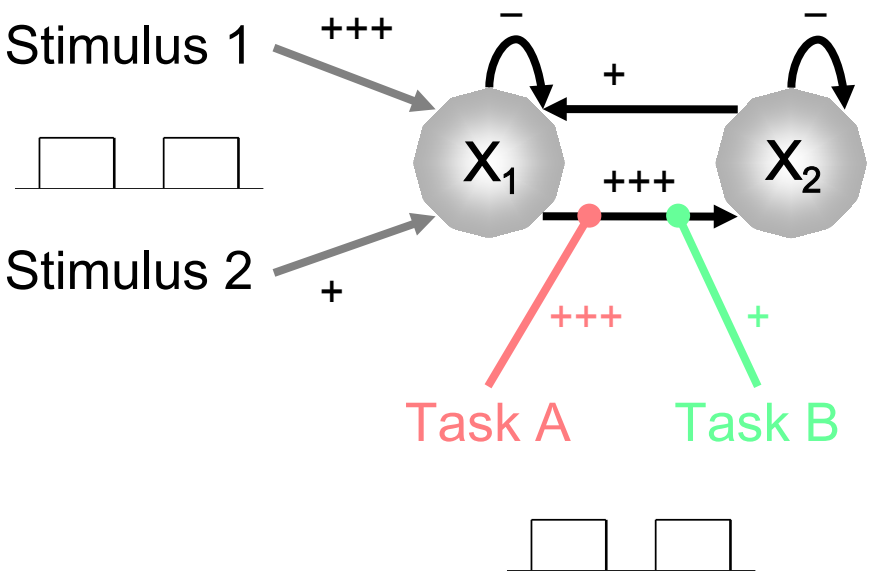
Let's assume that an SPM analysis shows a main effect of stimulus in X_1 and a stimulus × task interaction in X_2.

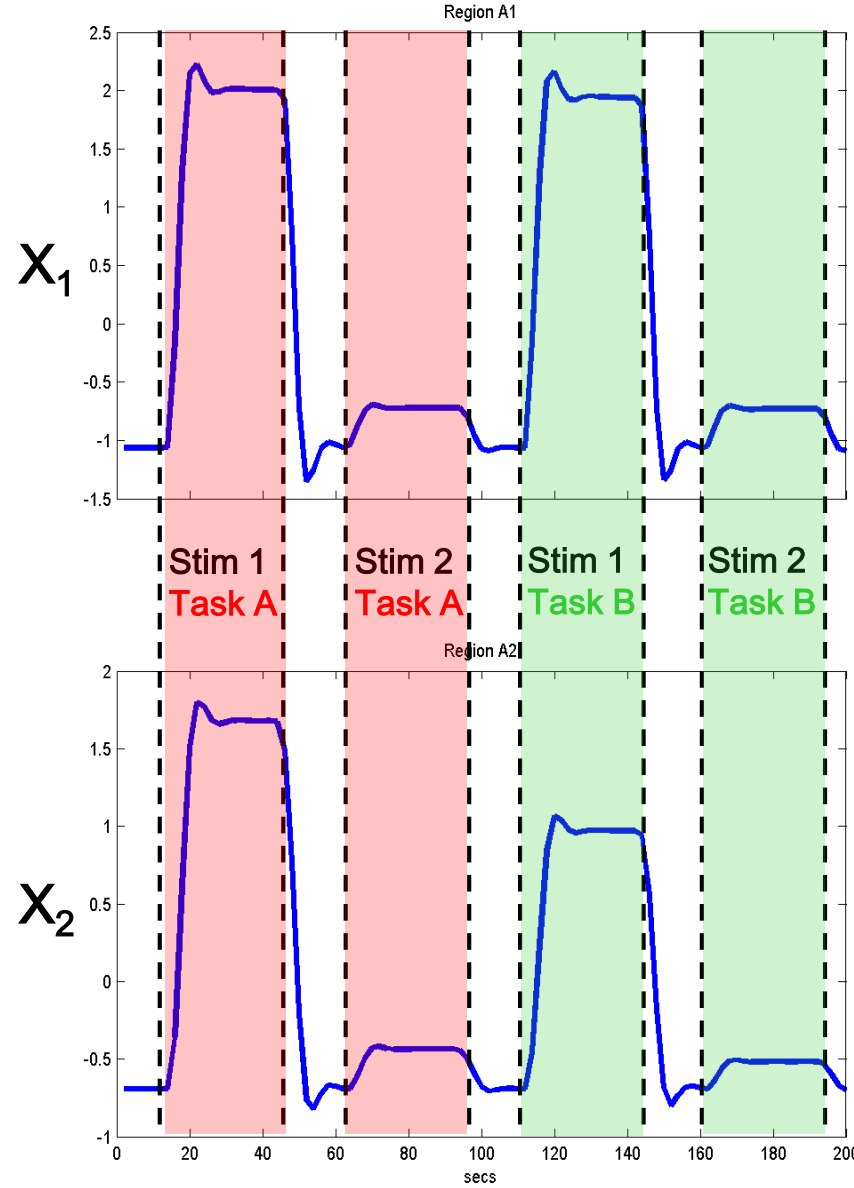
How do we model this using DCM?



DCM

Simulated data





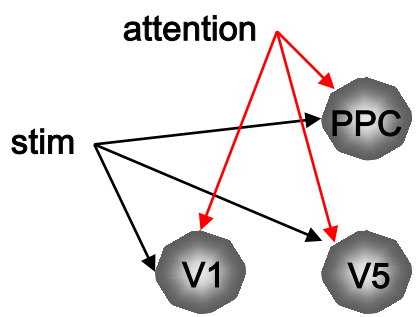
Stephan et al. 2007, *J. Biosci.*

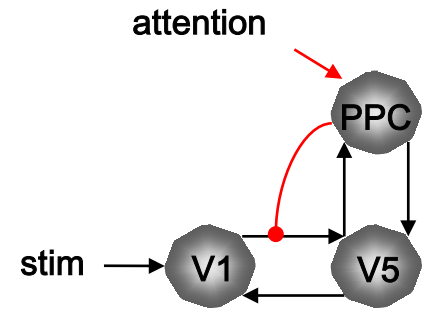
Note: GLM vs. DCM

DCM tries to model the same phenomena (i.e. local BOLD responses) as a GLM, just in a different way (via connectivity and its modulation).

No activation detected by a GLM
→ no motivation to include this region in a deterministic DCM.

However, a stochastic DCM could be applied despite the absence of a local activation.

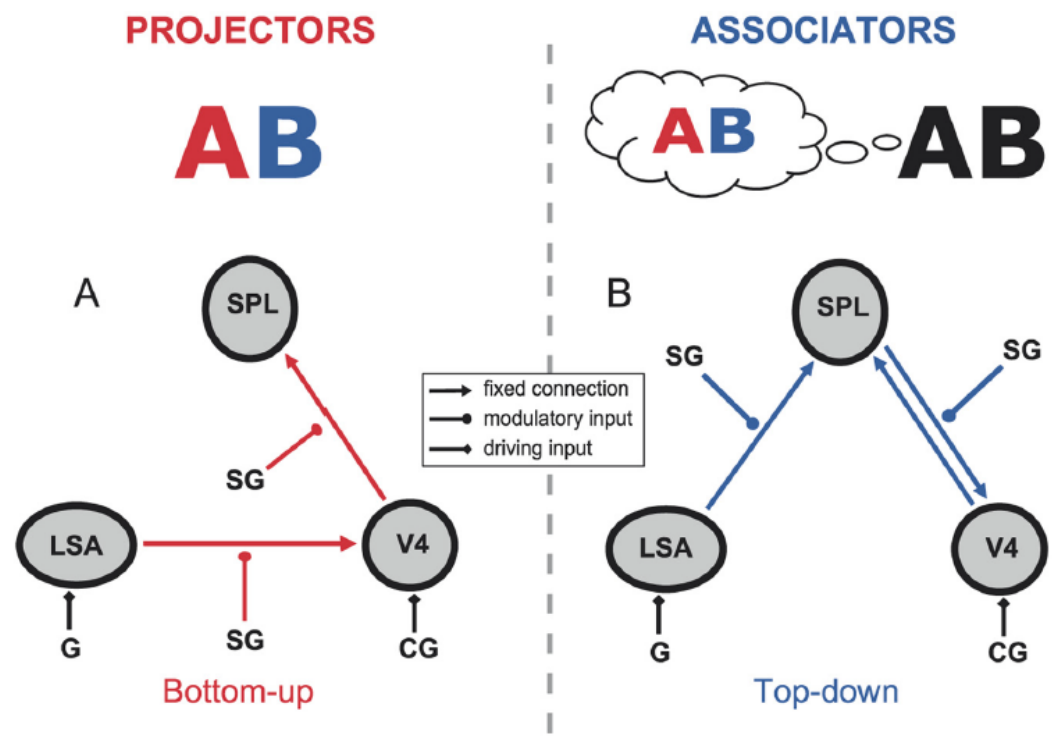




Stephan (2004) ***J. Anat.***

Model comparison: Synesthesia

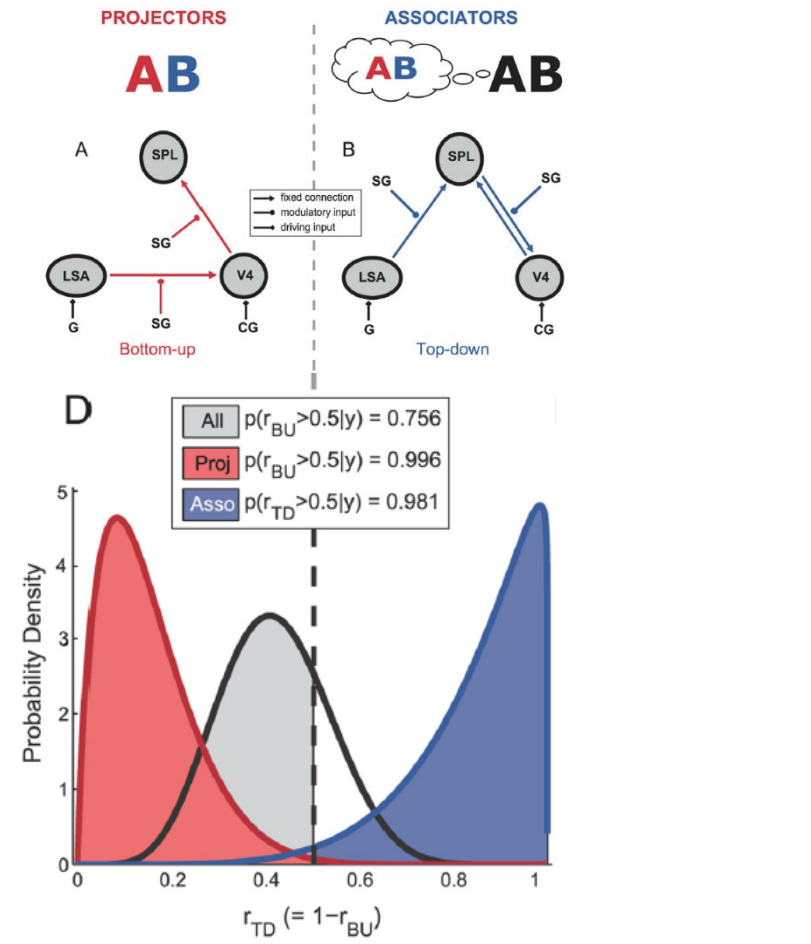
- “projector” synesthetes experience color externally co-localized with a presented grapheme
- “associators” report an internally evoked association



van Leeuwen et al., J Neurosci 2011

Model comparison: Synesthesia

- “projector” synesthetes experience color externally co-localized with a presented grapheme
- “associators” report an internally evoked association
- across all subjects: no evidence for either model
- but splitting into synesthesia types gives very strong evidence for bottom-up (projectors) and top-down (associators) mechanisms, respectively

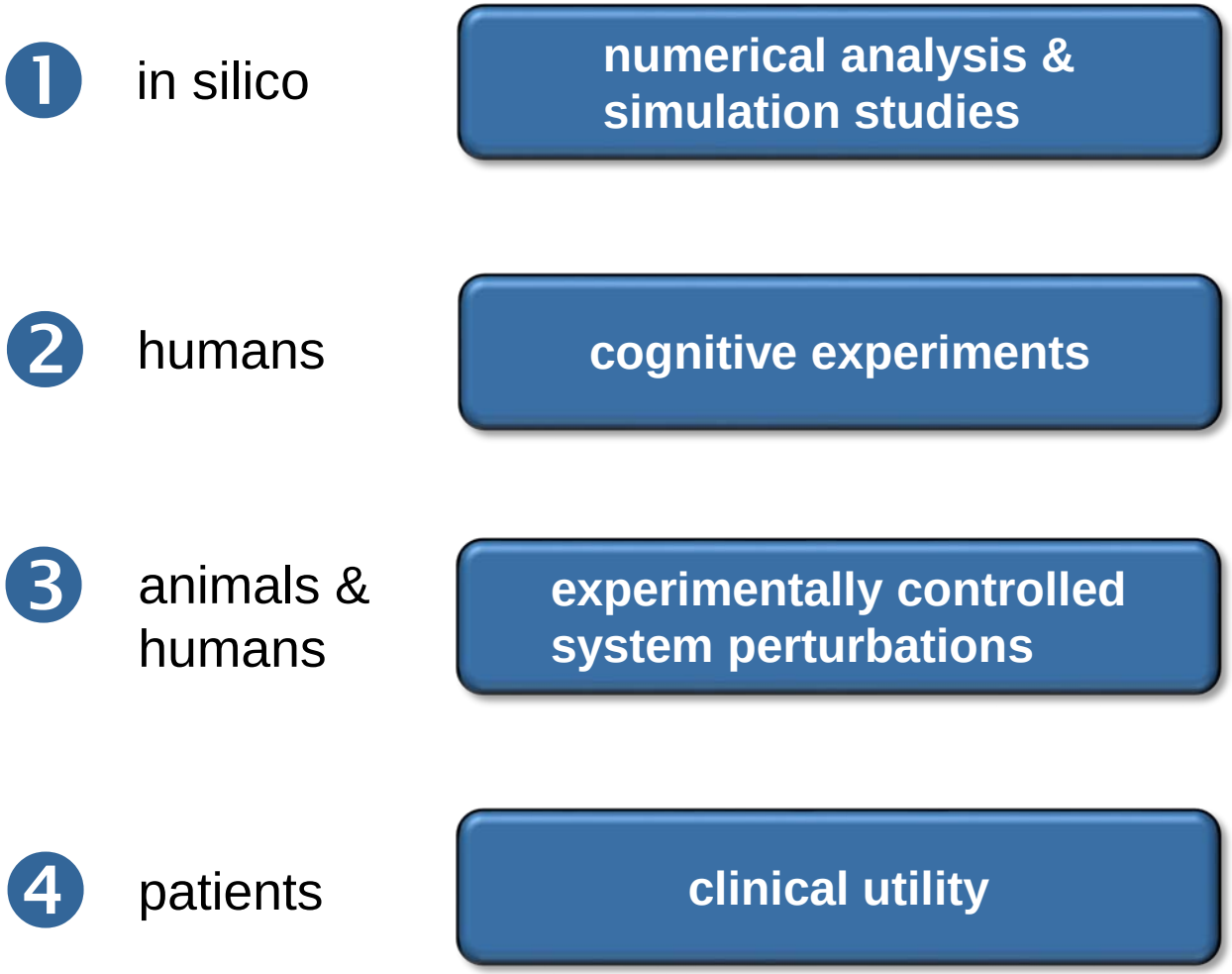


van Leeuwen et al., J Neurosci 2011

**“All models are wrong,
but some are useful.”**

George E.P. Box (1919-2013)

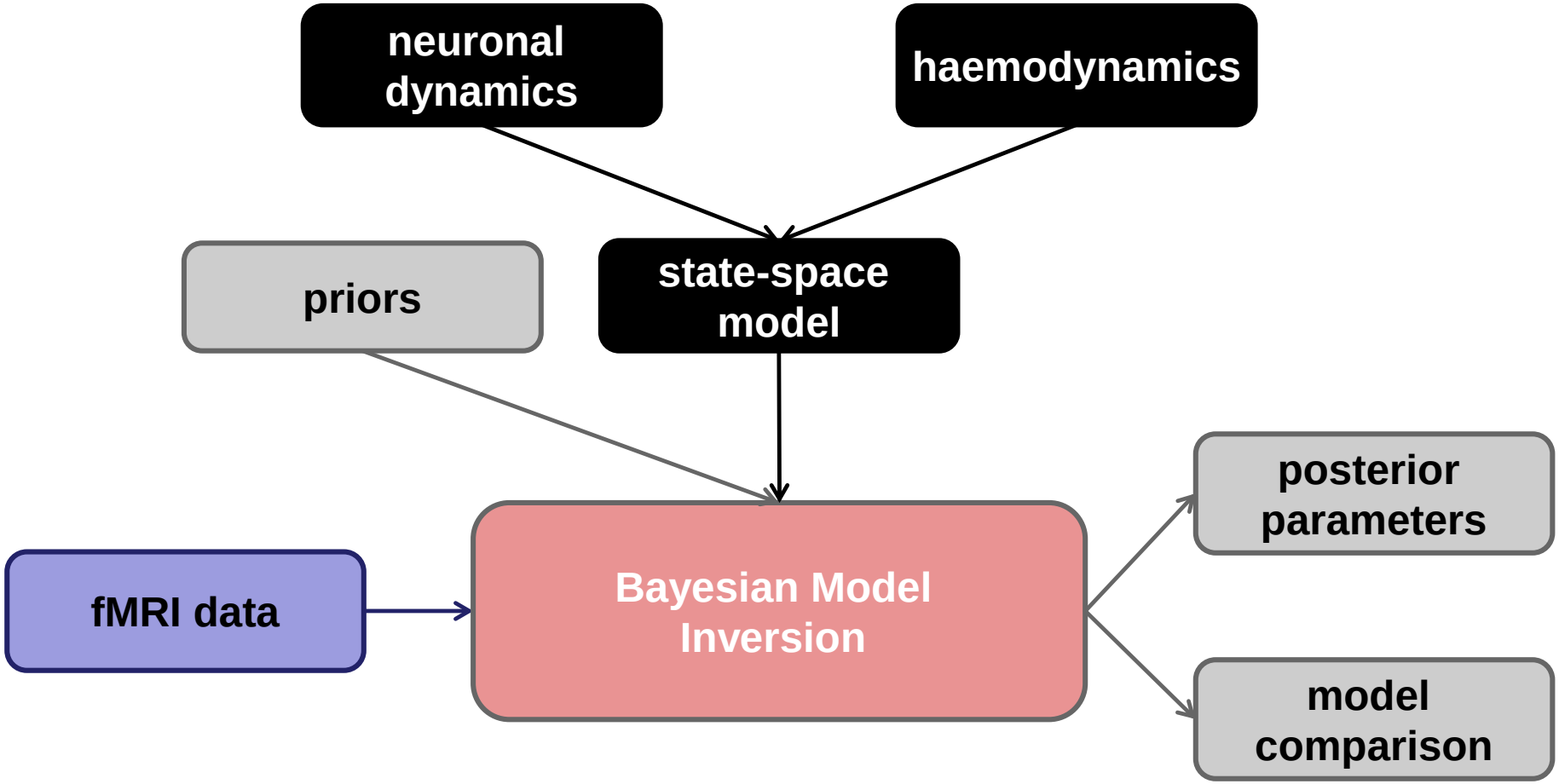
Hierarchical strategy for model validation



For DCM: >15 published validation studies (incl. 6 animal studies):

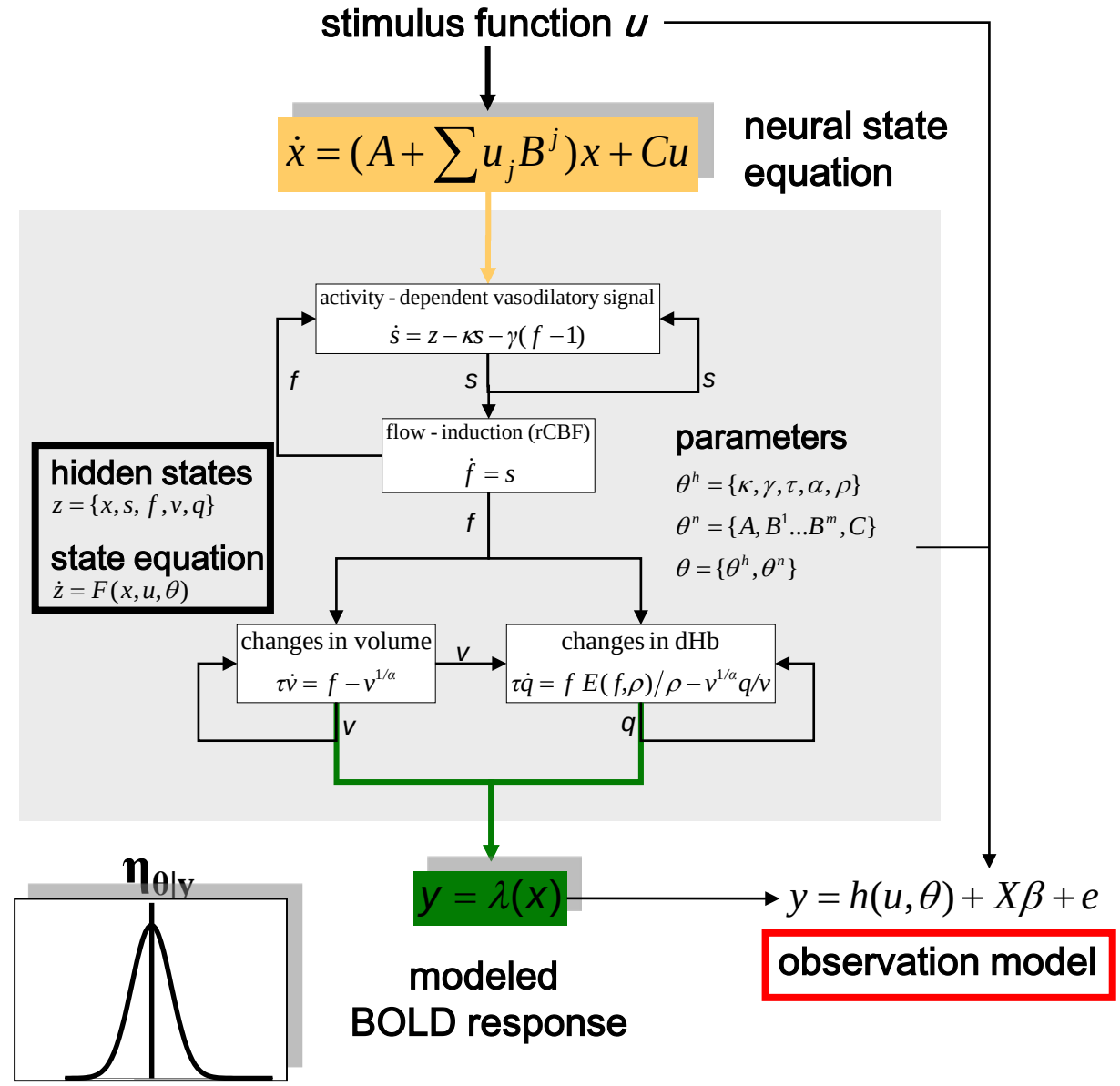
- infers site of seizure origin (David et al. 2008)
- infers primary recipient of vagal nerve stimulation (Reyt et al. 2010)
- infers synaptic changes as predicted by microdialysis (Moran et al. 2008)
- infers fear conditioning induced plasticity in amygdala (Moran et al. 2009)
- tracks anaesthesia levels (Moran et al. 2011)
- predicts sensory stimulation (Brodersen et al. 2010)

DCM – graphical overview



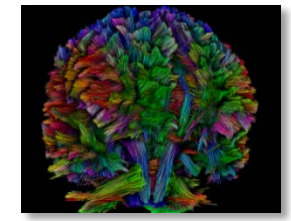
One slide summary

- Combining the neural and hemodynamic states gives the complete forward model.
- Observation model includes measurement error *e* and confounds *X* (e.g. drift).
- Bayesian inversion: parameter estimation variational Bayes or MCMC
- Result 1: A posteriori parameter distributions $p(\theta|y,m)$, characterised by mean $\eta_{\theta|y}$ and covariance $C_{\theta|y}$.
- Result 2: Estimate of model evidence $p(y|m)$.



Summary

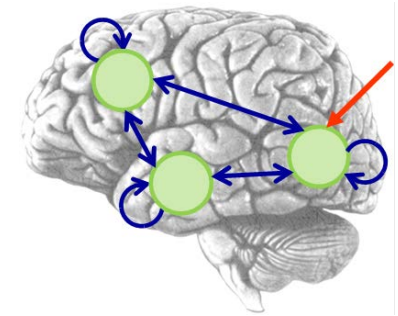
dwMRI



fMRI

$$p(y \mid \theta, m) \cdot p(\theta \mid m)$$

$$p(\theta \mid y, m)$$



1. **enforces mechanistic thinking: how could the data have been caused?**
2. **mechanistically explains observed effects (modulations).**
3. **generate synthetic data (observations) by sampling from the prior – can model explain certain phenomena at all?**
4. **inference about model structure: formal approach to disambiguating mechanisms → p(m|y)**
5. **inference about parameters → p(θ|y)**

Many thanks to Hanneke den Ouden, Andreea Diaconescu, Jean Daunizeau and Klaas Enno Stephan for some of the slides!

Thank you!

Useful references 1

- 10 Simple Rules for DCM (2010). Stephan et al. *NeuroImage 52*.
- The first DCM paper: Dynamic Causal Modelling (2003). Friston et al. *NeuroImage* 19:1273-1302.
- Physiological validation of DCM for fMRI: Identifying neural drivers with functional MRI: an electrophysiological validation (2008). David et al. *PLoS Biol.* 6 2683–2697
- Hemodynamic model: Comparing hemodynamic models with DCM (2007). Stephan et al. *NeuroImage* 38:387-401
- Nonlinear DCM:Nonlinear Dynamic Causal Models for FMRI (2008). Stephan et al. *NeuroImage* 42:649-662
- Two-state DCM: Dynamic causal modelling for fMRI: A two-state model (2008). Marreiros et al. *NeuroImage* 39:269-278
- Stochastic DCM: Generalised filtering and stochastic DCM for fMRI (2011). Li et al. *NeuroImage* 58:442-457.

Useful references 2

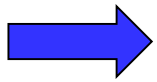
- Bayesian model comparison: Comparing families of dynamic causal models (2010). Penny et al. *PLoS Comp Biol.* 6(3):e1000709.
- A DCM for resting state fMRI (2014). Friston et al. *NeuroImage* 94:396–407
- A hemodynamic model for layered BOLD signals (2016). Heinzle et al. *NeuroImage* 125:556–570
- Stochastic Dynamic Causal Modelling of fMRI data: should we care about neural noise? (2012). Daunizeaue et al. *NeuroImage* 62:464-481.
- Optimizing experimental design for comparing models of brain function (2011). Daunizeau et al. *PLoS Comp Biol* 7(11):e1002280
- Dynamic Causal Modelling: a critical review of the biophysical and statistical foundations (2011). Daunizeau et al. *NeuroImage* 58:312-322.

DCM developments – for your reference

- **Nonlinear DCM for fMRI:** ***Could connectivity changes be mediated by another region?*** ***(Stephan et al. 2008, Neuroimage)***
- **Clustering DCM parameters:** ***Classify patients, or even find new sub-categories*** ***(Brodersen et al. 2011, Neuroimage)***
- **Embedding computational models in DCMs:** ***DCM can be used to make inferences on parametric designs like SPM*** ***(den Ouden et al. 2010, J Neuroscience)***
- **Integrating tractography and DCM:** ***Prior variance is a good way to embed other forms of information, test validity*** ***(Stephan et al. 2009, NeuroImage)***
- **Stochastic DCM:** ***Model resting state studies / background fluctuations*** ***(Li et al. 2011 Neuroimage, Daunizeau et al. Physica D 2009)***
- **Resting state DCM:** ***Model second order interactions directly*** ***(Friston et al. 2014, Neuroimage)***
- **DCM for layered BOLD:** ***Model high resolution fMRI to resolve layers*** ***(Heinzle et al. 2016, Neuroimage)***

Note: The evolution of DCM in SPM

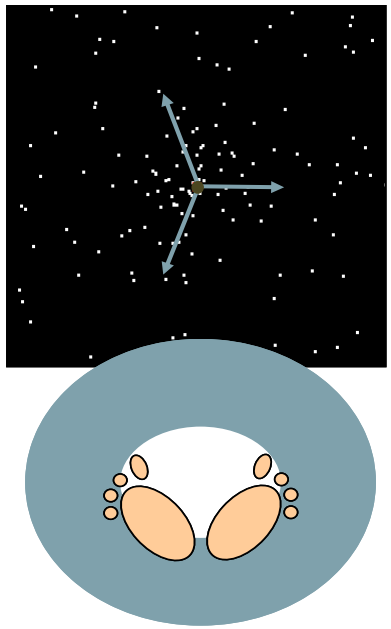
- DCM is not one specific model, but a framework for Bayesian inversion of dynamic system models
- The default implementation in SPM is evolving over time
  - better numerical routines for inversion
  - change in priors to cover new variants (e.g., stochastic DCMs, endogenous DCMs etc.)



To enable replication of your results, you should ideally state which SPM version (incl. release) you are using when publishing papers.
Matlab: [ver,release]=spm('ver');

Attention to motion in the visual system

Paradigm



Stimuli radially moving dots

Pre-Scanning

5 x 30s trials with 5 speed changes

Task - detect change in radial velocity

Scanning (no speed changes)

F A F N F A F N S ....

F - fixation

S - observe static dots + photic

N - observe moving dots + motion

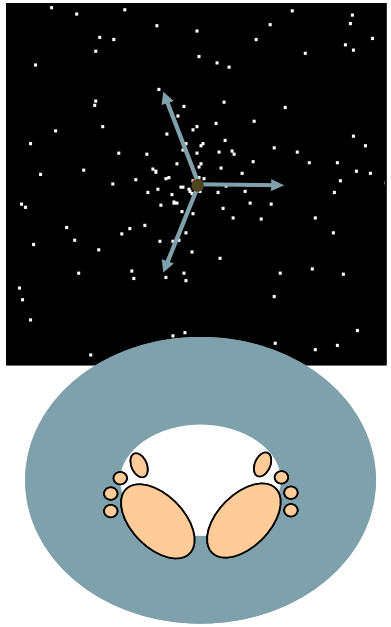
A - attend moving dots + attention

Parameters

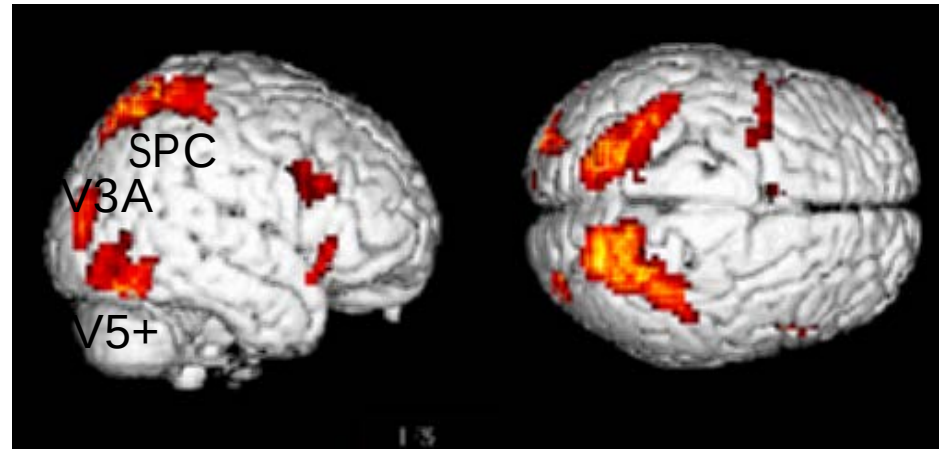
- blocks of 10 scans
- 360 scans total
- TR = 3.22 seconds

Attention to motion in the visual system

Paradigm

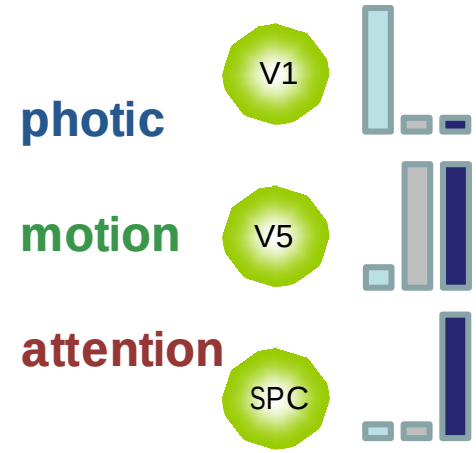


Results



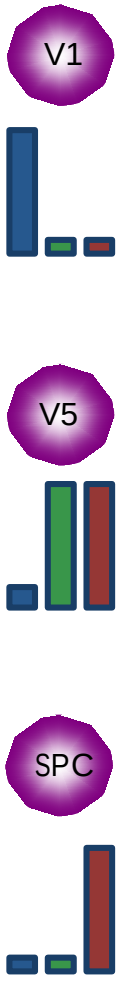
Attention – No attention

Büchel & Friston 1997, Cereb. Cortex
Büchel et al. 1998, Brain

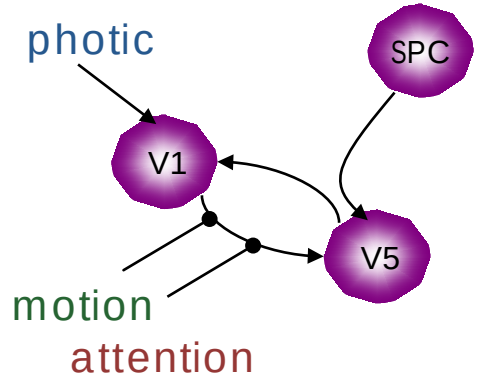


- fixation only
- observe static dots+ photic → V1
- observe moving dots + motion → V5
- task on moving dots + attention → V5 + parietal cortex

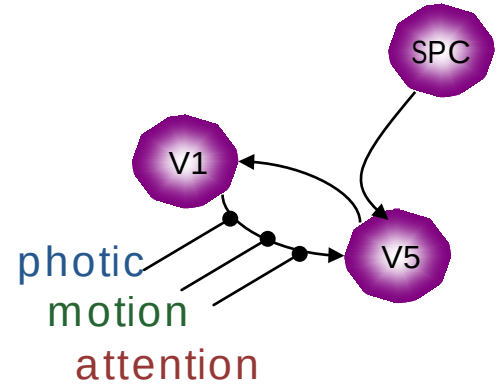
Quiz: can this DCM explain the data?



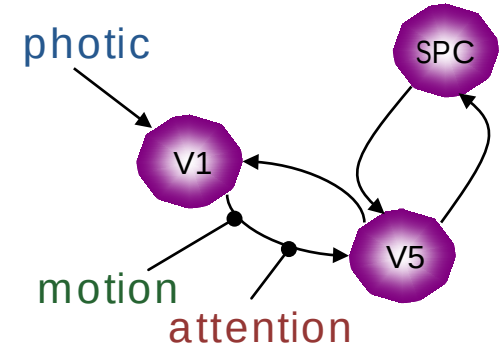
M1



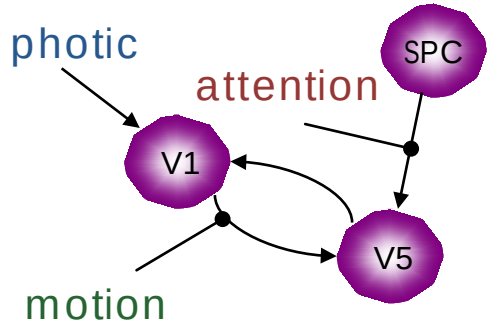
M2



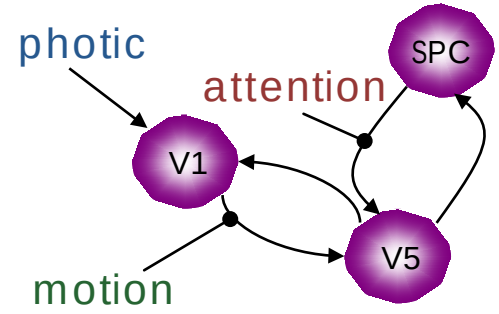
M3



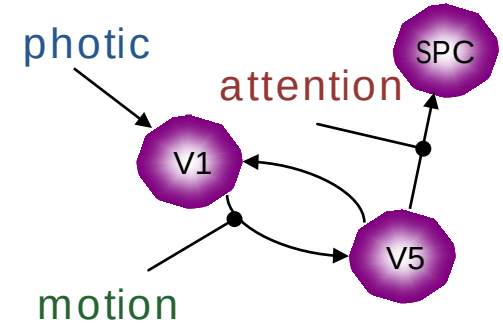
M4



M5

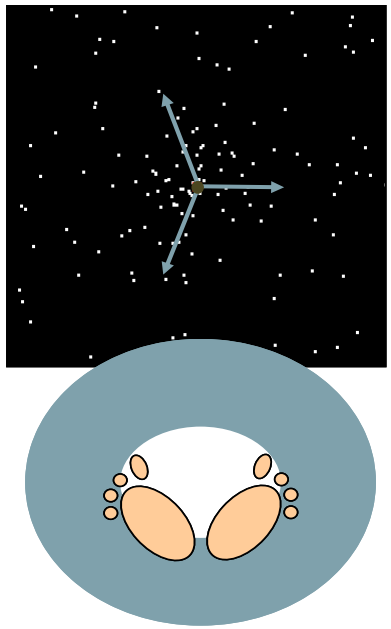


M6



Attention to motion in the visual system

Paradigm



Ingredients for a DCM

Specific hypothesis/question

Model: based on hypothesis

Timeseries: from the SPM

Inputs: from design matrix

DCM – GUI basic steps

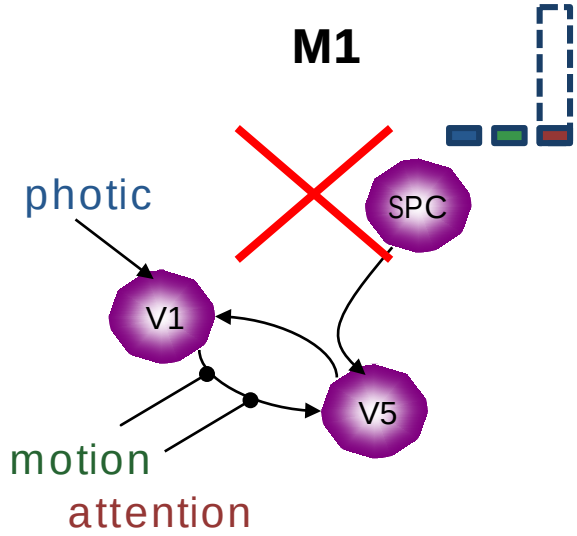
1 – Extract the time series (from all regions of interest)

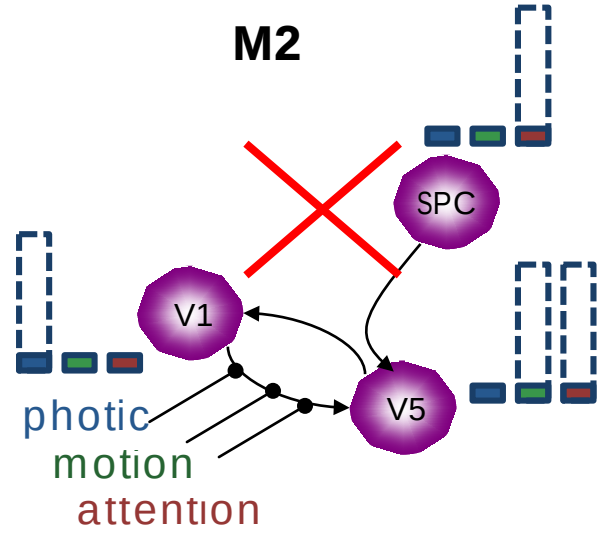
2 – Specify the model

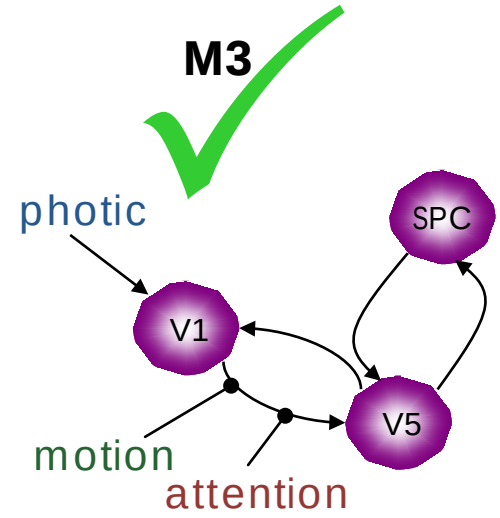
3 – Estimate the model

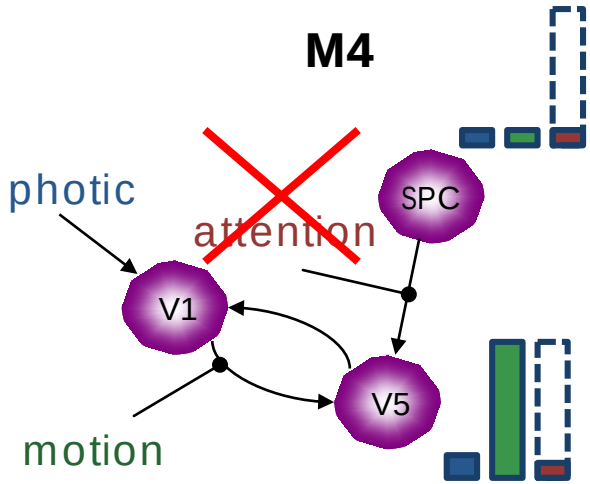
4 – Repeat steps 2 and 3 for all models in model space

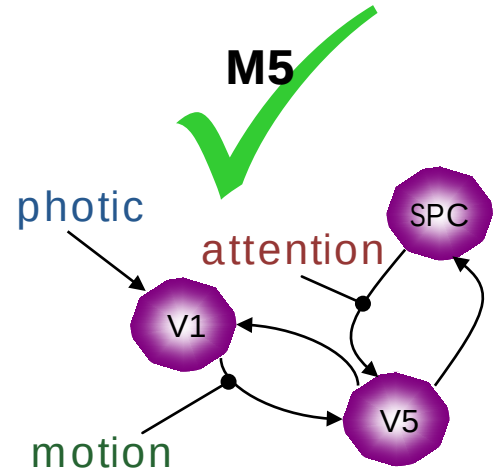
Quiz: can this DCM explain the data?

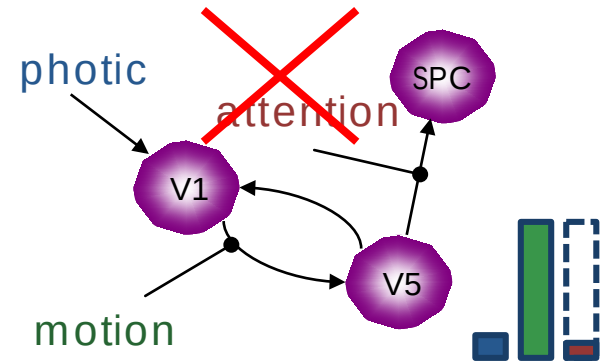




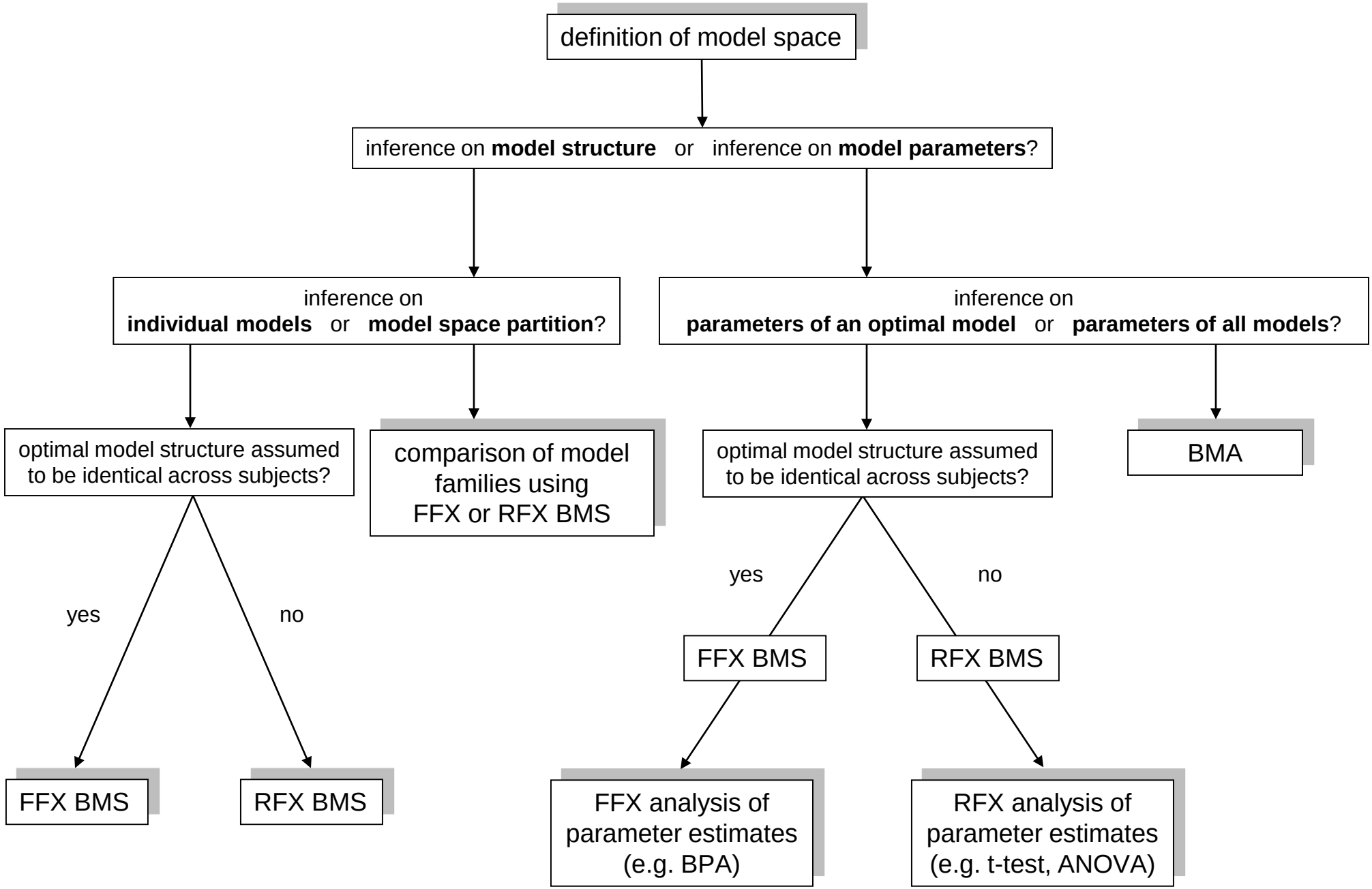








Additional material – not covered in lecture



Stephan et al. 2010, *NeuroImage*

Mean field assumption

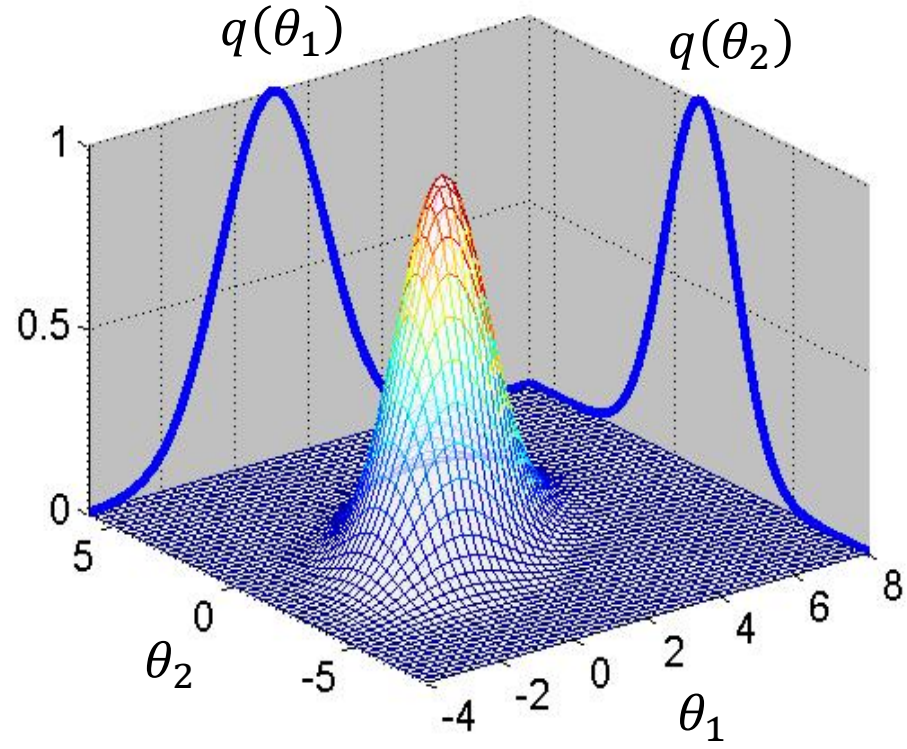
Factorize the approximate posterior $q(\theta)$ into independent partitions:

$$q(\theta) = \prod_i q_i(\theta_i)$$

where $q_i(\theta_i)$ is the approximate posterior for the i^{th} subset of parameters.

For example, split parameters and hyperparameters:

$$p(\theta, \lambda \mid y) \approx q(\theta, \lambda) = q(\theta) q(\lambda)$$



Jean Daunizeau, www.fil.ion.ucl.ac.uk/~jdaunize/presentations/Bayes2.pdf

VB in a nutshell (mean-field approximation)

❶ Neg. free-energy approx. to model evidence.

$$\ln p(y \mid m) = F + KL\left[q(\theta,\lambda), p(\theta,\lambda \mid y)\right]$$

$$F = \left\langle \ln p(y,\theta,\lambda) \right\rangle_q - KL\left[q(\theta,\lambda), p(\theta,\lambda \mid m)\right]$$

❷ Mean field approx.

$$p(\theta,\lambda \mid y) \approx q(\theta,\lambda) = q(\theta)q(\lambda)$$

❸ Maximise neg. free energy wrt. q = minimise divergence, by maximising variational energies

$$q(\theta) \propto \exp(I_\theta) = \exp\left[\left\langle \ln p(y,\theta,\lambda) \right\rangle_{q(\lambda)}\right]$$

$$q(\lambda) \propto \exp(I_\lambda) = \exp\left[\left\langle \ln p(y,\theta,\lambda) \right\rangle_{q(\theta)}\right]$$

❹ Iterative updating of sufficient statistics of approx. posteriors by gradient ascent.