Applications of Variational Bayes & DAGs in Neuroimaging

ECE 6504:
Advanced Topics in Machine Learning

Rosalyn Moran
rosalynj@vtc.vt.edu

Overview

1. Dynamics in Dynamic Causal Modeling

2. Graphical Model
- Variational Inversion
- Statistical Inference from VB

3. Examples
- Attention in the Human Brain
- Synesthesia

Dynamic Causal Modelling

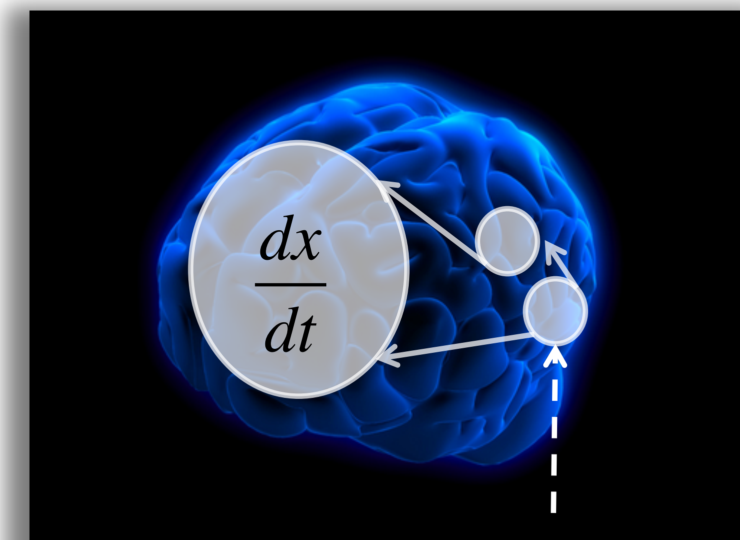
DCM *is not* intended for 'modelling'

DCM *is* an analysis framework for empirical data

DCM *uses* a times series to test mechanistic hypotheses

Hypotheses *are constrained* by the underlying dynamic generative (biological) model

Time Series

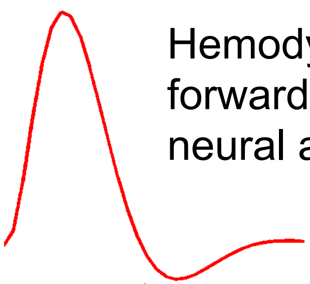


Friston et al 2003; Stephan et al 2008

Kiebel et al, 2006; Garrido et al, 2007

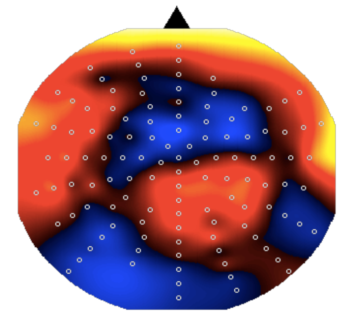
David et al, 2006; Moran et al, 2007

Dynamic Causal Modelling (DCM)



Hemodynamic
forward model:
neural activity→BOLD

Electromagnetic
forward model:
neural activity→EEG
MEG
LFP

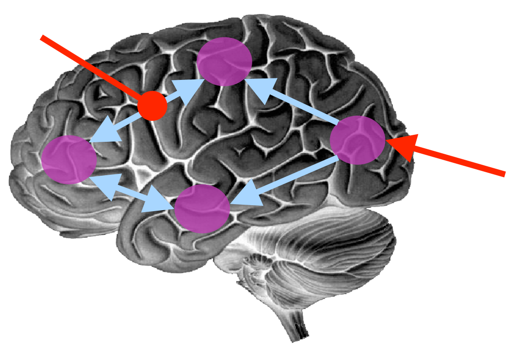


Neural state equation:

$$\frac{dx}{dt} = F(x,u,\theta)$$

fMRI

simple neuronal model
complicated forward model

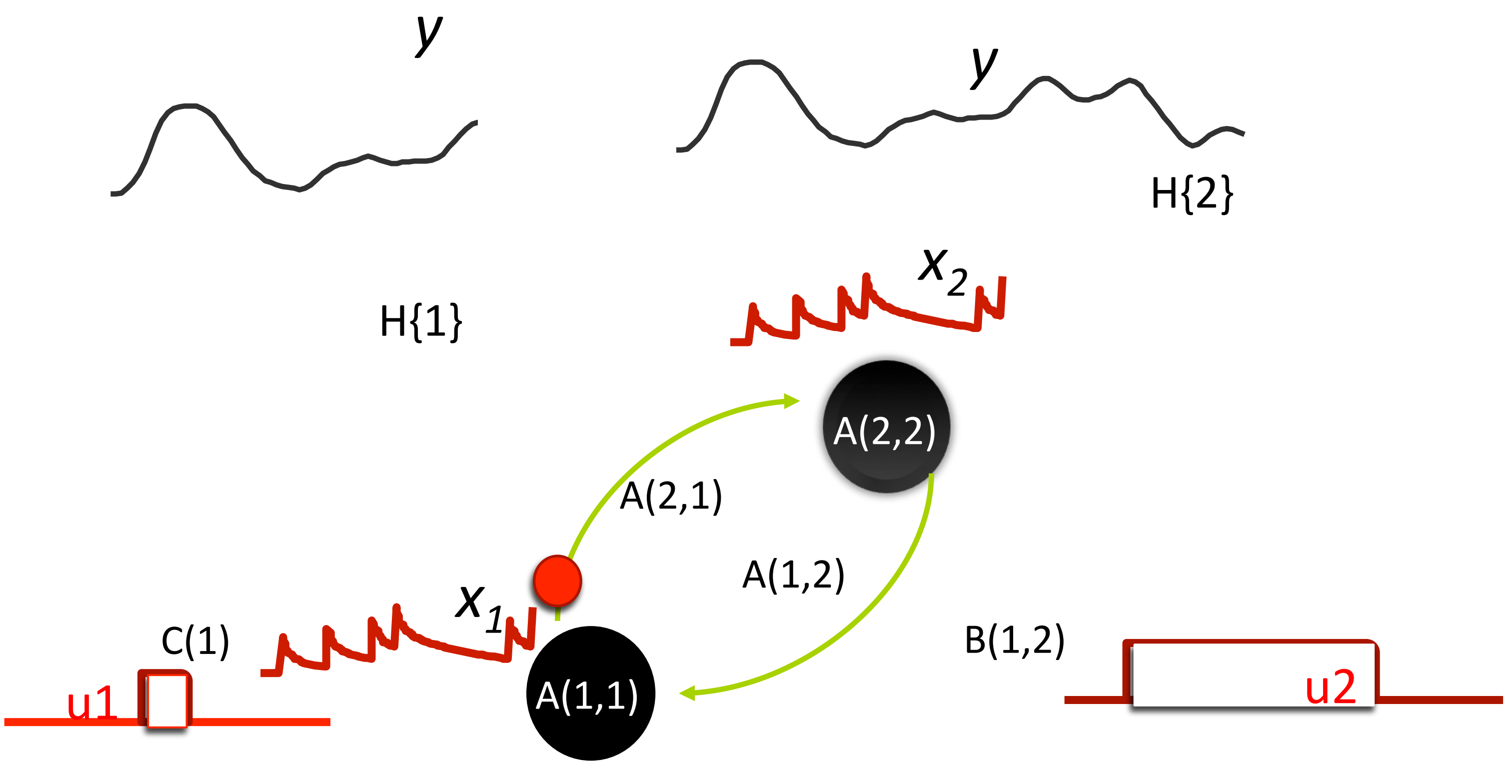


EEG/MEG

complicated neuronal model
simple forward model

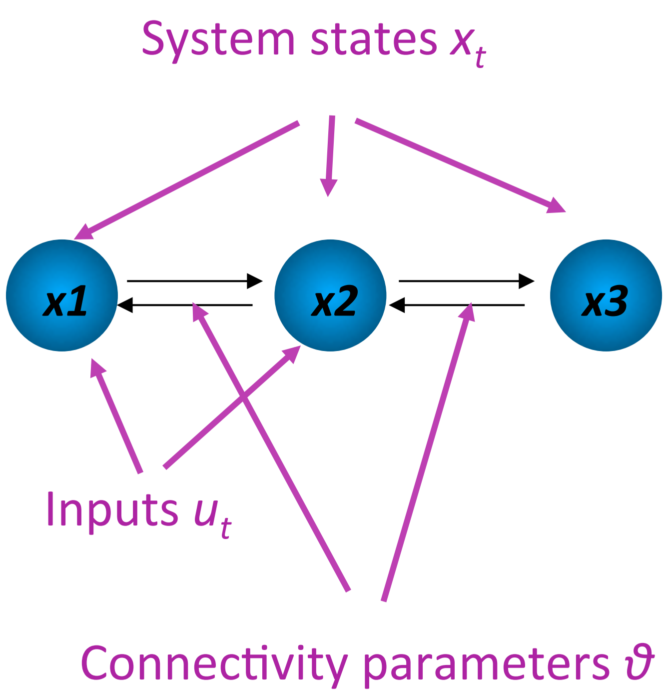
DCM for fMRI

$$\dot{x} = (A + uB)x + Cu$$



Neuronal model

Aim: model temporal evolution of a set of neuronal states x_t



State *changes* are dependent on:

- the current state *x*
- external inputs *u*
- its connectivity ϑ

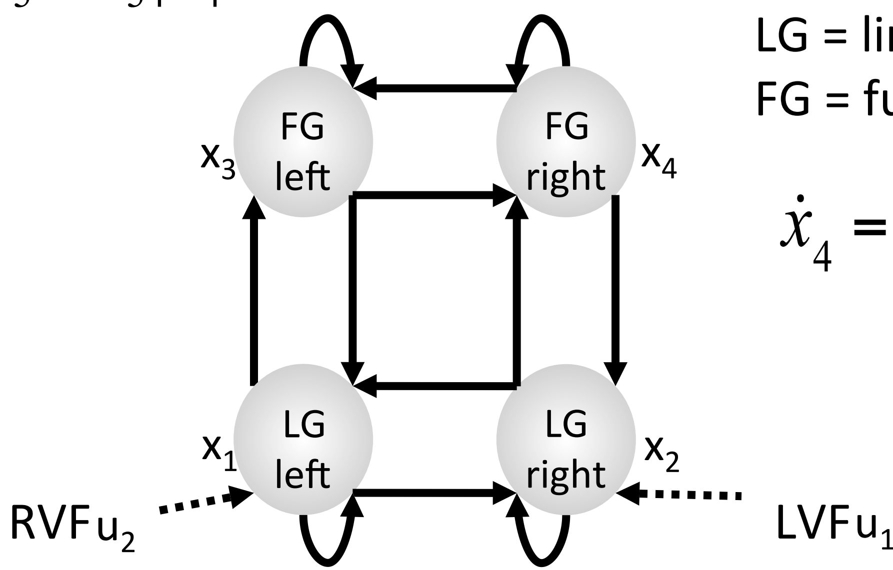
$$\frac{dx}{dt} = F(x, u, \theta)$$

Example: a linear model of interacting visual regions

Visual input in the visual field
- left (LVF)
- right (RVF)

$$\dot{x}_3 = a_{31}x_1 + a_{33}x_3 + a_{34}x_4$$

LG = lingual gyrus
FG = fusiform gyrus

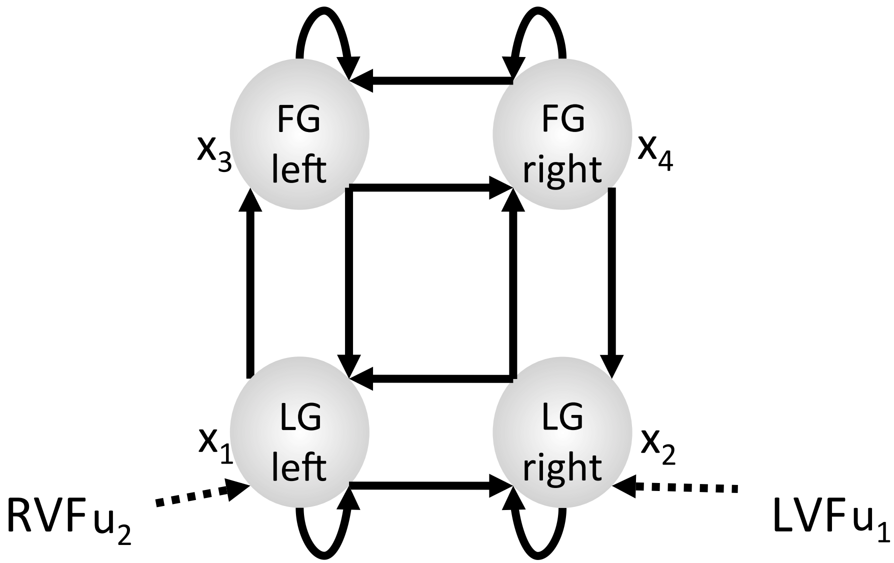


$$\dot{x}_4 = a_{42}x_2 + a_{43}x_3 + a_{44}x_4$$

$$\dot{x}_1 = a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + c_{12}u_2$$

$$\dot{x}_2 = a_{21}x_1 + a_{22}x_2 + a_{24}x_4 + c_{21}u_1$$

Example: a linear model of interacting visual regions

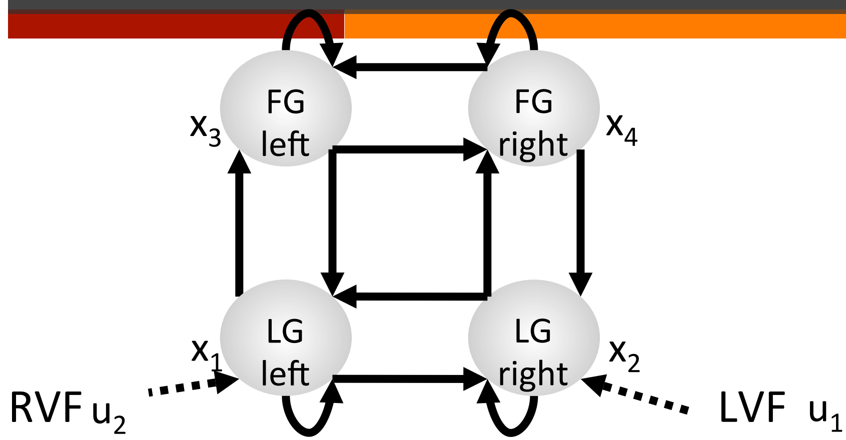


Visual input in the visual field
- left (LVF)
- right (RVF)

LG = lingual gyrus
FG = fusiform gyrus

$$\dot{x}_1 = a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + c_{12}u_2$$
$$\dot{x}_2 = a_{21}x_1 + a_{22}x_2 + a_{24}x_4 + c_{21}u_1$$
$$\dot{x}_3 = a_{31}x_1 + a_{33}x_3 + a_{34}x_4$$
$$\dot{x}_4 = a_{42}x_2 + a_{43}x_3 + a_{44}x_4$$

Example: a linear model of interacting visual regions



Visual input in the visual field
- left (LVF)
- right (RVF)

LG = lingual gyrus
FG = fusiform gyrus

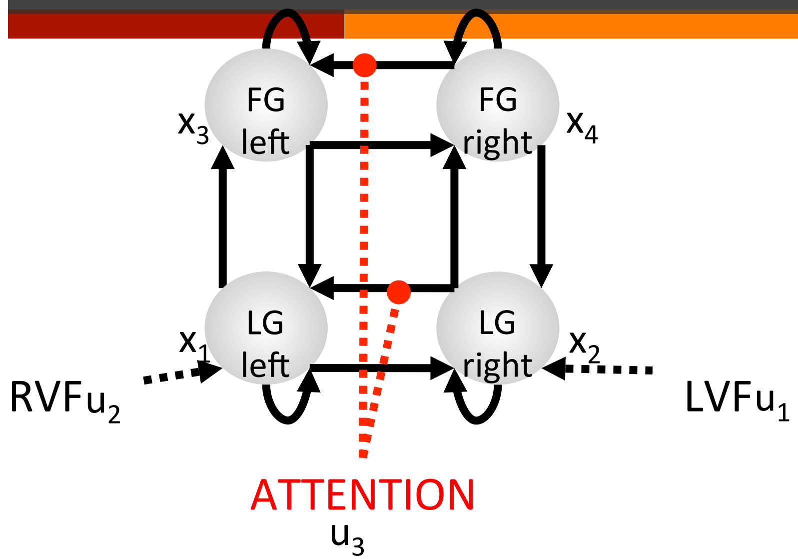
$$\dot{x} = Ax + Cu$$

$$\theta = \{A, C\}$$

state changes → effective connectivity → system state → input parameters → external inputs

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & 0 \\ a_{21} & a_{22} & 0 & a_{24} \\ a_{31} & 0 & a_{33} & a_{34} \\ 0 & a_{42} & a_{43} & a_{44} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + \begin{bmatrix} 0 & c_{12} \\ c_{21} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

Example: a linear model of interacting visual regions

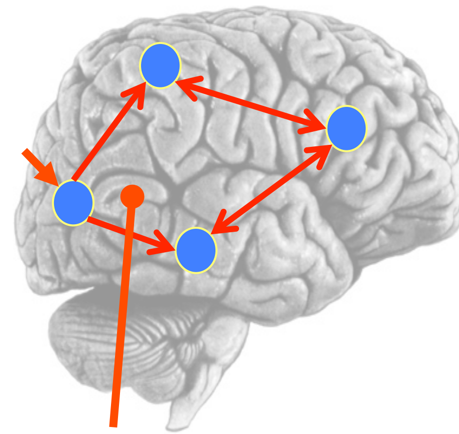


$$\dot{x} = (A + \sum_{j=1}^{m} u_j B^{(j)})x + Cu$$

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \end{bmatrix} = \left\{ \begin{bmatrix} a_{11} & a_{12} & a_{13} & 0 \\ a_{21} & a_{22} & 0 & a_{24} \\ a_{31} & 0 & a_{33} & a_{34} \\ 0 & a_{42} & a_{43} & a_{44} \end{bmatrix} + u_3 \begin{bmatrix} 0 & b_{12}^{(3)} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & b_{34}^{(3)} \\ 0 & 0 & 0 & 0 \end{bmatrix} \right\} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + \begin{bmatrix} 0 & c_{12} & 0 \\ c_{21} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$$

Deterministic Bilinear DCM

Simply a two-dimensional
taylor expansion (around x_0=0, u_0=0):



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

$$A = \left.\frac{\partial f}{\partial x}\right|_{u=0}$$

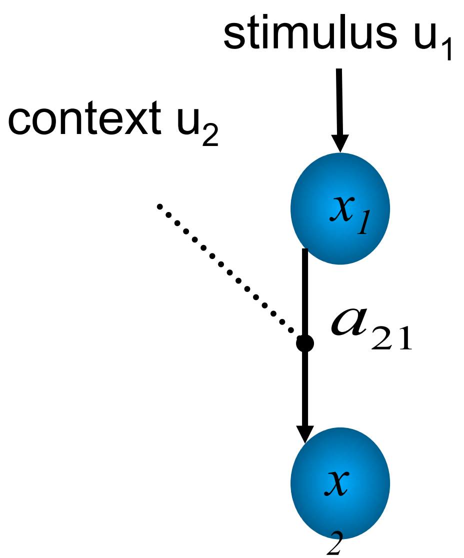
$$C = \left.\frac{\partial f}{\partial u}\right|_{x=0}$$

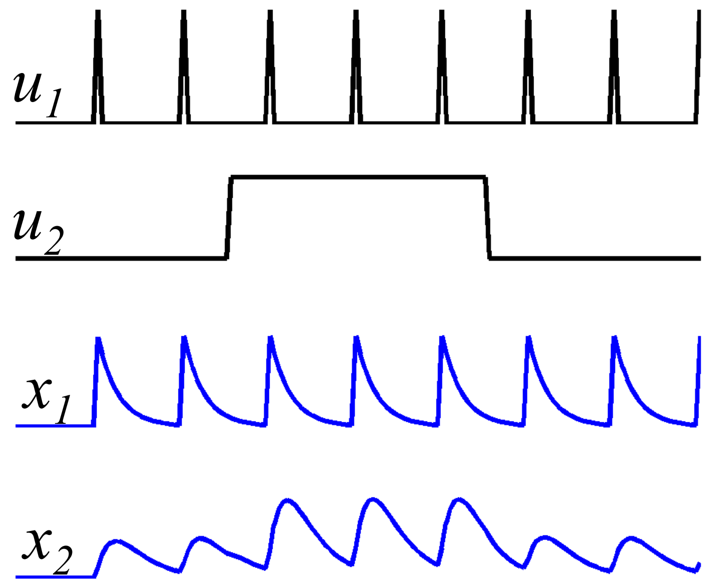
$$B = \frac{\partial^2 f}{\partial x \partial u}$$

Bilinear state equation:

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

Context-dependent enhancement

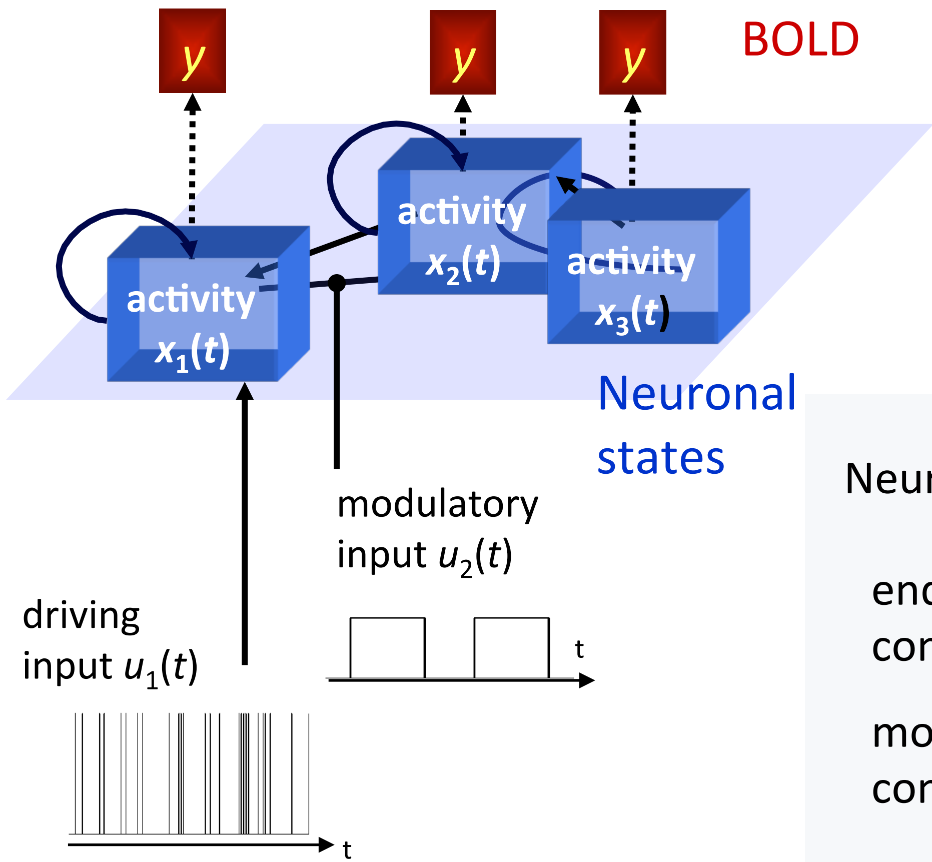


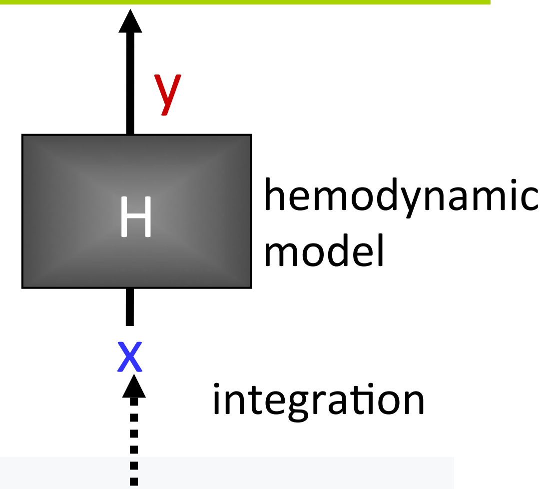


$$\dot{x} = Ax + u_2 B^{(2)} x + Cu$$

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} a_{11} & 0 \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + u_2 \begin{bmatrix} 0 & 0 \\ b_{21}^{(2)} & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} c_{11} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

DCM for fMRI: the full picture





Neural state equation $\dot{x} = (A + \sum u_j B^{(j)})x + Cu$

endogenous connectivity $A = \frac{\partial \dot{x}}{\partial x}$

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

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

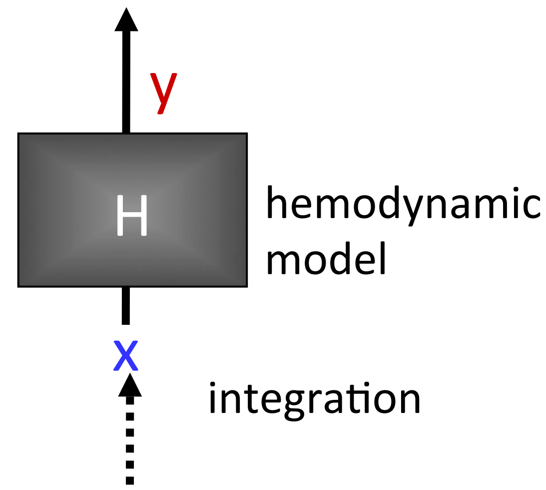
Stephan & Friston (2007), *Handbook of Brain Connectivity*

DCM: Neuronal and hemodynamic level

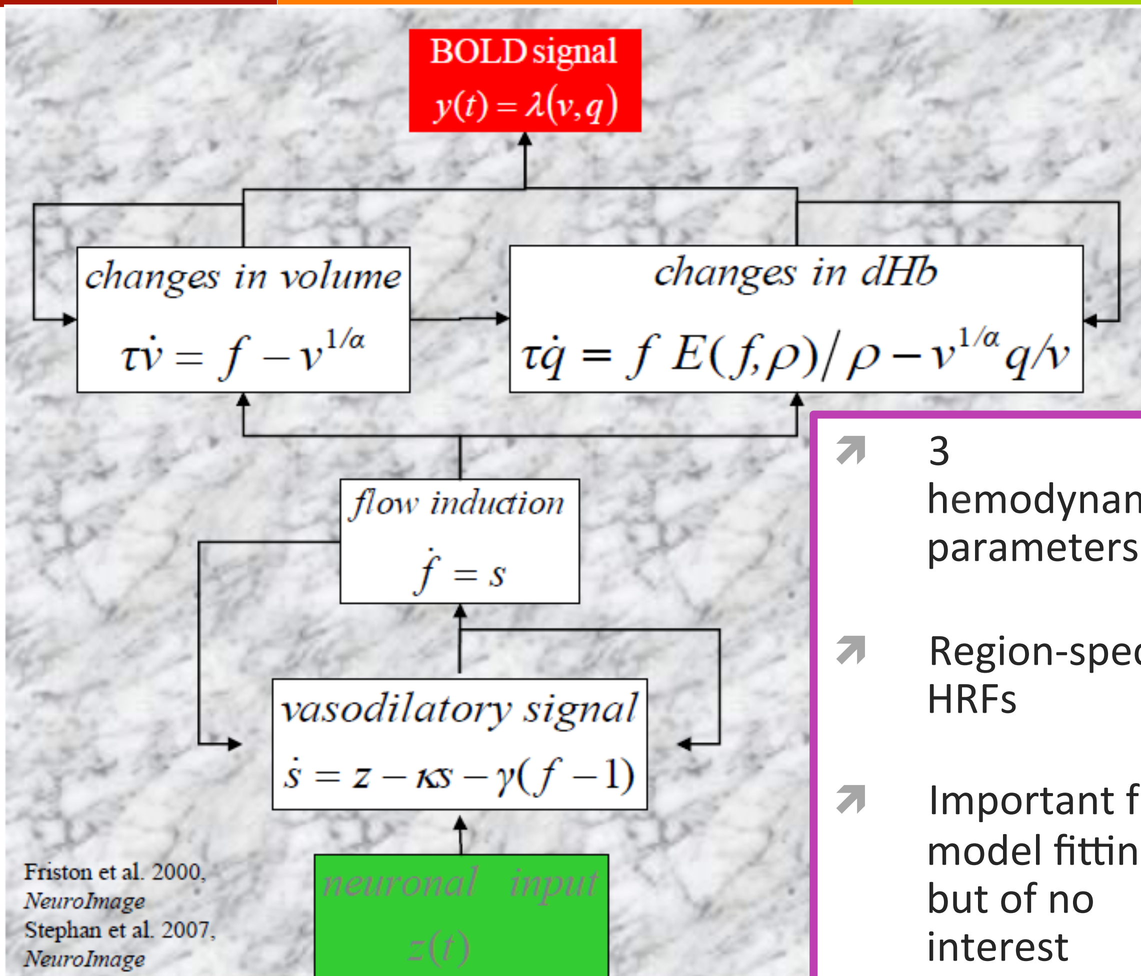
- Cognitive system is modelled at its underlying <u>neuronal level</u> (not directly accessible for fMRI).

- The modelled neuronal dynamics (**x**) are transformed into area-specific BOLD signals (**y**) by a hemodynamic model (λ).

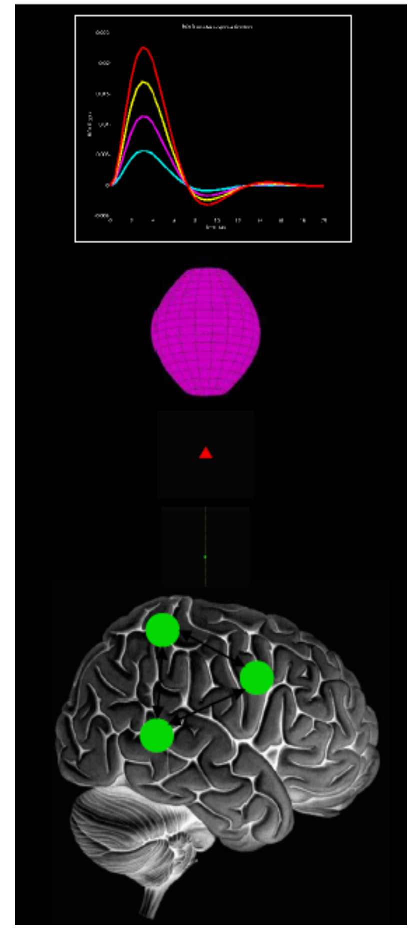
  - Overcomes regional variability at the hemodynamic level
  - DCM not based on temporal precedence at measurement level



The hemodynamic “Balloon” model



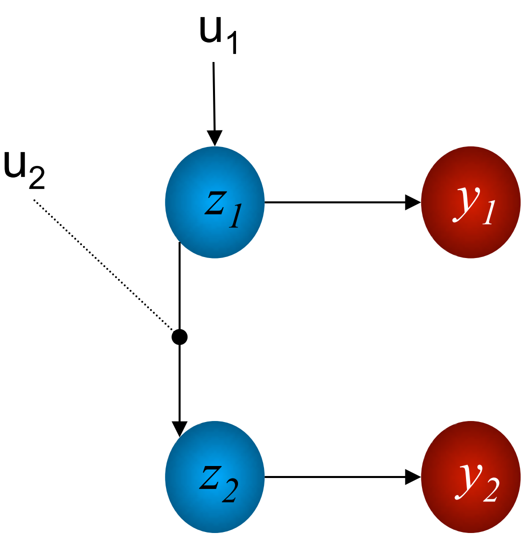
- 3 hemodynamic parameters
- Region-specific HRFs
- Important for model fitting, but of no interest

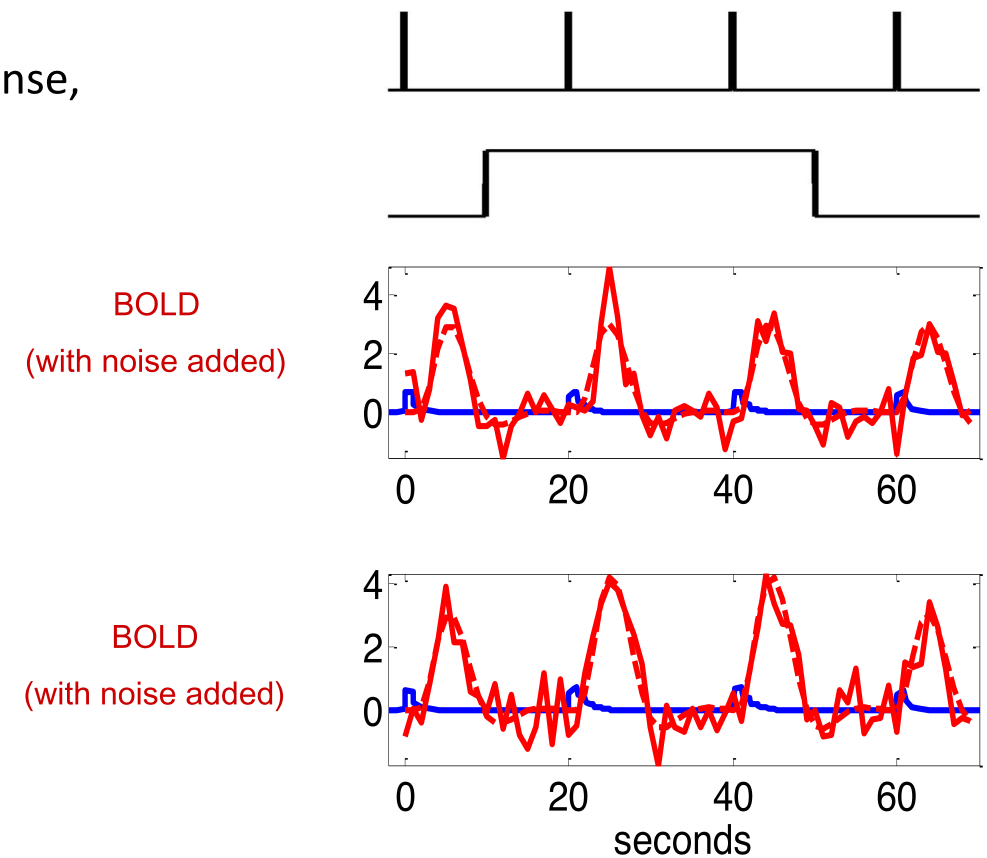


Hemodynamic model

y represents the simulated observation of the bold response, including noise, i.e.

$y = h(u,\vartheta)+e$





Z: neuronal activity
Y: BOLD response

Overview

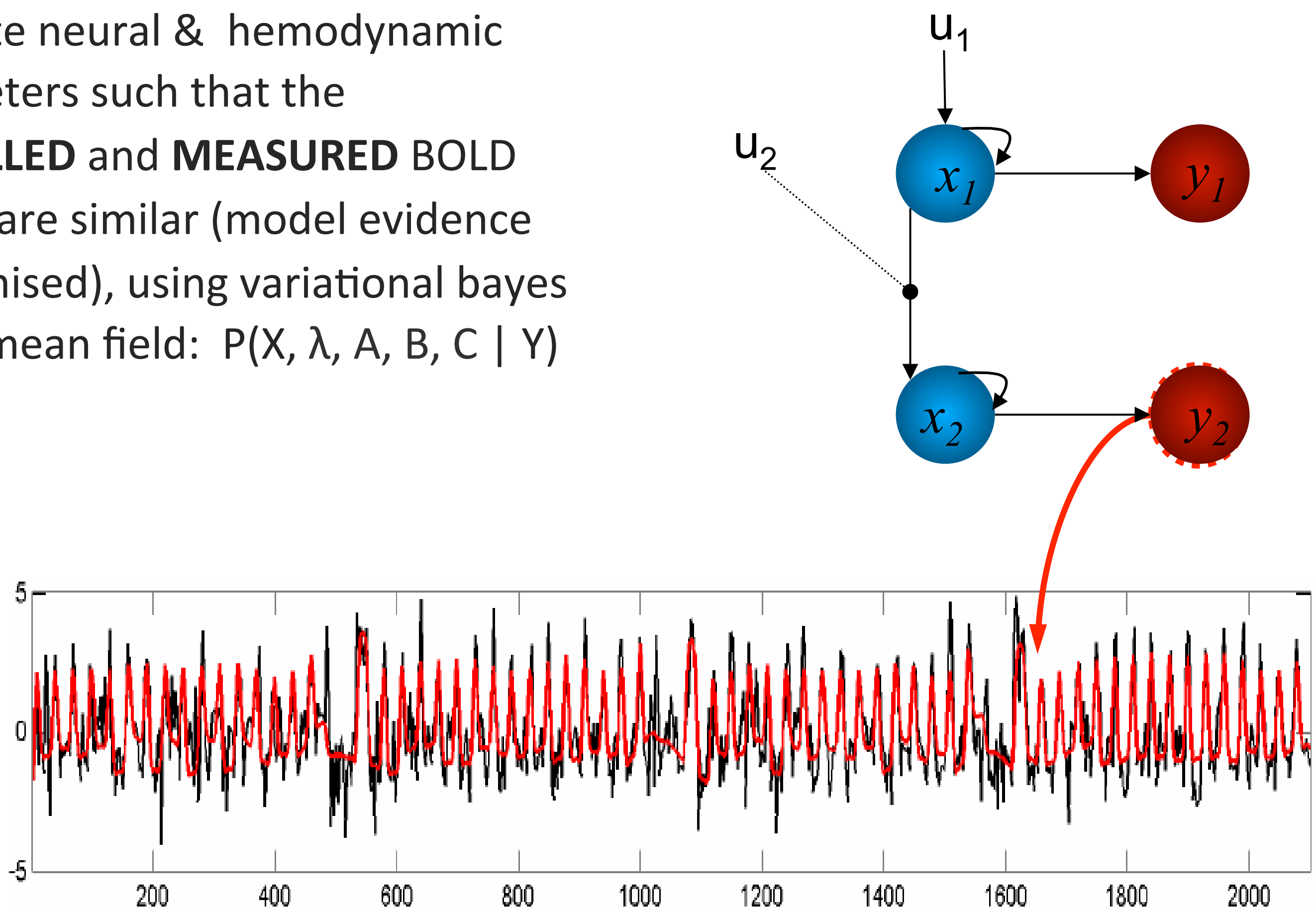
1. Dynamics in Dynamic Causal Modeling

2. Graphical Model
- **Variational Inversion**
- **Bayesian Statistical Inference from VB**

3. Examples
- Attention in the Human Brain
- Synesthesia

Parameter estimation: Bayesian inversion

Estimate neural & hemodynamic parameters such that the **MODELLED** and **MEASURED** BOLD signals are similar (model evidence is optimised), using variational bayes under mean field: P(X, λ, A, B, C | Y)



Recall from Tuesday

Main Issues in PGMs

VB: A procedure to do inference:
That implicitly 'does double duty' in Directed Graphs!

- **Representation**
  - How do we store $P(X_1, X_2, \ldots, X_N)$
  - What does my model mean/imply/assume? (Semantics)
- **Inference**
  - How do I answer questions/queries with my model, such as
  - Marginal Estimation: $P(X_5 \mid X_1, X_4)$
  - Most Probable Explanation: argmax $P(X_1, X_2, \ldots, X_N)$
- **Learning**
  - How do we learn parameters and structure of $P(X_1, X_2, \ldots, X_N)$ from data
  - What is the right model for my data?

Key Results for VB

- Approximate Inference using constrained optimization

- Where: The approximation arises from constructing an approximating distribution over X: *q(X)* which is closest in *p(X)* "in the KL sense"

- Derived a cost function Which can be maximized

$$F = \sum_{\phi} \langle \ln \phi \rangle_q + H[q]$$

- And is equivalent to minimizing KL(*q*|*p*)

$$F = \ln Z - KL(q \mid p)$$

- Z: Partition Function; a normalization function equal to the probability of the evidence in directed graphs

Key Result for Mean-Field, Structured VB

- The structured variational approach aims to optimize *F* over a *coherent* distribution *q* (ie. giving a proper joint distribution), at the expense of capturing all the information in *p*.

- Assume the approximating or proposal density factorizes over groups of parameters - where this factorization is *a relaxation* (a superspace) of the space of true marginals.

- Approximate *q* using a factorization

$$q(X) = \prod_i q(x_i)$$

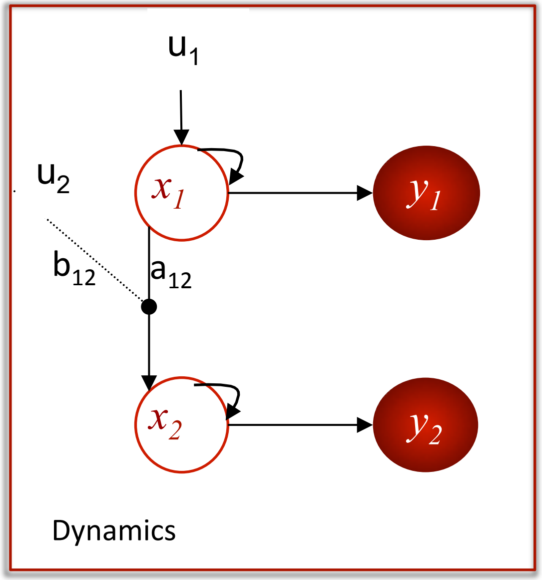
- Found iterative update equations for *q* using fixed point solutions

$$q(x_i) = \frac{\exp[I(x_i)]}{Z}$$

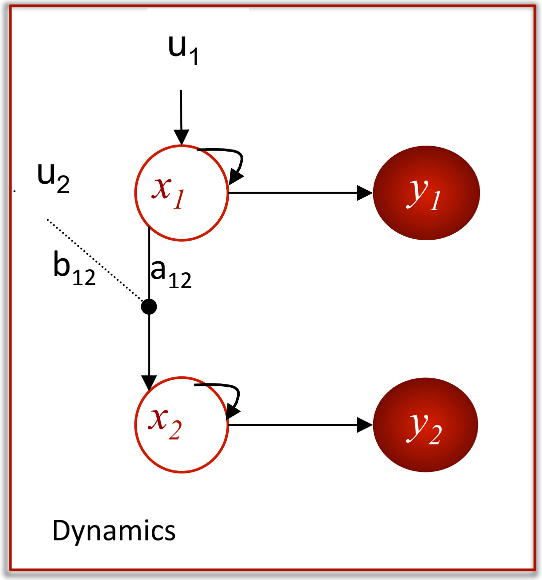
- *F* is a guaranteed lower bound on *ln(Z)*

$$F = \ln Z - KL(q \mid p)$$

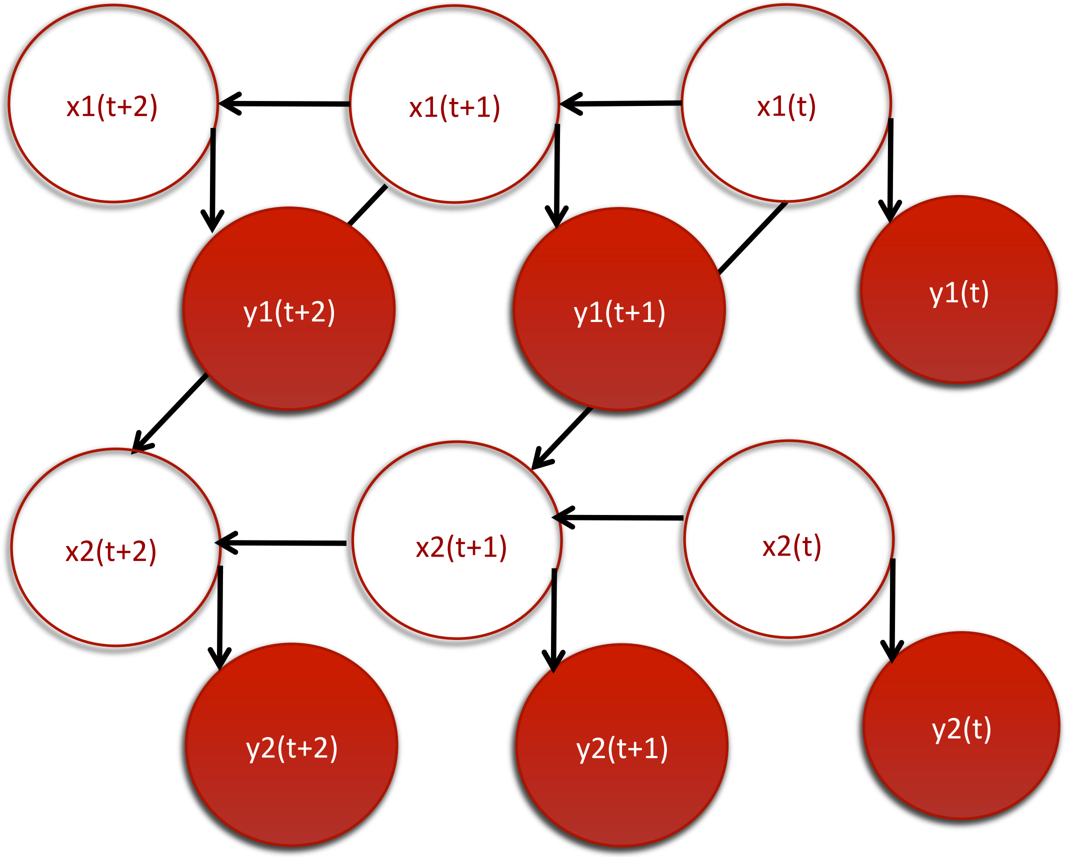
DCM: Probabilistic Graphical Model Representation



DCM: Probabilistic Graphical Model Representation

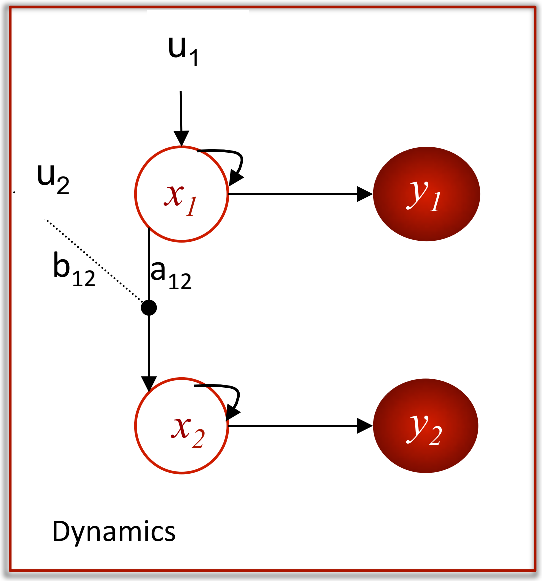


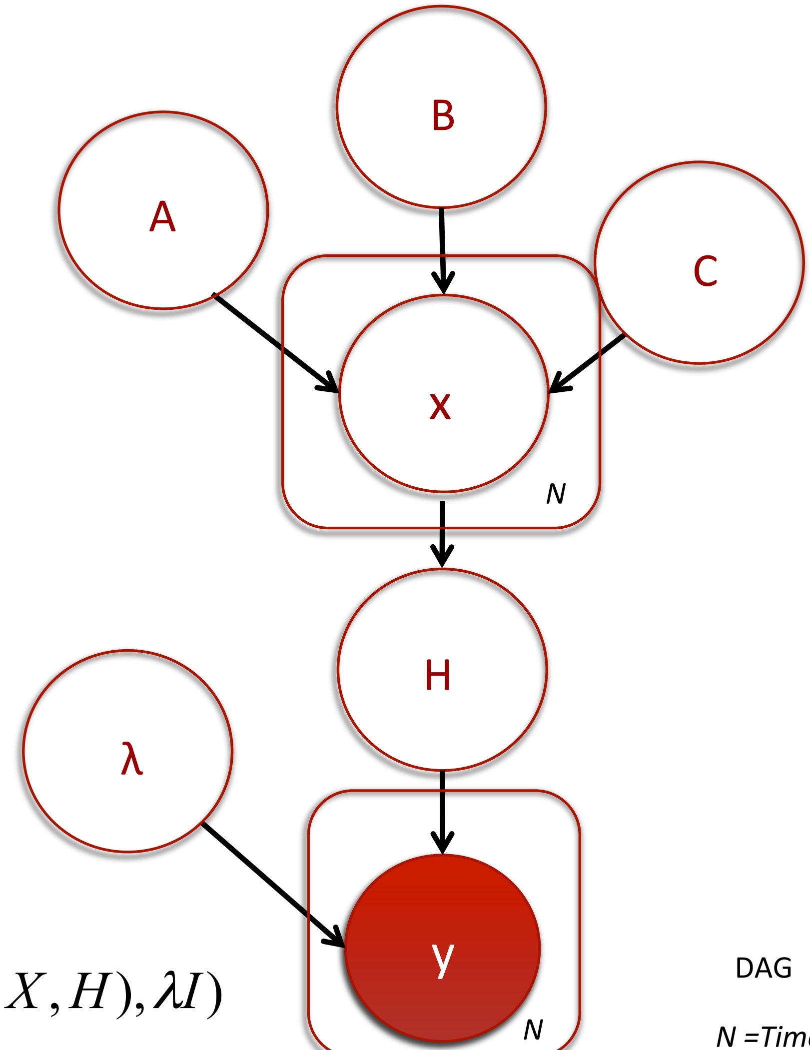
Causal Links expressed through implicit delays, which makes the graph a Directed Acyclic Graph



DAG

DCM: Probabilistic Graphical Model Representation



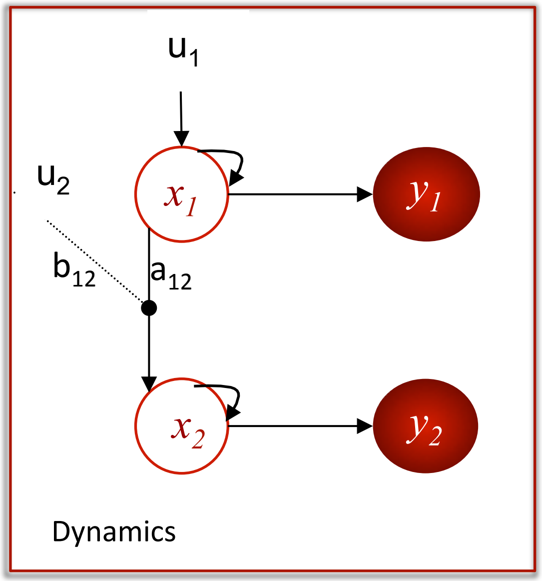


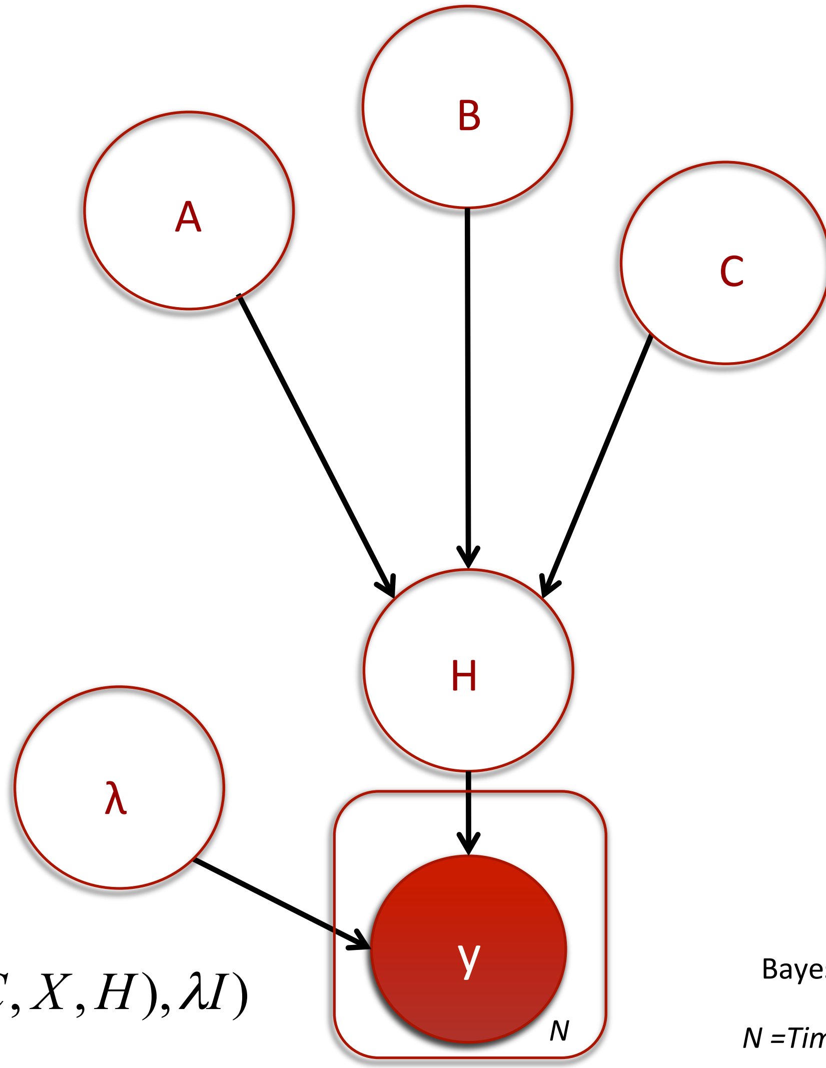
$$p(y|A,B,C,X,H) \rightarrow N(f(A,B,C,X,H),\lambda I)$$

DAG

N =Time steps x # Regions

DCM: Probabilistic Graphical Model Representation



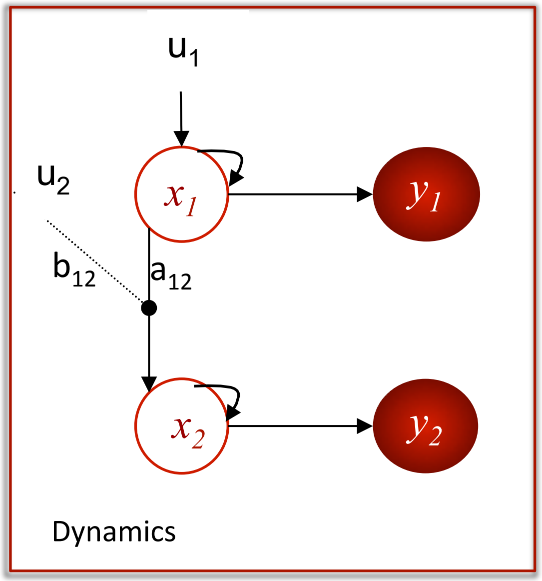


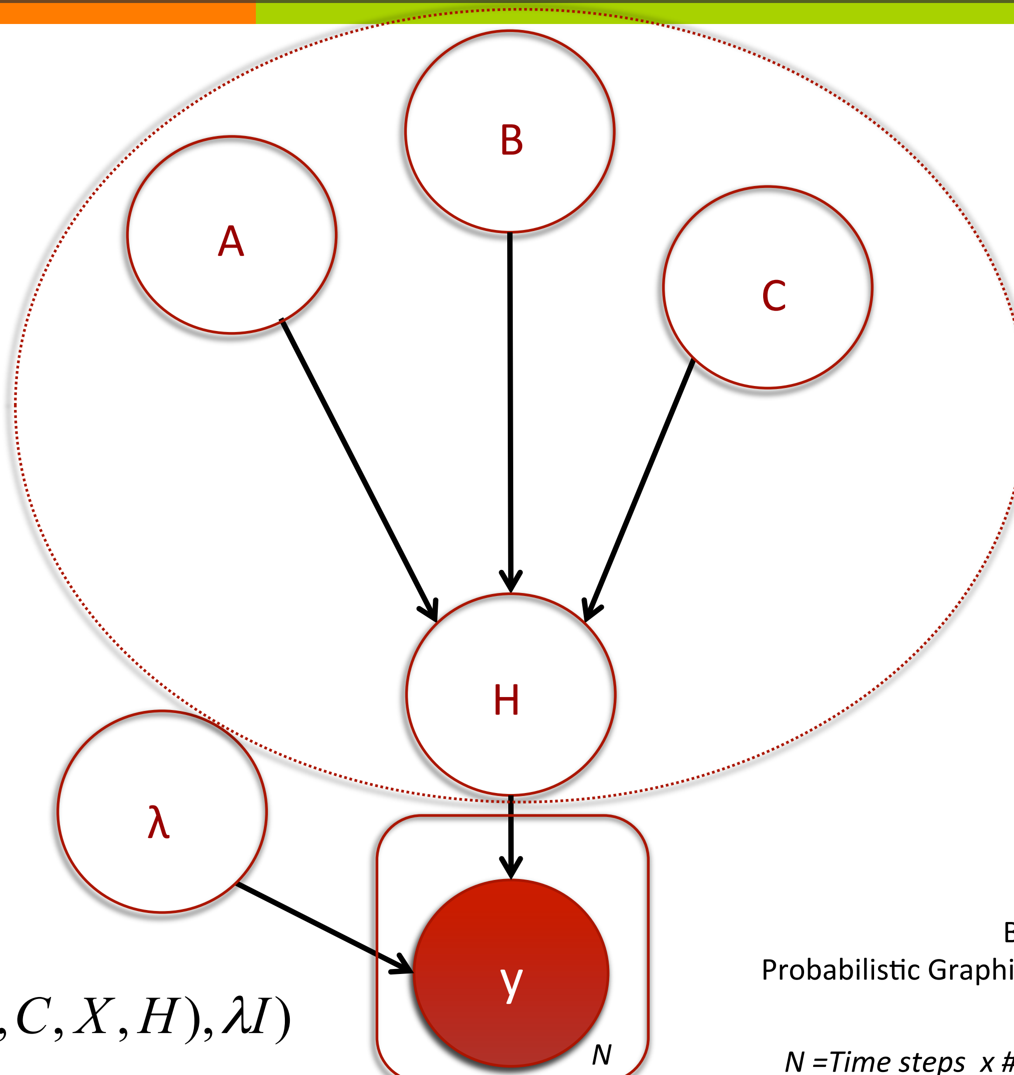
$$p(y|A,B,C,X,H) \rightarrow N(f(A,B,C,X,H),\lambda I)$$

Bayes Net: PGM

N =Time steps x # Regions

DCM: Probabilistic Graphical Model Representation



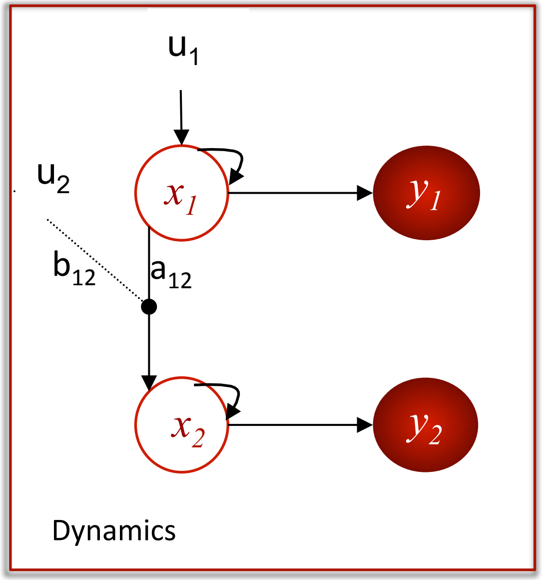


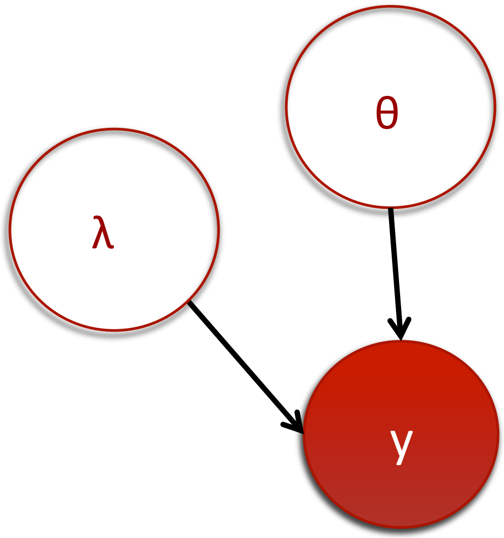
$$p(y|A,B,C,X,H) \rightarrow N(f(A,B,C,X,H),\lambda I)$$

Bayes Net:
Probabilistic Graphical Model

N =Time steps x # Regions

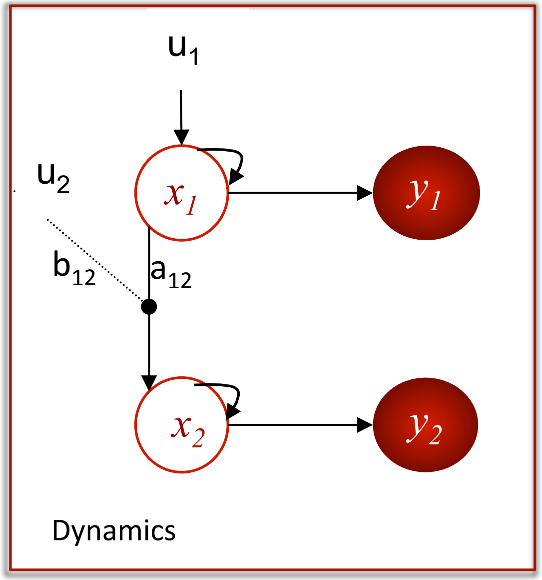
DCM: Probabilistic Graphical Model Representation

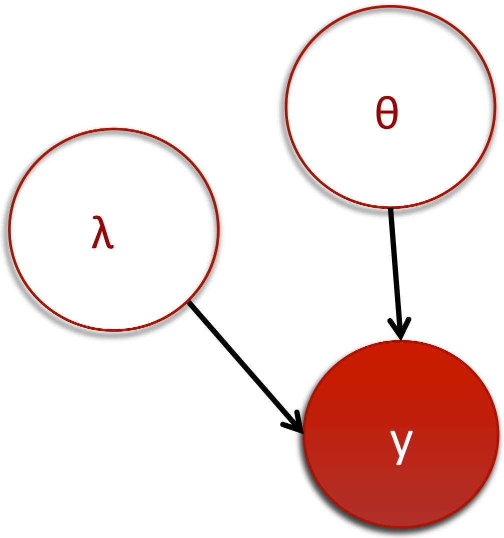




Goal: Find the set of latent variables θ, given y: **p(θ|y)**
Ie. inference or
Query for the marginal distribution of the connectivity parameters given data, marginalized w.r.t noise parameter

DCM: Probabilistic Graphical Model Representation

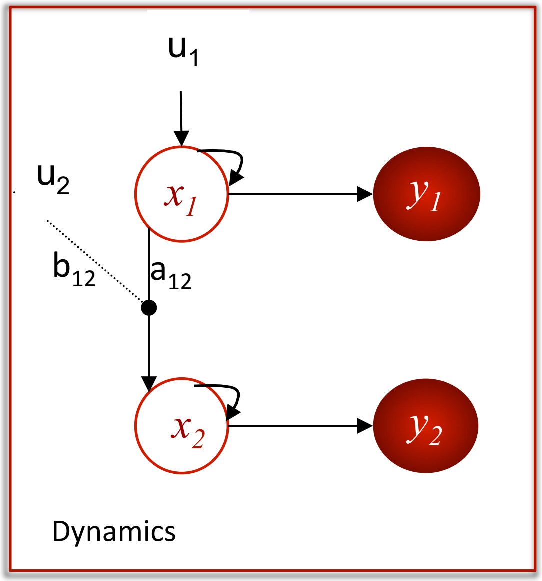


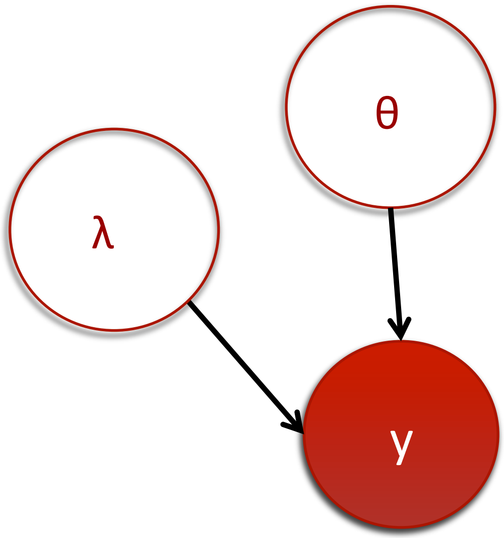


Goal: Find the set of latent variables θ, given y: **p(θ|y)**
Ie. inference or
Query for the marginal distribution of the connectivity parameters given data, marginalized w.r.t noise parameter

Given this type of graph we know: $p(\theta,\lambda|y) = \frac{p(\theta)p(\lambda)p(y|\theta,\lambda)}{p(y)}$ $\quad \theta \not\perp\!\!\!\perp \lambda \mid y$

DCM: Probabilistic Graphical Model Representation





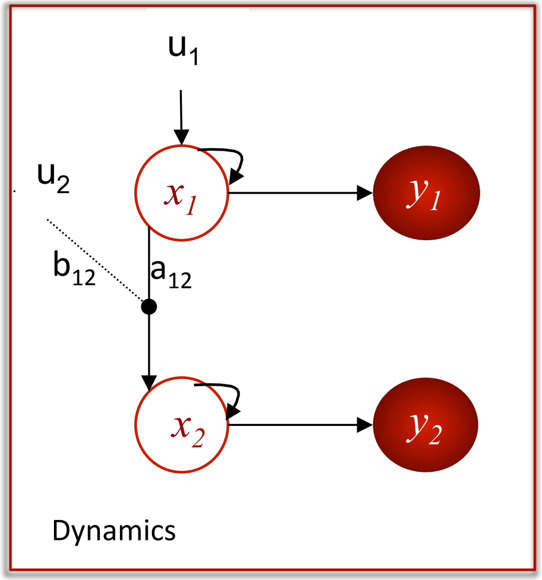
Goal: Find the set of latent variables θ, given y: **p(θ|y)**
Ie. inference or
Query for the marginal distribution of the connectivity parameters given data, marginalized w.r.t noise parameter

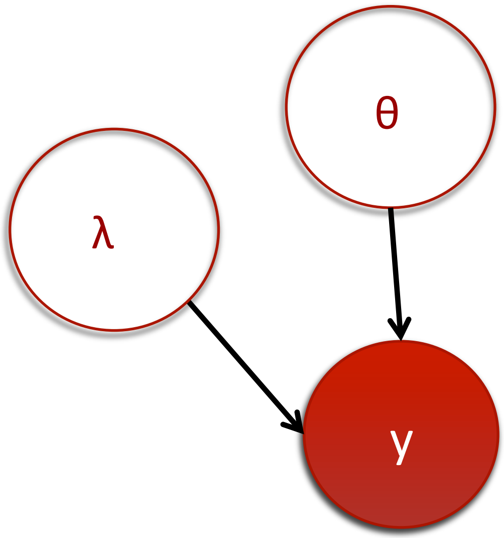
Given this type of graph we know: $p(\theta,\lambda|y) = \dfrac{p(\theta)p(\lambda)p(y|\theta,\lambda)}{p(y)}$ and $\theta \not\perp\!\!\!\perp \lambda \mid y$

But Employ Approximating Density q,
Using the mean field structure:

Where:

DCM: Probabilistic Graphical Model Representation





Goal: Find the set of latent variables θ, given y: **p(θ|y)**
Ie. inference or
Query for the marginal distribution of the connectivity parameters given data, marginalized w.r.t noise parameter

Given this type of graph we know:

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

But Employ Approximating Density q,
Using the mean field structure:

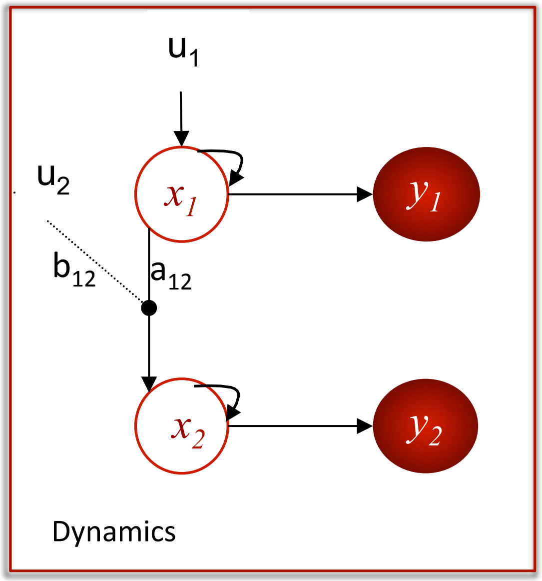
$$p(\theta,\lambda|y) = q(\theta|y)q(\lambda|y)$$

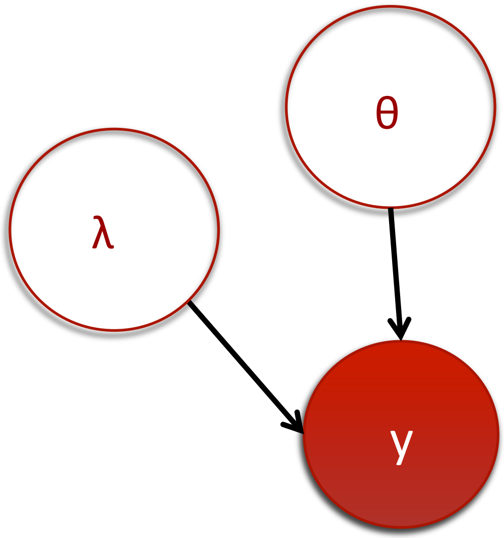
Where:

$$q(\theta|y) \rightarrow N(\mu,\Sigma)$$

$$q(\lambda|y) \rightarrow N(0,\lambda I)$$

DCM: Probabilistic Graphical Model Representation





Goal: Find the set of latent variables θ, given y: **p(θ|y)**
Ie. inference or
Query for the marginal distribution of the connectivity parameters given data, marginalized w.r.t noise parameter

Given this type of graph we know:

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

But Employ Approximating Density q,
Using the mean field structure:

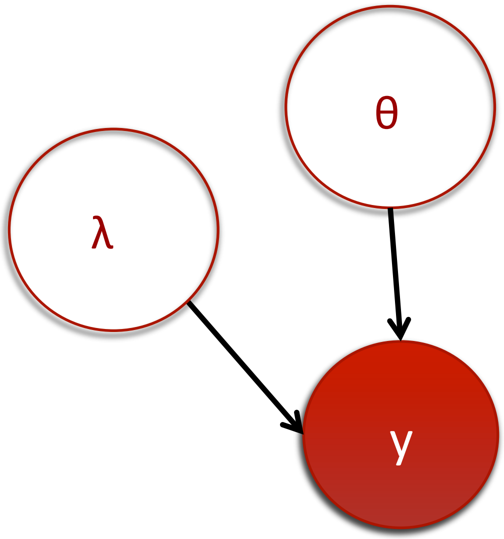
$$p(\theta,\lambda|y) = q(\theta|y)q(\lambda|y)$$

Where:

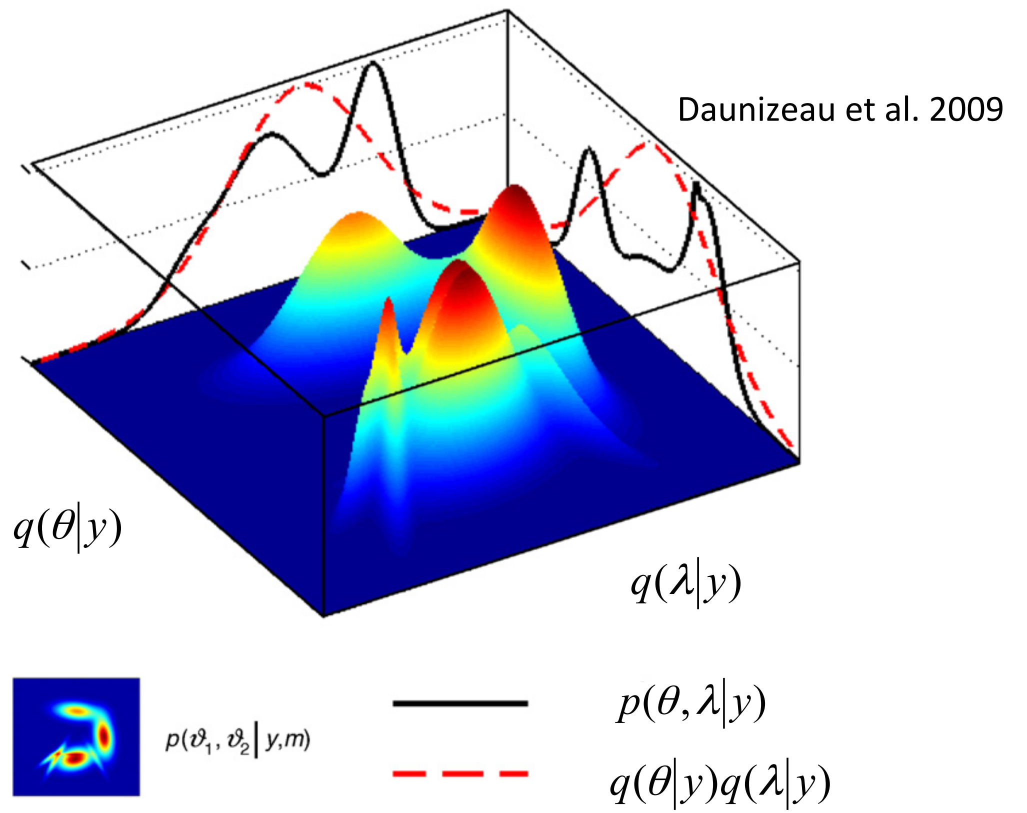
$$q(\theta|y) \rightarrow N(\mu,\Sigma)$$

$$q(\lambda|y) \rightarrow N(0,\lambda I)$$

DCM: Probabilistic Graphical Model Representation



Goal: Find the set of latent variables θ, given y,



- Assuming Independence of parameters & hyperparameters
- And a Gaussian form on the PDF

VB with a mean-field approximation

❶ Free-energy approx. to model evidence.

$$F = \langle \ln p(y,\theta,\lambda) \rangle_q - KL(q(\theta,\lambda \mid y) \| p(\theta,\lambda \mid y))$$

❷ Mean field approx.

$$p(\theta,\lambda|y) = q(\theta|y)q(\lambda|y)$$

❸ Fixed point solutions for two factors

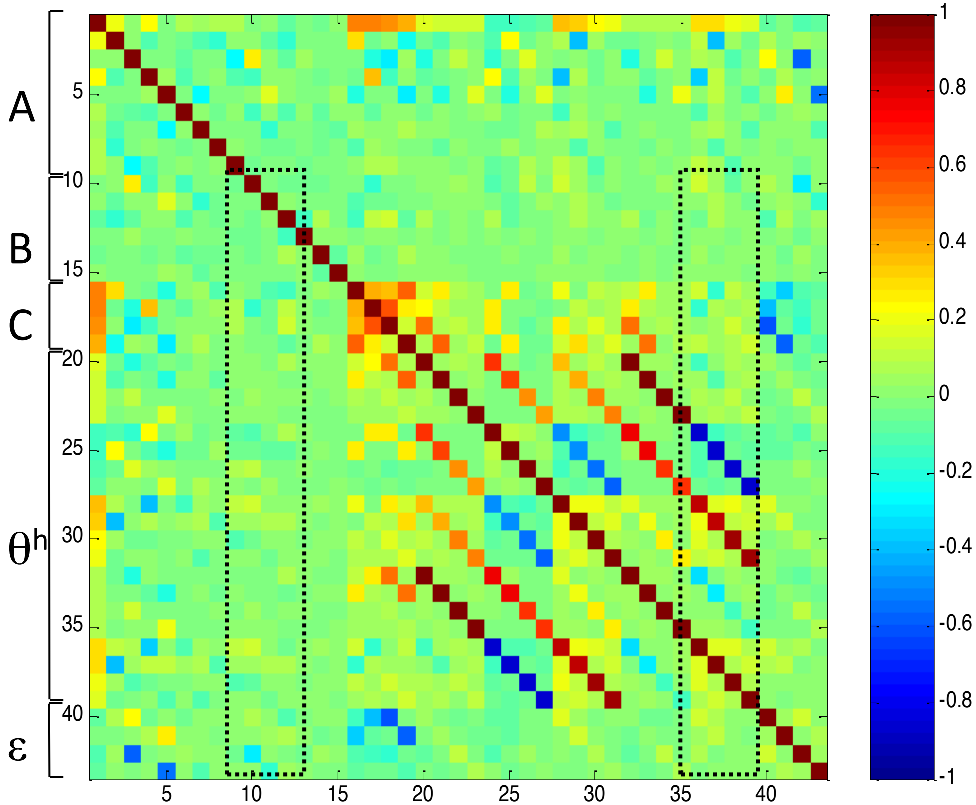
$$q(\theta) \propto \exp(I_\theta) = \exp\left[\langle \ln p(y,\theta,\lambda) \rangle_{q(\lambda)}\right]$$

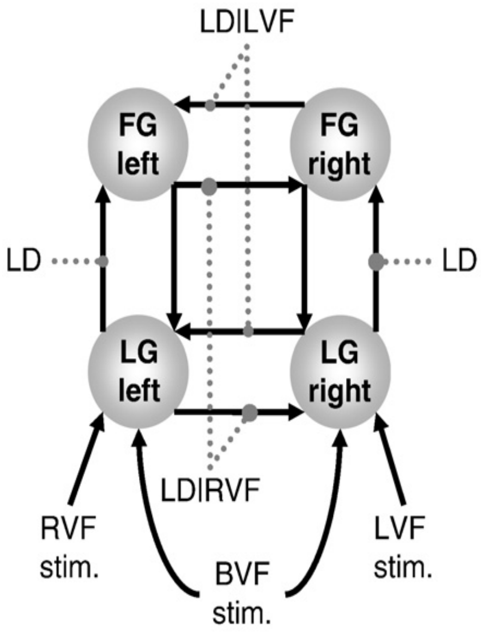
$$q(\lambda) \propto \exp(I_\lambda) = \exp\left[\langle \ln p(y,\theta,\lambda) \rangle_{q(\theta)}\right]$$

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

How independent are neural and hemodynamic parameter estimates?

$$q(\theta|y) \rightarrow N(\mu, \Sigma)$$





Stephan et al. (2007) *NeuroImage*

Roadmap inversion

Regional responses

Specify generative forward model
(with prior distributions of parameters)

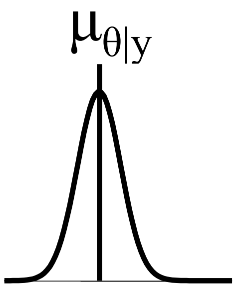
Variational Expectation-Maximization algorithm

<u>Iterative procedure:</u>

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

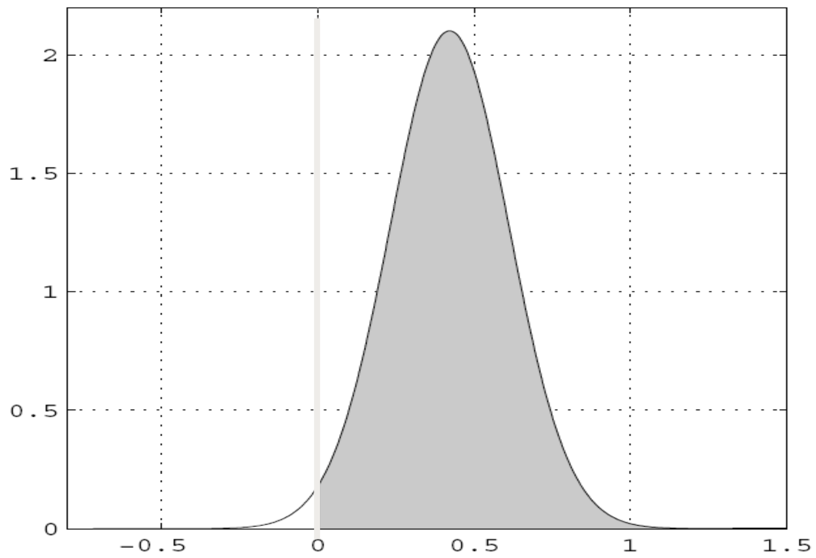
1. Gaussian posterior distributions of parameters $p(\theta \mid y, m)$

2. Model evidence $p(y \mid m)$



Inference about DCM parameters: Bayesian single subject analysis

- Gaussian assumptions about the posterior distributions of the parameters
- posterior probability that a certain parameter (or contrast of parameters) is above a chosen threshold γ:
- By default, γ is chosen as zero – the prior ("does the effect exist?").



Inference about DCM parameters: Bayesian parameter averaging

FFX group analysis

- Likelihood distributions from different subjects are independent
- Under Gaussian assumptions, this is easy to compute
- Simply 'weigh' each subject's contribution by your certainty of the parameter

group posterior covariance

individual posterior covariances

$$\Sigma^{-1}_{\theta|y_1,\ldots,y_N} = \sum_{i=1}^{N} \Sigma^{-1}_{\theta|y_i}$$

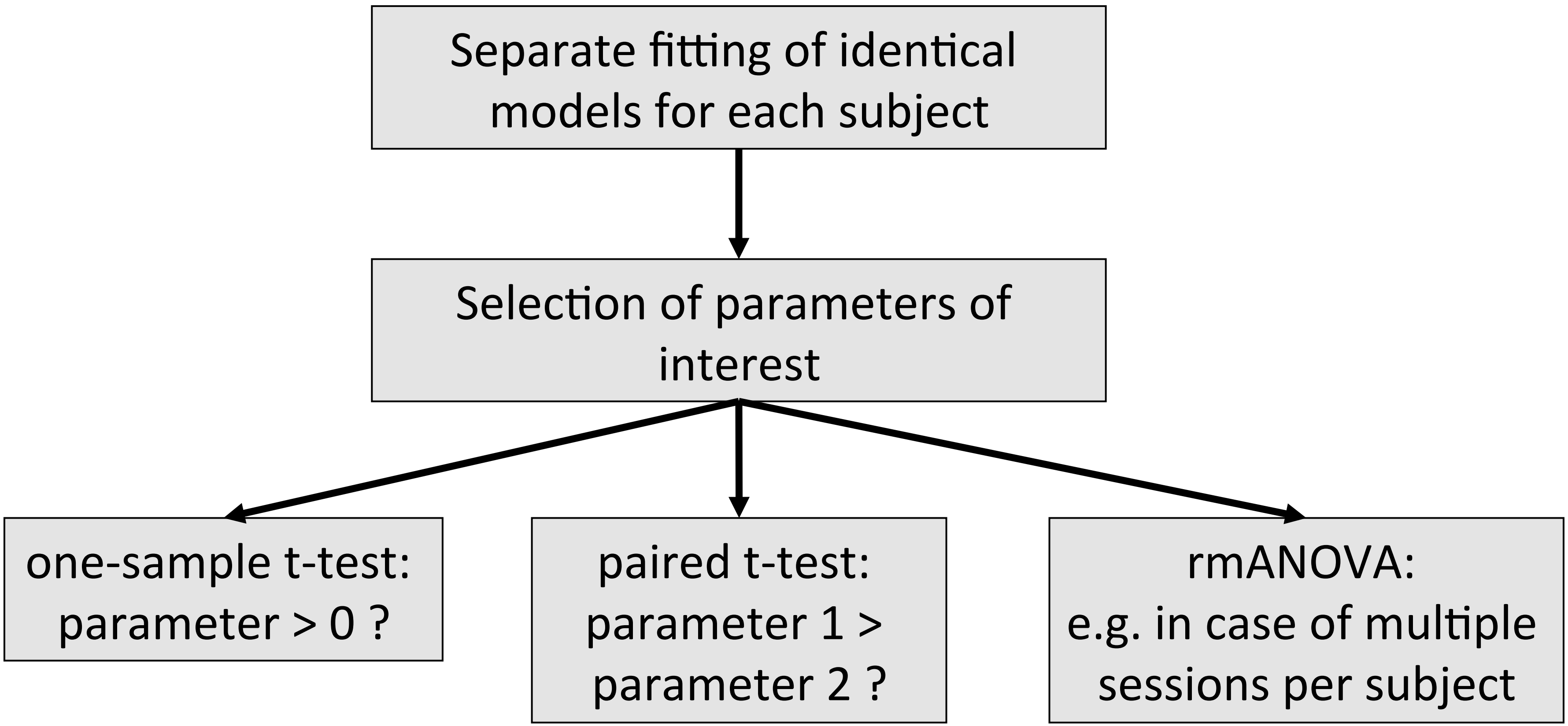
$$\mu_{\theta|y_1,\ldots,y_N} = \left(\sum_{i=1}^{N} \Sigma^{-1}_{\theta|y_i} \mu_{\theta|y_i} \right) \Sigma_{\theta|y_1,\ldots,y_N}$$

group posterior mean

individual posterior covariances and means

Inference about DCM parameters: RFX analysis (frequentist)

- 'Summary Statistic Approach'



Inference about models: Bayesian model comparison

- Prior / instead of to inference on parameters
- Which of various mechanisms / models best explains my data
- Use model evidence

→ accounts for <u>both</u> accuracy and complexity of the model

→ allows for inference about structure (generalisability) of the model

Fixed Effects Model selection via

log Group Bayes factor:

$$BF_{1,2} = \sum_k \ln p(y|m_1) - \sum_k \ln p(y|m_2)$$

Random Effects Model selection

via Model probability:

$$p(r \mid y, \alpha)$$

$$\langle r_k \rangle_q = \alpha_k / (\alpha_1 + \ldots + \alpha_K)$$

Bayes factors

For a given dataset, to compare two models, we compare their evidences.

$$B_{12} = \frac{p(y \mid m_1)}{p(y \mid m_2)}$$

Kass & Raftery 1995, *J. Am. Stat. Assoc.*

Kass & Raftery classification:

B_{12}	$p(m_1 \mid y)$	Evidence
1 to 3	50-75%	weak
3 to 20	75-95%	positive
20 to 150	95-99%	strong
≥ 150	≥ 99%	Very strong

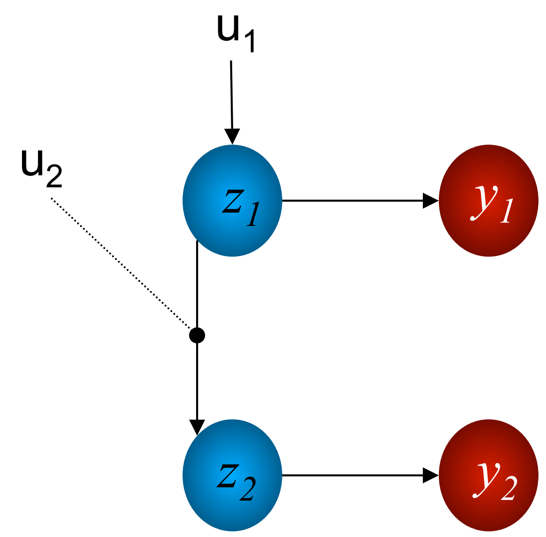
or their log evidences

$$\ln(B_{12}) \approx F_1 - F_2$$

Ketamine modulates:

1. All extrinsic connections,
2. Intrinsic NMDA and
3. Inhibitory / Modulatory processes (one of the red arrows) : use log bayes factors

Bayesian Model Comparison



One other way to view F!!

$$F = \log p(y \mid m) - KL\left[q(\theta), p(\theta \mid y, m)\right]$$

Accuracy - Complexity

$$KL\left[q(\theta), p(\theta \mid m)\right]$$
$$= \frac{1}{2}\ln\left|\Sigma_\theta\right| - \frac{1}{2}\ln\left|\Sigma_{\theta|y}\right| + \frac{1}{2}\left(\mu_{\theta|y} - \mu_\theta\right)^T \Sigma_\theta^{-1}\left(\mu_{\theta|y} - \mu_\theta\right)$$

The complexity term of *F* is higher

the more independent the prior parameters (↑ effective DFs)

the more dependent the posterior parameters

the more the posterior mean deviates from the prior mean

Overview

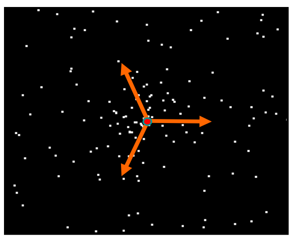
1. Dynamics in Dynamic Causal Modeling

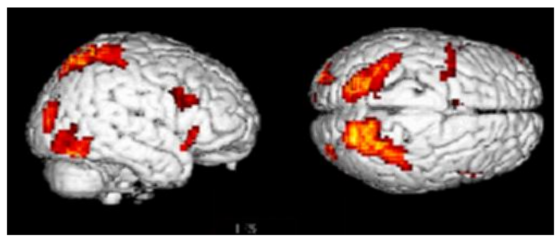
2. Graphical Model
- Variational Inversion
- Statistical Inference from VB

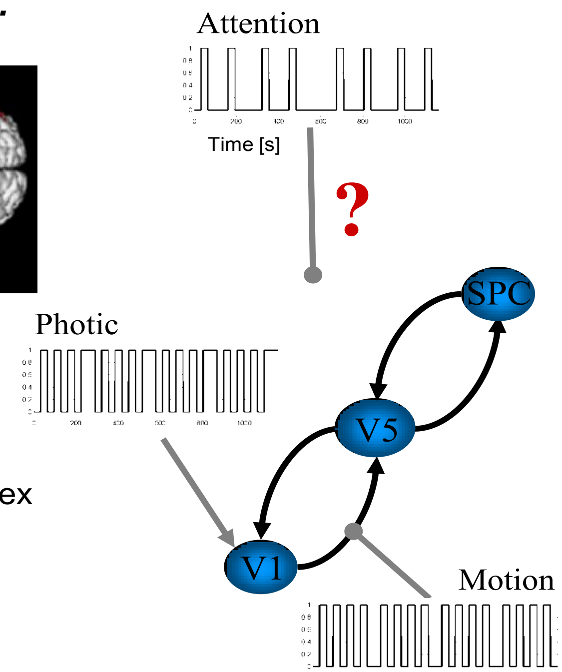
3. Examples
- Attention in the Human Brain
- Synesthesia

Example: Attention to motion

We used this model to assess the site of ***attention modulation*** during *visual motion processing* in an fMRI paradigm reported by *Büchel & Friston.*



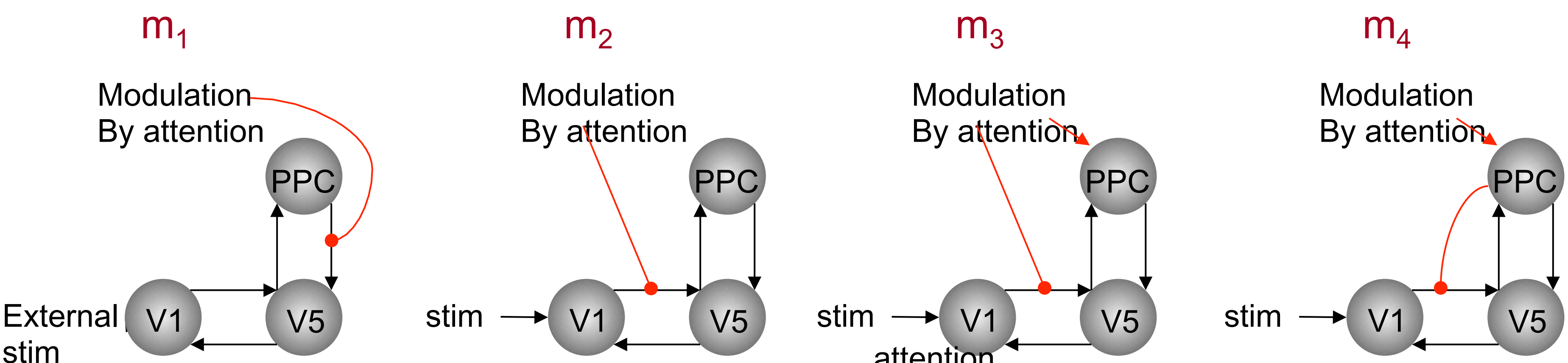




- fixation only
- observe static dots + photic → V1
- observe moving dots + motion → V5
- task on moving dots + attention → V5 + parietal cortex

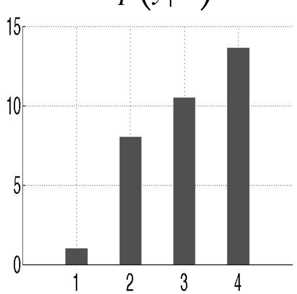
Friston et al. (2003) *NeuroImage*

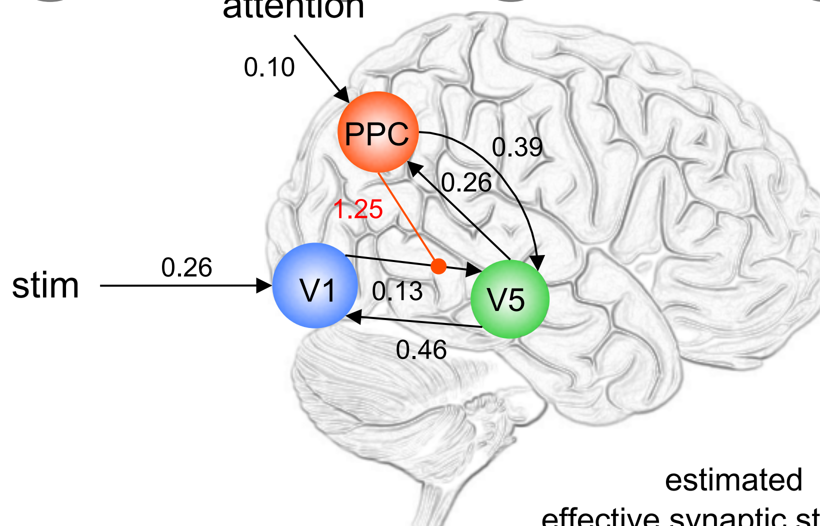
Bayesian model selection



models marginal likelihood

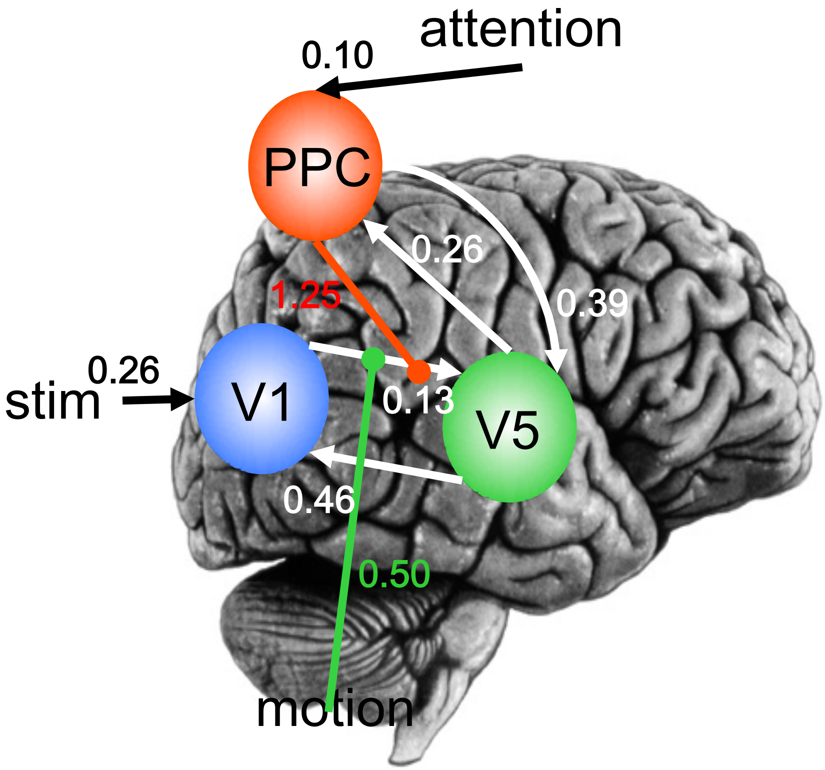
$\ln p(y|m)$

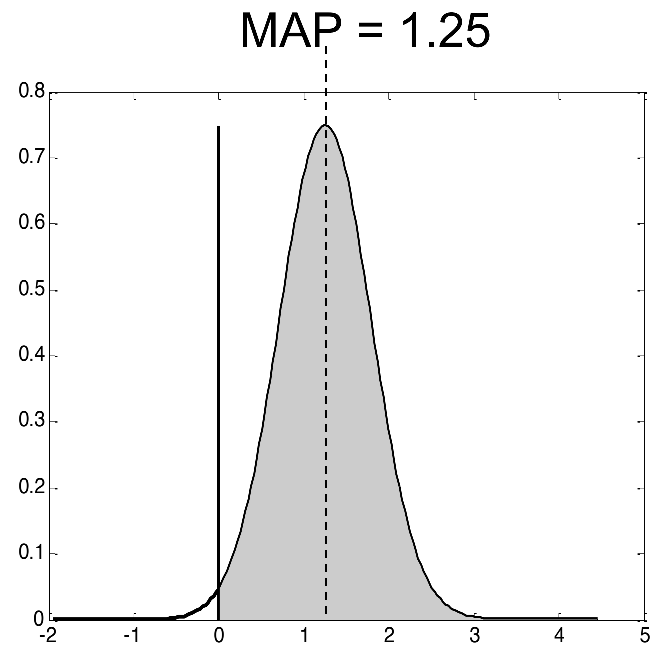




estimated
effective synaptic strengths
for best model (m_4)

Parameter inference

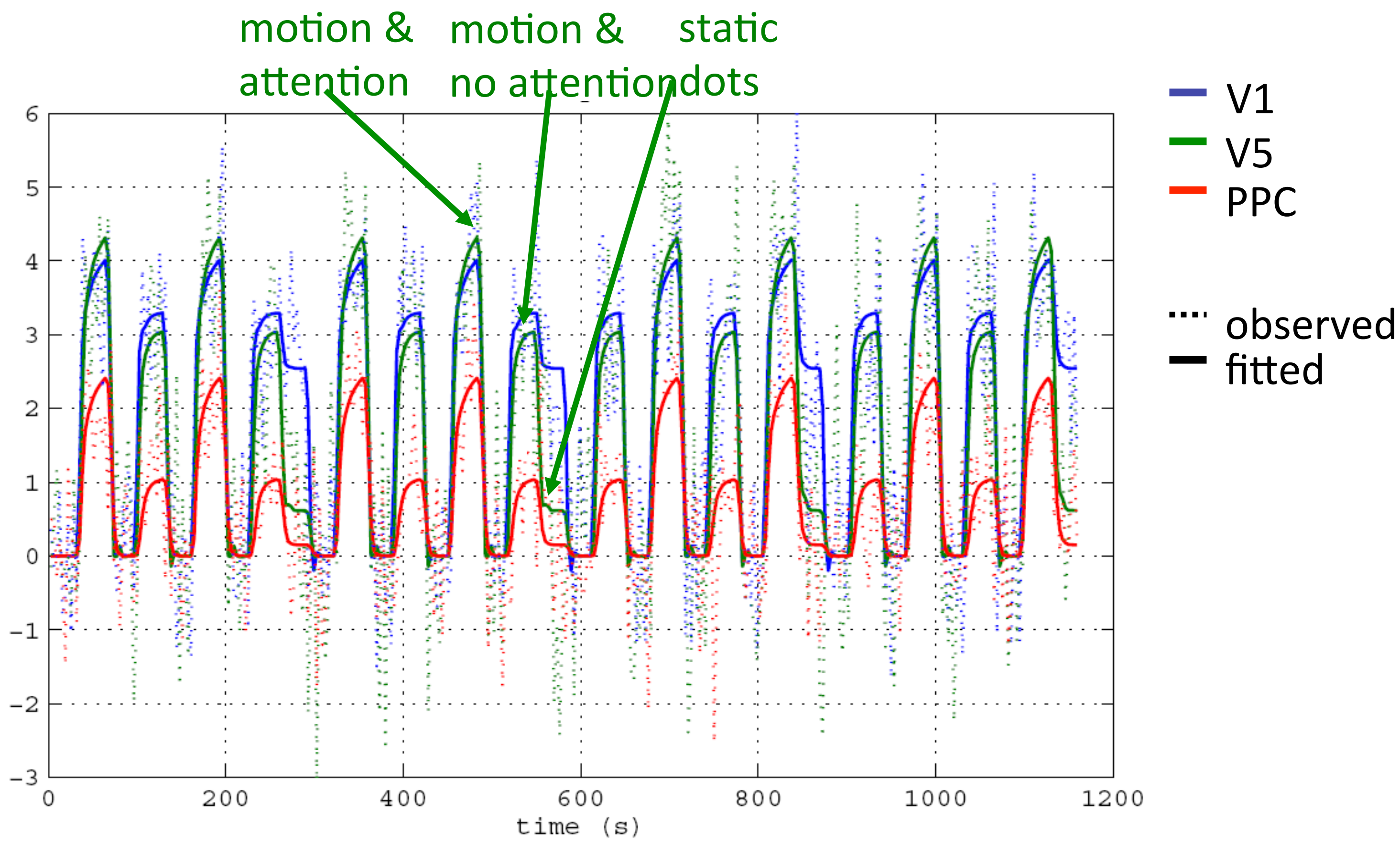




$$p(D^{PPC}_{V5,V1} > 0 \mid y) = 99.1\%$$

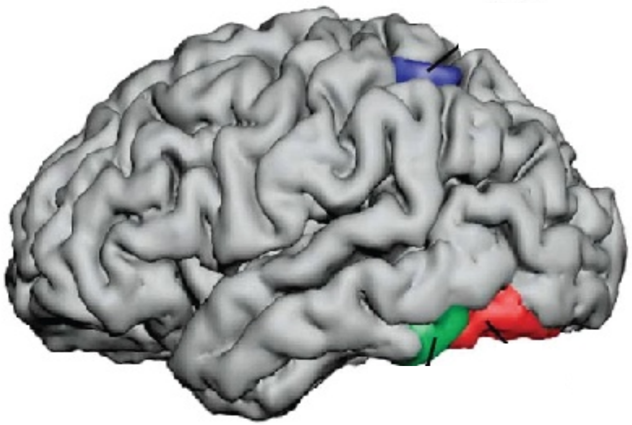
Stephan et al. 2008, *NeuroImage*

Data fits


motion & attention
motion & no attention
static dots
V1
V5
PPC
observed
fitted
6
5
4
3
2
1
0
−1
−2
−3
0
200
400
600
800
1000
1200
time (s)

Example 2: Brain Connectivity in Synesthesia

- Specific sensory stimuli lead to unusual, additional experiences
- Grapheme-color synesthesia: **color**
- Involuntary, automatic; stable over time, prevalenc
- Potential cause: aberrant **cross-activation** between
  - grapheme encoding area
  - color area V4
  - superior parietal lobule (SPL)

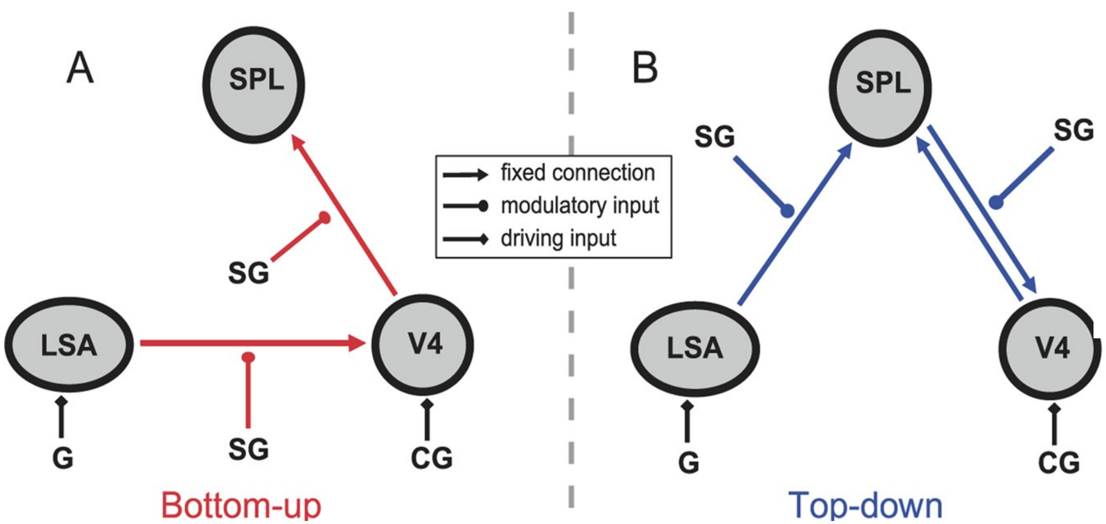


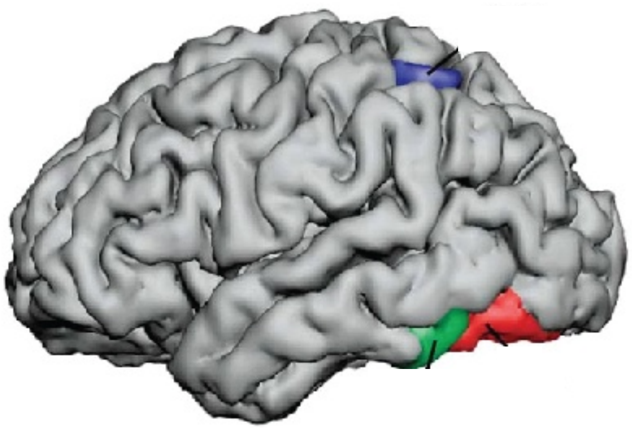
Hubbard, 2007

Can changes in effective connectivity explain synesthesia activity in V4?

DCM of Synesthesia

Models



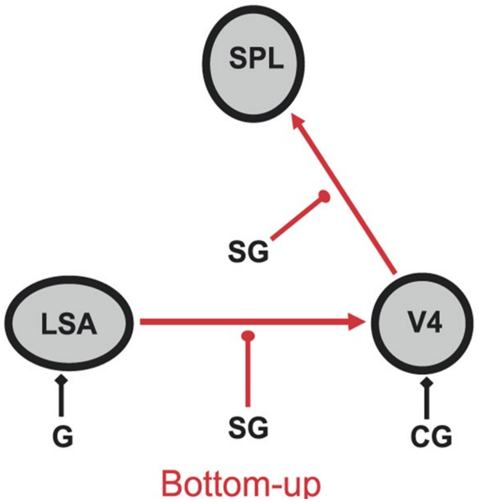


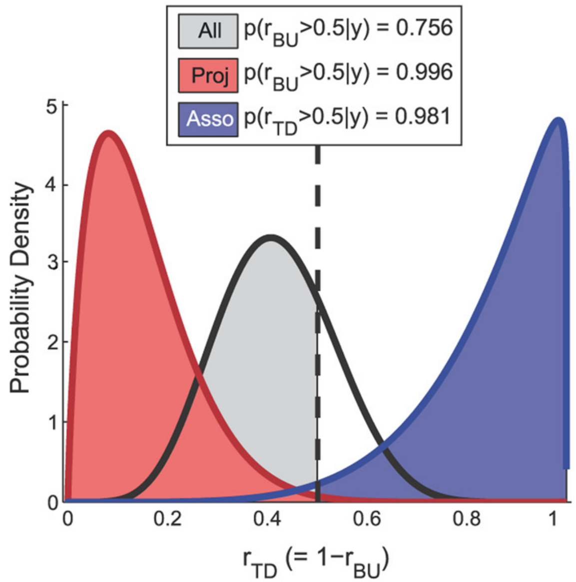
Hubbard, 2007

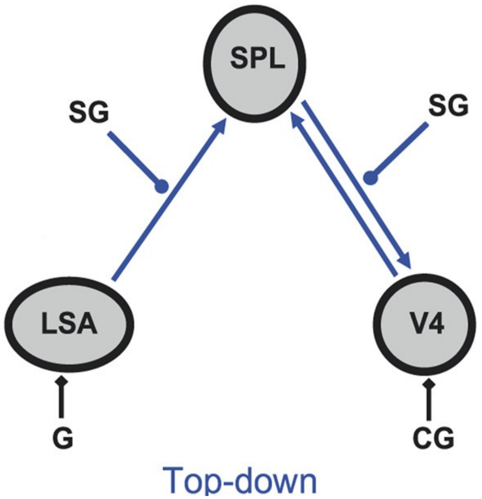
Van Leeuwen, den Ouden, Hagoort (2011) *JNeurosci*

DCM of Synesthesia

Model Evidence: $F \leq Z$

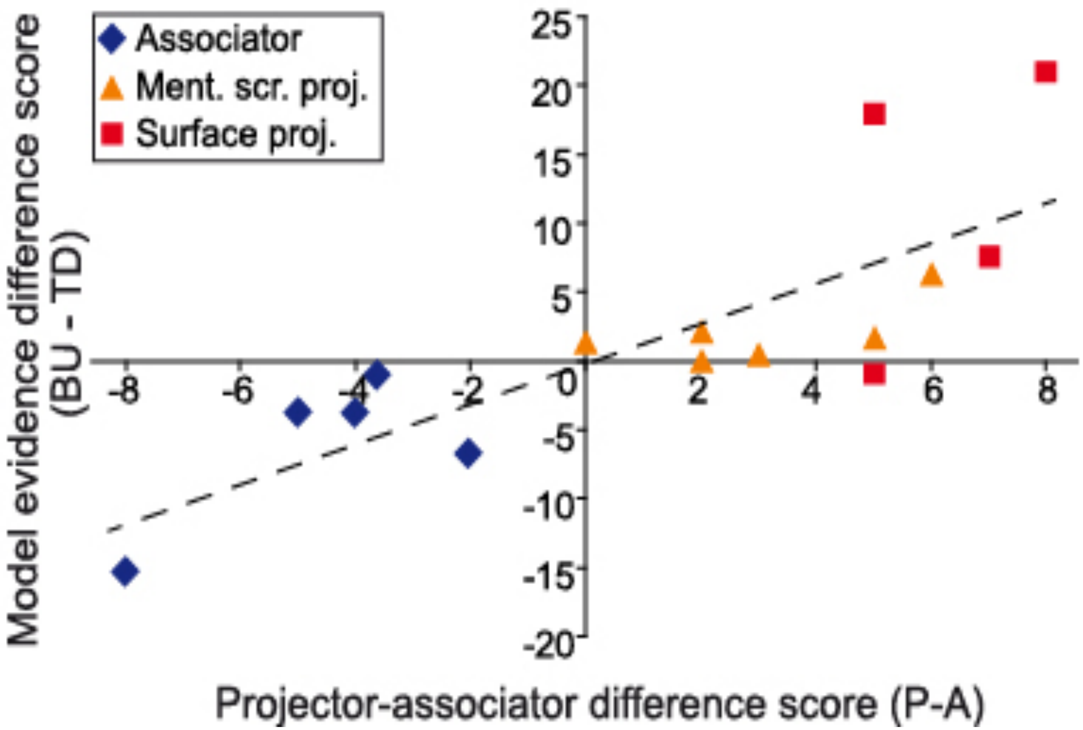






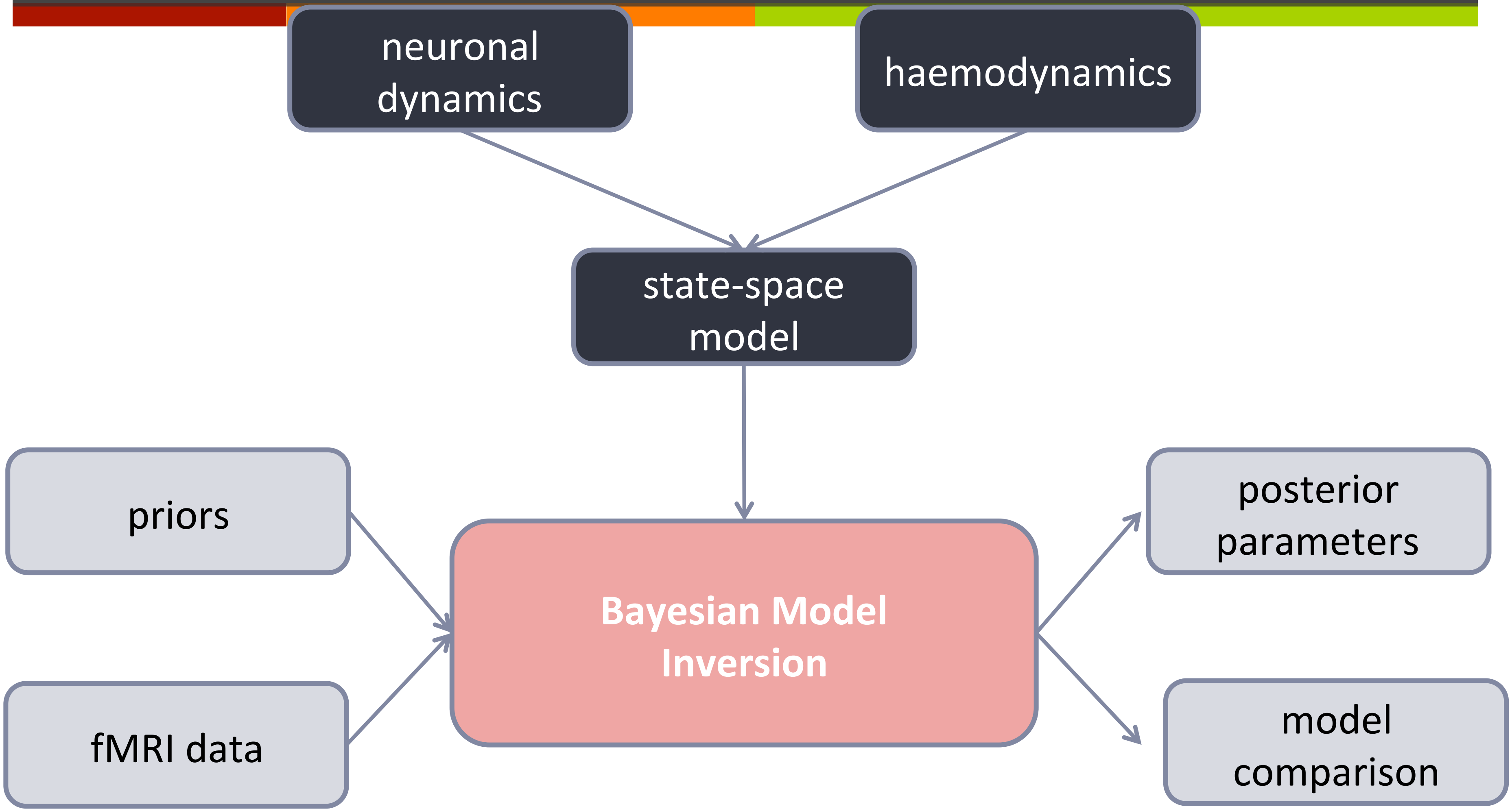
Van Leeuwen, den Ouden, Hagoort (2011) *JNeurosci*

Relative model evidence predicts sensory experience



Van Leeuwen, den Ouden, Hagoort (2011) *JNeurosci*

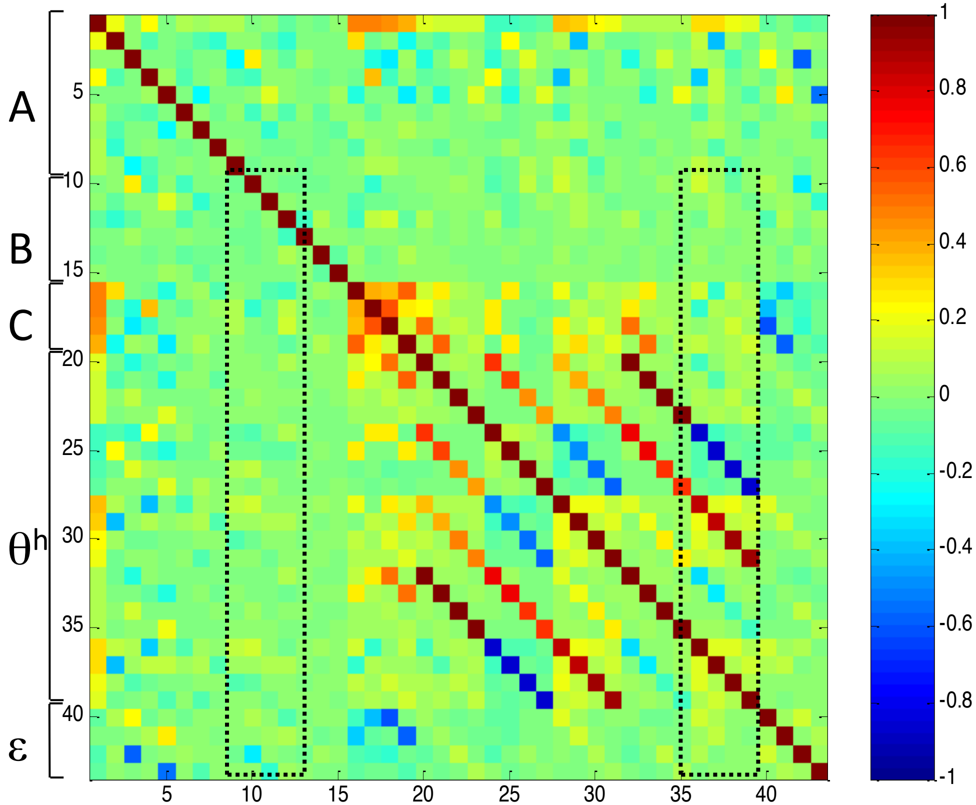
DCM Roadmap

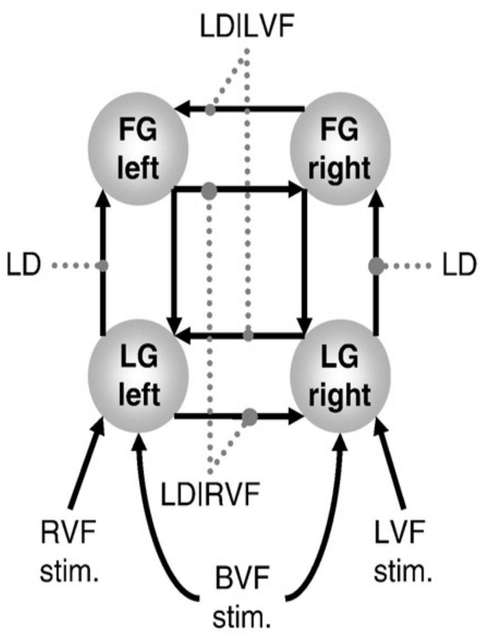

neuronal dynamics
haemodynamics
state-space model
priors
fMRI data
Bayesian Model Inversion
posterior parameters
model comparison

Some useful references

- 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.
- Bayesian model comparison: Comparing families of dynamic causal models (2010). Penny et al. *PLoS Comput Biol.* 6(3):e1000709.

How independent are neural and hemodynamic parameter estimates?





Stephan et al. (2007) *NeuroImage*