Bayesian inference

J. Daunizeau

Brain and Spine Institute, Paris, France
Wellcome Trust Centre for Neuroimaging, London, UK

Overview of the talk

1 Probabilistic modelling and representation of uncertainty

1.1 Bayesian paradigm

1.2 Hierarchical models

1.3 Frequentist versus Bayesian inference

2 Notes on Bayesian inference

2.1 Variational methods (ReML, EM, VB)

2.2 Family inference

2.3 Group-level model comparison

3 SPM applications

3.1 aMRI segmentation

3.2 Decoding of brain images

3.3 Model-based fMRI analysis (with spatial priors)

3.4 Dynamic causal modelling

Overview of the talk

1 Probabilistic modelling and representation of uncertainty

1.1 Bayesian paradigm

1.2 Hierarchical models

1.3 Frequentist versus Bayesian inference

2 Notes on Bayesian inference

2.1 Variational methods (ReML, EM, VB)

2.2 Family inference

2.3 Group-level model comparison

3 SPM applications

3.1 aMRI segmentation

3.2 Decoding of brain images

3.3 Model-based fMRI analysis (with spatial priors)

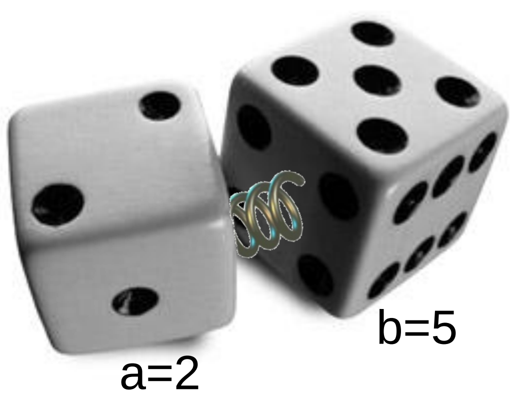
3.4 Dynamic causal modelling

Bayesian paradigm

probability theory: basics

***Degree of** **plausibility** desiderata*:
- should be represented using real numbers (D1)
- should conform with intuition (D2)
- should be consistent (D3)

• normalization: $$\sum_a P(a) = 1$$

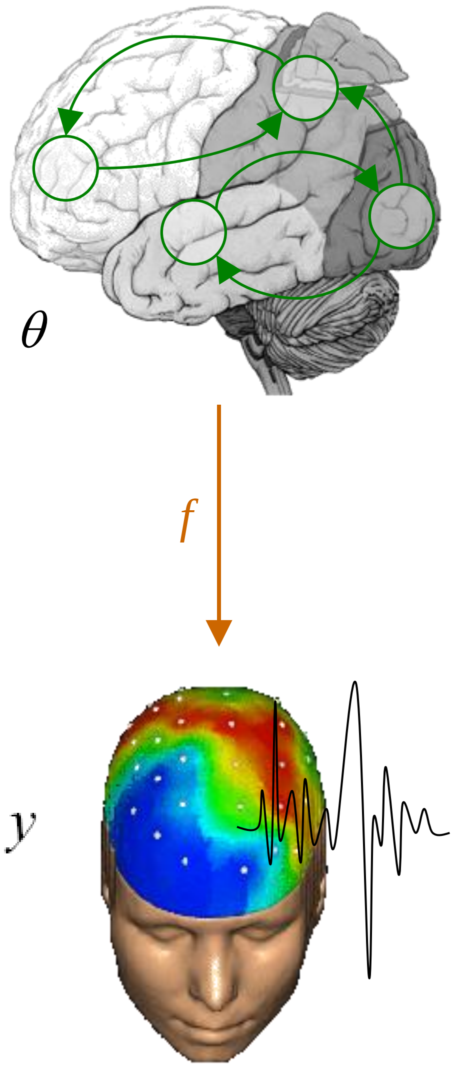


• marginalization: $$P(b) = \sum_a P(a,b)$$

• conditioning :
(*Bayes rule*)
$$P(a,b) = P(a|b)\,P(b) = P(b|a)\,P(a)$$

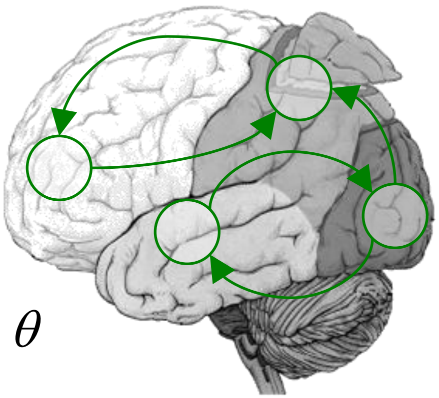
Bayesian paradigm

deriving the likelihood function



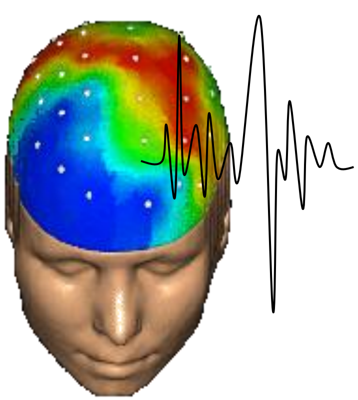
Bayesian paradigm

likelihood, priors and the model evidence



θ

generative model m

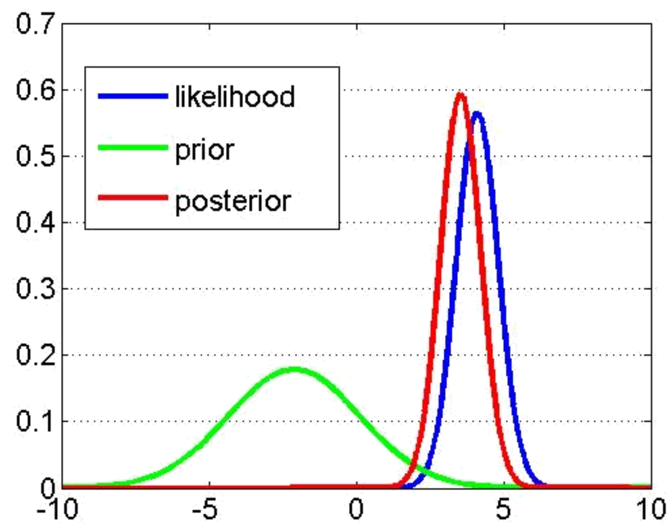


y

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

Prior: $p(\theta|m)$

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



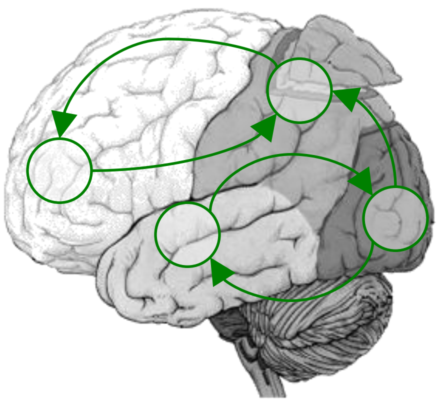
Bayesian paradigm

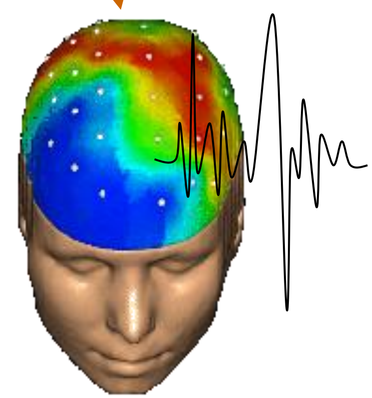
forward and inverse problems

forward problem

$$p\left(y \middle| \vartheta, m\right)$$

likelihood





posterior distribution

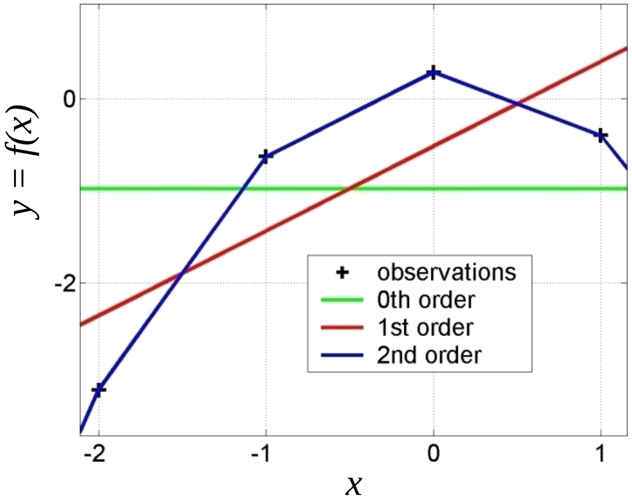
$$p\left(\vartheta \middle| y, m\right)$$

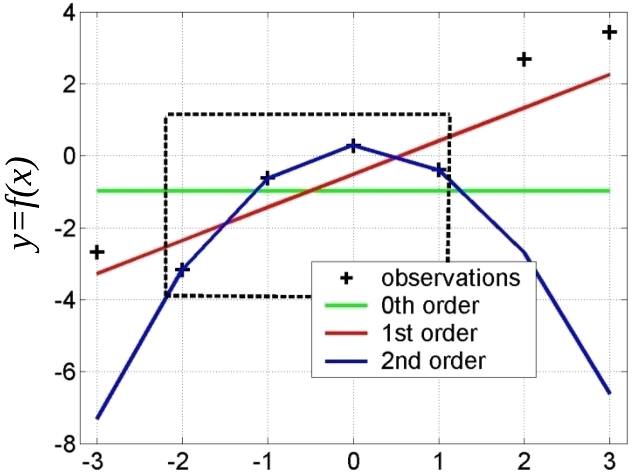
inverse problem

Bayesian paradigm

model comparison

Principle of parsimony :
« plurality should not be assumed without necessity »

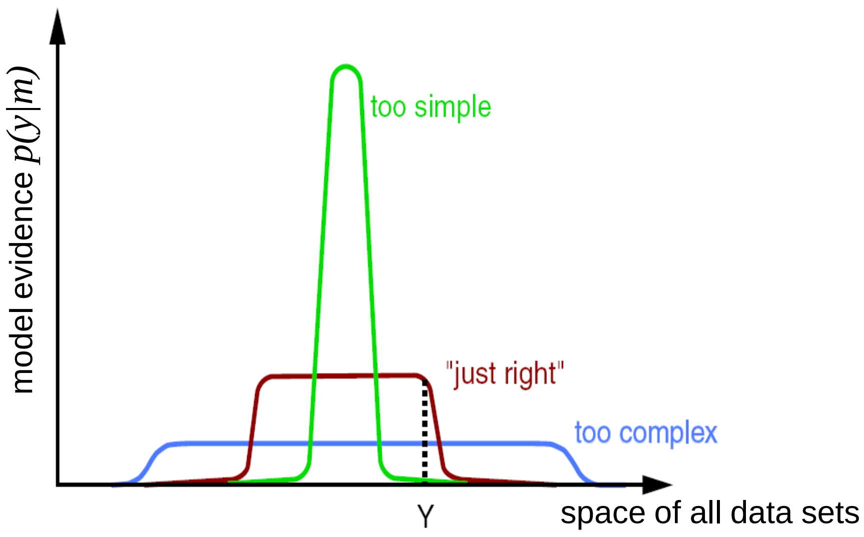




Model evidence:

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

"Occam's razor" :

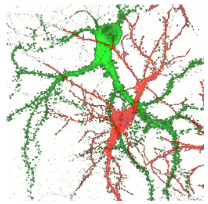


Hierarchical models

principle

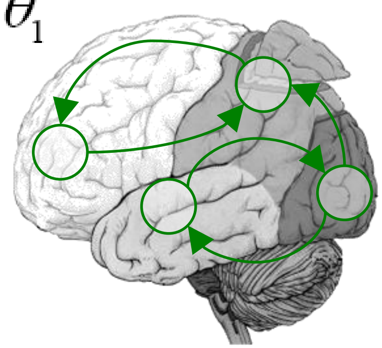
$\vdots$

θ_2



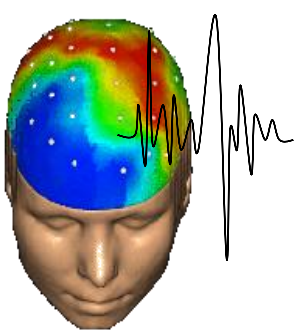
$$p(\theta_2 | \theta_3, m)$$

θ_1



$$p(\theta_1 | \theta_2, m)$$

y



$$p(y | \theta_1, m)$$

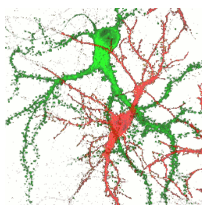
inference

causality

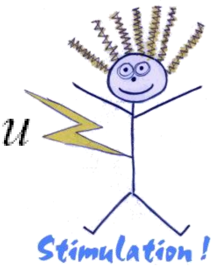
Hierarchical models

directed acyclic graphs (DAGs)

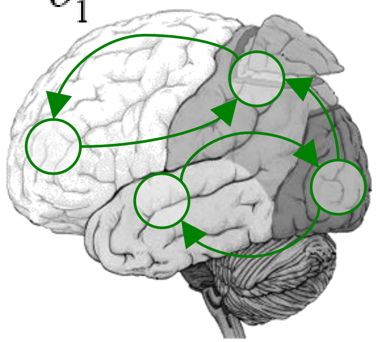
θ_2



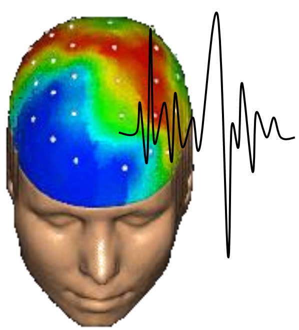
u

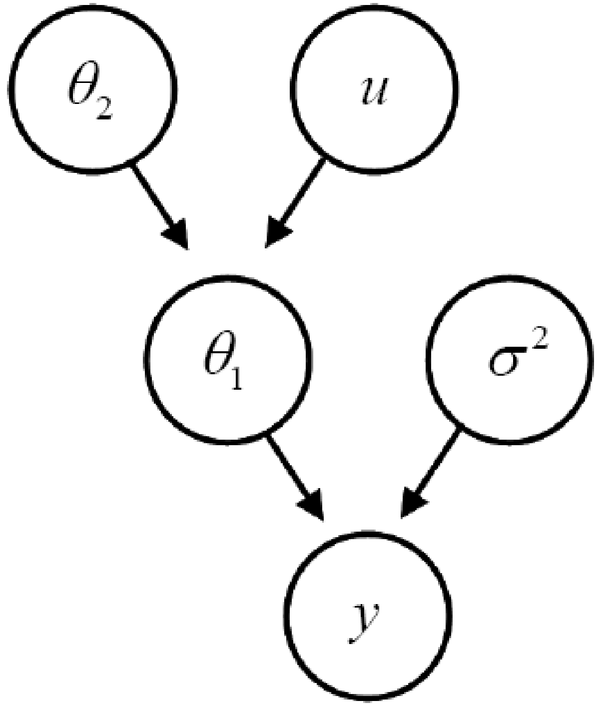


θ_1



y





$$p\left(\theta_1 \middle| \theta_2, u, m\right)$$

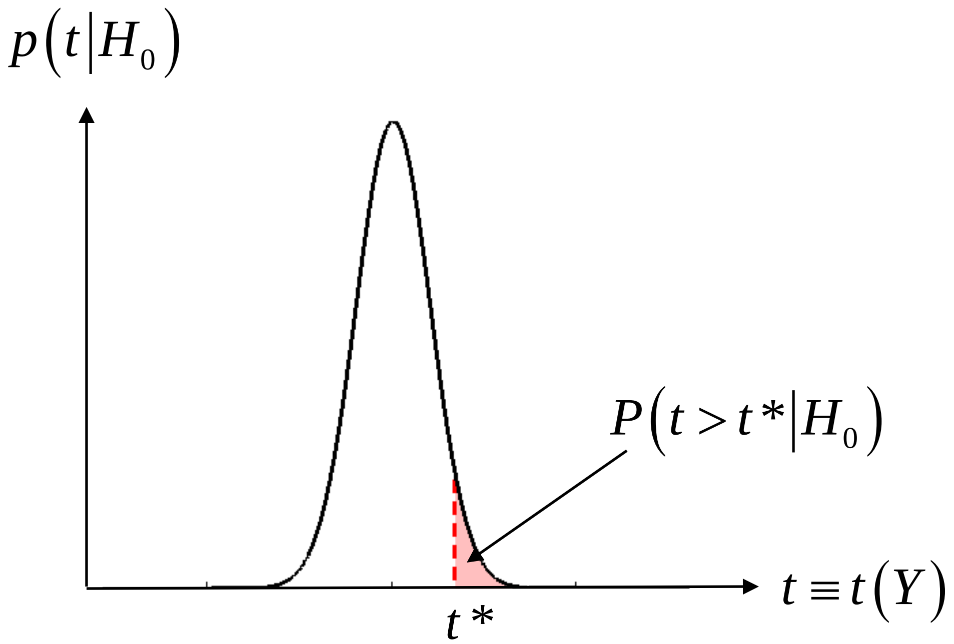
$$p\left(y \middle| \theta_1, \sigma^2, m\right)$$

$$p\left(\theta \middle| m\right) = \prod_j p\left(\theta_j \middle| par\left(\theta_j\right), m\right)$$

Frequentist versus Bayesian inference

a (quick) note on hypothesis testing

- define the null, e.g.: $H_0 : \theta = 0$



- estimate parameters (obtain test stat.)

- apply decision rule, i.e.:

  if $P\left(t > t^* \middle| H_0\right) \leq \alpha$ then reject H0

classical (null) hypothesis testing

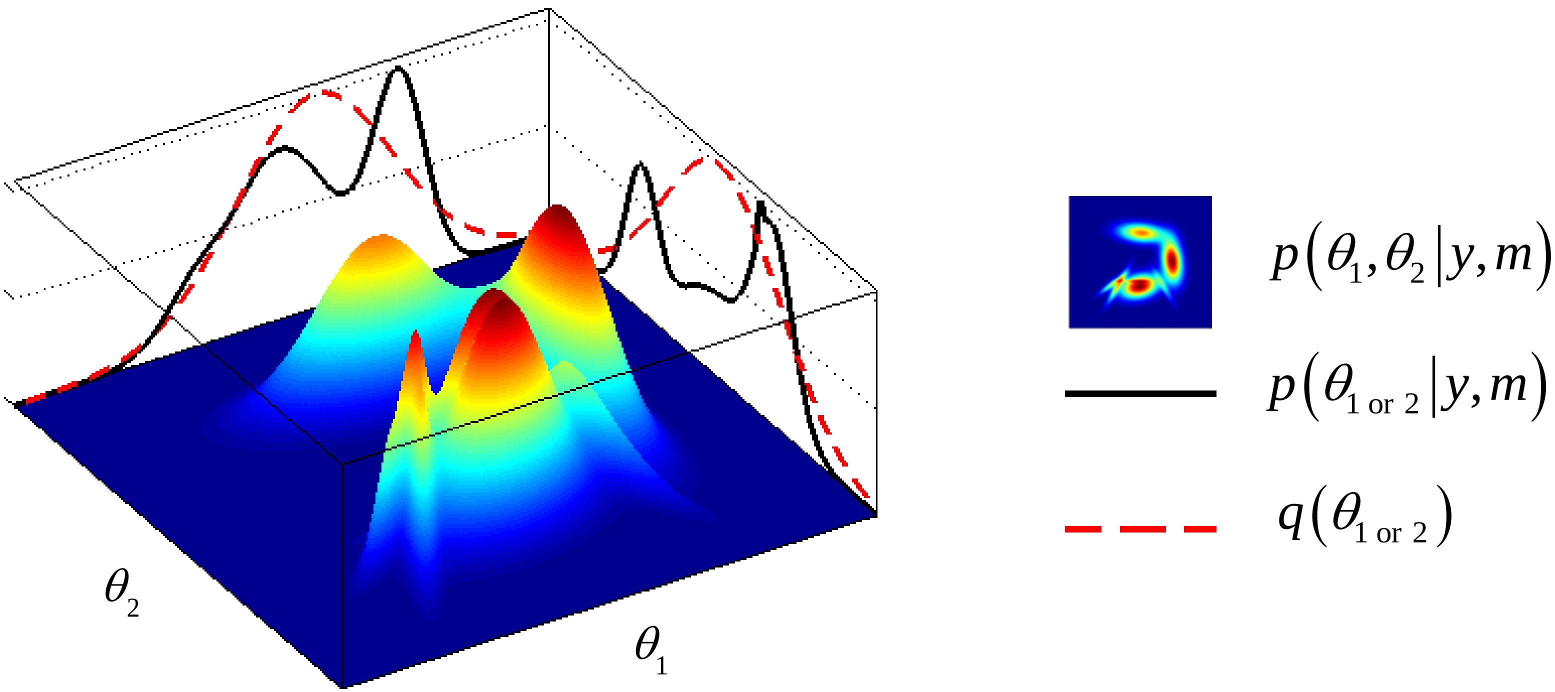
Overview of the talk

Variational methods

VB / EM / ReML

$$\ln p(y|m) = \underbrace{\left\langle \ln p(\theta, y|m) \right\rangle_q + S(q)}_{\text{free energy } F(q)} + D_{KL}\left(q(\theta); p(\theta|y,m)\right)$$

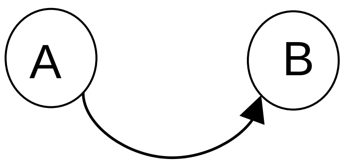
→ **VB** : maximize the free energy $F(q)$ w.r.t. the approximate posterior $q(\theta)$ under some (e.g., *mean field, Laplace*) simplifying constraint



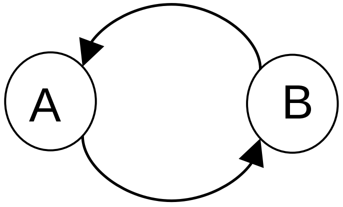
Family-level inference

trading inference resolution against statistical power

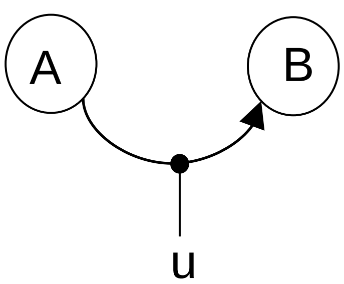
$P(m_1|y) = 0.04$



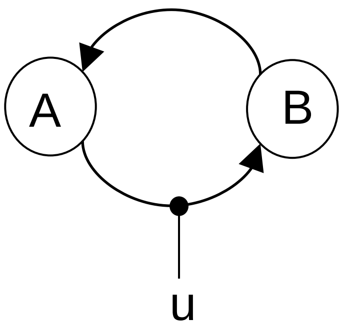
$P(m_2|y) = 0.25$



$P(m_2|y) = 0.01$



$P(m_2|y) = 0.7$



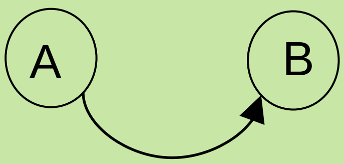
model selection error risk:

$$P(e=1|y) = 1 - \max_m P(m|y)$$
$$= 0.3$$

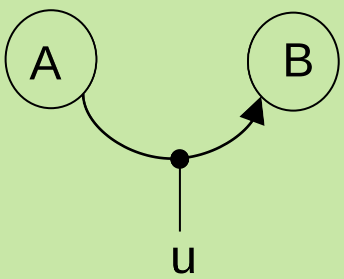
Family-level inference

trading inference resolution against statistical power

P(m_1|y) = 0.04

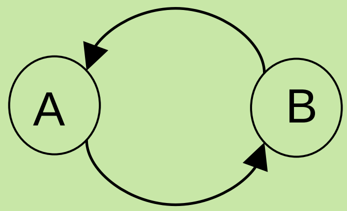


P(m_2|y) = 0.01

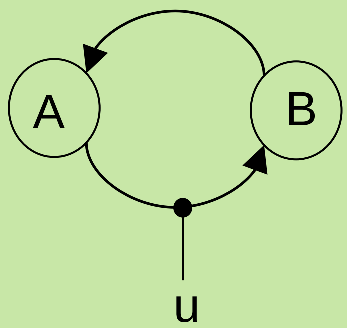


P(f_1|y) = 0.05

P(m_2|y) = 0.25



P(m_2|y) = 0.7



P(f_2|y) = 0.95

model selection error risk:

$$P\left(e=1|y\right)=1-\max_{m} P\left(m|y\right)$$
$$=0.3$$

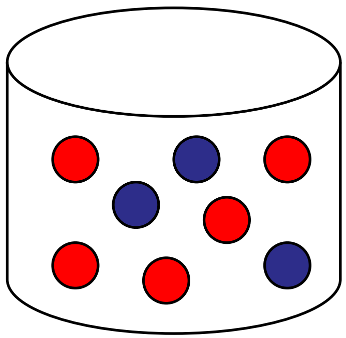
family inference
(pool statistical evidence)

$$P\left(f|y\right)=\sum_{m\in f} P\left(m|y\right)$$

$$P\left(e=1|y\right)=1-\max_{f} P\left(f|y\right)$$
$$=0.05$$

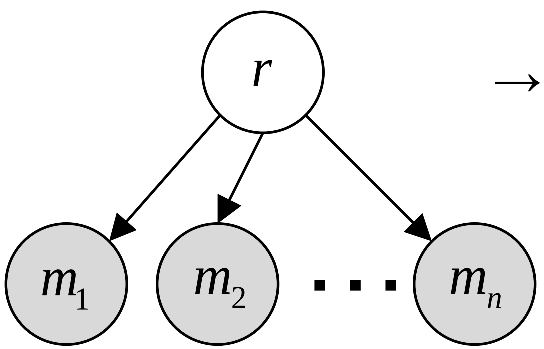
Group-level model comparison

preliminary: Polya's urn



$$\begin{cases} m_i = 1 & \rightarrow i^{\text{th}} \text{ marble is blue} \\ m_i = 0 & \rightarrow i^{\text{th}} \text{ marble is purple} \end{cases}$$

r = proportion of blue marbles in the urn



$\rightarrow$ (binomial) probability of drawing a set of n marbles:

$$p(m|r) = \prod_{i=1}^{n} r^{m_i} (1-r)^{1-m_i}$$

Thus, our belief about the proportion of blue marbles is:

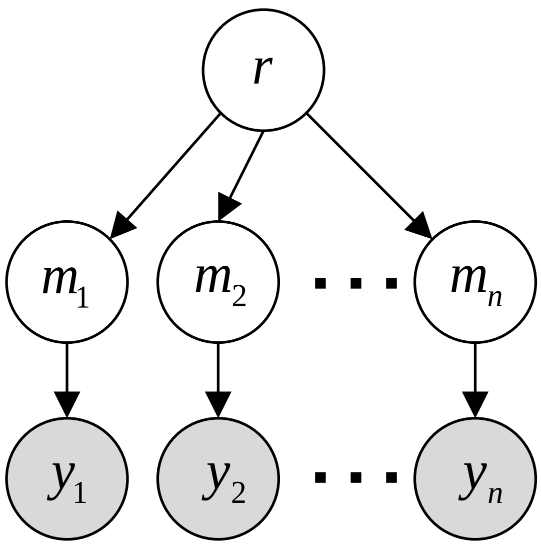
$$p(r|m) \propto p(r) \prod_{i=1}^{n} r^{m_i} (1-r)^{1-m_i} \quad \overset{p(r)\propto 1}{\Rightarrow} \quad E[r|m] = \frac{1}{n}\sum_{i=1}^{n} m_i$$

Group-level model comparison

what if we are colour blind?

At least, we can measure how likely is the i^{th} subject's data under each model!

$p(y_1|m_1)$ $p(y_2|m_2)$ $p(y_i|m_i)$ $p(y_n|m_n)$



$$p(r,m|y) \propto p(r)\prod_{i=1}^{n} p(y_i|m_i)\,p(m_i|r)$$

Our belief about the proportion of models is:

$$p(r|y) = \sum_m p(r,m|y)$$

Exceedance probability: $\varphi_k = P\left(r_k > r_{k'\neq k} \mid y\right)$

Overview of the talk

1 Probabilistic modelling and representation of uncertainty

1.1 Bayesian paradigm

1.2 Hierarchical models

1.3 Frequentist versus Bayesian inference

2 Notes on Bayesian inference

2.1 Variational methods (ReML, EM, VB)

2.2 Family inference

2.3 Group-level model comparison

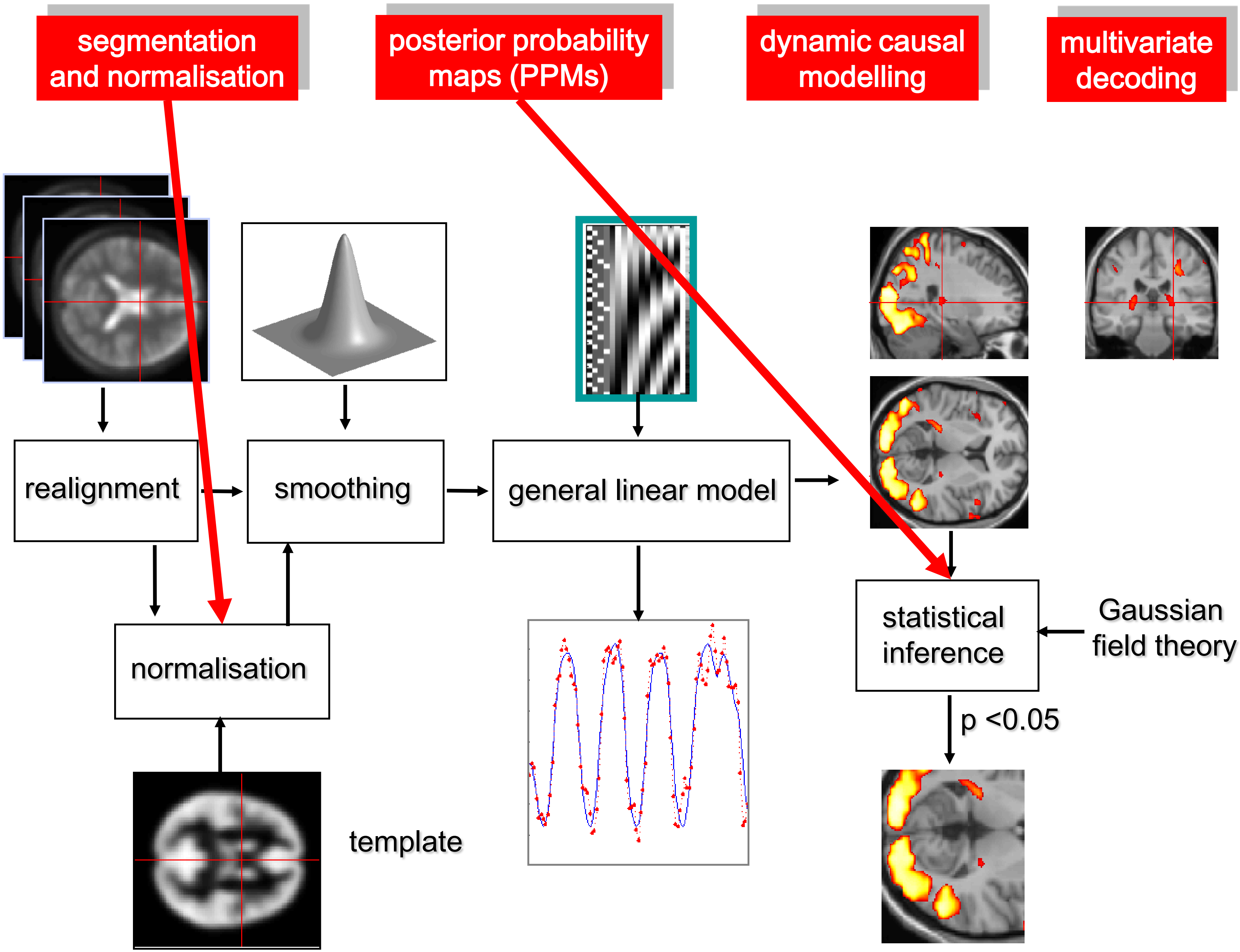
3 SPM applications

3.1 aMRI segmentation

3.2 Decoding of brain images

3.3 Model-based fMRI analysis (with spatial priors)

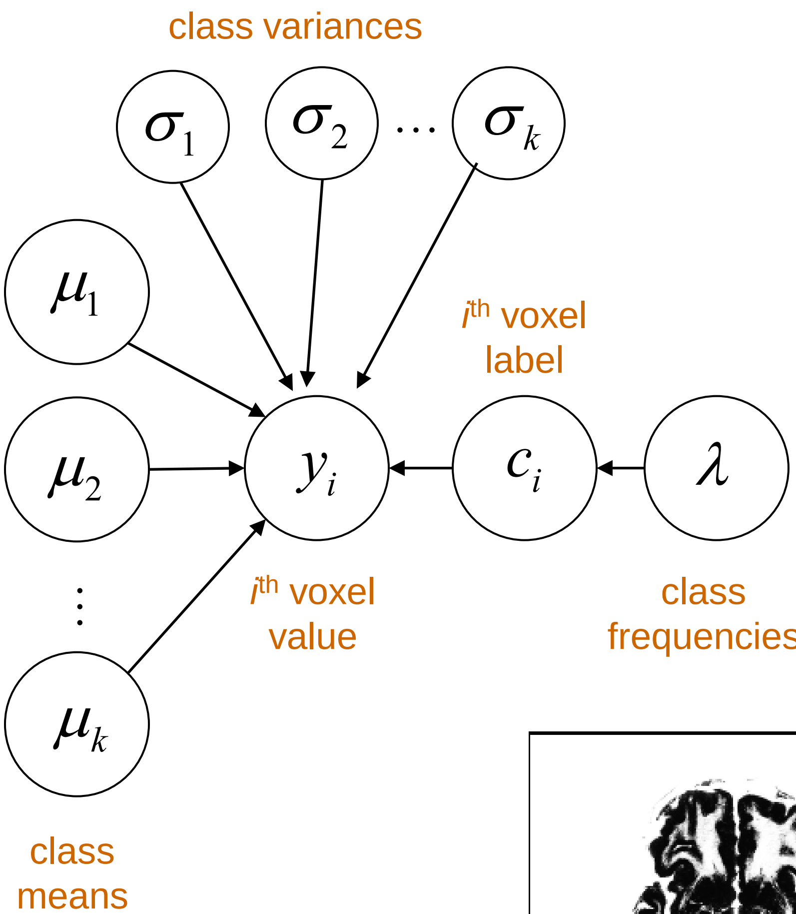
3.4 Dynamic causal modelling

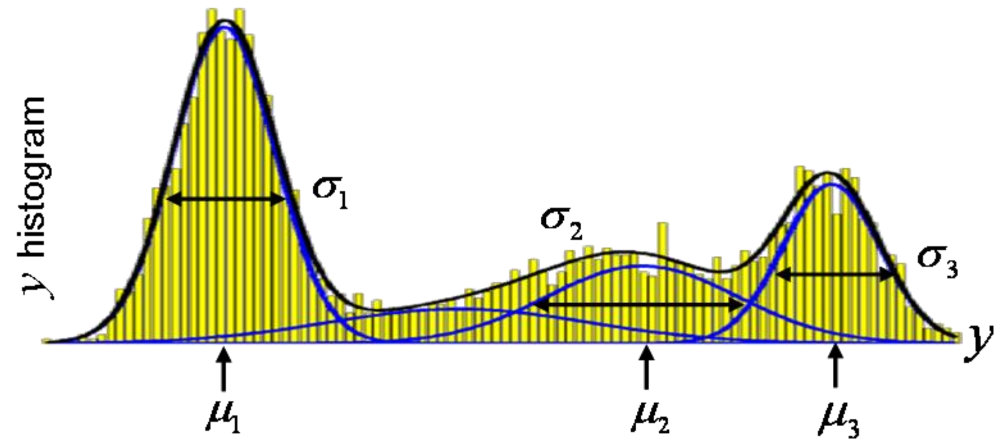


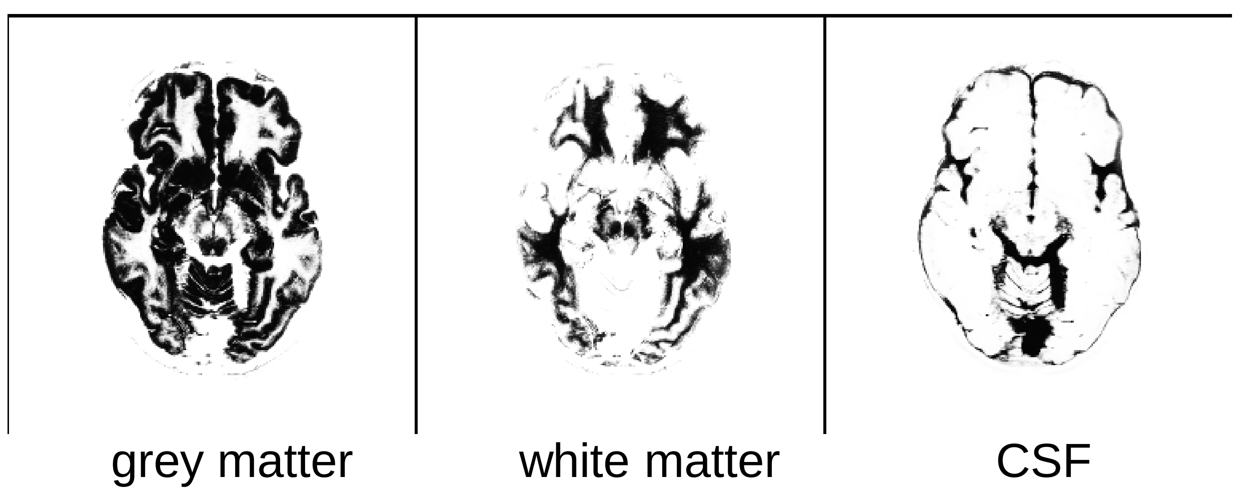
segmentation and normalisation
posterior probability maps (PPMs)
dynamic causal modelling
multivariate decoding
realignment
smoothing
general linear model
normalisation
template
statistical inference
Gaussian field theory
p <0.05

aMRI segmentation

mixture of Gaussians (MoG) model

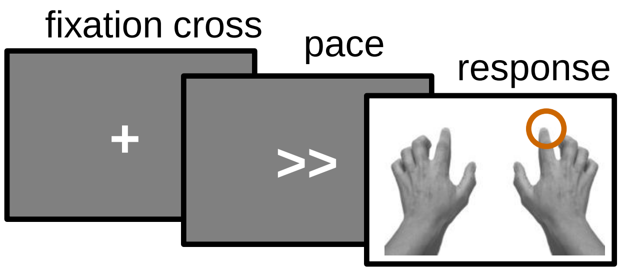


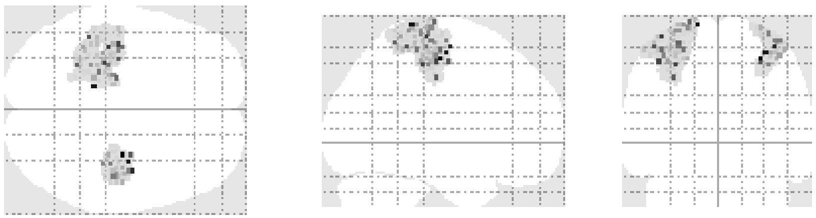




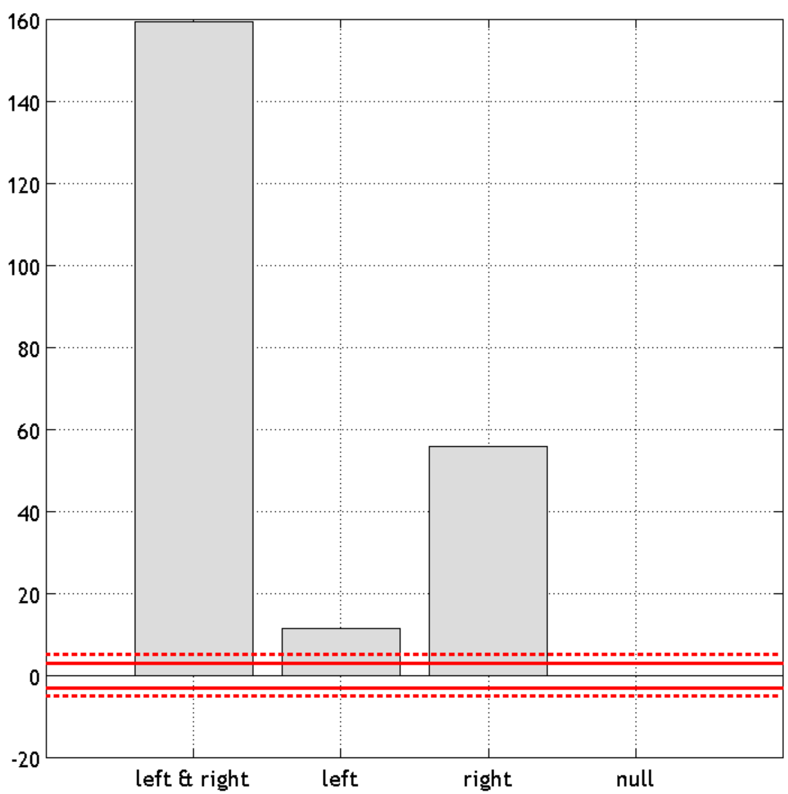
Decoding of brain images

recognizing brain states from fMRI

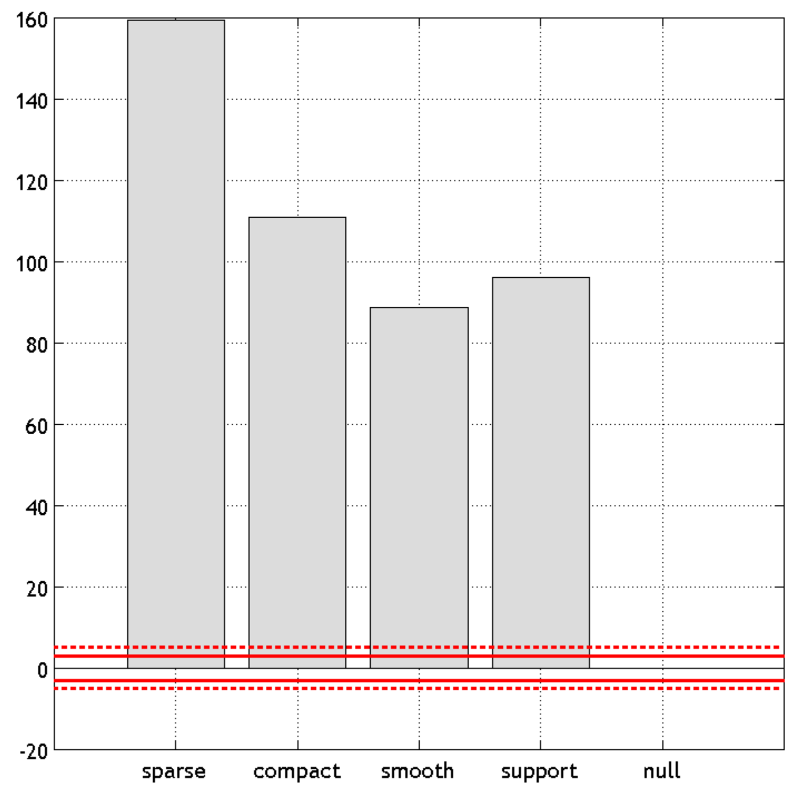




log-evidence of X-Y sparse mappings:
effect of lateralization

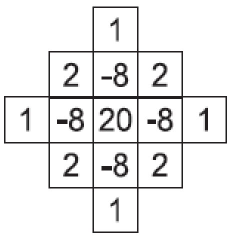


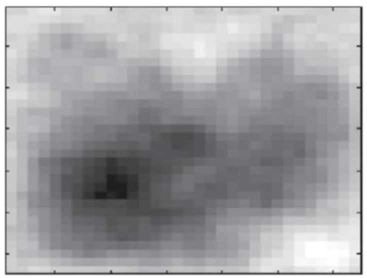
log-evidence of X-Y bilateral mappings:
effect of spatial deployment



fMRI time series analysis

spatial priors and model comparison





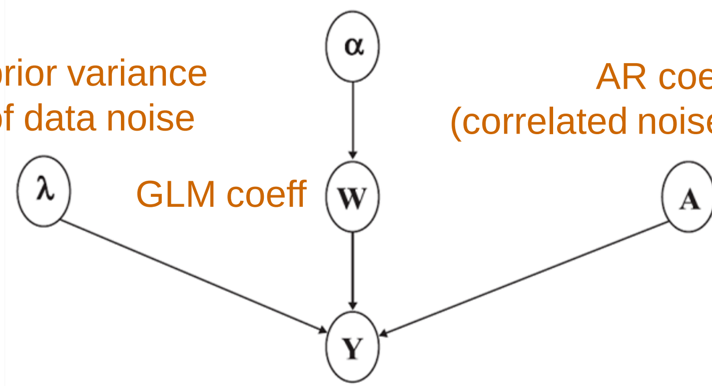
prior variance of GLM coeff

prior variance of data noise

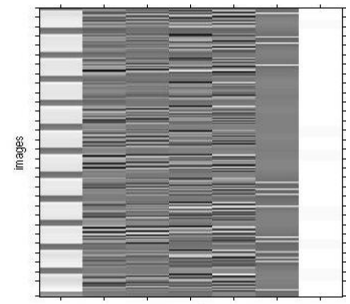
AR coeff (correlated noise)

GLM coeff

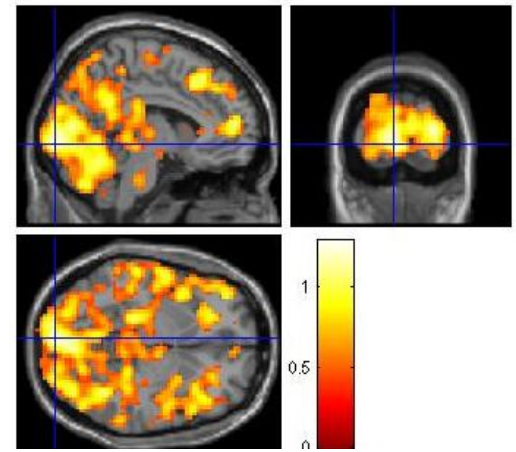
fMRI time series



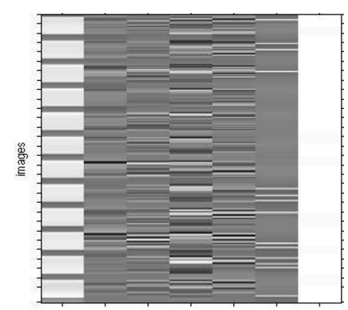
short-term memory design matrix (X)



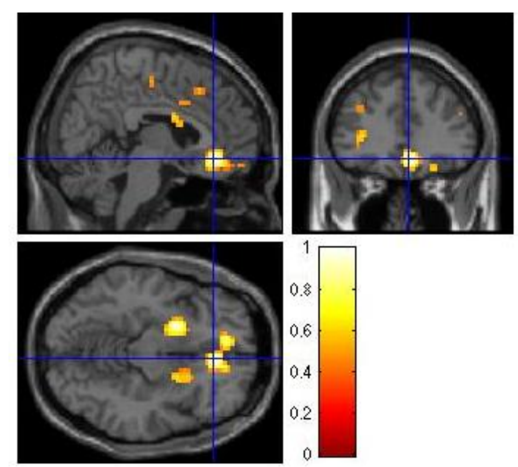
PPM: regions best explained by short-term memory model



long-term memory design matrix (X)

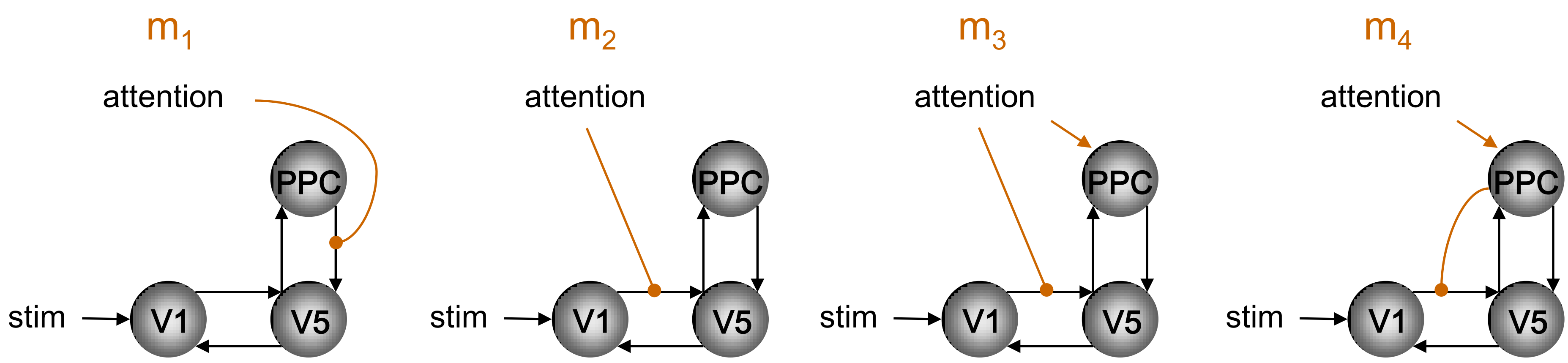


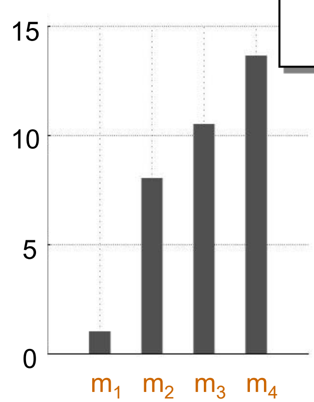
PPM: regions best explained by long-term memory model



Dynamic Causal Modelling

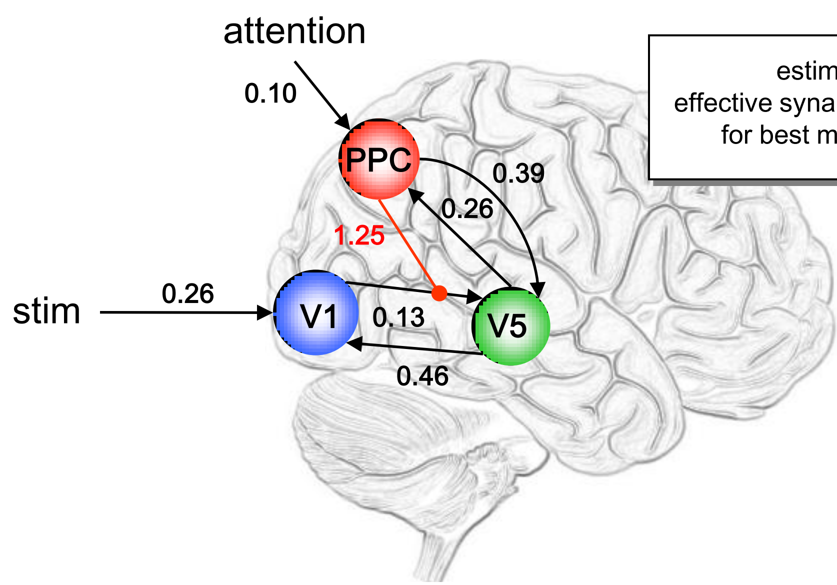
network structure identification





models marginal likelihood

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



estimated
effective synaptic strengths
for best model (m_4)

I thank you for your attention.

A note on statistical significance

lessons from the Neyman-Pearson lemma

- Neyman-Pearson lemma: the likelihood ratio (or Bayes factor) test

$$\Lambda = \frac{p\left(y \middle| H_1\right)}{p\left(y \middle| H_0\right)} \geq u$$

is the most powerful test of size $\alpha = p\left(\Lambda \geq u \middle| H_0\right)$ to test the null.

- what is the threshold u, above which the Bayes factor test yields a error I rate of 5%?

ROC analysis