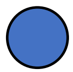
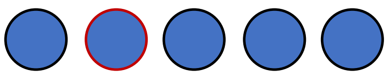
Bayesian Model Selection and Averaging

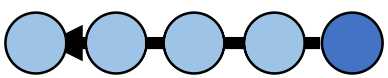
SPM for MEG/EEG Course

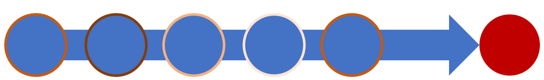
Ulrich Stoof

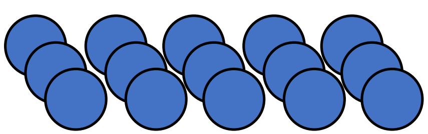
Contents



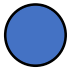




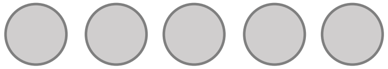




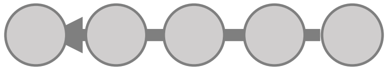
Contents



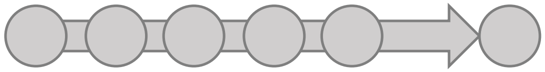
Introduction and model inversion
Dynamic Causal Modelling (DCM)



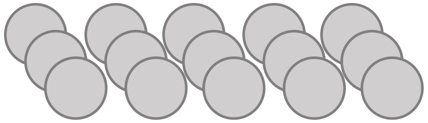
Comparing models
Bayesian Model Comparison and Selection (BMS)



Rapidly evaluating models
Bayesian Model Reduction (BMR)

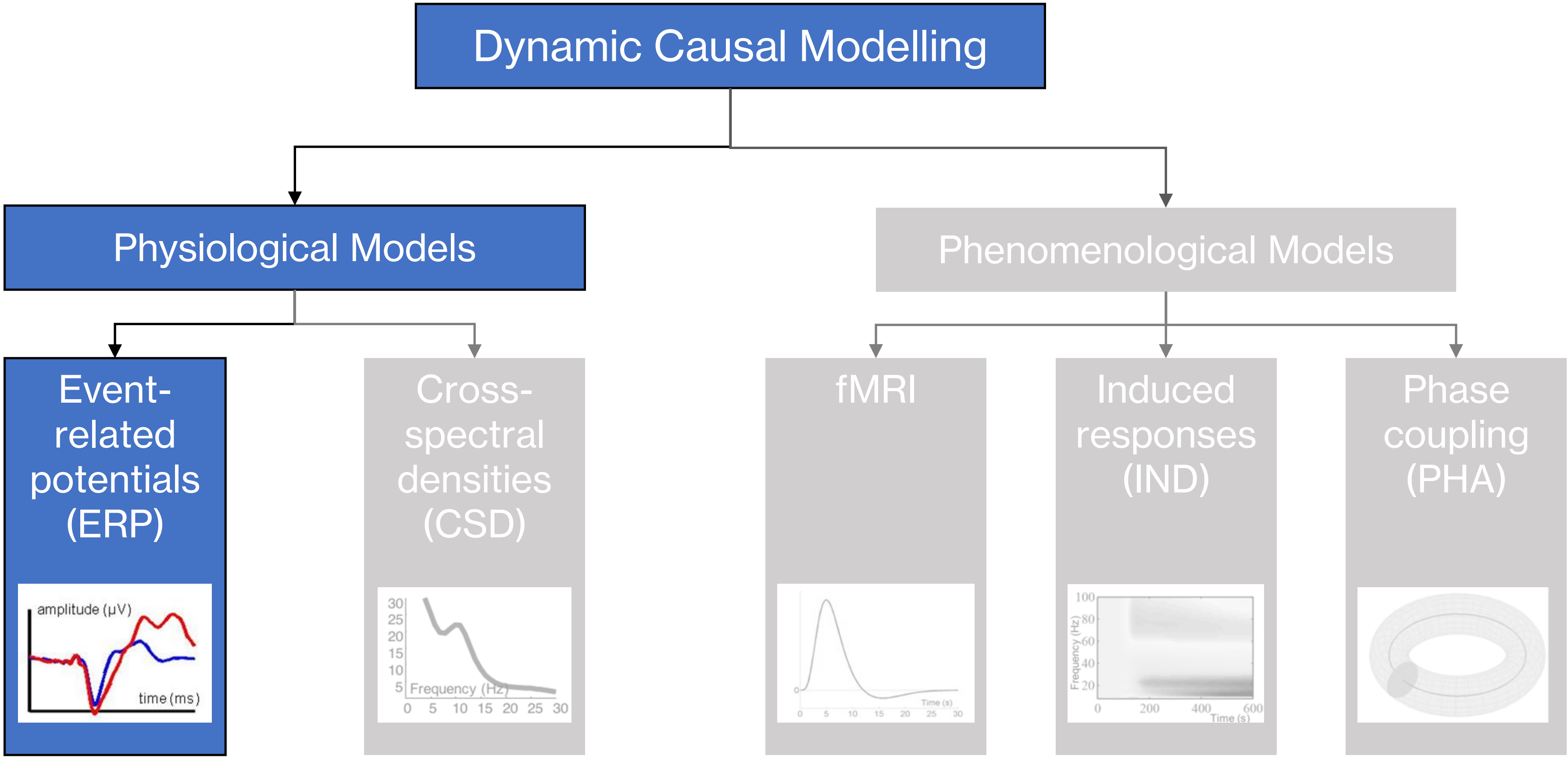


Investigating the parameters
Bayesian Model Averaging (BMA)



Multi-subject analysis
Parametric Empirical Bayes (PEB)

DCM for ERP – Examples in this Presentation



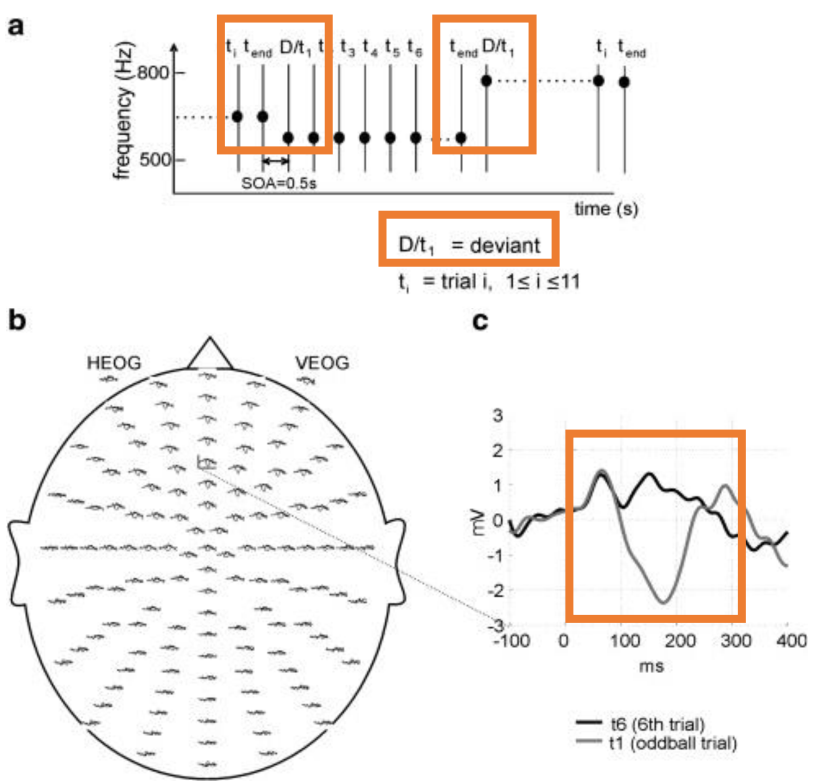
Adapted from Bernadette van Wijk, SPM for MEG/EEG Course, Principles of Dynamic Causal Modelling

Mismatch Negativity (MMN) and Roving Paradigm

Design and responses elicited in a roving paradigm

Adapted from Garrido et al. (2008), Figure 1

Garrido et al. 2008, doi.org/10.1016/j.neuroimage.2008.05.018

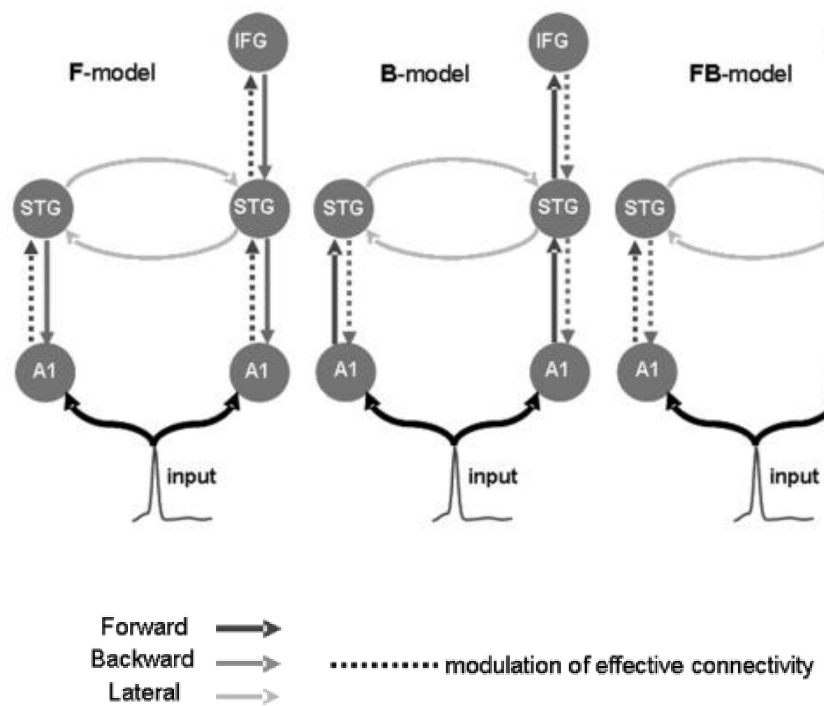


- MMN is an event-related potential (ERP) component evoked by detectable violations in acoustic regularity
- Roving paradigms are characterised by sporadic frequency changes of a repeating tone

Model specification in a MMN paradigm

Copied from Garrido et al. (2007), Figure 1

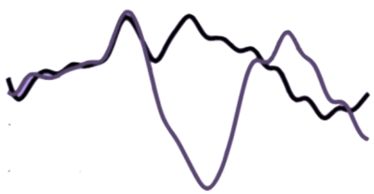
Garrido et al. 2007, doi.org/10.1016/j.neuroimage.2007.03.014



A1 - left and right primary auditory cortex
STG - left and right superior temporal gyrus
IFG - right inferior frontal gyrus

Bayesian Framework, Forward and Inverse Problems

Data Y



Forward Problem

Inverse Problem

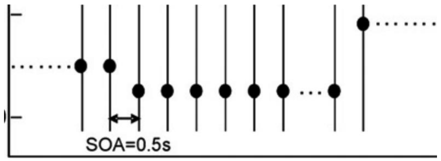
$p(Y|\theta, m)$
Likelihood

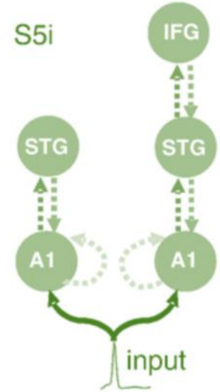
$p(Y|m)$
Evidence

Dynamic states

$p(\theta|Y, m)$
Posterior

Input





Parameters θ

Model m

DCM rests on estimating certain conditional probabilities

Garrido et al. (2008), doi.org/10.1016/j.neuroimage.2008.05.018

Adapted from Richard Rosch, SPM for MEG/EEG Course, Bayesian Model Selection and Averaging

DCM Structure, Symbols used in this Presentation

model m

```
>> spm_dcm_fit
```

Inverted model

prior probability $p(\theta|m)$

```
>> DCM.M.pE
>> DCM.M.pC
```

data Y

```
>> DCM.xY
```

$p(\theta|Y,m)$ posterior probability

```
>> DCM.Ep
>> DCM.Cp
```

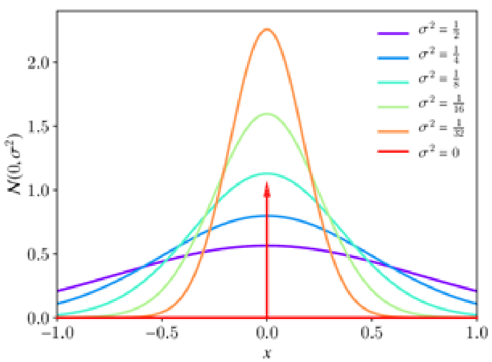
$F \approx \log p(Y|m)$ Free energy approximates model evidence

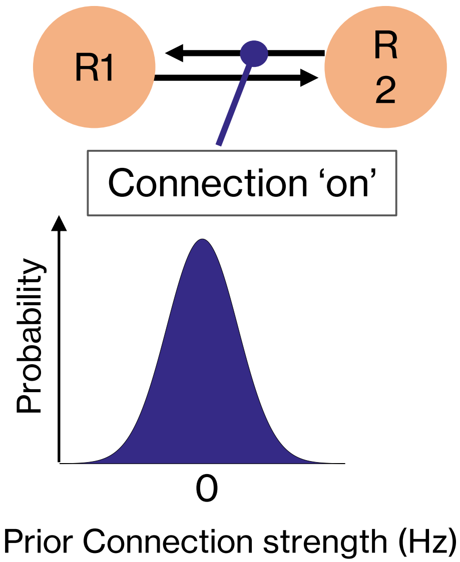
```
>> DCM.F
```

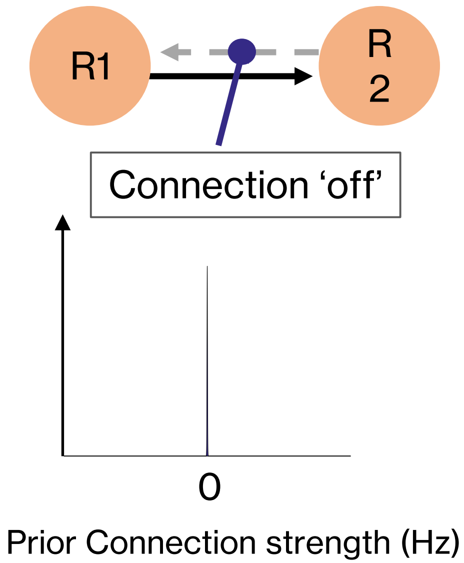
DCM analysis provides a score for model likelihood and posterior parameter estimates

Priors determine Model Structure and Solutions

prior probability $p(\theta|m)$



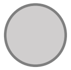




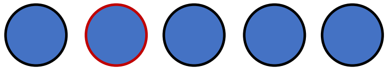
Priors restrict parameters to a specific search space to achieve realistic (interpretable) solutions

Adapted from Richard Rosch, SPM for MEG/EEG Course, Bayesian Model Selection and Averaging

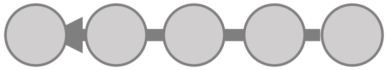
Contents



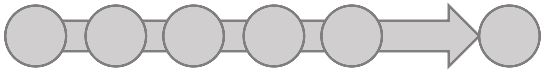
Introduction and model inversion
Dynamic Causal Modelling (DCM)



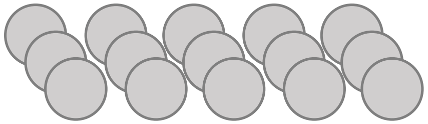
Comparing models
Bayesian Model Comparison and Selection (BMS)



Rapidly evaluating models
Bayesian Model Reduction (BMR)



Investigating the parameters
Bayesian Model Averaging (BMA)



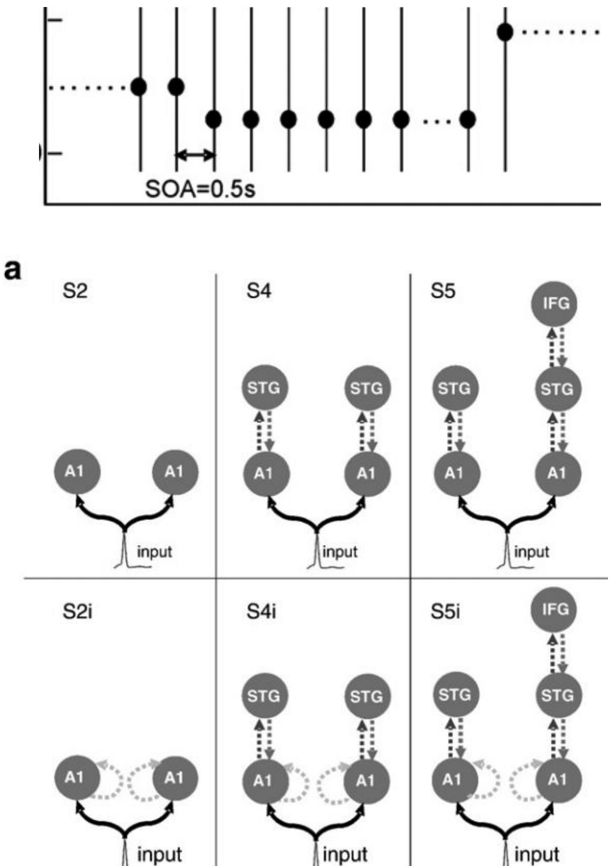
Multi-subject analysis
Parametric Empirical Bayes (PEB)

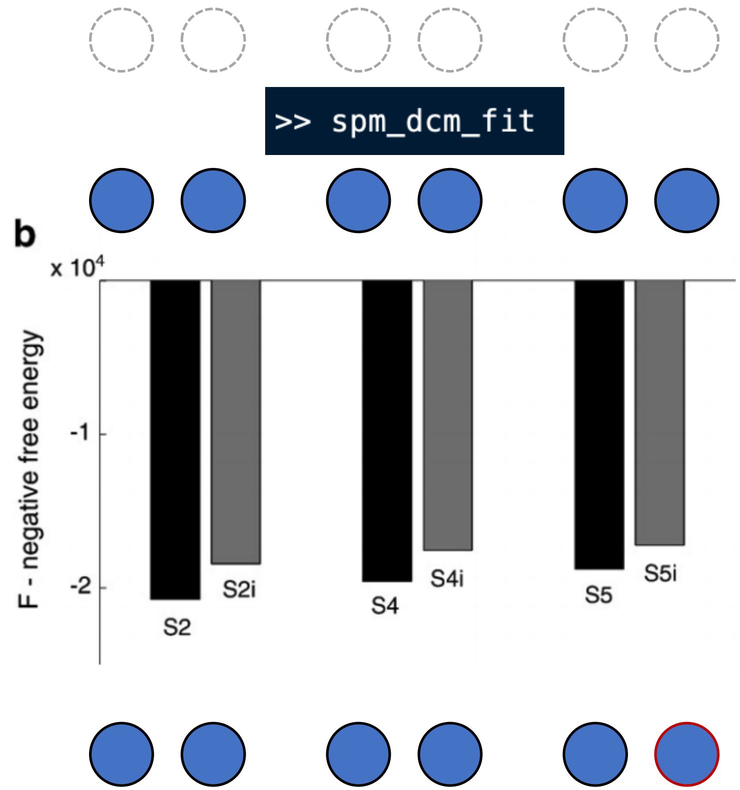
Inverting Models individually and Free Energy Scoring

The functional anatomy of the MMN: A DCM study of the roving paradigm

Marta I. Garrido [a,c,*], Karl J. Friston [a], Stefan J. Kiebel [a], Klaas E. Stephan [a], Torsten Baldeweg [b], James M. Kilner [a]

[a] Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London, UK
[b] Developmental Cognitive Neuroscience, Institute of Child Health, University College London, UK
[c] Department of Psychology, University of California, Los Angeles, USA





Models can be scored against each use using free energy

Garrido et al. (2008), doi.org/10.1016/j.neuroimage.2008.05.018

Adapted from Richard Rosch, SPM for MEG/EEG Course, Bayesian Model Selection and Averaging

Model Comparison using log Bayes Factor

Bayes factor

$$B_{\mathrm{ij}} = \frac{p(Y|m=i)}{p(Y|m=j)}$$

Free energy approximates
log model evidence

$$F \approx \log p(Y|m)$$

Log Bayes factor is approximately the differences of free energies

Interpretation of Bayes factors

Table 1
Interpretation of Bayes factors

B_{ij}	$p(m = i\|y)$ (%)	Evidence in favor of model i
1–3	50–75	Weak
3–20	75–95	Positive
20–150	95–99	Strong
≥150	≥99	Very strong

Bayes factors can be interpreted as follows. Given candidate hypotheses i and j, a Bayes factor of 20 corresponds to a belief of 95% in the statement 'hypothesis i is true'. This corresponds to strong evidence in favor of i.

Copied from Raftery et al. (1995)

$$\log B_{\mathrm{ij}} = \log\, p(Y|m=i) - \log\, p(Y|m=j) \approx \boldsymbol{F_i} - \boldsymbol{Fj}$$

Bayes factor can be interpreted as evidence for a model / hypothesis, e.g., log B > 3 suggests strong evidence and a posterior probability of 95%

The Bayes factor helps to convert model free energies into measures of evidence

Kass and Raftery (1995), doi.org/10.1080/01621459.1995.10476572

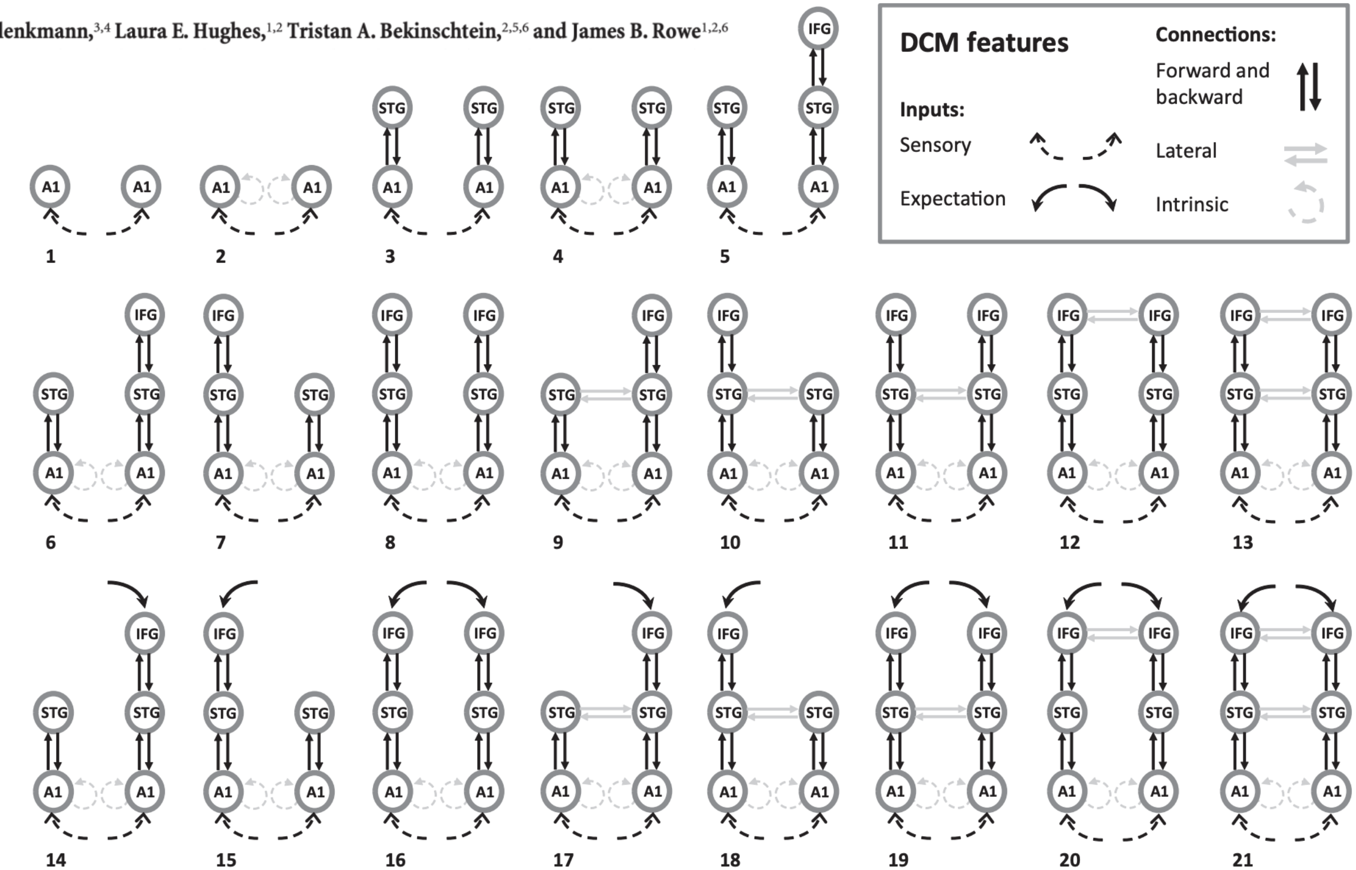
Adapted from Richard Rosch, SPM for MEG/EEG Course, Bayesian Model Selection and Averaging

Evaluating large Model Spaces

Behavioral/Cognitive

Hierarchical Organization of Frontotemporal Networks for the Prediction of Stimuli across Multiple Dimensions

Holly N. Phillips,[1,2] Alejandro Blenkmann,[3,4] Laura E. Hughes,[1,2] Tristan A. Bekinschtein,[2,5,6] and James B. Rowe[1,2,6]



Phillips et al. (2015), doi.org/10.1523/JNEUROSCI.5095-14.2015

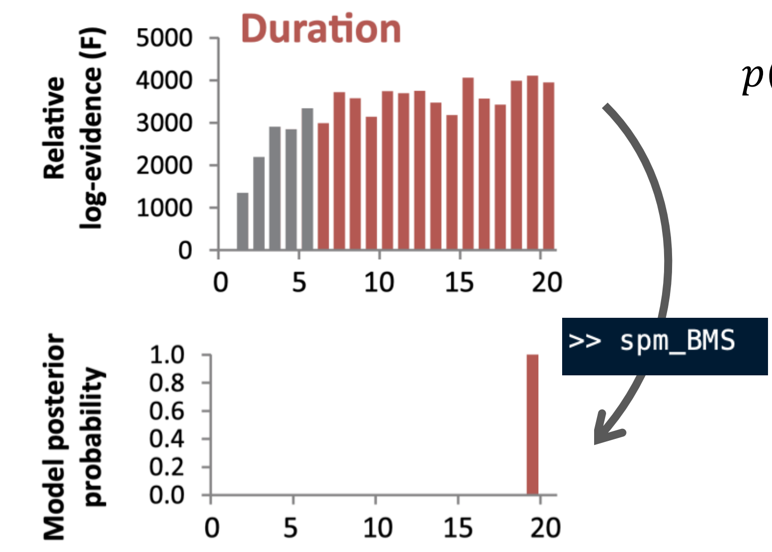
Adapted from Richard Rosch, SPM for MEG/EEG Course, Bayesian Model Selection and Averaging

Bayesian Model Selection based on Posterior Probabilities

Behavioral/Cognitive

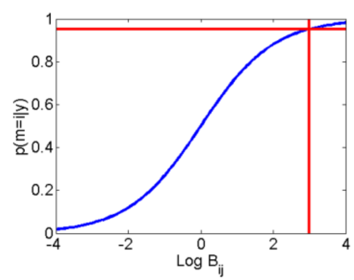
Hierarchical Organization of Frontotemporal Networks for the Prediction of Stimuli across Multiple Dimensions

Holly N. Phillips,[1,2] Alejandro Blenkmann,[3,4] Laura E. Hughes,[1,2] Tristan A. Bekinschtein,[2,5,6] and James B. Rowe[1,2,6]



$$p(m = i|Y) = \frac{p(Y|m = i)}{p(Y)}$$

$$= \overset{*}{} \frac{1}{1 + \exp -\log(Bij)}$$



*

$$\begin{aligned} p(m = i|y) &= \frac{p(y|m = i)}{p(y)} \\ &= \frac{p(y|m = i)}{p(y|m = i) + p(y|m = j)} \\ &= \frac{1}{1 + \frac{p(y|m = j)}{p(y|m = i)}} \\ &= \frac{1}{1 + B_{ji}} \\ &= \frac{1}{1 + \exp(\log(B_{ji}))} \\ &= \frac{1}{1 + \exp(-\log(B_{ij}))} \end{aligned}$$

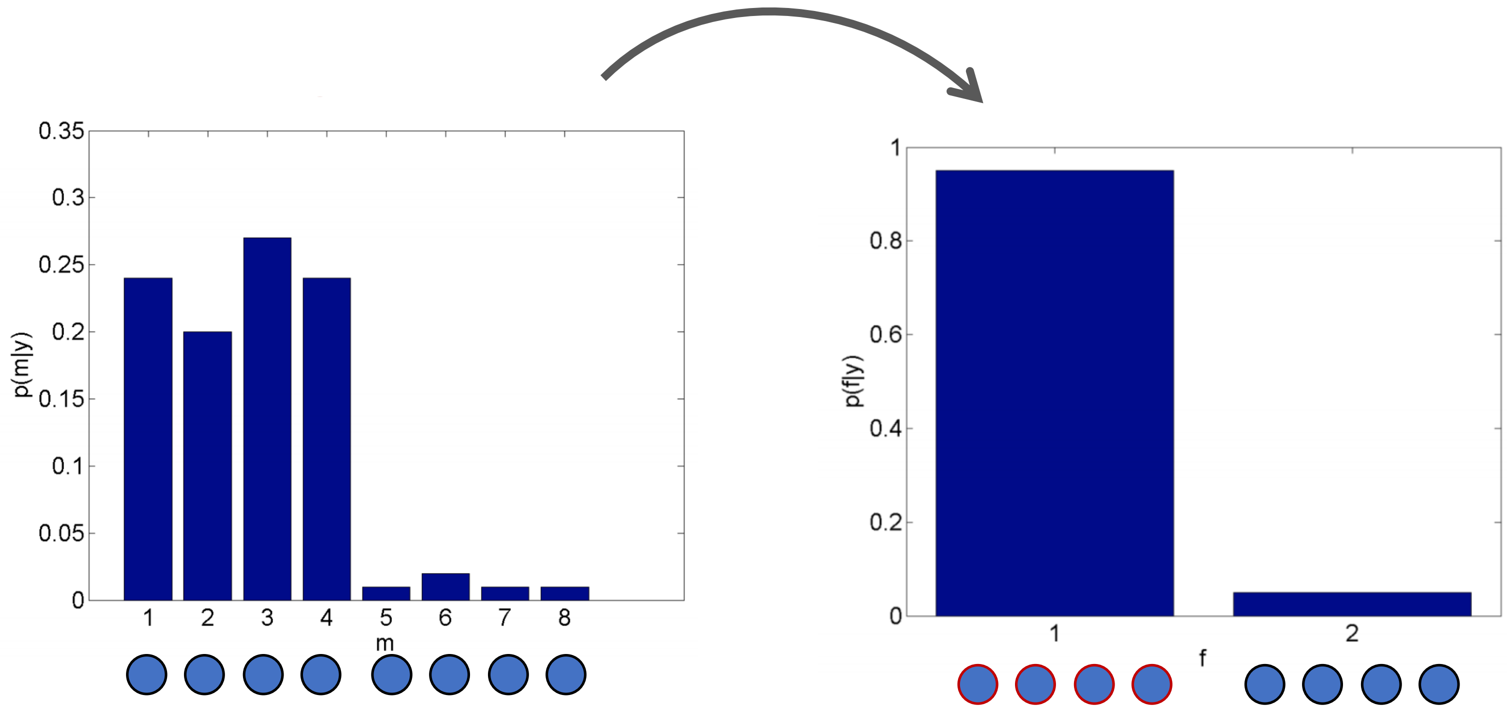
Transformation of Bayes factors into posterior probabilities using Bayes rule

Phillips et al. (2015), doi.org/10.1523/JNEUROSCI.5095-14.2015

Adapted from Richard Rosch, SPM for MEG/EEG Course, Bayesian Model Selection and Averaging

Avoiding Evidence Dilution / Structuring Model Space

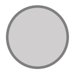
$$p(f|Y) = \sum_{m \in S_f} p(m|Y)$$



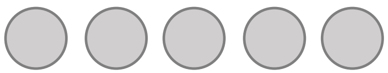
Structuring the model space by grouping models into families helps to avoid evidence dilution

Adapted from Will Penny, SPM for fMRI Course, DCM Advanced - Part I: Model Selection

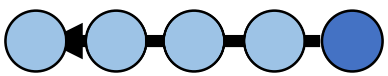
Contents



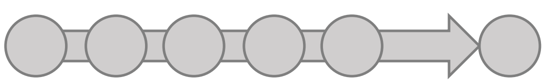
Introduction and model inversion
Dynamic Causal Modelling (DCM)



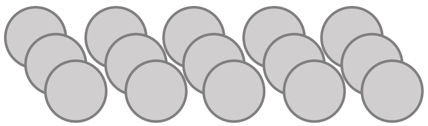
Comparing models
Bayesian Model Comparison and Selection (BMS)



Rapidly evaluating models
Bayesian Model Reduction (BMR)

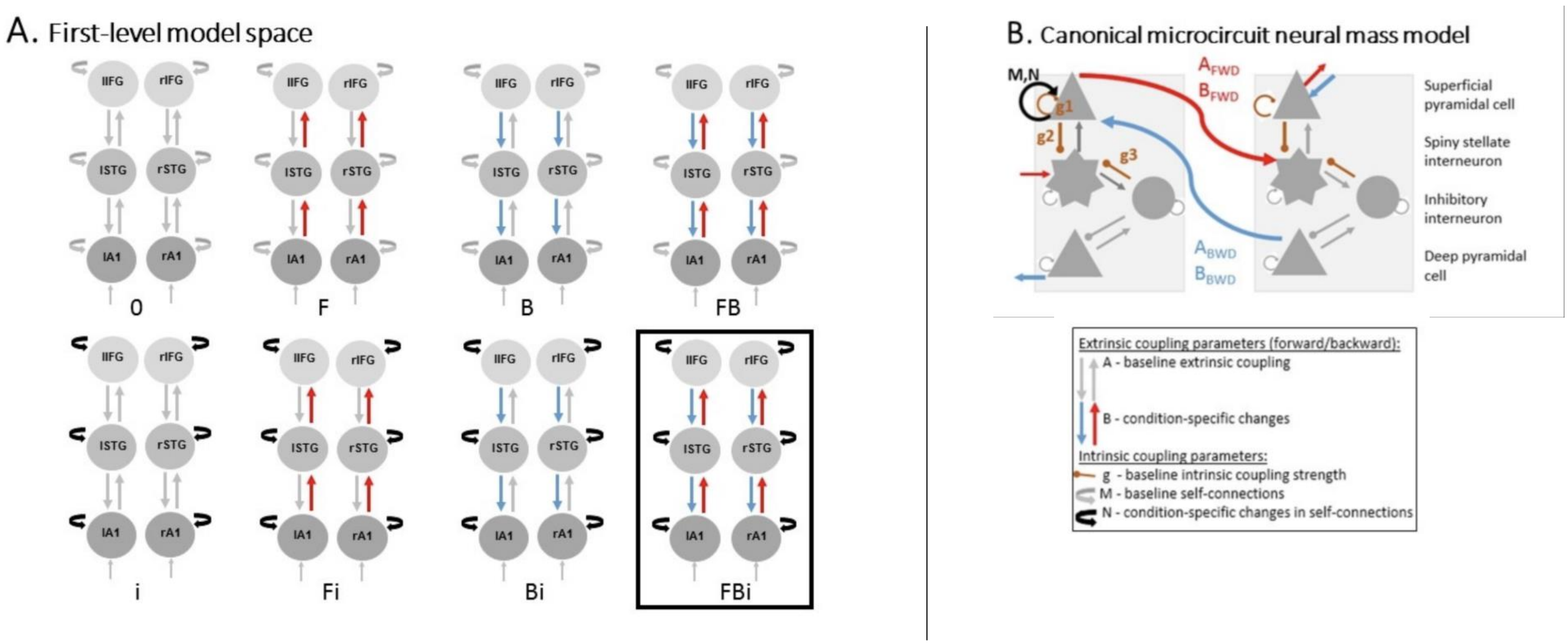


Investigating the parameters
Bayesian Model Averaging (BMA)



Multi-subject analysis
Parametric Empirical Bayes (PEB)

Evaluating complex, multi-level Model Spaces



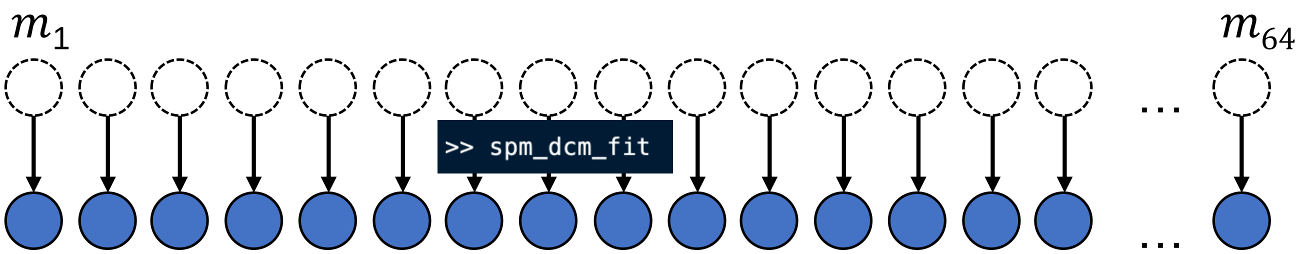
Example of complex model space with 64 models: 8 between and 8 within regional variations

Fitzgerald et al. (2019), doi.org/10.1101/768846

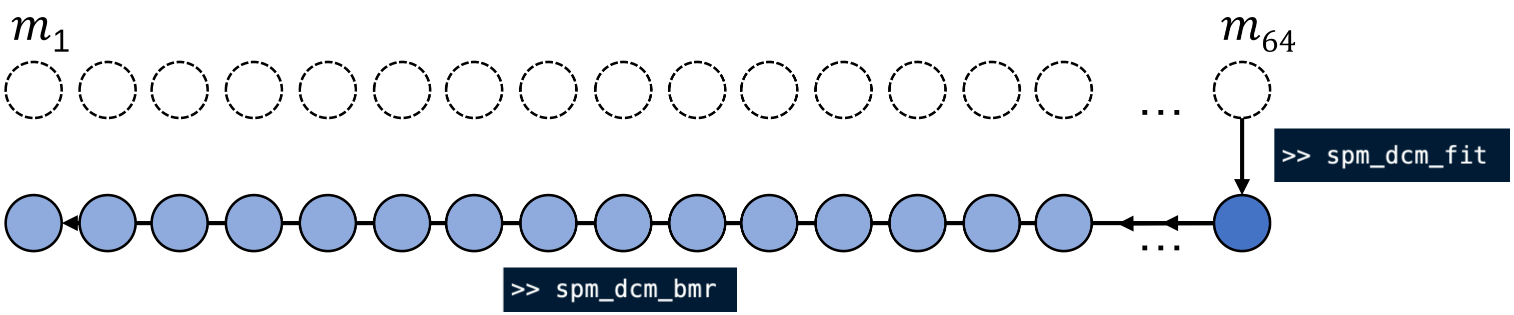
Adapted from Richard Rosch, SPM for MEG/EEG Course, Bayesian Model Selection and Averaging

Bayesian Model Reduction (BMR) Procedure

Individual model inversions



Bayesian model reduction

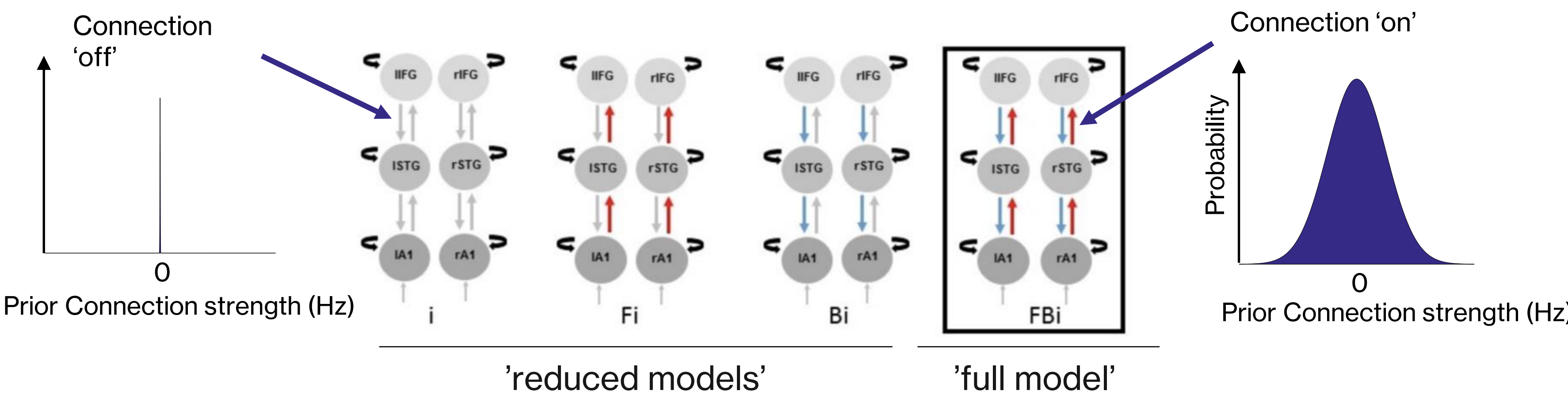


BMR allows estimation of parameters and free energy from a single inverted (full) model

Adapted from Richard Rosch, SPM for MEG/EEG Course, Bayesian Model Selection and Averaging

Theoretical Basis for BMR

These (approximate) equalities mean one can evaluate the posterior and evidence of any reduced model, given the posteriors of the full model. In other words, $F[\tilde{P}(\theta):P(\theta)] \approx \ln \tilde{P}(y)$ allows us to skip the optimization of the reduced posterior $\tilde{Q}(\theta)$ and use the optimized posterior of the full model to compute the evidence (and posterior) of the reduced model directly.

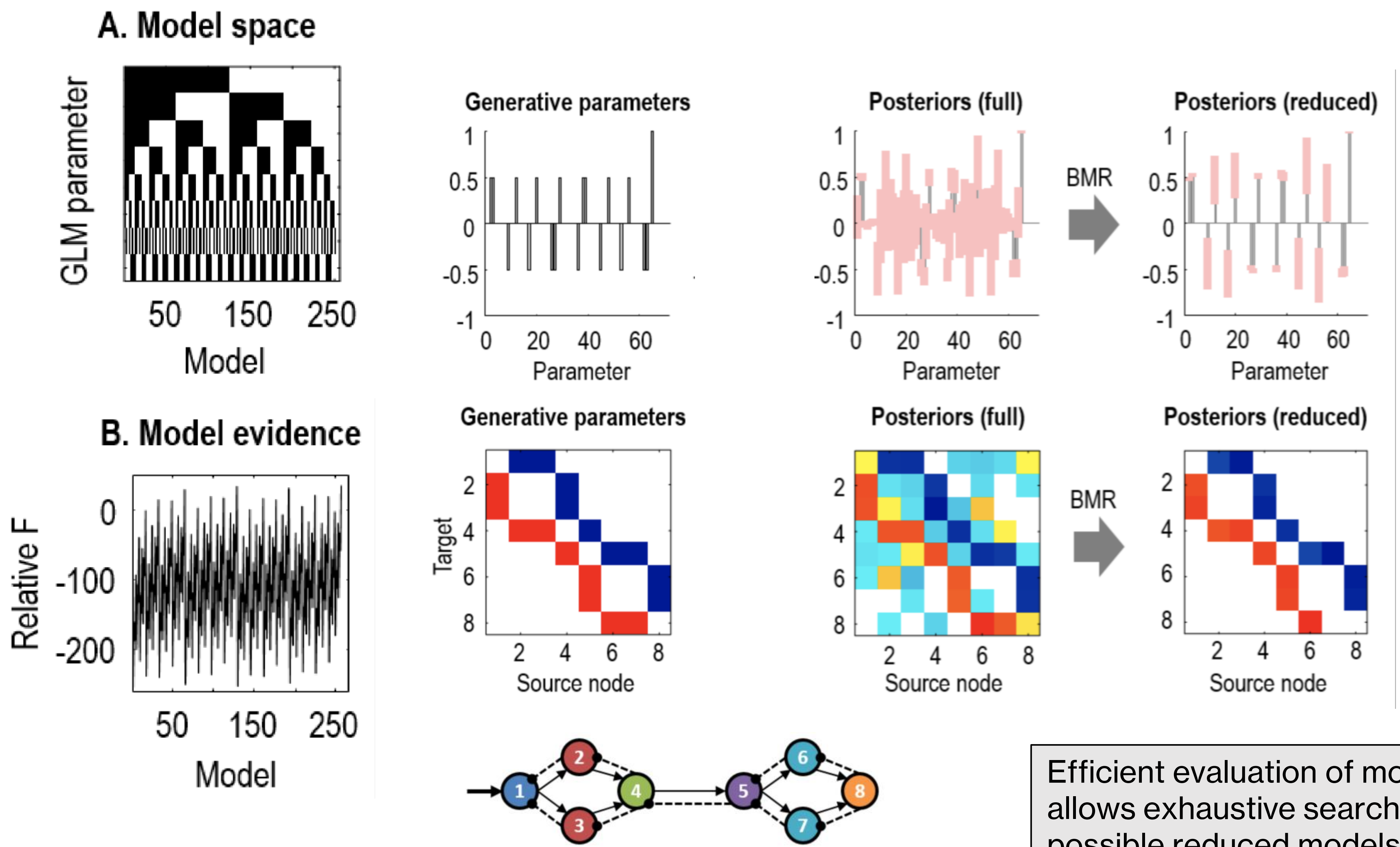


BMR requires an inverted 'full model' and a set of structurally identical 'reduced models'

Friston et al. (2018), arxiv.org/pdf/1805.07092.pdf

Adapted from Richard Rosch, SPM for MEG/EEG Course, Bayesian Model Selection and Averaging

BMR enables Exploration of entire Model Spaces

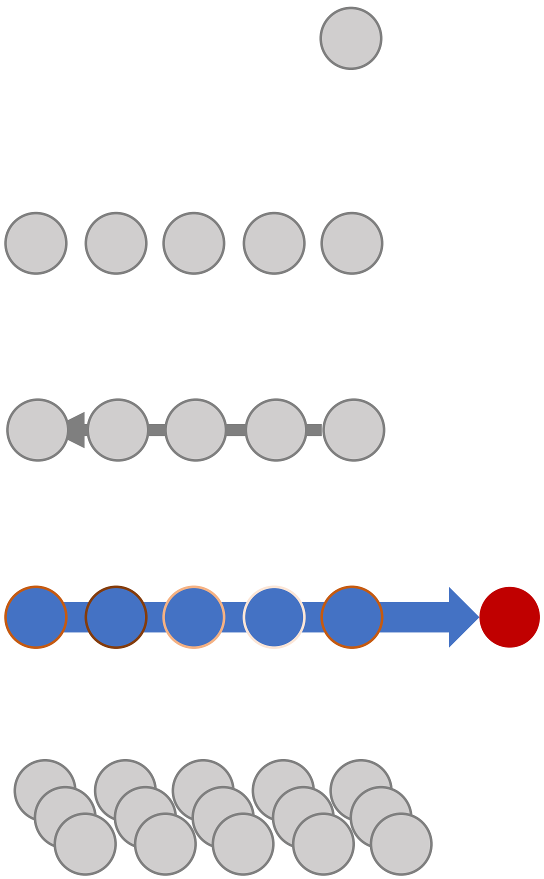


Efficient evaluation of models allows exhaustive search over all possible reduced models

Friston et al. (2018), arxiv.org/pdf/1805.07092.pdf

Adapted from Richard Rosch, SPM for MEG/EEG Course, Bayesian Model Selection and Averaging

Contents



Parameter Comparison under Model Structure Uncertainty

Schizophrenia Research
Volume 135, Issues 1–3, March 2012, Pages 23-27

Abnormal intrinsic and extrinsic connectivity within the magnetic mismatch negativity brain network in schizophrenia: A preliminary study

D. Dima [a], S. Frangou [a], L. Burge [b], S. Braeutigam [c], A.C. James [b, d]

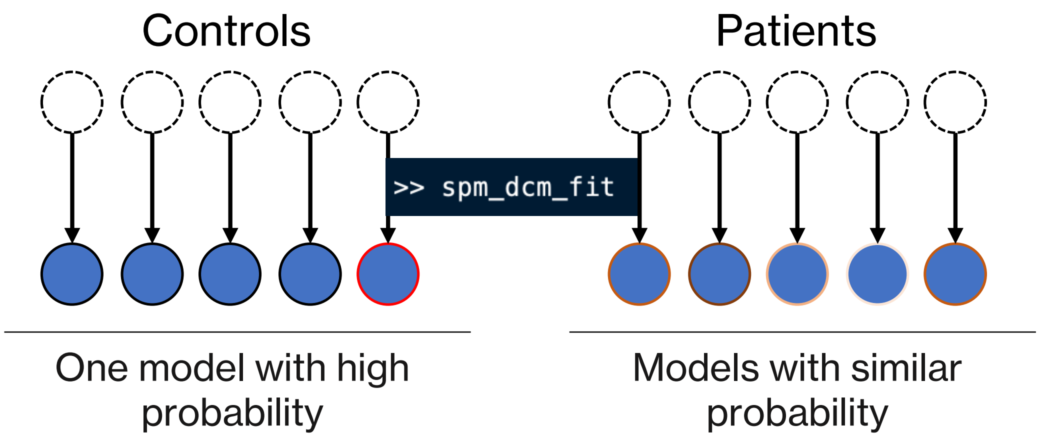
Show more

+ Add to Mendeley Share Cite

https://doi.org/10.1016/j.schres.2011.12.024 Get rights and content

3.2. Model comparison

As anticipated (Garrido et al., 2008) in healthy controls the Combination model outperformed both the Forward and Backward models with exceedance probability of 89%. The Combination model assumes that the MMN response emerged from changes in all bidirectional extrinsic as well as in intrinsic connections. In patients the optimal model included modulation via the MMN of the intrinsic connections and only the forward connections, which was the Forward model. The exceedance probability for this model was 44%, surpassing the exceedance probabilities of the other two tested models.



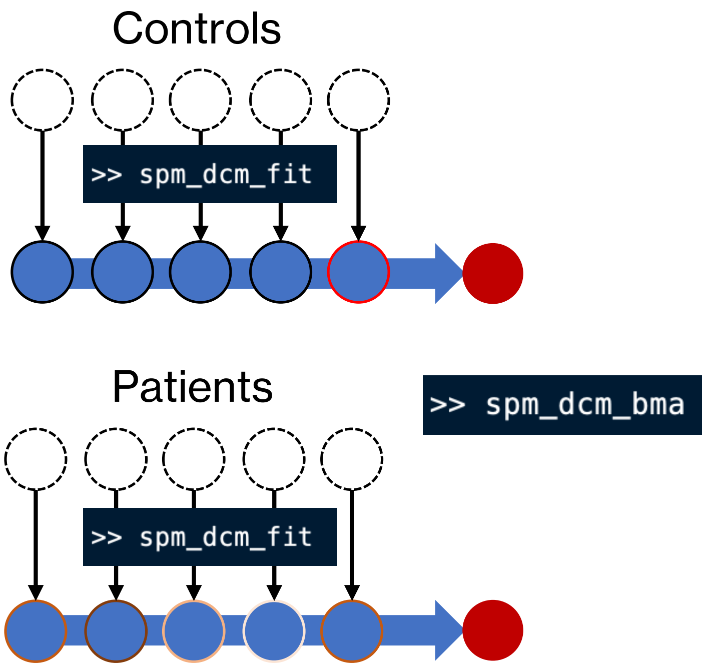
Dima et al. (2012), doi.org/10.1016/j.schres.2011.12.024

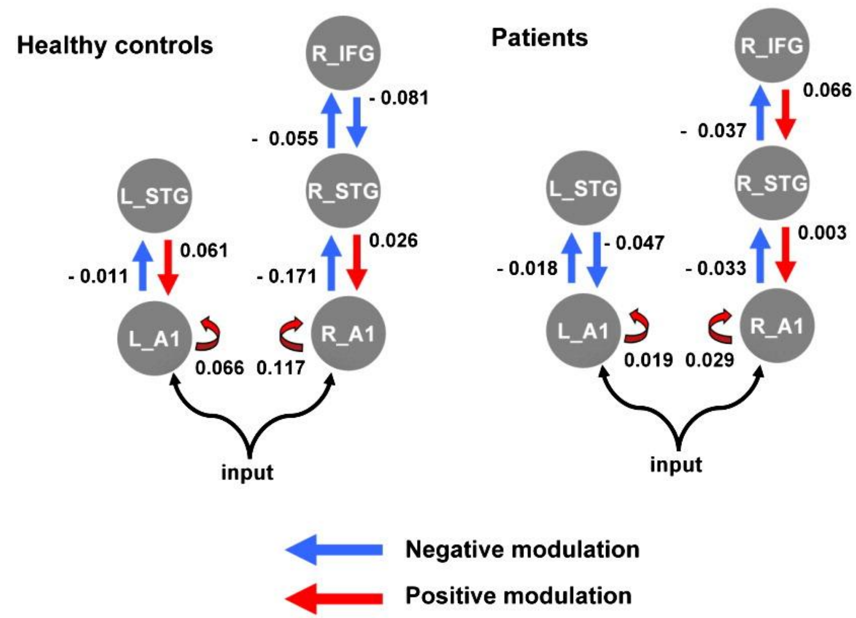
Adapted from Richard Rosch, SPM for MEG/EEG Course, Bayesian Model Selection and Averaging

Bayesian Model Averaging (BMA)

$$p(\theta|Y) = \sum_m p(\theta|m,Y) \quad p(m|Y)$$

Parameter posterior prob. | Parameter probability | Model probability





Estimate posterior densities over parameters across different models (using sampling)

Dima et al. (2012), doi.org/10.1016/j.schres.2011.12.024

Adapted from Richard Rosch, SPM for MEG/EEG Course, Bayesian Model Selection and Averaging

Contents

Parametric Empirical Bayes (PEB) – Example

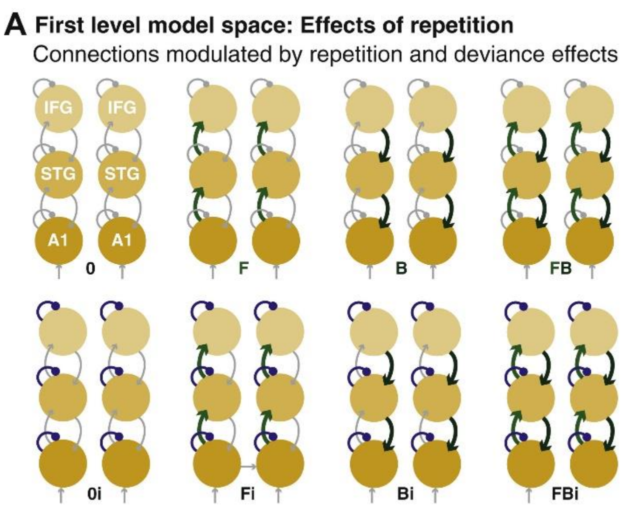

Biological Psychiatry: Cognitive Neuroscience and Neuroimaging
Volume 4, Issue 2, February 2019, Pages 140-150

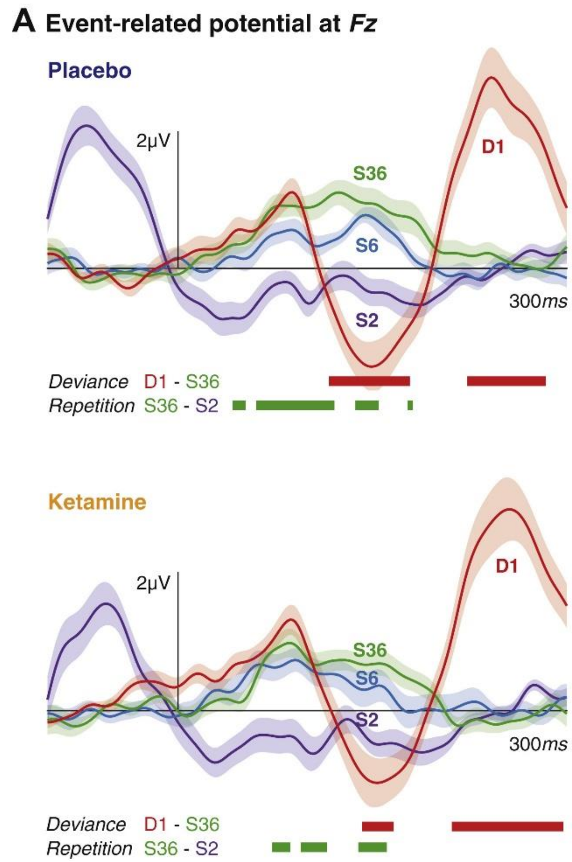
Archival Report

Selective Prefrontal Disinhibition in a Roving Auditory Oddball Paradigm Under N-Methyl-D-Aspartate Receptor Blockade

Richard E. Rosch [a, b], Ryszard Auksztulewicz [a, c], Pui Duen Leung [a], Karl J. Friston [a], Torsten Baldeweg [b]

Show more





Many scientific questions centre on group comparisons in connectivity parameters

Rosch et al. (2019), doi.org/10.1016/j.bpsc.2018.07.003, GitHub: Ketamine_DCM

Adapted from Richard Rosch, SPM for MEG/EEG Course, Bayesian Model Selection and Averaging

First and Second Level Modelling

$$\theta^{(2)} = \eta + \varepsilon^{(3)}$$

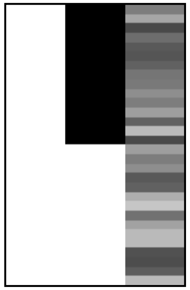
Priors on second level parameters

Second level

$$\theta^{(1)} = \Gamma^{(2)}(\theta^{(2)}) + \varepsilon^{(2)}$$

Between-subject error

Second level (linear) model

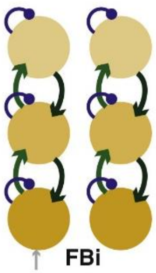


First level

$$y = \Gamma_i^{(1)}(\theta^{(1)}) + \varepsilon^{(1)}$$

Measurement noise

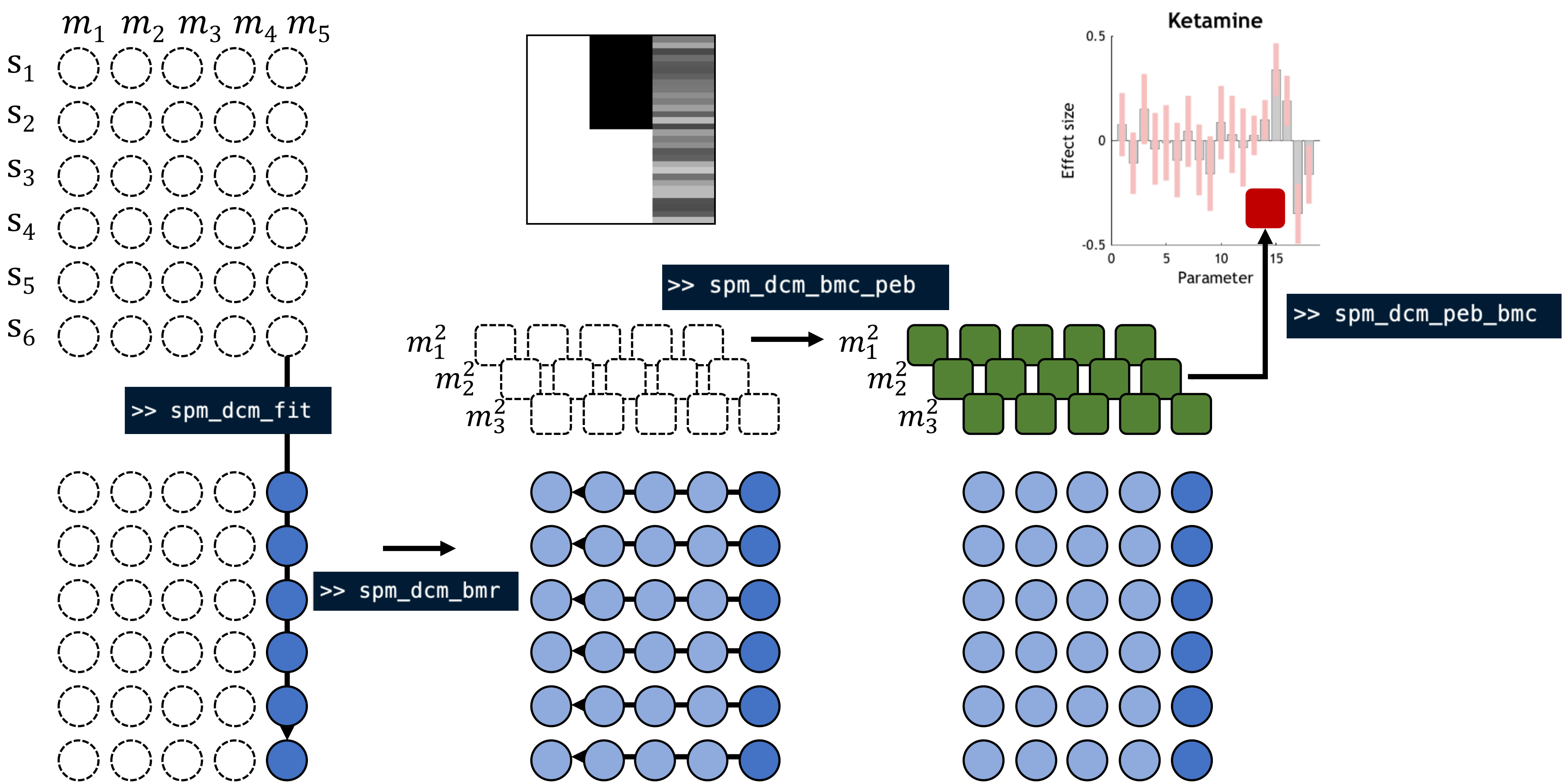
DCM for subject i



PEB is a hierarchical modelling approach incorporating first and second level

Adapted from Richard Rosch and Peter Zeidman, SPM for MEG/EEG Course, Bayesian Model Selection and Averaging

Modelling Steps: DCM to reduced PEB




```
[BMA,BMR] = spm_dcm_peb_bmc(PEB,models)
[BMC,PEB] = spm_dcm_bmc_peb(DCM,M,field)
[PEB,P]   = spm_dcm_peb(P,M,field)
```

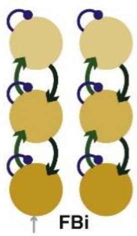
Modelling first and second level within a Bayesian context using SPM procedures and functions

Rosch et al. (2019), doi.org/10.1016/j.bpsc.2018.07.003, GitHub: Ketamine_DCM

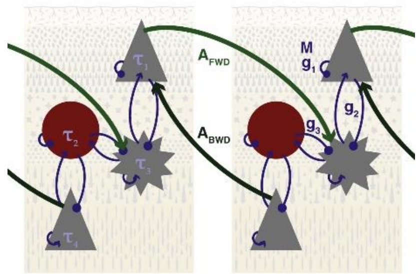
Adapted from Richard Rosch, SPM for MEG/EEG Course, Bayesian Model Selection and Averaging

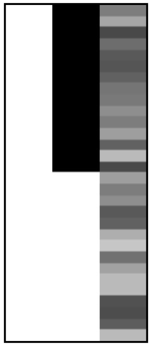
PEB Example: Effect of ketamine on (Intrinsic) Connectivity

1st Level
DCM

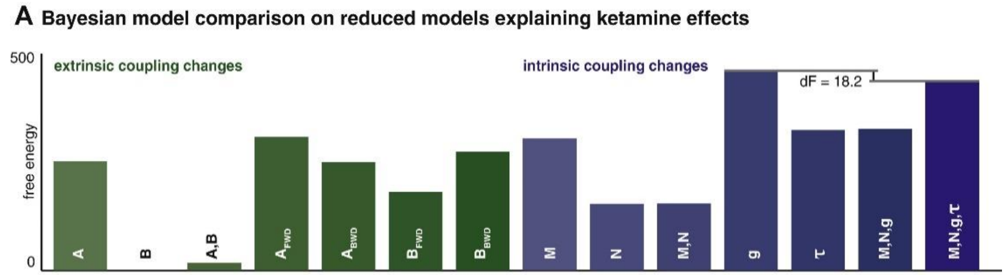


2nd Level
PEB Design Matrix

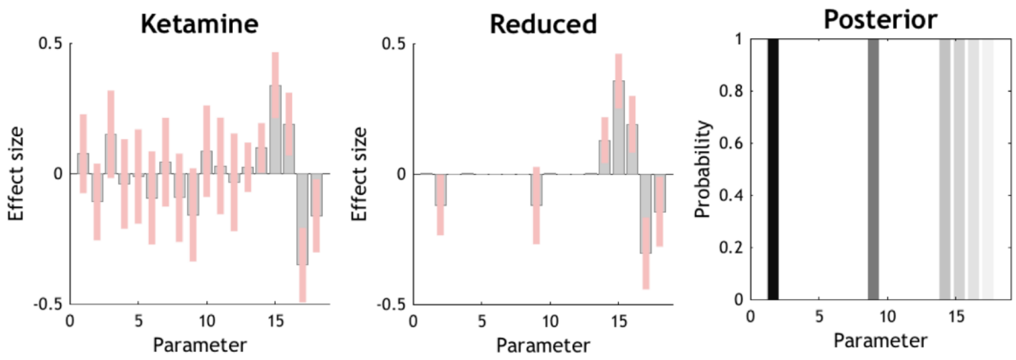




2nd Level
Model Comparison



2nd Level
Model Average



Rosch et al. (2019), doi.org/10.1016/j.bpsc.2018.07.003, [GitHub: Ketamine_DCM]()

Adapted from Richard Rosch, SPM for MEG/EEG Course, Bayesian Model Selection and Averaging

PEB Advantages and Applications

- Conveys uncertainty about parameters from the subject level to the group level
- Can improve first level parameters estimates
- Can be used to ...
  - ... compare specific reduced PEB models (switching off combinations of group-level parameters)
  - ... or to search over nested models (BMR)
- Prediction (leave-one-out cross validation)

References and additional Material

Additional Resources

Will Penny's advanced DCM lecture slides
Penny: DCM advanced, SPM Course Slides

Lecture by Stefan Frässle on Bayesian model selection and averaging
Fraessle: BMS and BMA

Tutorial for PEB by Peter Zeidman
Zeidman: DCM-PEB Example

PEB Paper (Friston et al., 2015)
Bayesian model reduction and empirical Bayes for group (DCM)

10 Simple Rules for Group studies before PEB (Stephan et al., 2010)
Ten simple rules for dynamic causal modeling

Worked example using PEB with code by Natalie Adams
Adams: PEB Example

GLM of Connectivity Parameters

$$\theta^{(1)} = X\theta^{(2)} + \epsilon^{(2)}$$

← Unexplained between-subject variability

↑ Design matrix (covariates)

↑ Group level parameters