MRC Cognition and Brain Sciences Unit

UNIVERSITY OF CAMBRIDGE

A Brief History of fMRI in Cognitive Neuroscience

Richard (Rik) Henson

rik.henson@mrc-cbu.cam.ac.uk

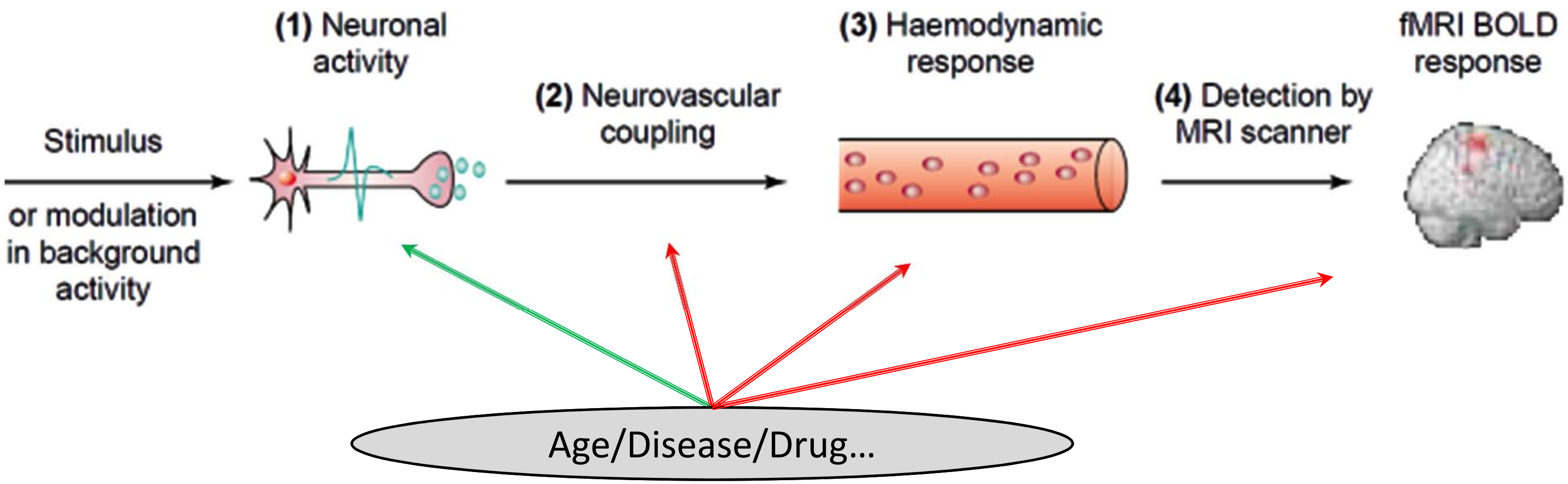
fMRI

- Functional MRI **contrast**:
  - BOLD
  - VASO
  - …

- Functional MRI **sequence**:
  - 2D Gradient-Echo (GE) Echo-Planar Imaging (EPI)
  - 2D Spin-Echo (SE) EPI
  - 3D Gradient-Echo EPI
  - Multi-echo EPI
  - Multi-slice EPI
  - ….

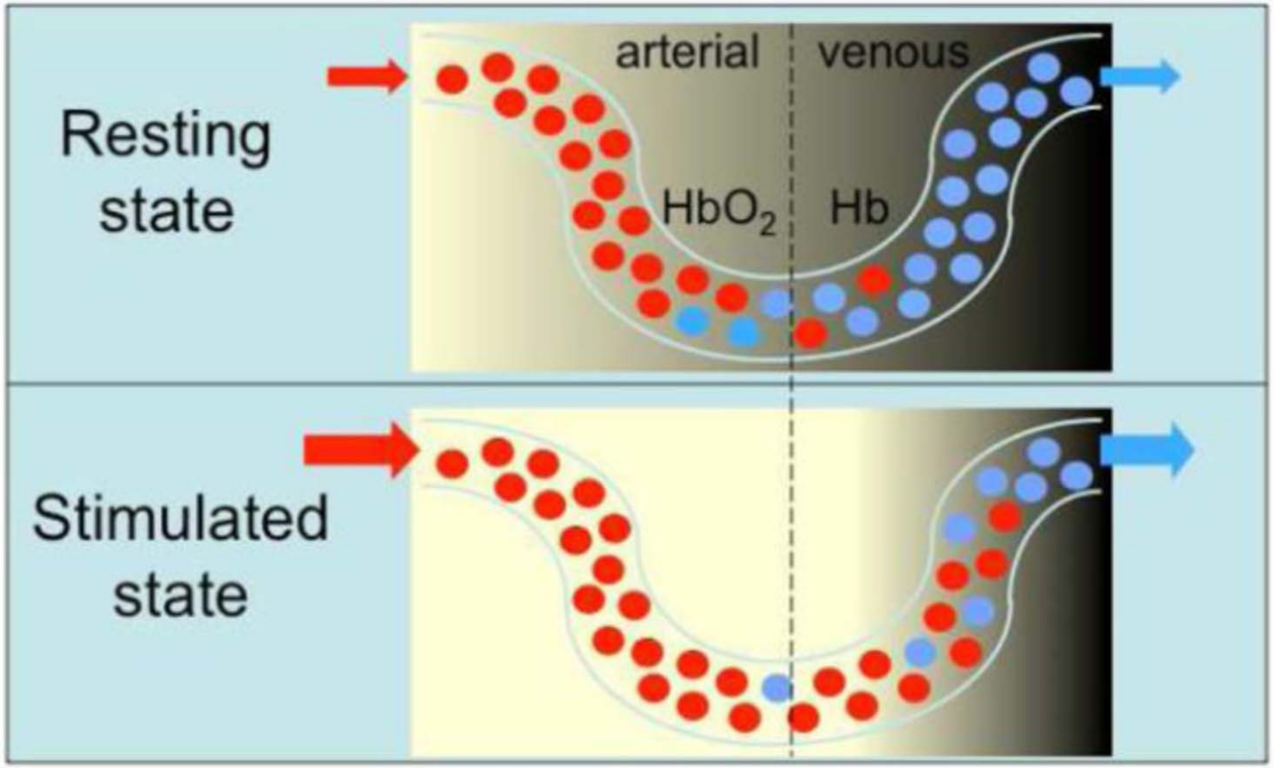
The BOLD contrast

- Complex (nonlinear) function of neural activity, blood flow, blood volume and blood oxygenation (Ogawa et al, 1990, PNAS)



Arthurs & Boniface, 2002, TINS

The BOLD contrast



Oxygenated blood cells
HbO2

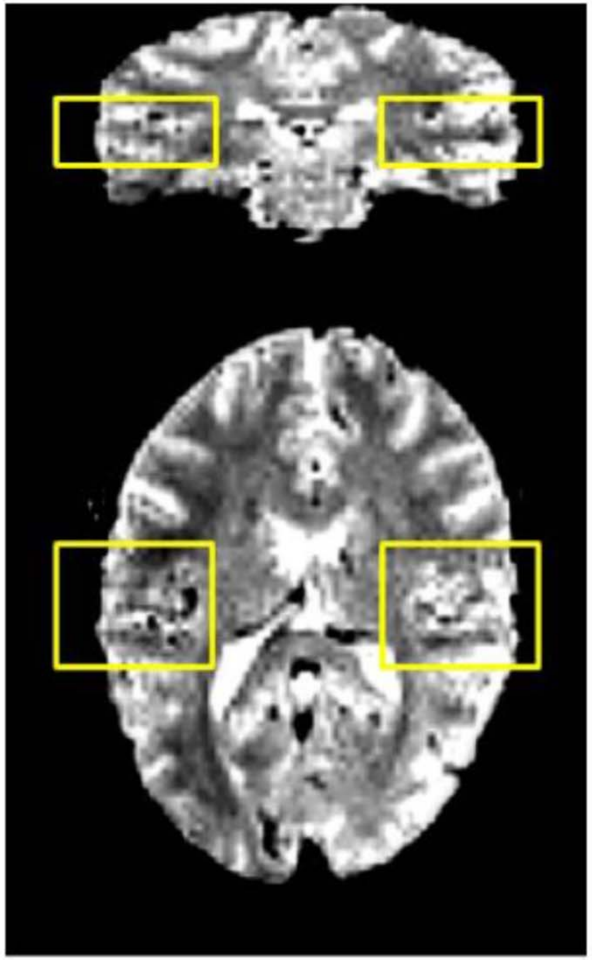
De-oxygenated blood cells
Hb

→ dampens the MRI signal

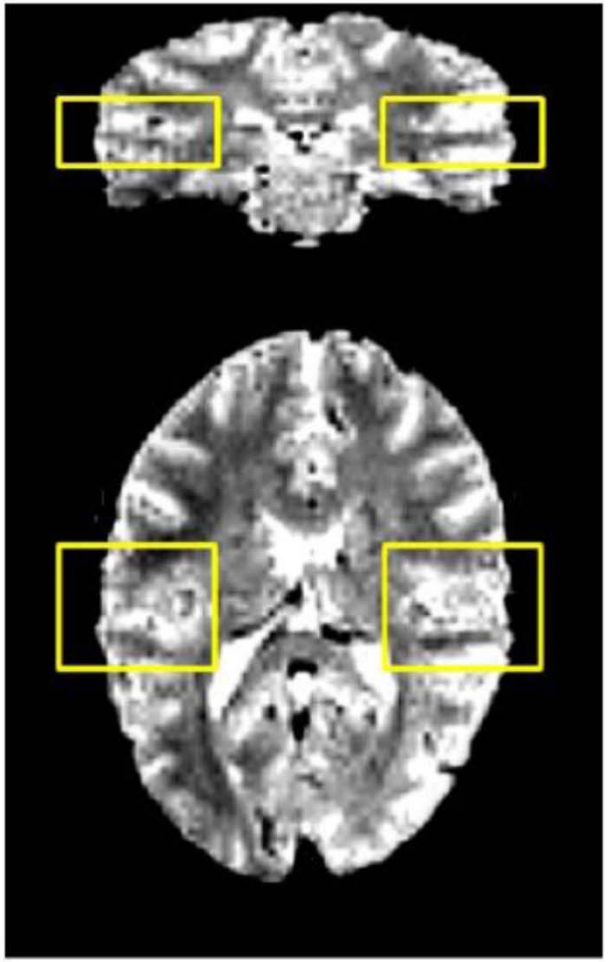
Glover (2011) Neurosurg. Clin. N. Am.

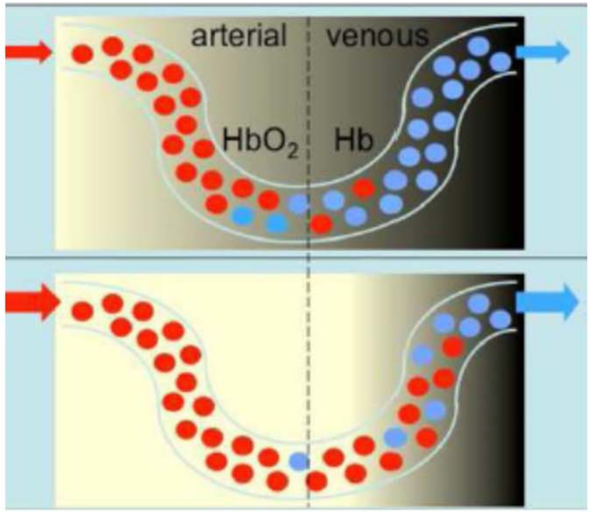
The BOLD contrast

Baseline

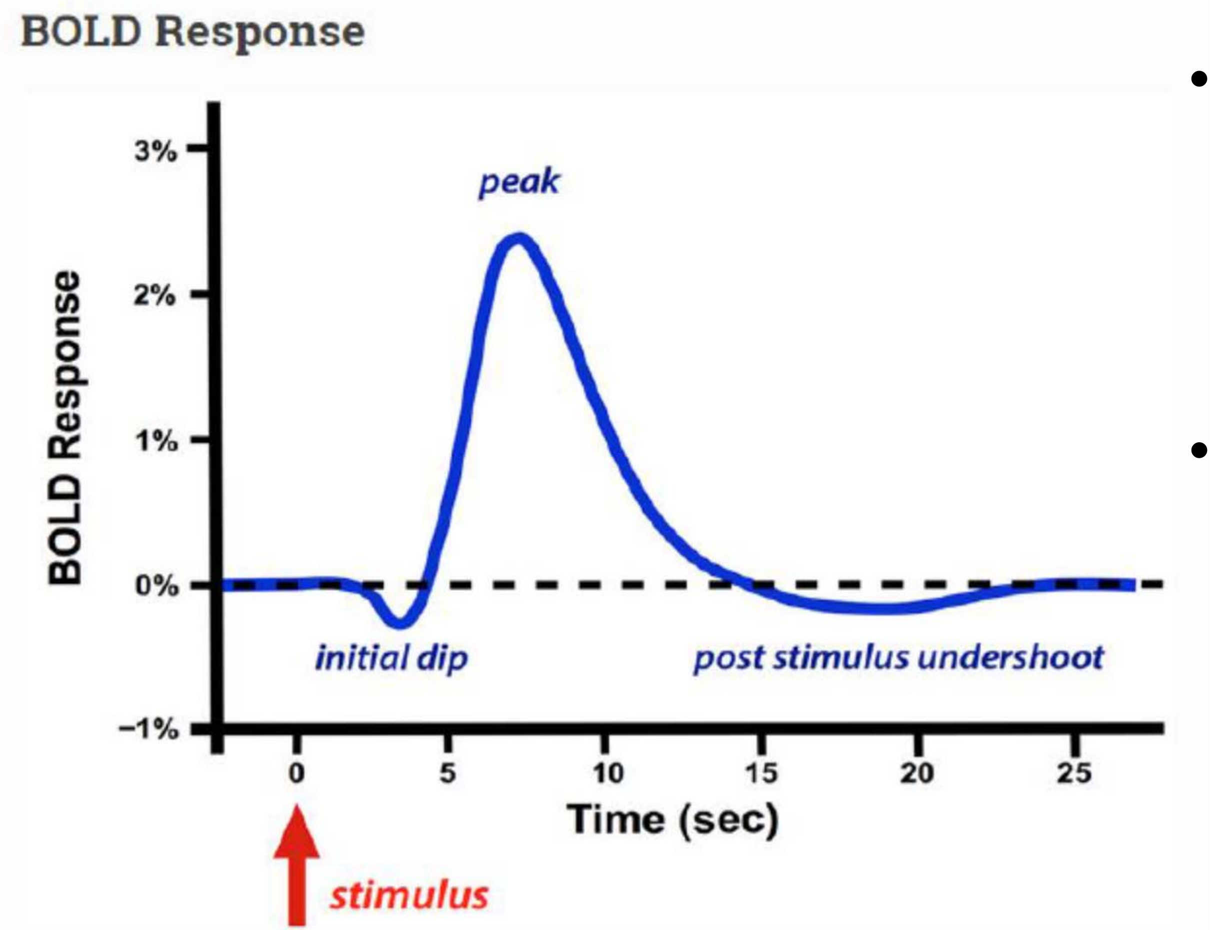


Neural Activity





The Hemodynamic Response Function (HRF) (BOLD impulse response function)



- Varies across brain region, individual, state (eg, caffeine)…
- Roughly linear when stimulus onset asynchrony (SOA) > ~2s

Applications

- Activation Analysis
  - Brain Mapping (mass univariate)
  - Subtraction logic (pure insertion)
  - Blocked versus Intermixed (Epoch vs Event-related)
  - Functional localisation; Forward and Reverse Inference

- Functional Connectivity
  - Matrix factorisation (Independent Component Analysis)
  - ROI-based connectomes, graph theory
  - Task-based (effective) connectivity

- Pattern (Information) Analysis
  - Multi-voxel pattern analysis (MVPA)
  - Representational Similarity Analysis (RSA)

Theme: Fusiform Face Area (FFA)

Activation Analysis

- Early fMRI studies averaged BOLD signal change across (anatomical) Regions Of Interest (ROIs)

- Given possible ROI functional hetereogeneity, subsequent studies tested every single voxel in the image -> “brain mapping” (“mass univariate analysis”)

- More recent studies have reverted back to ROI analyses, but using more sophisticated (eg functionally-defined) ROIs (“parcellations”)

Mass Univariate Analysis ("Brain Mapping")

- Mostly uses the General Linear Model (GLM):

$$y(t,v) = X(t,p) \times \beta(p) + \varepsilon(t,v)$$

y (measured) = BOLD signal in voxel *v* at time *t* (TR/volume)

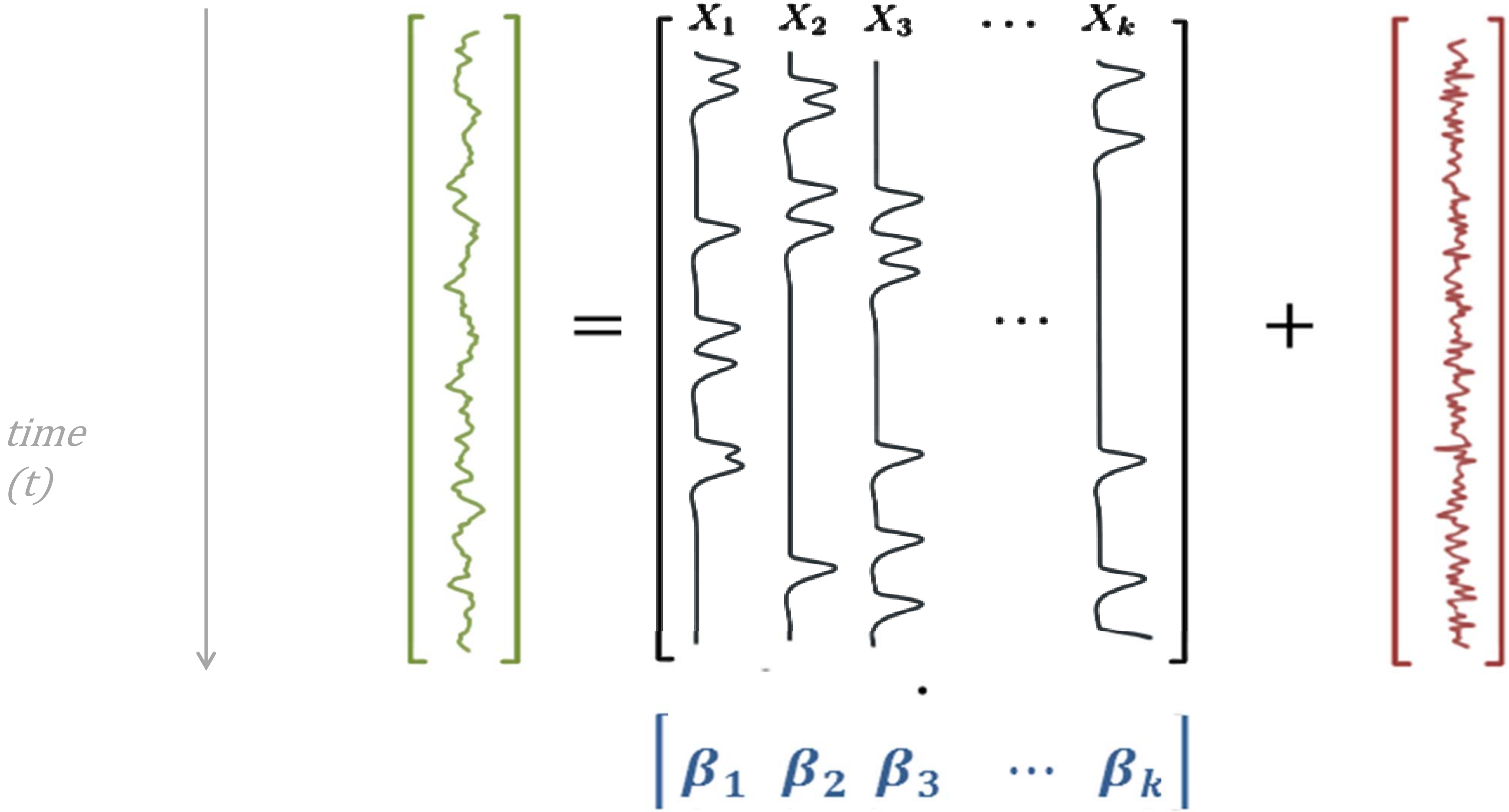
X (specified) = "Design matrix" coding predictions at each *t* according to each experimental condition *p*

β (estimated) = "Parameters" or "Betas" (weights) for each condition *p*

ε (estimated) = residual error in voxel *v* at time *t*

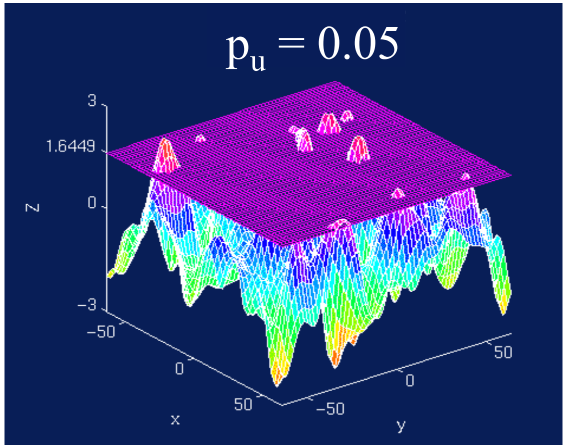
- Uses an "encoding" model, where predict BOLD signal from experimental conditions (cf "decoding" models later)

Mass Univariate Analysis
(“Brain Mapping”)



Mass Univariate Analysis
("Brain Mapping")

- Images sometimes smoothed (e.g., by 8x8x8mm Gaussian, if voxel sizes 3x3x3mm) – can improve statistics (matched filter theorem / central limit theorem)
- Doing statistical tests on ~100,000s of voxels, so risk of false positives (using frequentist p-values), though voxels not independent (smoothed)
- Many methods developed for correcting for multiple comparisons across voxels, e.g. "Random Field Theory", or "Permutation Testing" (of the Maximum T-value, Cluster-size, Cluster-mass…etc)
- Over-arching aim is **"localisation"** (of function)



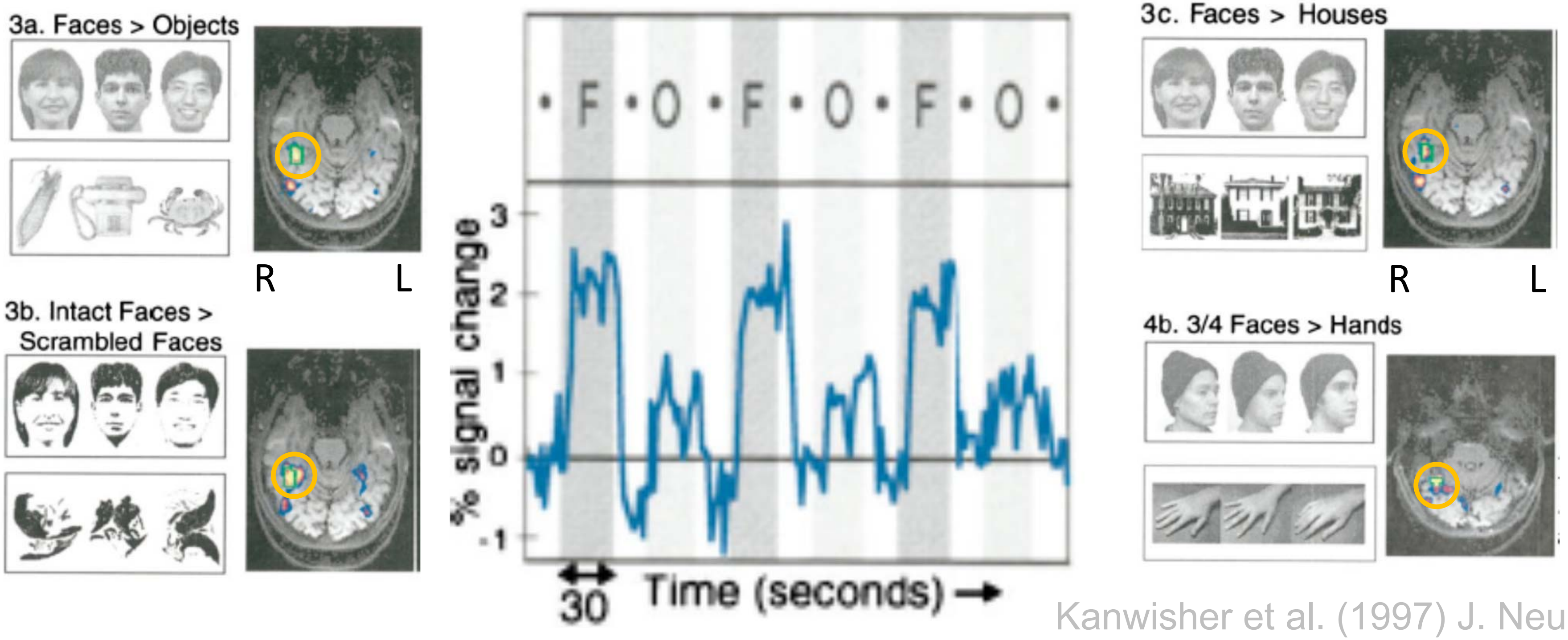
Subtraction Logic & Blocked Designs

- Since absolute value of BOLD signal is arbitrary, fMRI studies used Donder's "subtraction logic", i.e., to isolate function F, compare activity in a condition with F with a condition matched in every way except F
- This makes assumption of "Pure Insertion", i.e, that adding F to a condition does not change anything else, which can be violated…
- …for example, Price et al. (1997, Human Brain Mapping) showed that BOLD response to objects vs colors in some brain regions depends whether you are naming them, or passively viewing them
- (a stimulus x task interaction, or failure of pure insertion; "conjunctions" of simple effects in each task can address to some extent)

- Although fMRI could acquire an image in 2-4secs, while PET took 60+ seconds, early fMRI studies mimicked PET by averaging activity over 10s of seconds ("blocked" designs)

Mass Univariate Analysis
("Brain Mapping")

- For example, discovery of the "Fusiform Face Area" (FFA)



Kanwisher et al. (1997) J. Neuro.

- Found for a variety of control stimuli, but again, difficult to match for every difference, e.g, eye-movement differences / center-periphery
Levy et al. (2001) Nat. Neuro

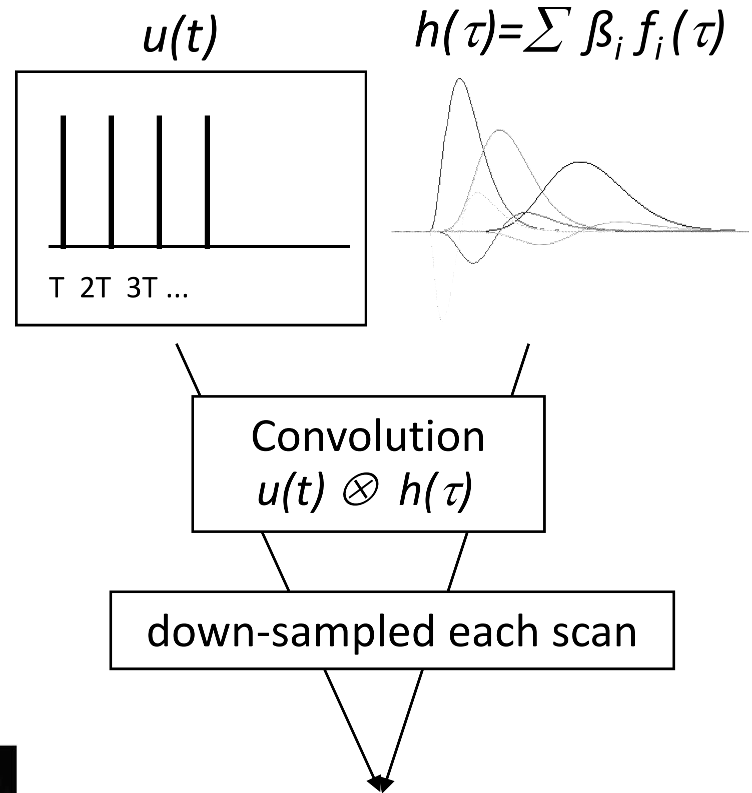
Subtraction Logic & Blocked Designs

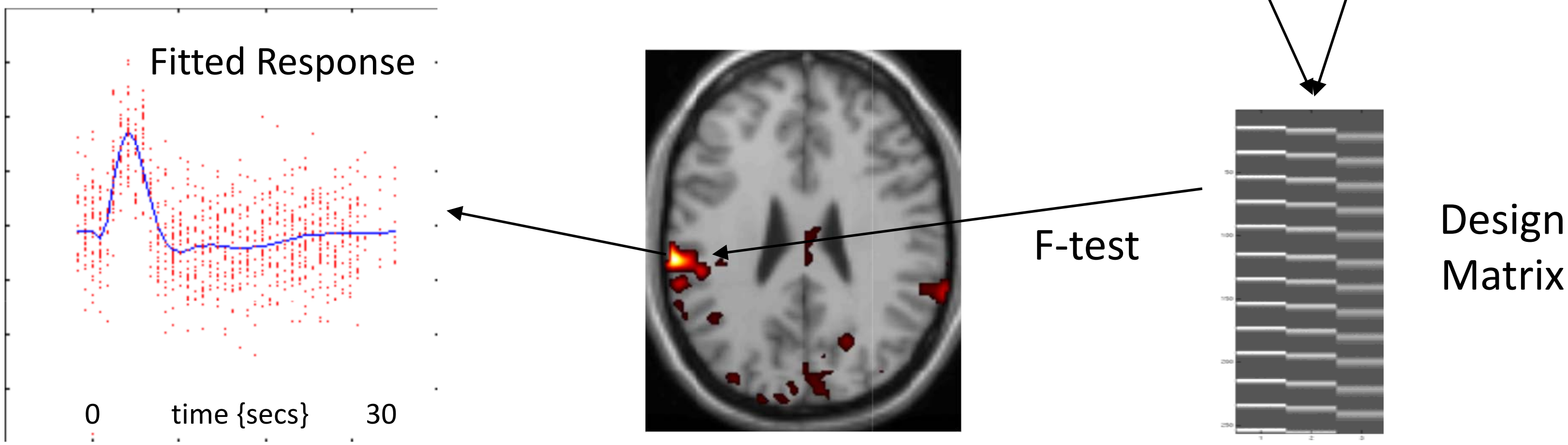
- Interpretation of Blocked designs can be particularly difficult, e.g., differences between two conditions in their mean activity across ~30secs could be due to attention / predictability of each stimulus…
- … if could estimate activity to individual trials, then can randomise their order to remove effects of predictability, etc

- Early such "event-related" designs waited ~30secs between trials, to allow BOLD response to return to baseline (much like ERP analyses)…
- … but if responses to successive trials summate in a linear fashion, then can model the overlap expected from shorter times between trials…
- … effectively using the GLM to "deconvolve" the mean response to trials of the same type (condition)

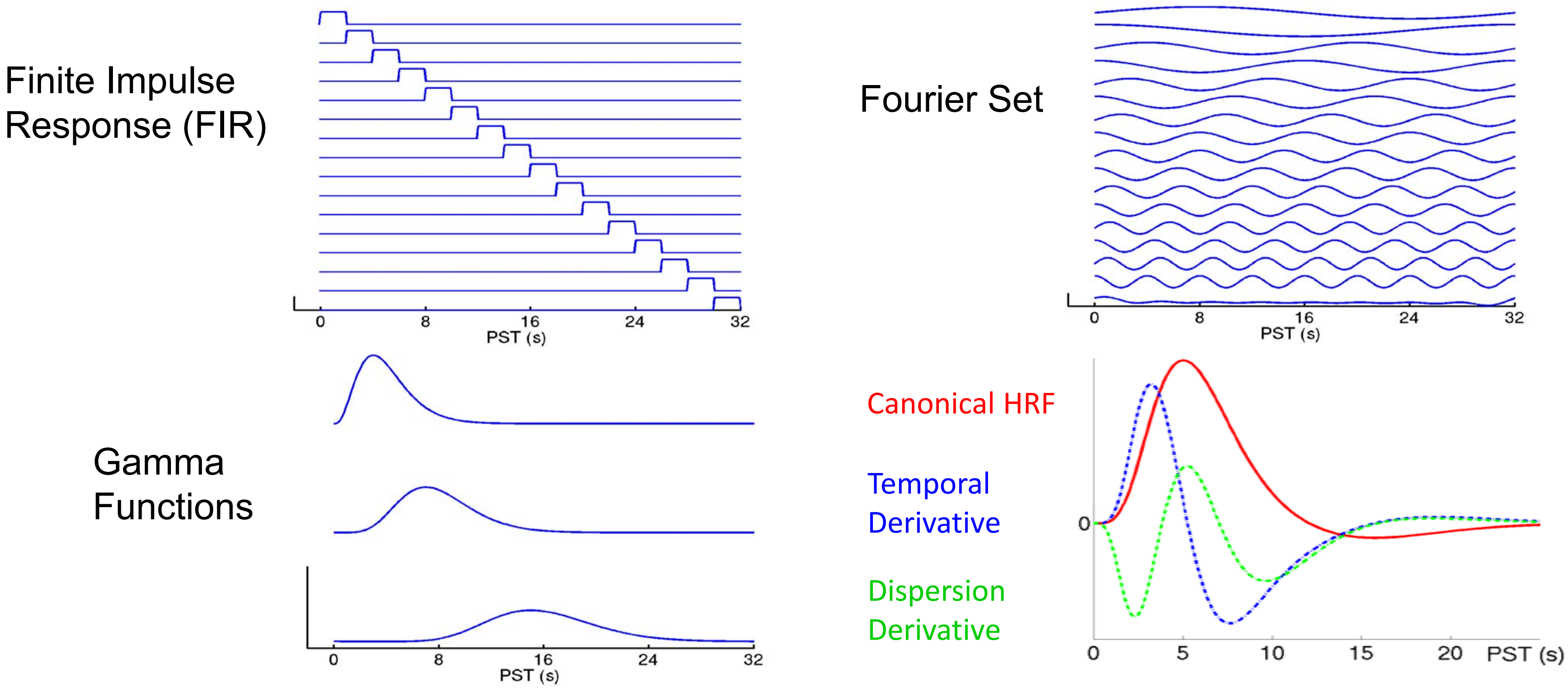
Deconvolution & HRF variability

- Noted earlier that HRF can differ across brain regions / individuals, e.g., owing to differences in vasculature
- One can allow for this variability by using a "temporal basis set", such as Gamma functions, or temporal derivatives of a "canonical" HRF





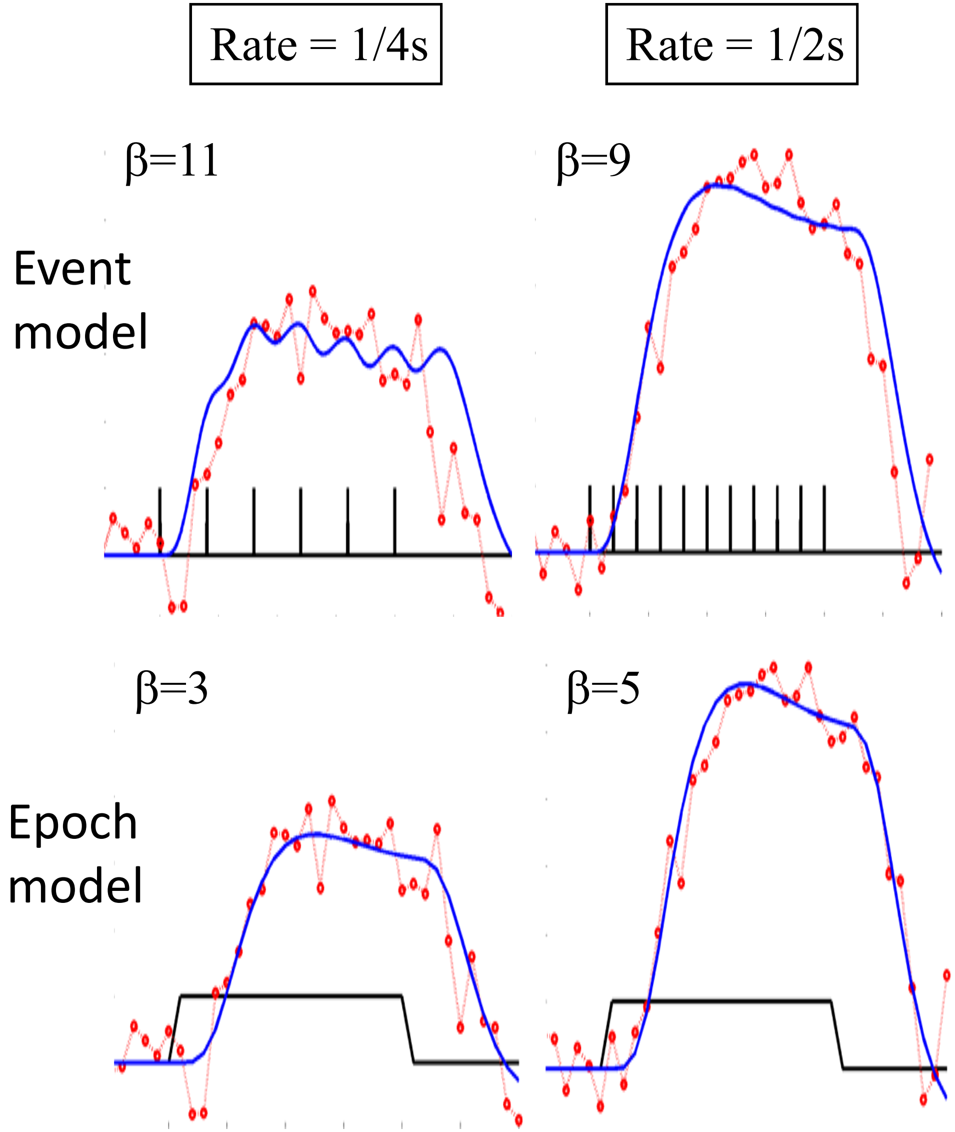
Temporal Basis Sets



- Could estimate HRF for each person (from known brain region during known task)… but would not handle between-region variance
- Could fit HRF with nonlinear (iterative) model, but computationally expensive at every voxel…
- In practice, many people stick with a single “canonical HRF” which probably explains ~80% variance in most regions/people (Henson et al, 2024, HBM)

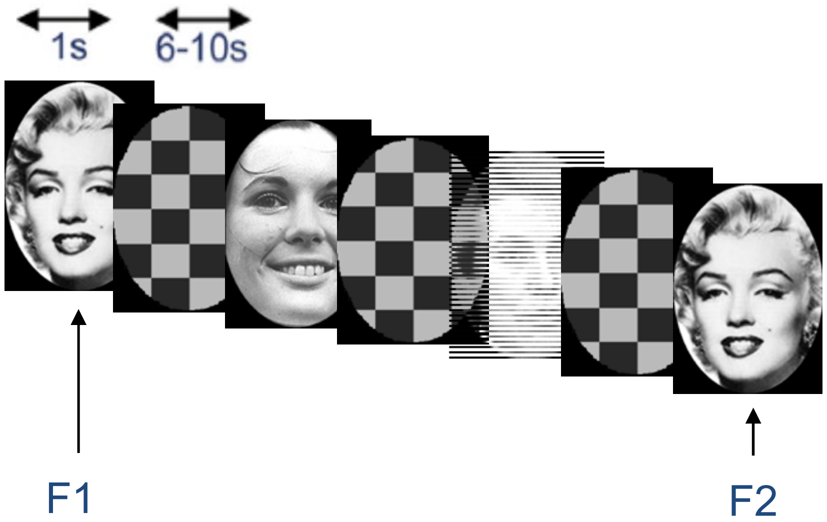
Some terminology

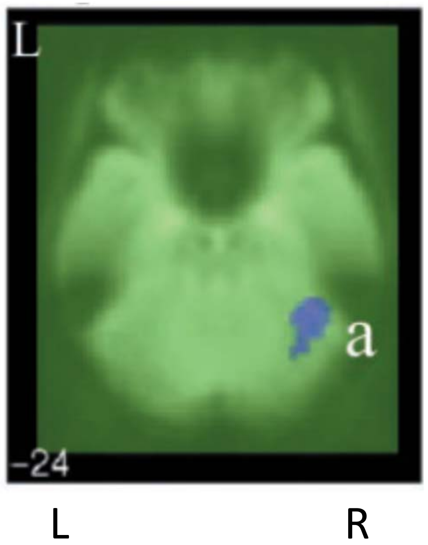
- Conditions can be *blocked* or *intermixed*…
- … neural activity can be modelled as *events* or *epochs*
- (event = delta function of duration 0; epoch = top-hat function with duration > 0)
- An event vs epoch model affects interpretation of parameters:
  - For event model, β is response *per trial*
  - For epoch models, β is response *per block*

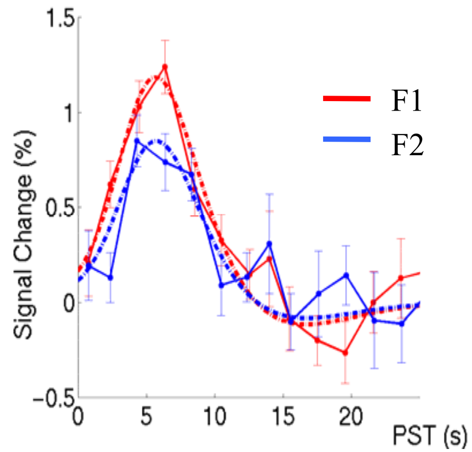


Intermixed, Trial-based Designs

- Randomly intermixed initial and repeated presentations of faces (blocking repetition could affect attention, etc)
- The FFA shows adaptation to repeated presentations (of familiar faces)





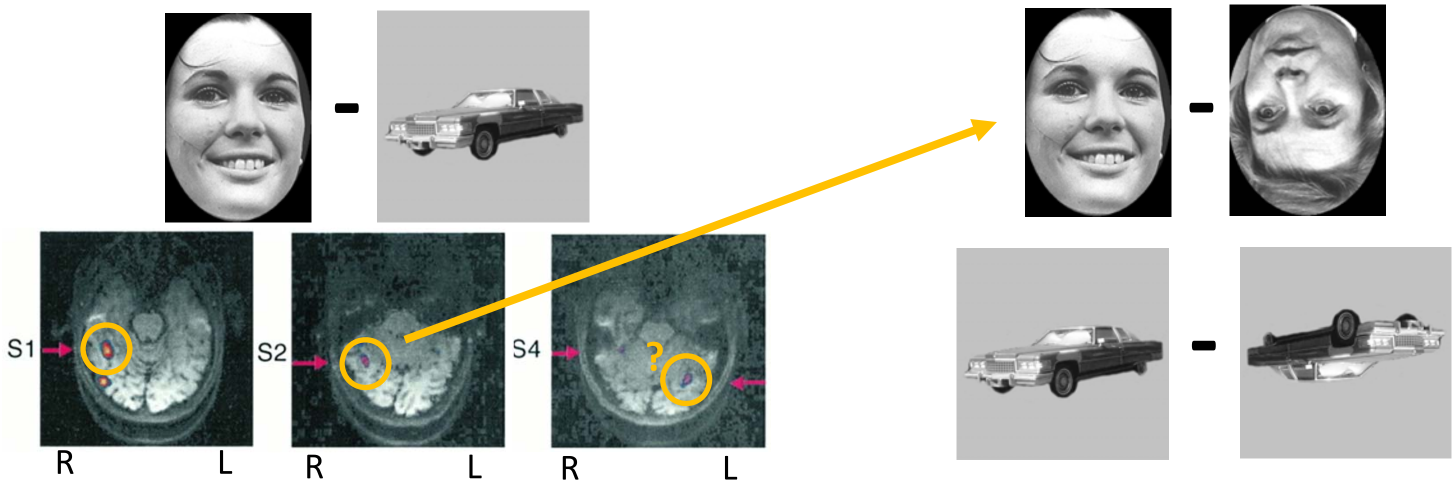


Henson et al. (2000) Science

- Other advantages of event-related fMRI:
  - When timing of events can only be indicated by participants (e.g., spontaneous transitions of perceptually ambiguous stimuli)
  - When the type of event can only be determined by participant's response (e.g, correct vs incorrect trials, or subsequently remembered vs forgotten…)
  - When events cannot be blocked (e.g, oddball trials)

Functional Localisers

- A separate session/run/scan ("localiser") to identify, e.g, "face-responsive" regions like FFA in each person…
- …followed by further investigation of properties of those functionally-defined ROIs in a main experiment

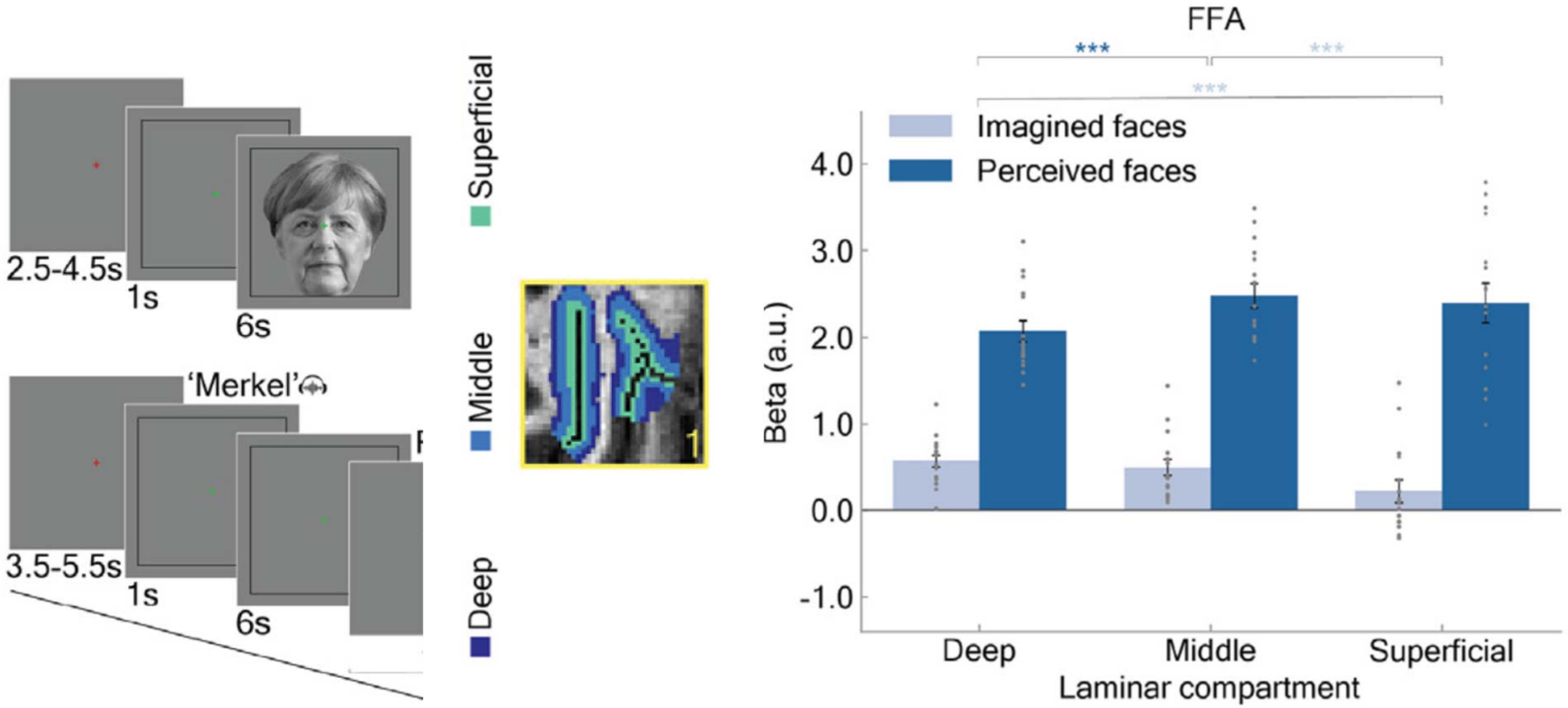


- Good for allow anatomical variability across participants (but to what extent; how much location variance is measurement noise; face-responsive regions also seen in smoothed, anatomically-registered images…)
- Problematic to identify ROI (statistically) in some participants
- (Often missed opportunity to examine factorial interactions)
- Important to match other factors, such as task and any effect of time/fatigue/habituation (and match localiser stimuli/tasks across studies)

Friston et al. vs. Saxe et al. (2006) Neuroimage

Laminar fMRI

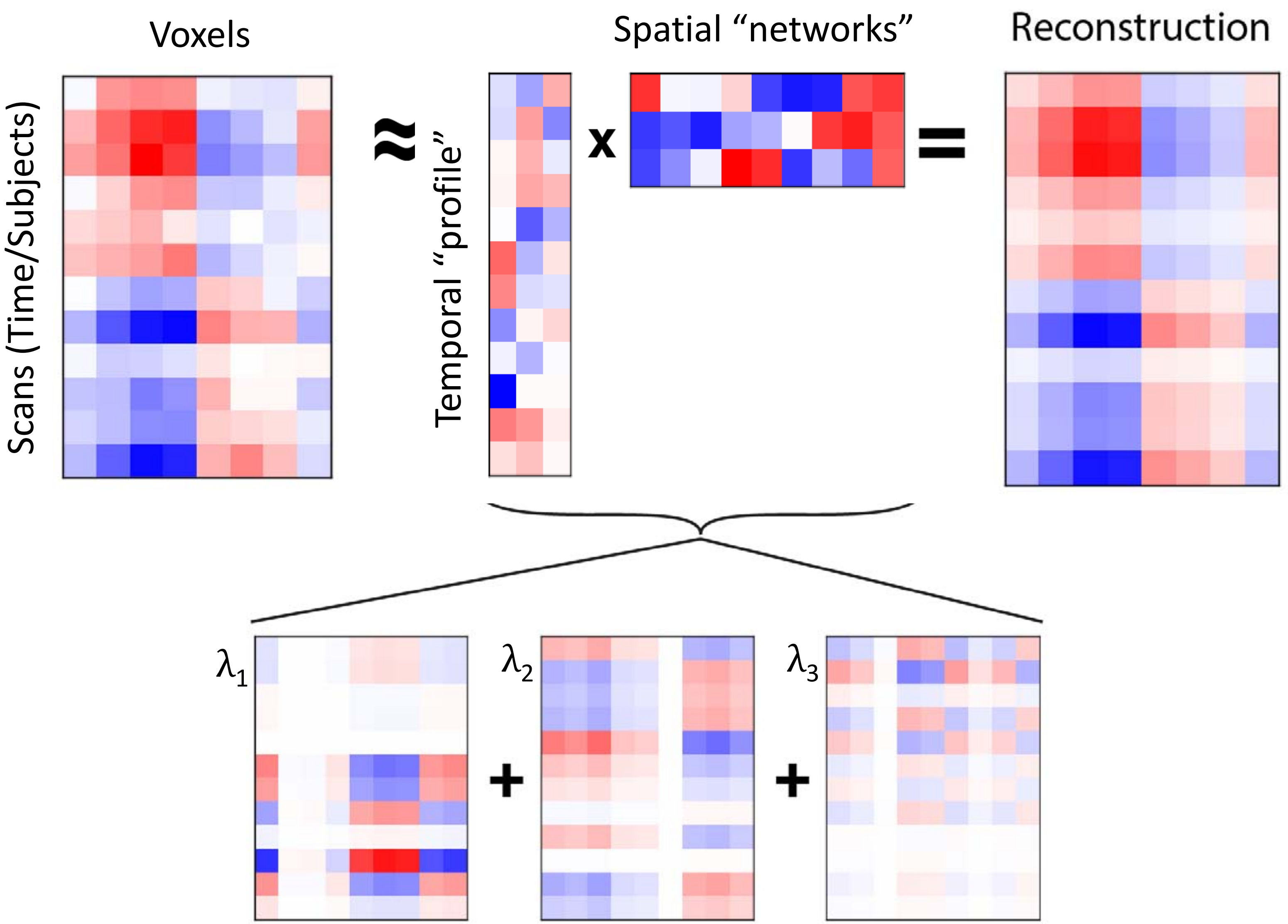
- High-field (eg 7T) fMRI offers higher spatial resolution (<1mm voxels), distinguishing layers (laminae) within cortex (typically 2-3mm thick), e.g, deep vs middle vs superficial
- Top-down connections tend to target the deep and superficial layers, whereas bottom-up connections preferentially target the middle layer
- Perceived vs Imagined faces should activate different layers?



- (Also used for layer-specific MVPA)

Carricarte et al (2024) iSci.

Matrix Factorisation
(e.g. Singular-Value Decomposition, SVD)

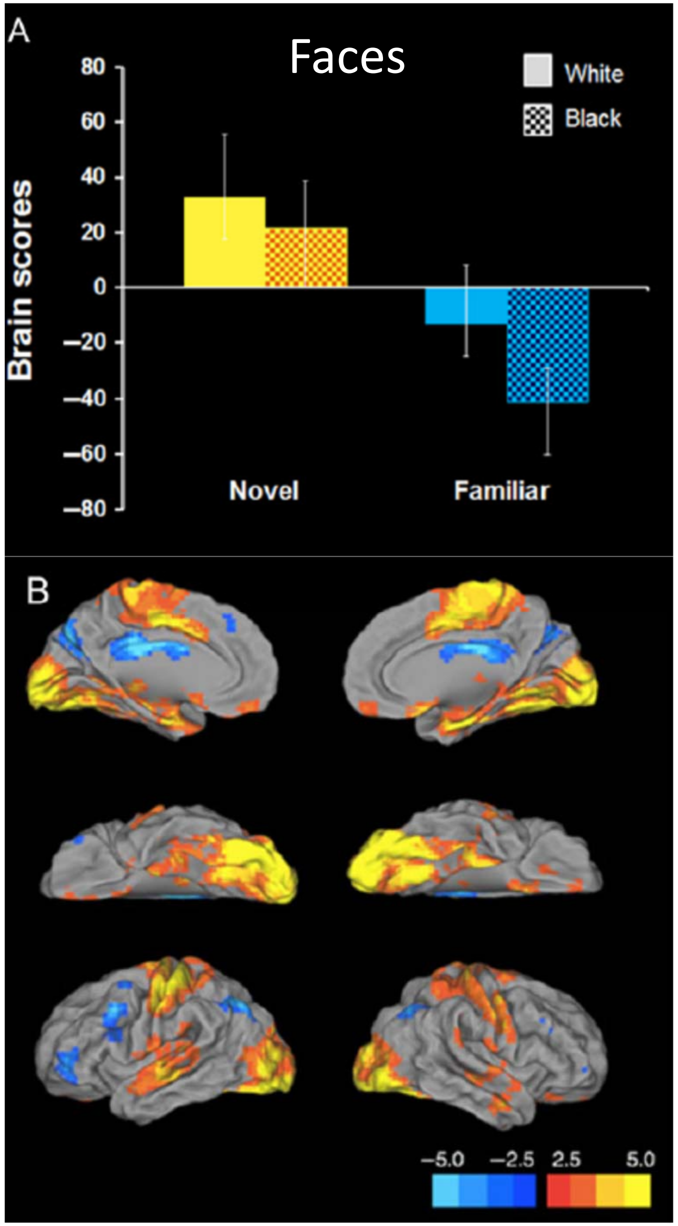


http://alexhwilliams.info/itsneuronalblog/2016/03/27/pca/

Multi-variate Analysis

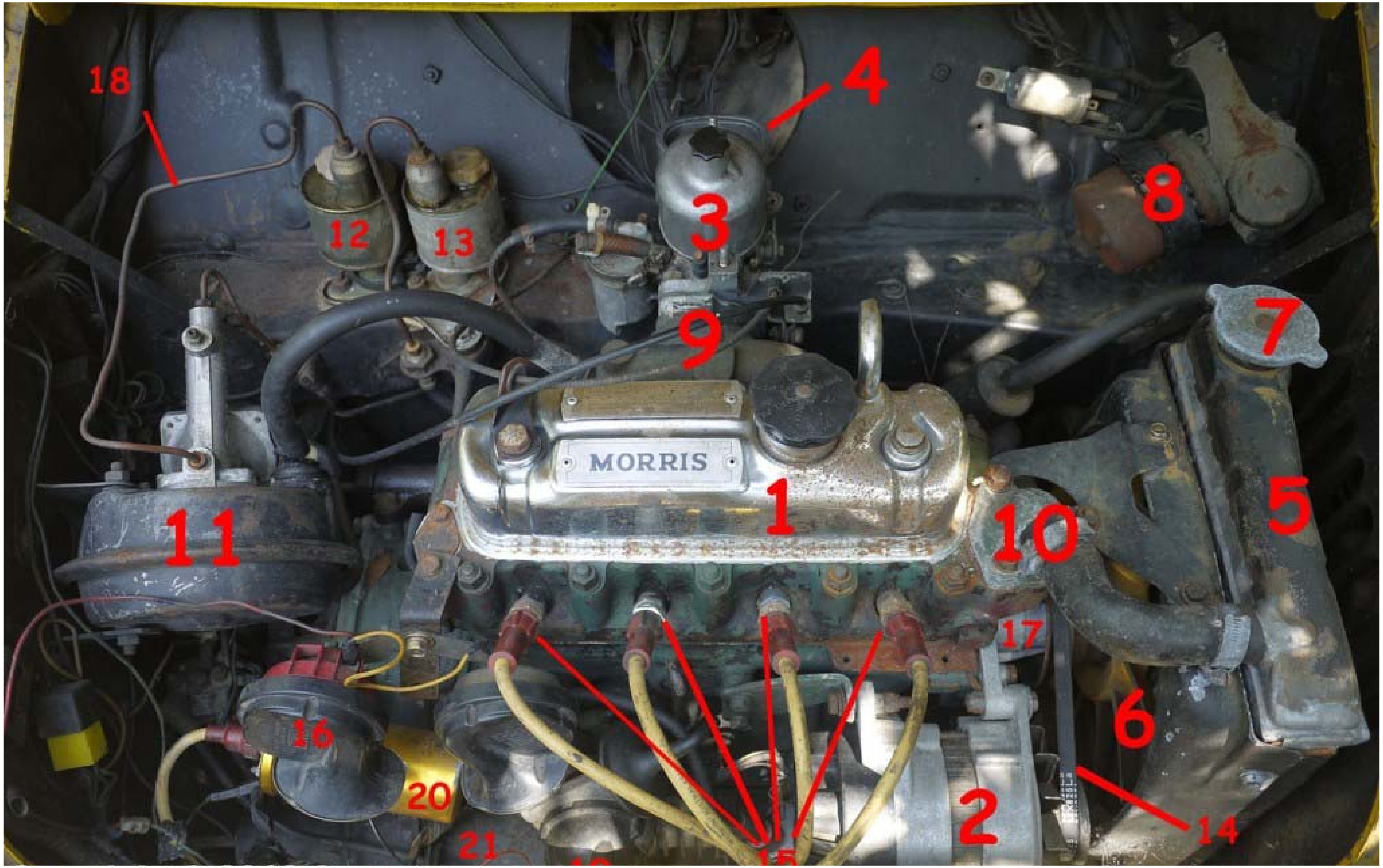
- Early studies decomposed images using Principal Component Analysis (PCA = *svd(Y'Y)*)…
- …or Partial Least Squares (PLS = *svd(X'Y)*) or Canonical Correlation Analysis (CCA)…
- …or Independent Component Analysis (ICA), and correlate with predictors in X

- These are more sensitive, and appropriate if functions are performed by Networks of ROIs rather than individual ROIs…
- … but care that difficult to interpret role of individual ROIs



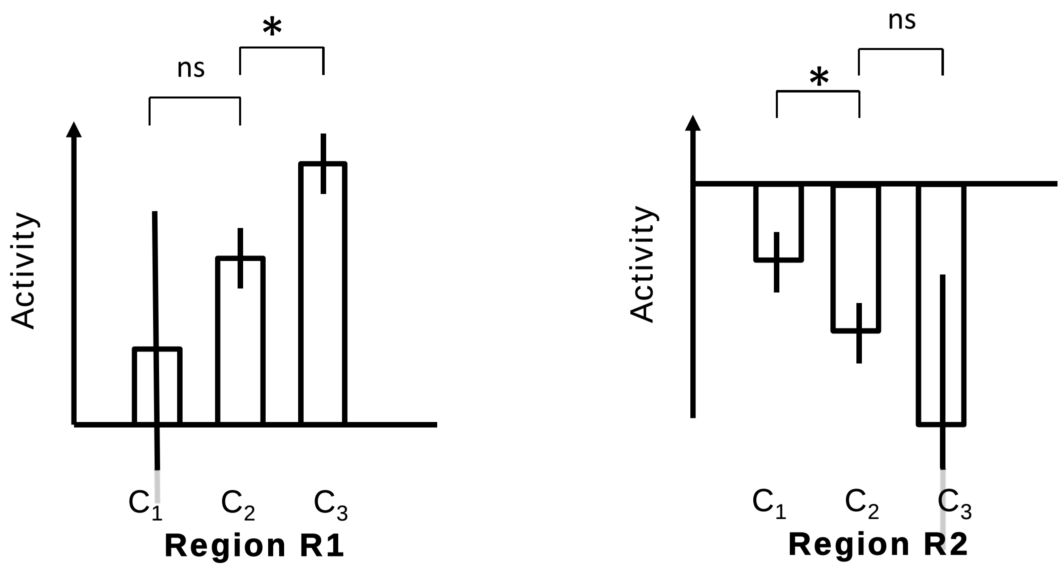
Cloutier et al. (2017) Cereb. Cortex

Are Functions “localised” in Brain?

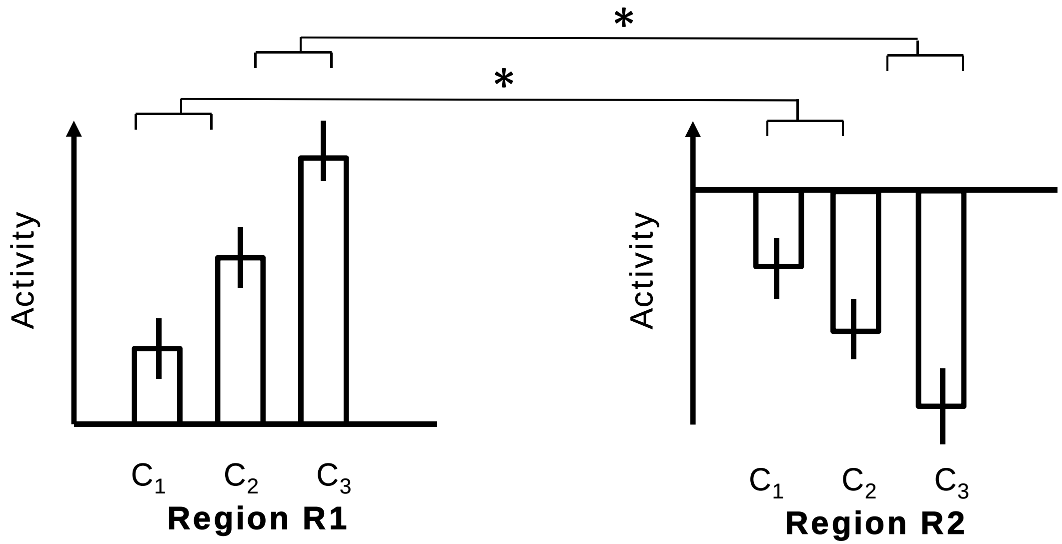


Functional Specialisation
("Imagers Fallacy")

Simple effects within each ROI not sufficient

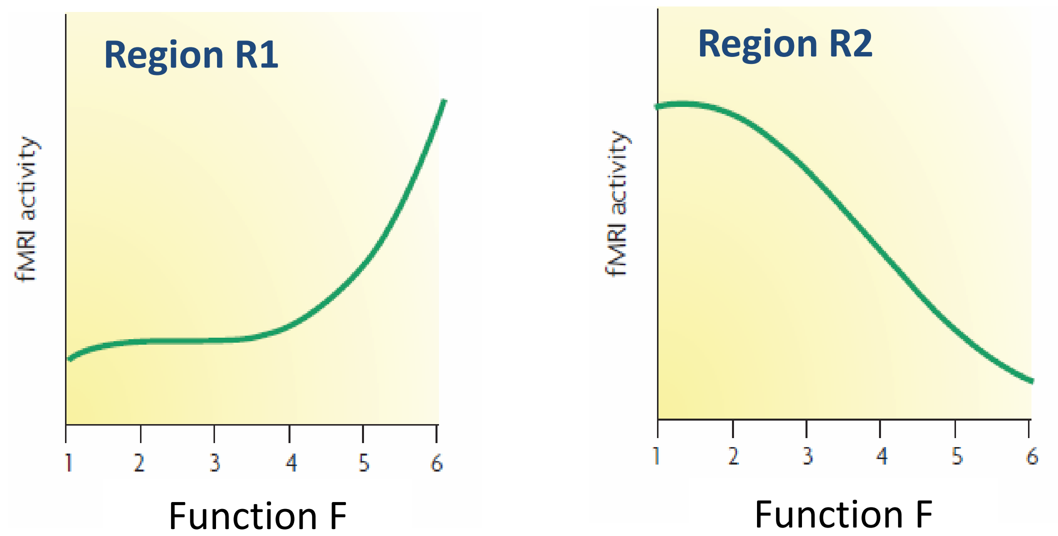


Need interaction between Region and Condition…?

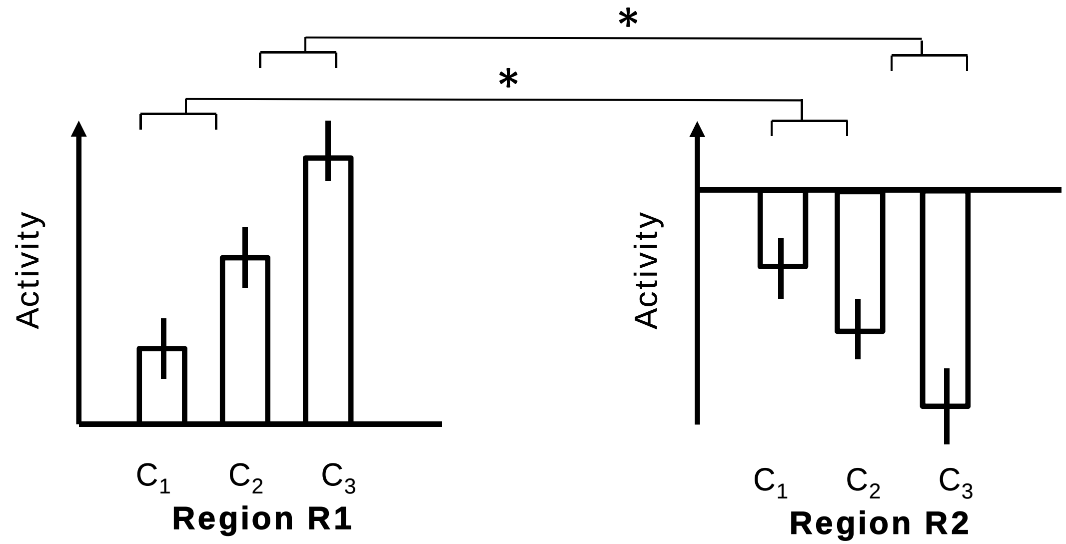


Henson (2006) Trends Cog. Sci.

Functional Specialisation
Nonlinear (but monotonic) Function-> BOLD



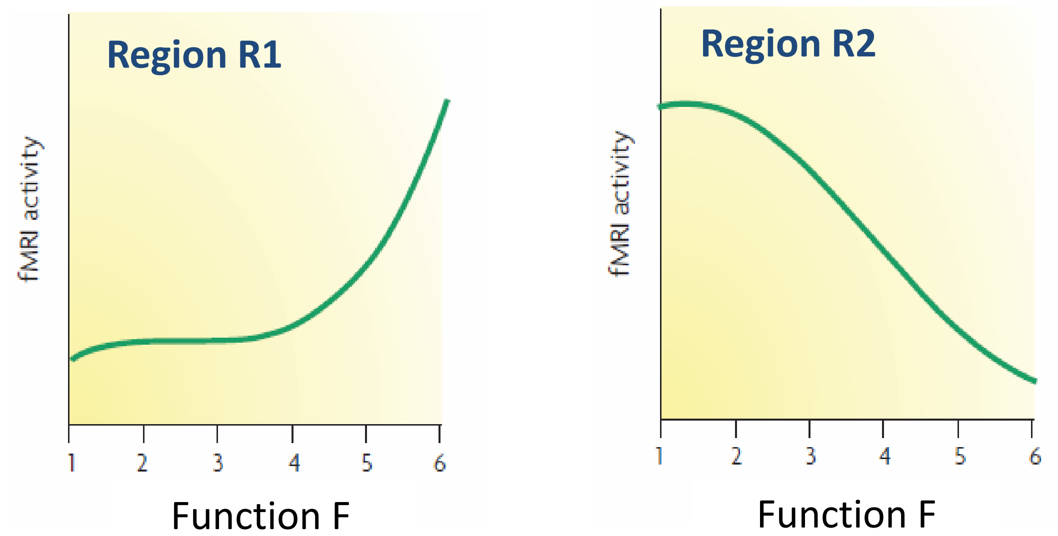
Squire et al (2007) Nat. Rev. Neuro.



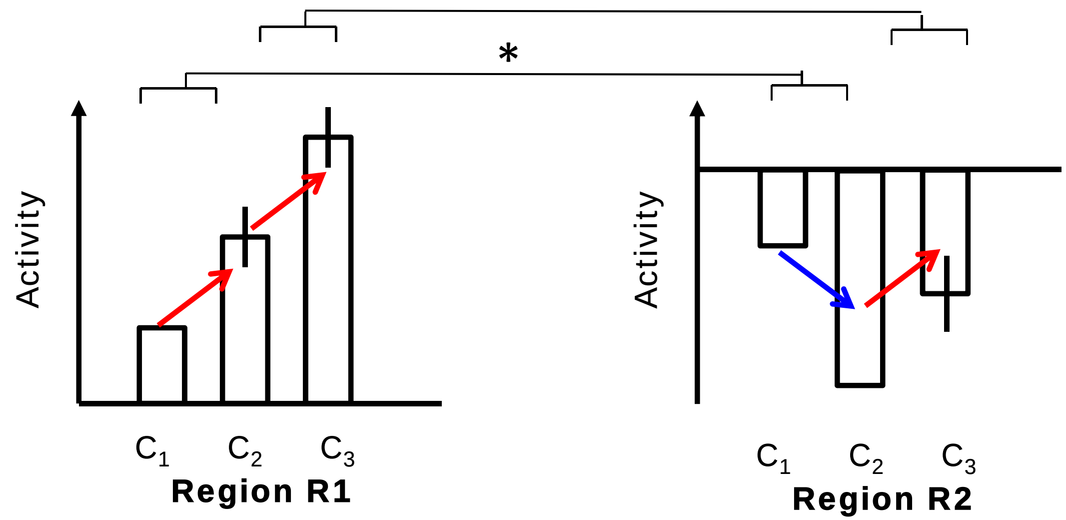
Need interaction between Region and Condition…?

Henson (2006) Trends Cog. Sci.

Functional Specialisation
"Reversed Association"



Squire et al (2007) Nat. Rev. Neuro.



…need a "reversed association"…

("state-trace" analysis)

Henson (2006) Trends Cog. Sci.

Functional Specialisation (“Forward Inference”)

- Finding a reversed association between 2+ Regions and 3+ Conditions implies more than one underlying Function (assuming only a monotonic relationship between Function and Activation)…
- … so can rule out single-function accounts
- This does not require a one-to-one mapping between function and region, unlike “Reverse Inference”….

Functional Specialisation
Dangers of "Reverse Inference"

- If, Region R activated when comparing Conditions C_1 and C_2 …
- … and, in previous studies, Region R associated with Function F …
- … then, Conditions C_1 and C_2 also differ in Function F

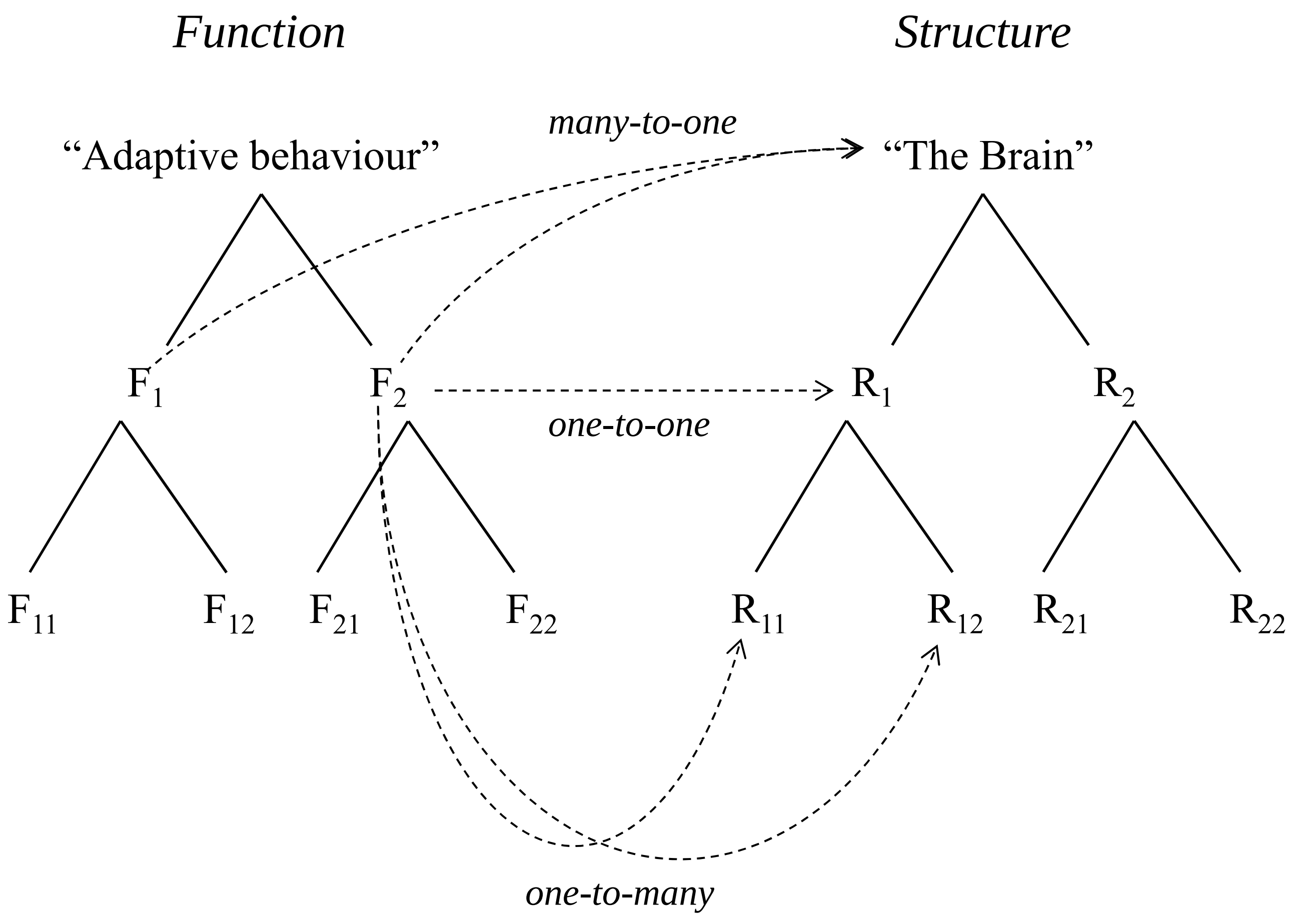
- But assumes a one-to-one mapping between Region and Function

$$p(F|R) = p(R|F)p(F) \,/\, (\, p(R|F)p(F) + p(R|{\sim}F)p({\sim}F)\,)$$

- If Region R could be associated with other functions (in other studies, ie $p(R|{\sim}F) \sim= 0$), then probability of Function F occurring in your study is reduced…
- … and meta-analyses often do show same Region R in multiple different contexts

Poldrack (2006) Trends Cog. Sci.

Function-Structure Mappings

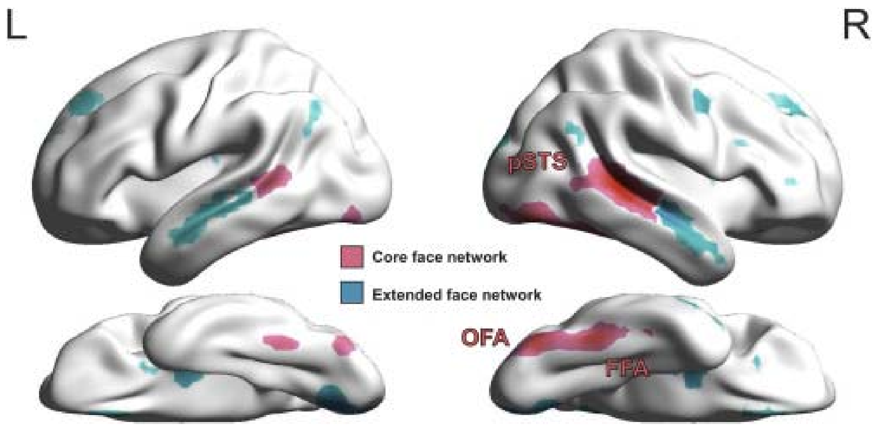


Henson (2005) QJEP

- (Same problem if replace Regions with Networks N_1, N_2…)

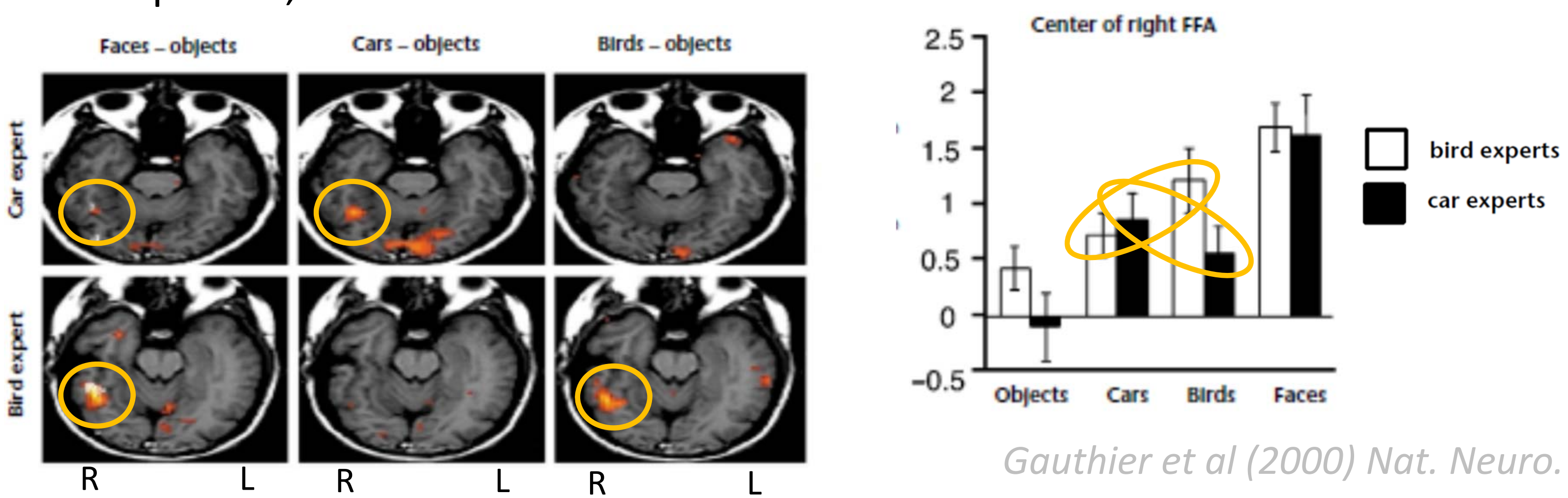
Function-Structure Mappings

- As an example of one-to-many mapping, comparing Faces to non-Face stimuli can activate not only FFA, but also OFA, posterior STS, medial PFC, Amygdala…



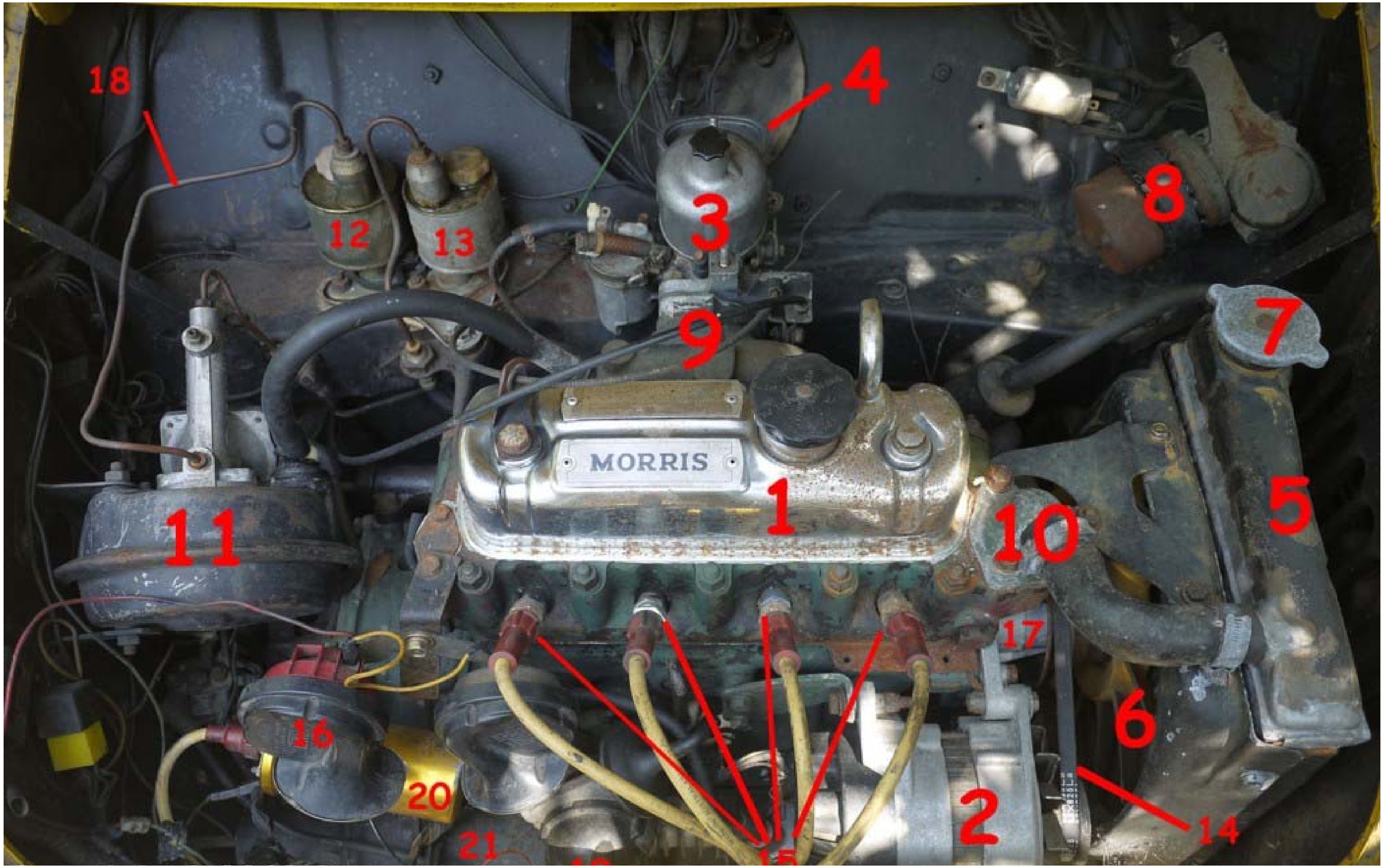
Gobbini & Haxby (2006) Neuropsychologia

- As an example of many-to-one mapping, others have associated FFA activation with visual expertise, even with non-face stimuli..



Gauthier et al (2000) Nat. Neuro.

Are Functions "localised" in Brain?



Applications

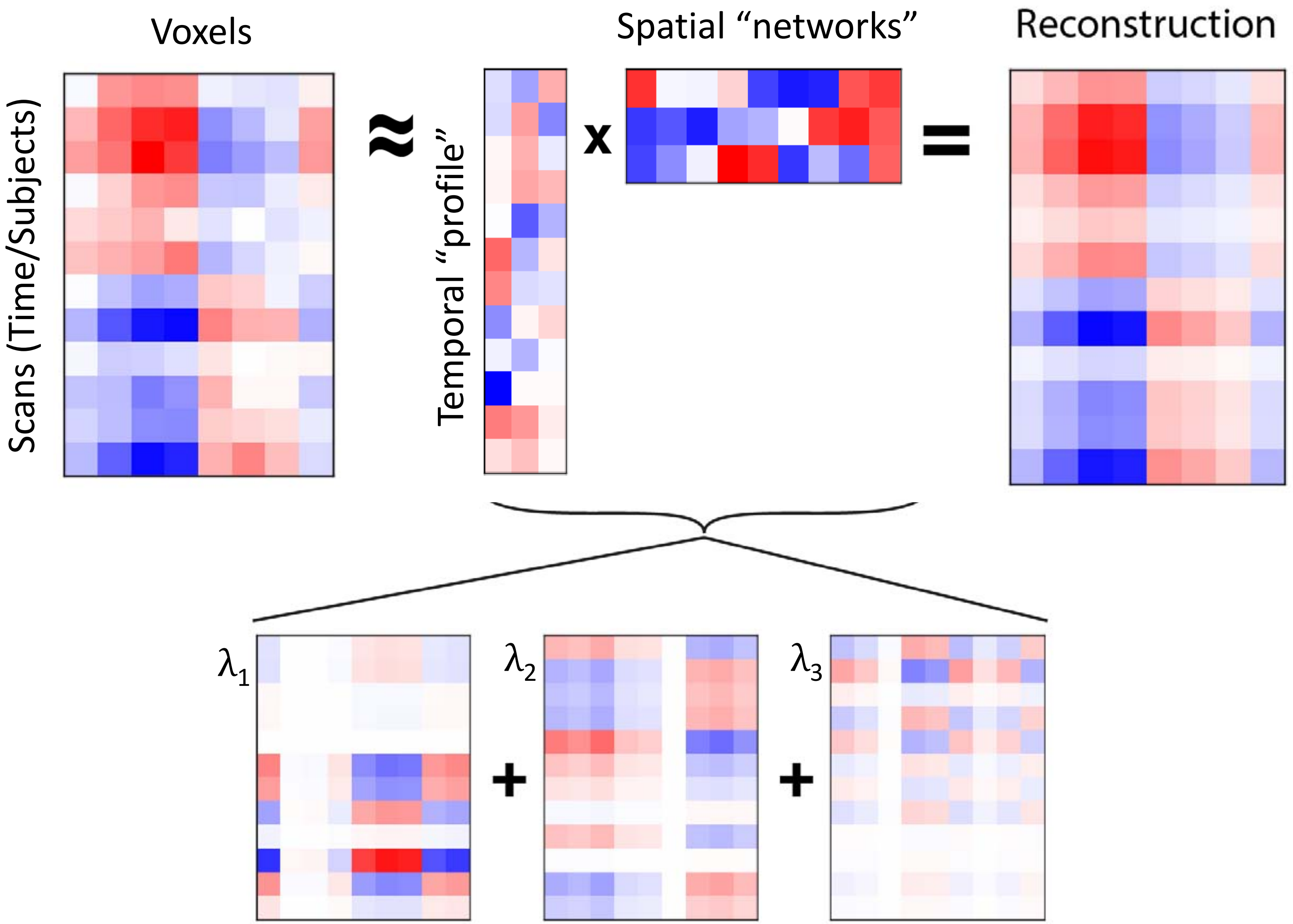
- Activation Analysis
  - Brain Mapping (mass univariate)
  - Subtraction logic (pure insertion)
  - Blocked versus Intermixed (Epoch vs Event-related)
  - Functional localisation; Forward and Reverse Inference

- Functional Connectivity
  - Matrix factorisation (Independent Component Analysis)
  - ROI-based connectomes, graph theory
  - Task-based (effective) connectivity

- Pattern (Information) Analysis
  - Multi-voxel pattern analysis (MVPA)
  - Representational Similarity Analysis (RSA)

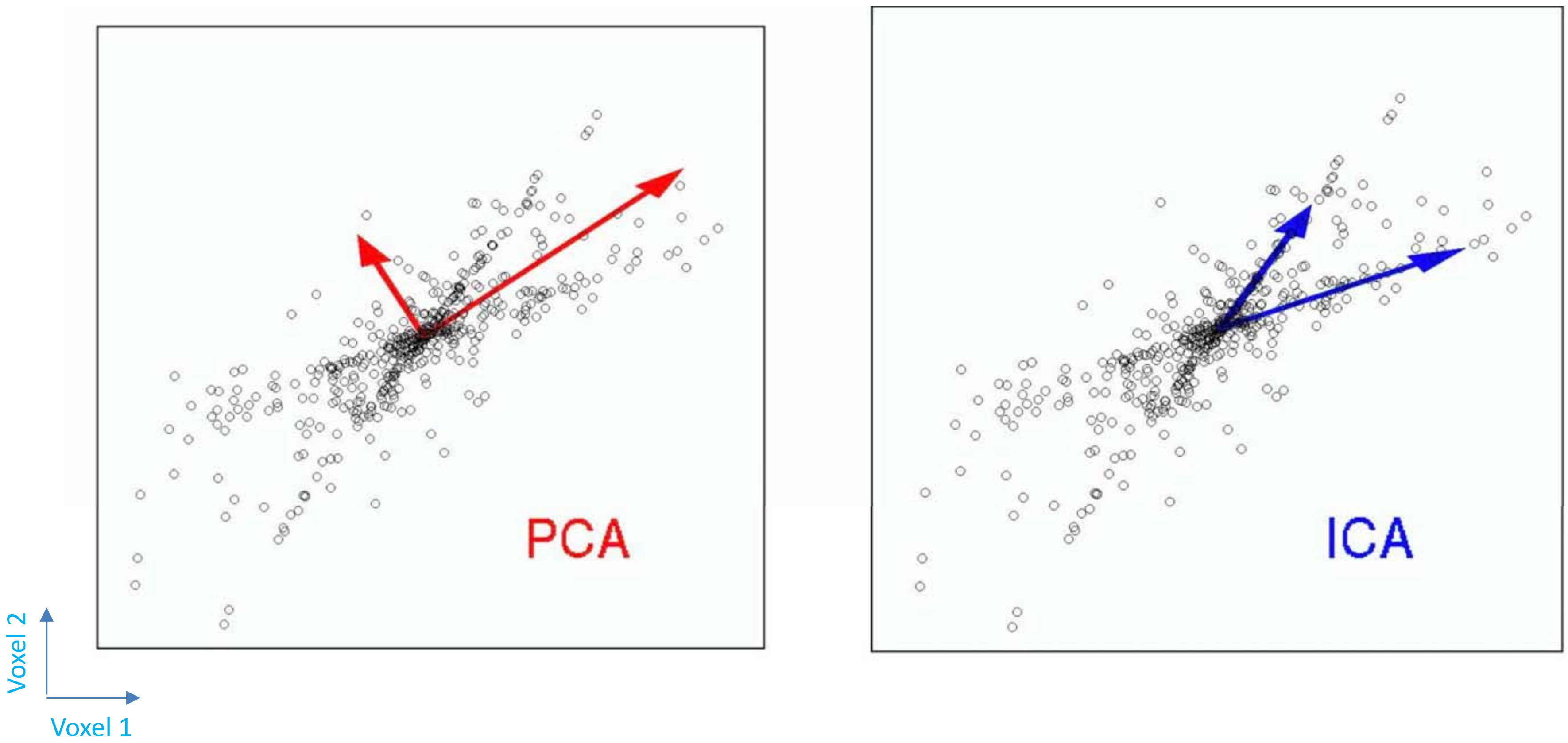
Theme: Fusiform Face Area (FFA)

Matrix Factorisation



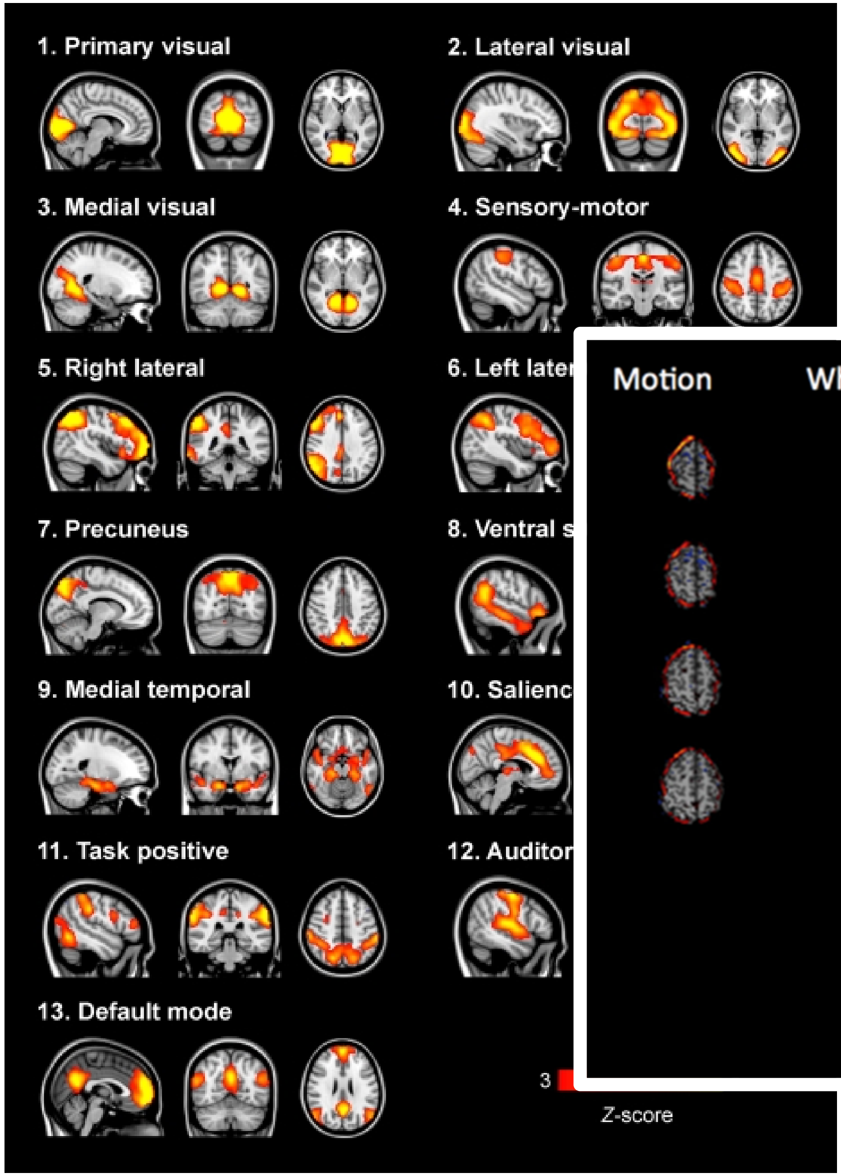
http://alexhwilliams.info/itsneuronalblog/2016/03/27/pca/

PCA vs ICA

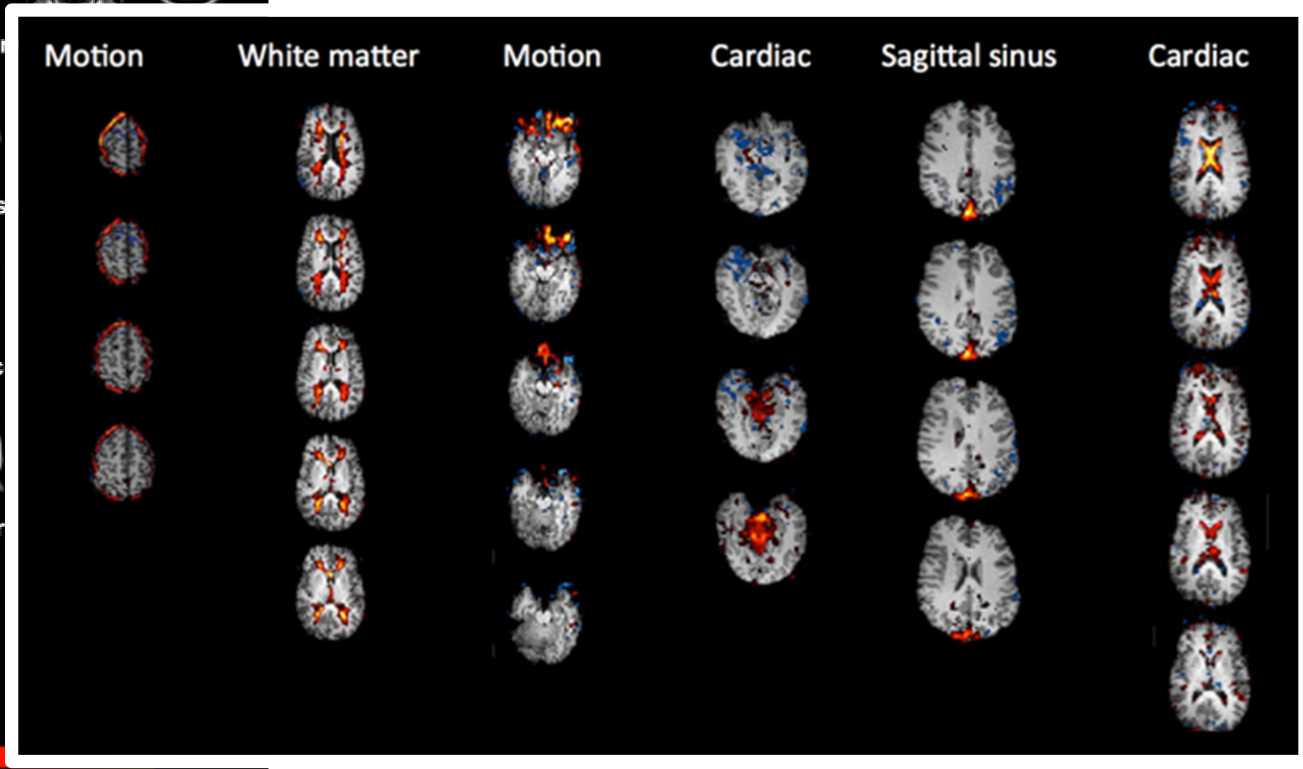


- ICA assumes components are independent (usually across space with fMRI, but could be across time) and has been shown more effective than PCA in (re)producing characteristic resting-state networks (RSNs), and separating signal from noise

ICA RSNs

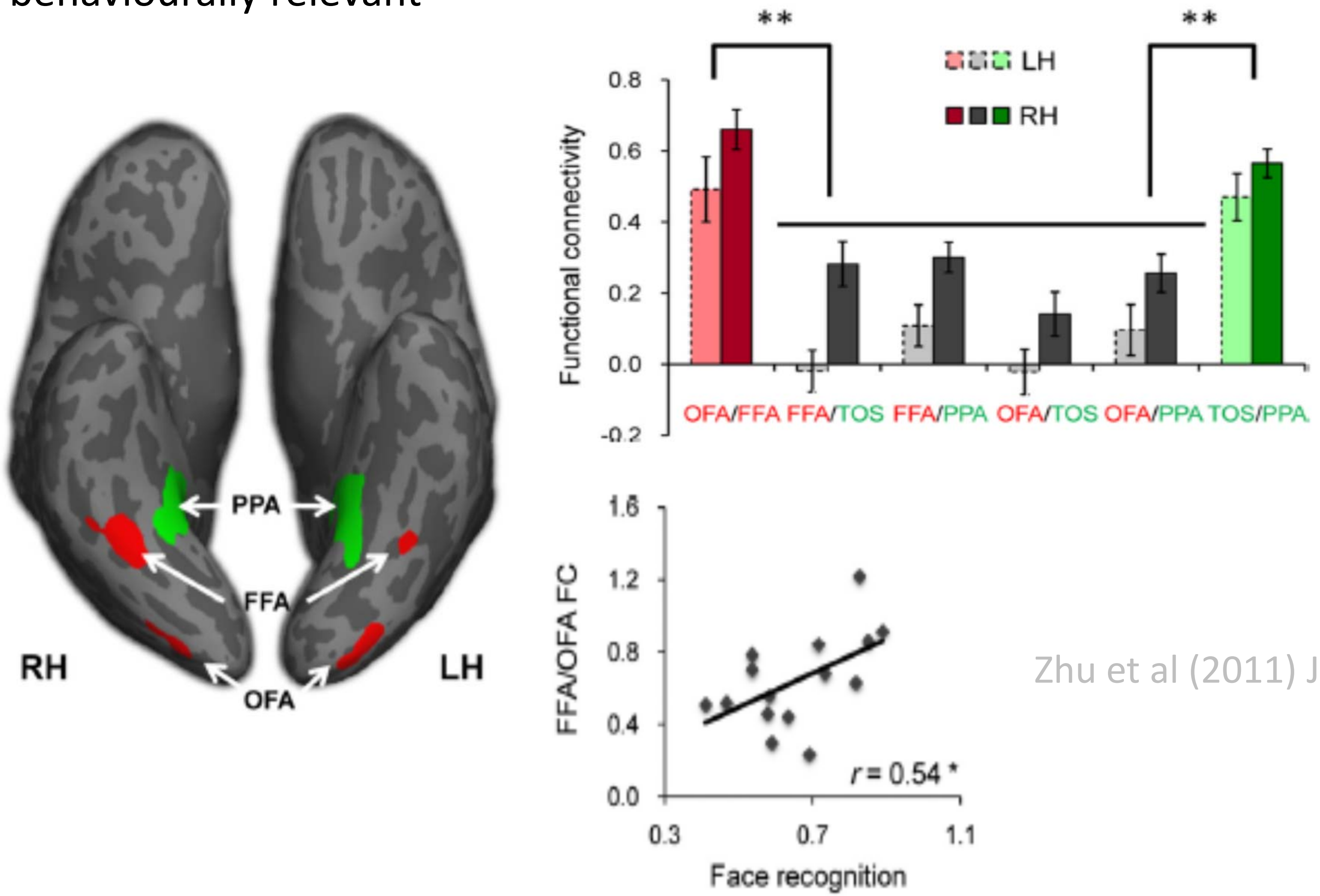


ICA Noise components



Seed(ROI)-based Connectivity

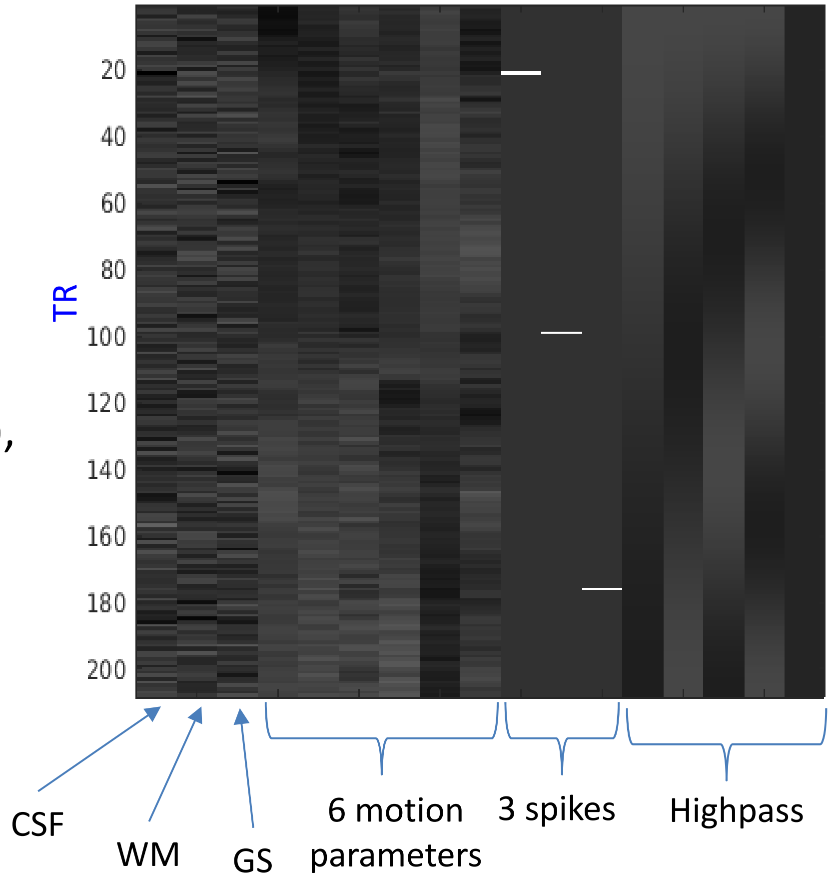
- Resting-state functional connectivity between FFA and OFA is behaviourally relevant



Zhu et al (2011) J. Neuro.

Motion & Physiological Artefacts

- Lowpass filter (eg, 0.1 Hz) to remove high-freq motion since HRF slow? But loose many dfs…
- Highpass filter (eg, 0.01 Hz) to remove low-freq aliased physiological noise
- Covary WM/CSF/Global Signal (latter contentious)
- “Scrubbing” (removing volumes with high motion), but ignores temporal autocorrelation…
- … so rather than remove data, regress out motion parameters, including derivatives, second-order expansions (Volterra expansion), etc
- Spikes = extreme values (high order expansion) – model as separate regressors



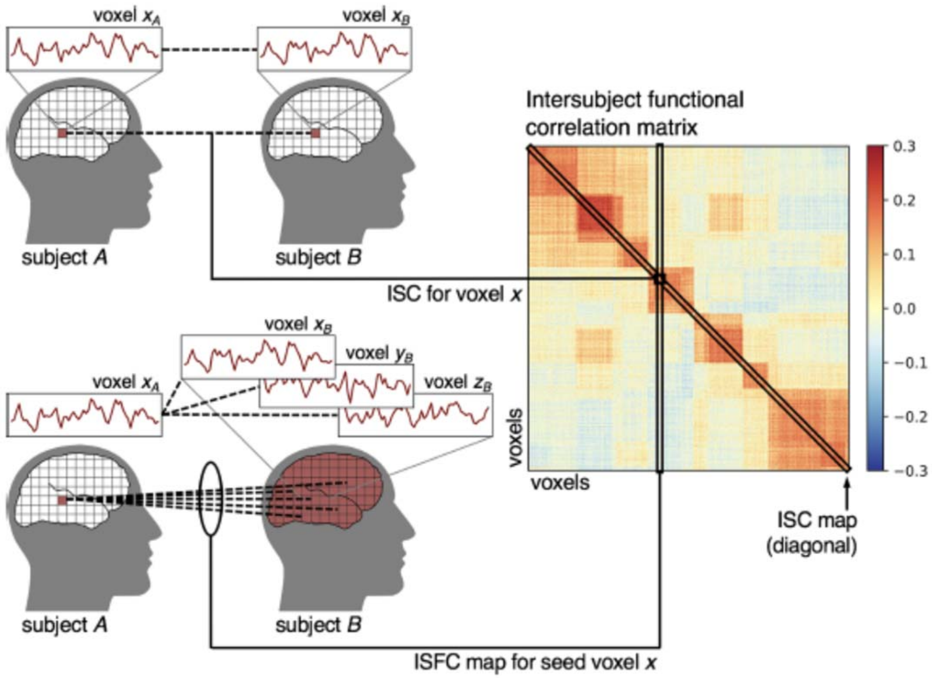
Resting-state, Movie-watching…

- Resting-state (eyes open or closed) is easy for all participants (even patients who might struggle with many tasks), and often assumed to reflect a trait component of people

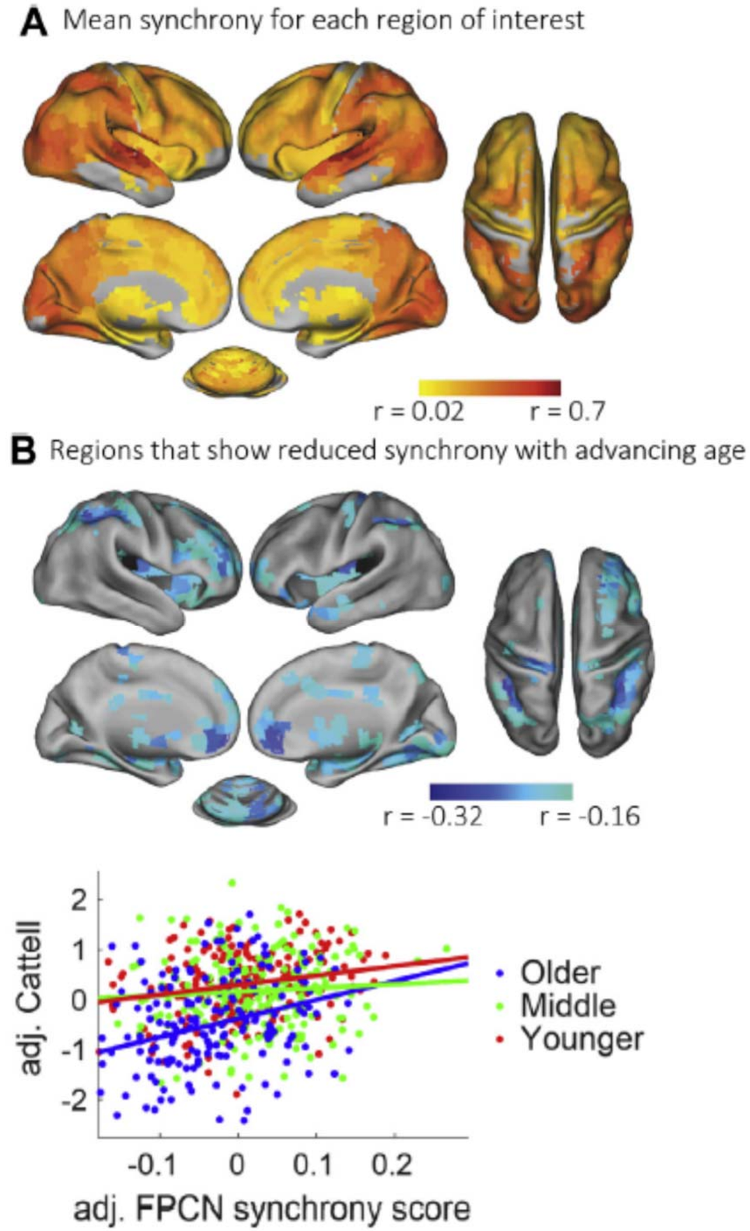
- However, no control over what people are doing cognitively (e.g., day-dreaming), i.e., time-series cannot be compared across people

- Movie-watching is also easy for everyone, and:
  - 1) Can assume similar time-series across people watching same movie ("inter-subject correlation")

Inter-Subject Correlations



Nastase et al (2019) Soc. Cog. Aff. Neuro.



Geerligs & Campbell (2018) Neurobio. Age.

Resting-state, Movie-watching…

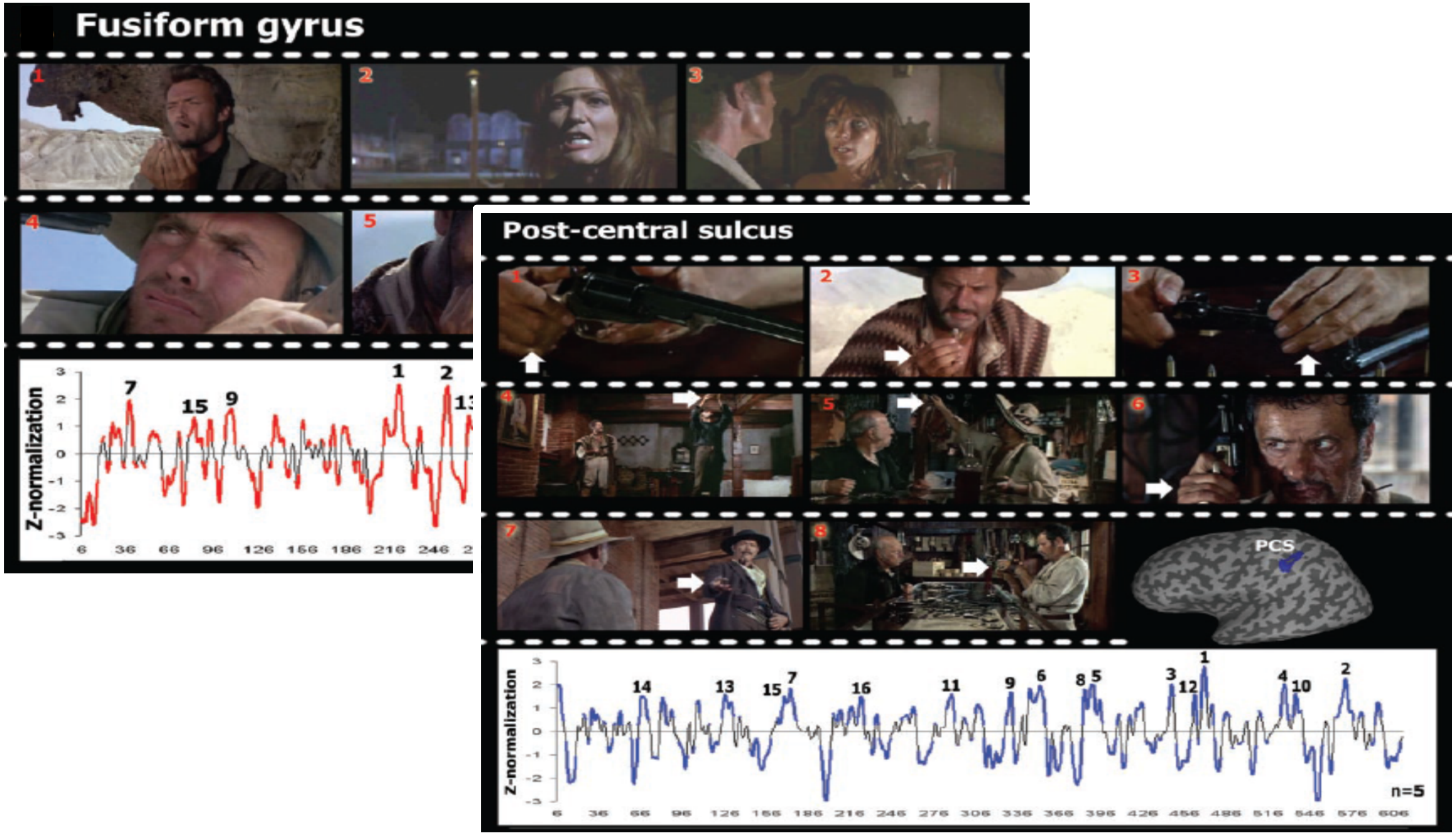
- Resting-state (eyes open or closed) is easy for all participants (even patients who might struggle with many tasks), and often assumed to reflect a trait component of people

- However, no control over what people are doing cognitively (e.g., day-dreaming), i.e., time-series cannot be compared across people

- Movie-watching is also easy for everyone, and:
  - 1) Can assume similar time-series across people watching same movie ("inter-subject correlation")
  - 2) Can examine common activation/patterns at certain points in movie (e.g. whenever a face present) – more "naturalistic" than many tasks

Reverse Correlation

- Confirm expected function in more naturalistic context…

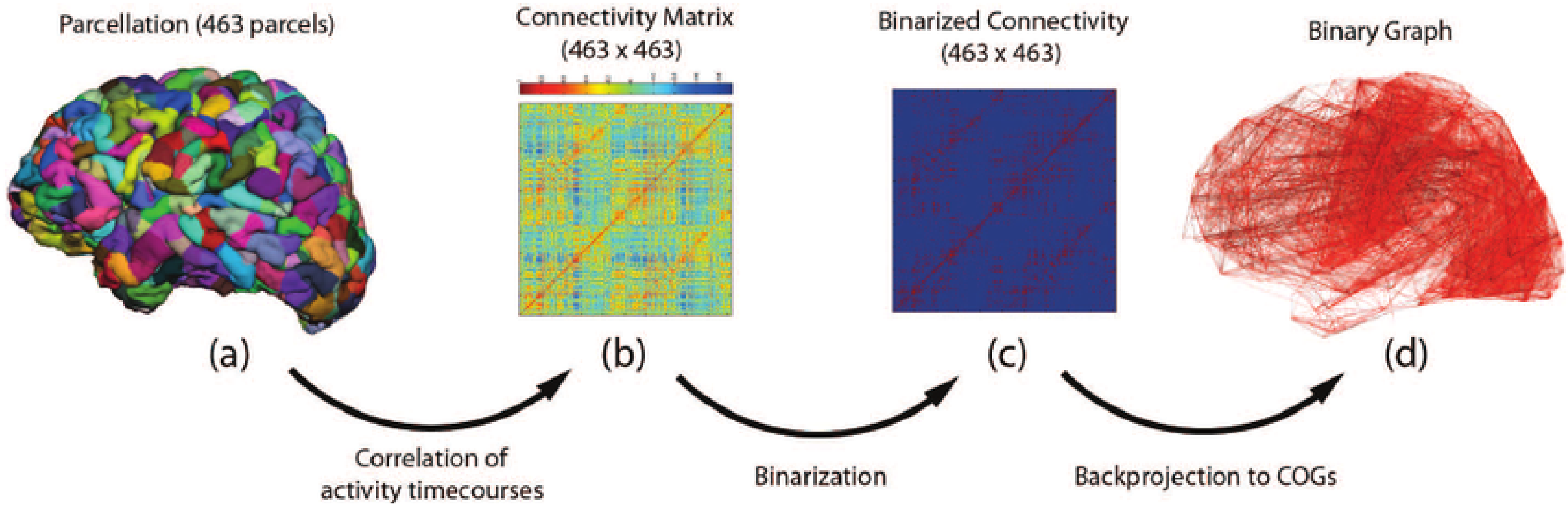


- …potentially "discover" new function of some regions

Hasson et al. (2004) Science

ROI-to-ROI Connectomes

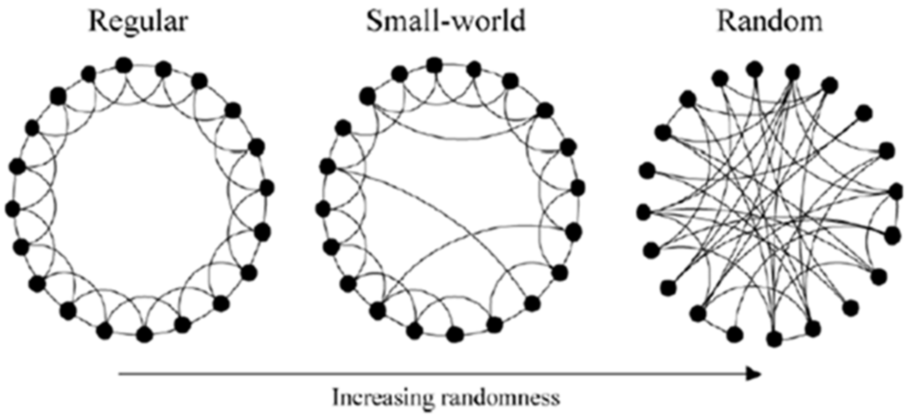
- All pairwise connections between ROIs (parcels), e.g., from a structural or functional atlas

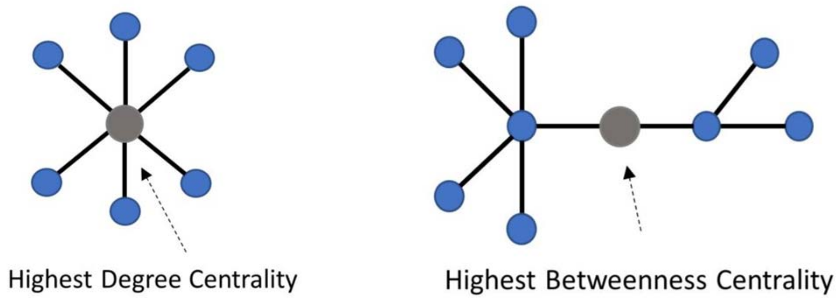


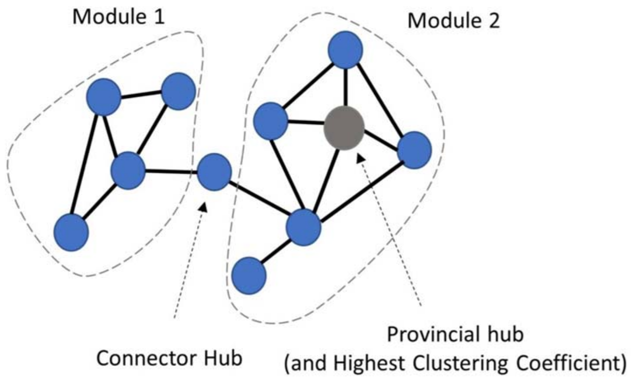
Bottger et al (2013) IEEE

Network / Graph Analysis

- Connectome can be summarised by one or more graph-theoretic measures, e.g., efficiency, clustering, small-worldness…



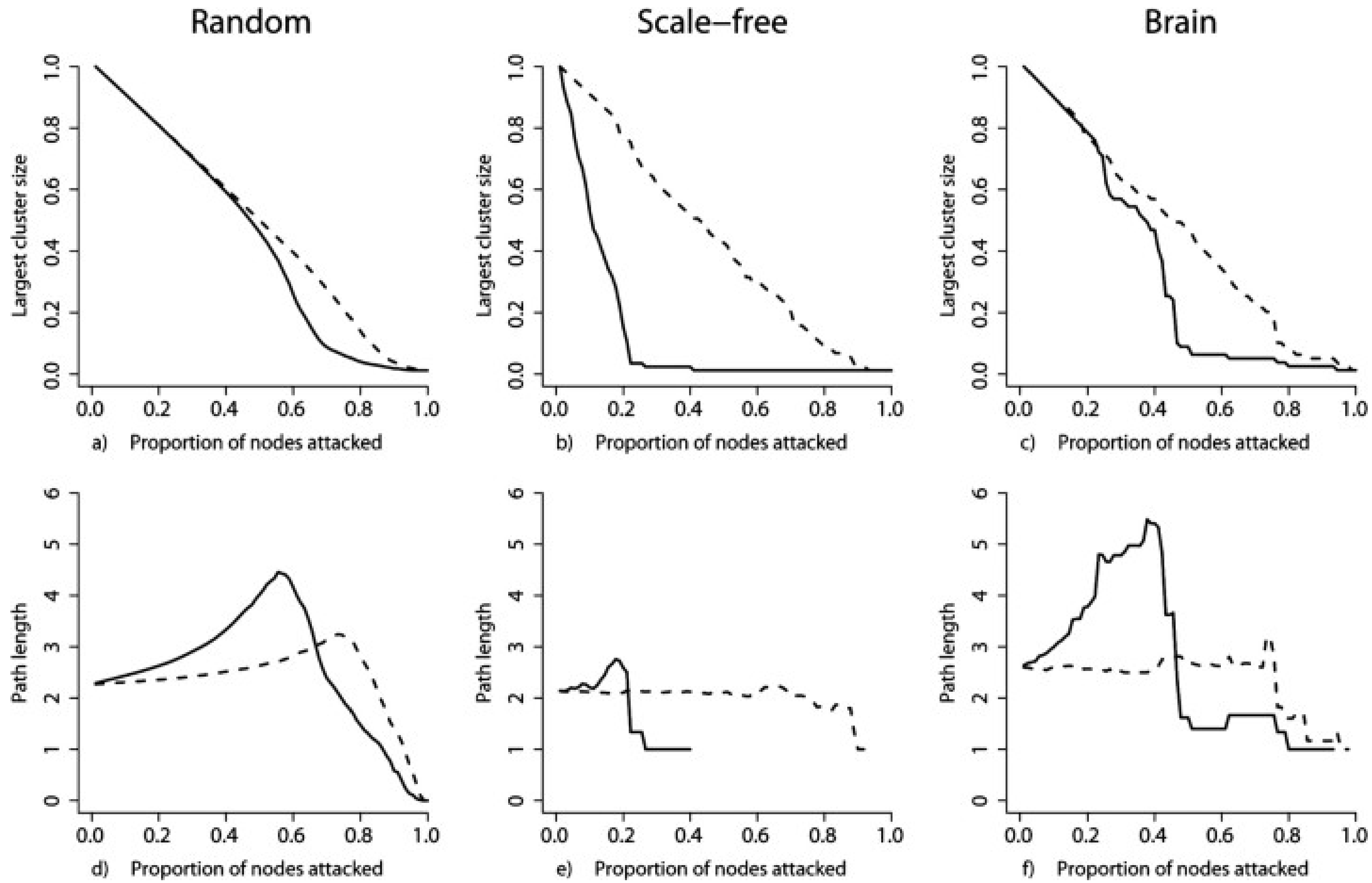




- Graph metrics abstracted from meanings of nodes/connections, so can compare fMRI networks with white-matter DTI networks, gene transcription networks, etc.

Network / Graph Analysis

- Rather than simply using graph theory to *describe* networks, we can simulate processes like neuro-degeneration or brain development, and match to graph-metrics, in order to understand *mechanisms*



Achard et al. (2006) J. Neuro.

Other types of Functional Connectivity

- **Multi-dimensional** (pattern-based) connectivity

  Basti et al., 2020, Neuroimage

- **Dynamic** connectivity (e.g moving windows or Hidden Markov Models)

  Hutchison et al. 2013 Neuroimage

Task-Based Connectivity

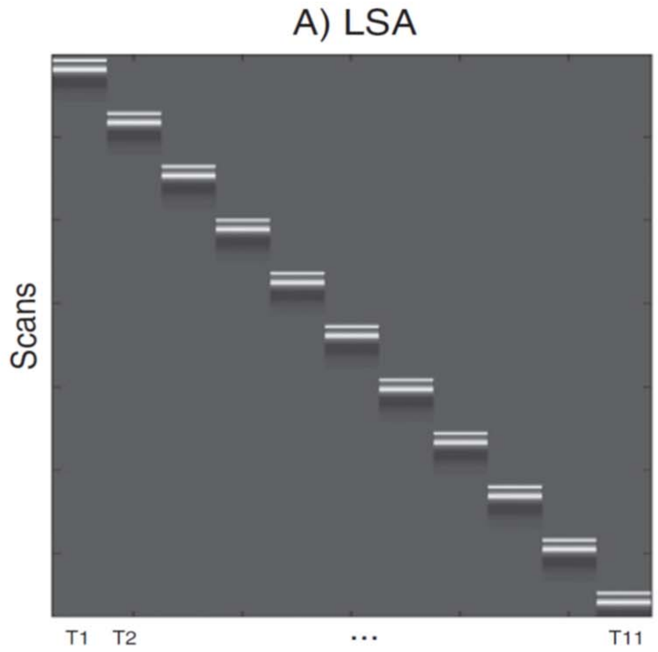
- Another way to minimise confounds is to compare functional connectivity across two or more tasks/conditions. Two main approaches:

I. Beta-Series Correlation/Regression (BSR): correlate Betas (trial-wise parameter estimates from single-trial model) across ROIs, and compare correlation coefficients/regression slopes across conditions

II. Psycho-Physiological Interactions (PPIs): construct a model (GLM) of fMRI timeseries with regressors for 1) conditions ("psychological"), 2) timeseries from one ROI (seed) ("physiological") and 3) the interaction between 1+2 (the key PPI term), and test significance of interaction term

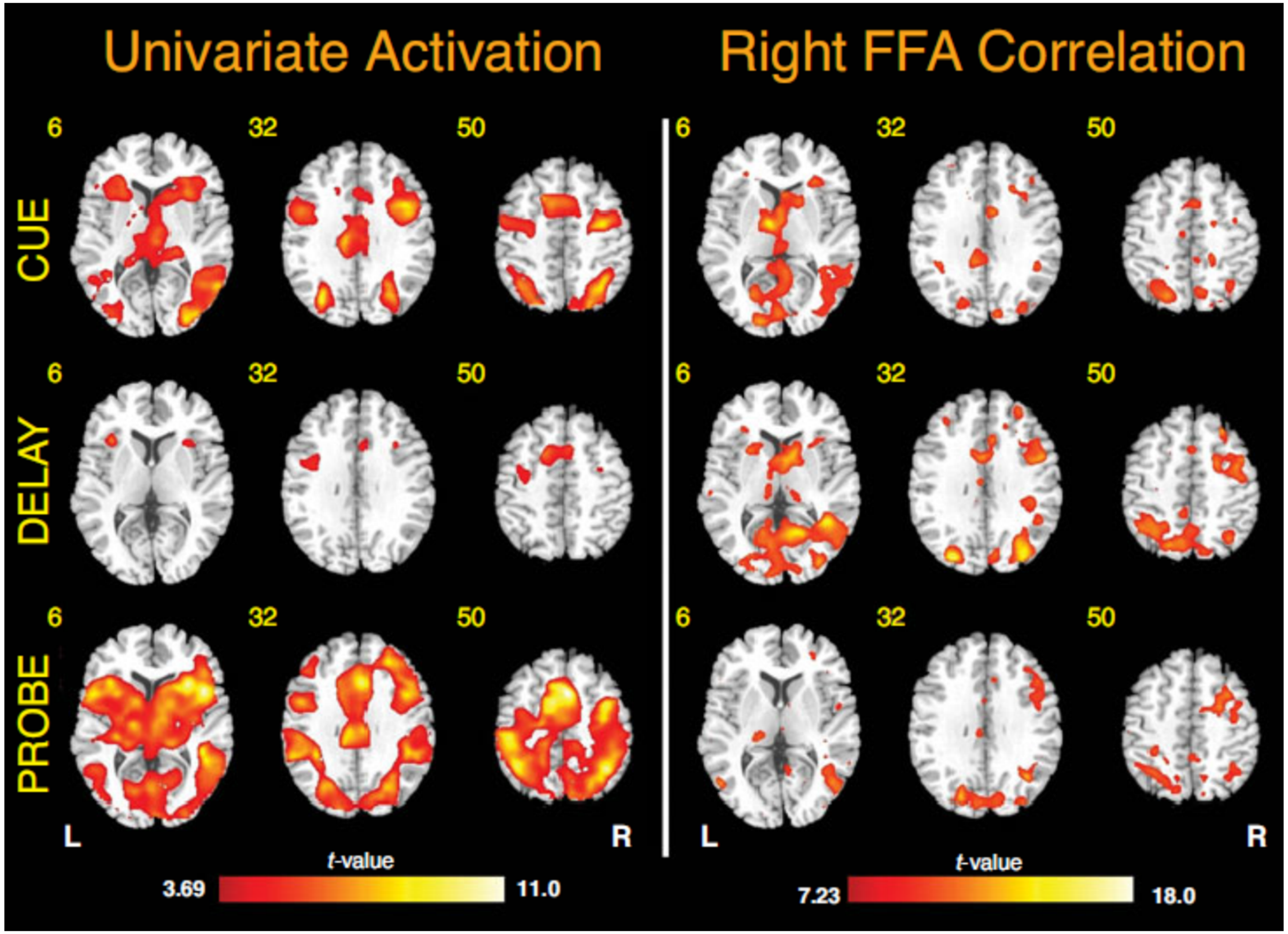
Beta-Series Regression (BSR)

- Estimate a separate Beta for each individual trial (as long as SOA > TR)



- Then correlate Betas (rather than TRs) across ROIs….

Beta-Series Regression (BSR)



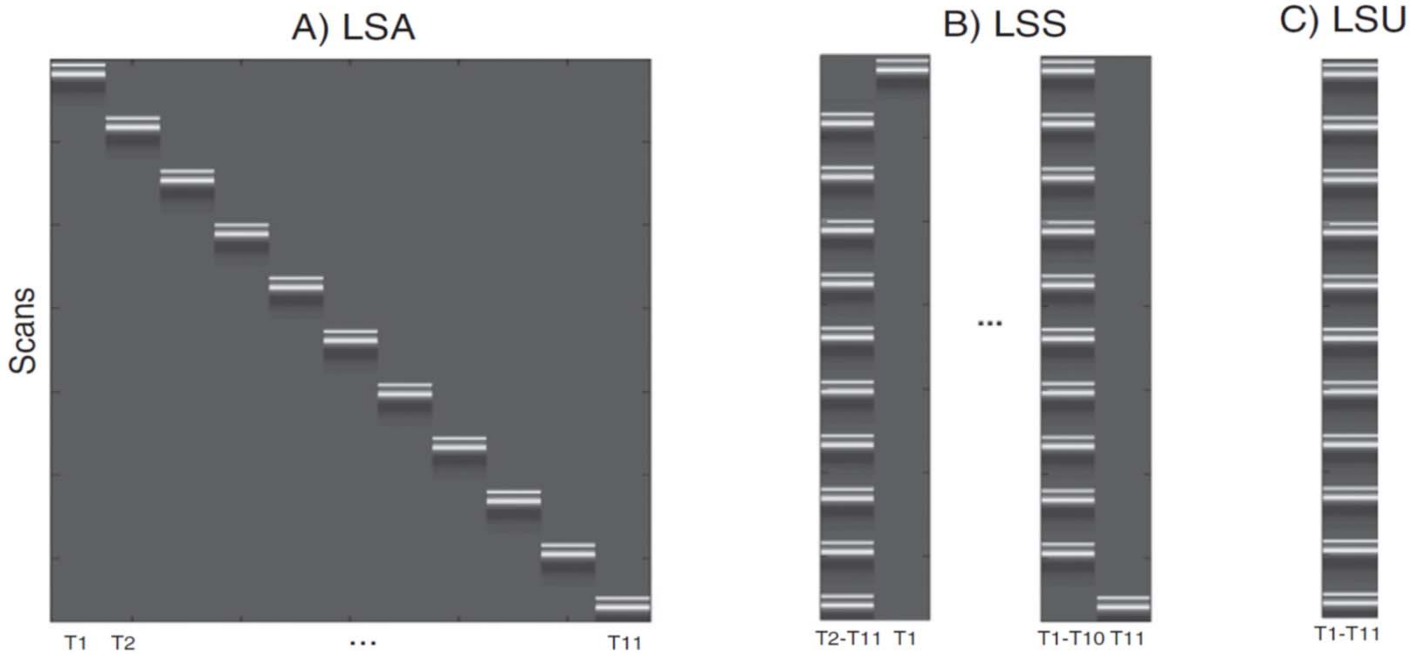
- FFA Beta Correlation reveals similarities, as well as differences, with standard activation analysis

Gazzaley et al. (2004) Cog. Aff. Beh. Neuro.

Beta-Series Regression (BSR)

- Some trial-to-trial correlation could be due to global variability in attention; safer to compare regression slopes across two or more conditions

- One problem with BSR is how to estimate single-trial responses for event-related designs with short SOAs:



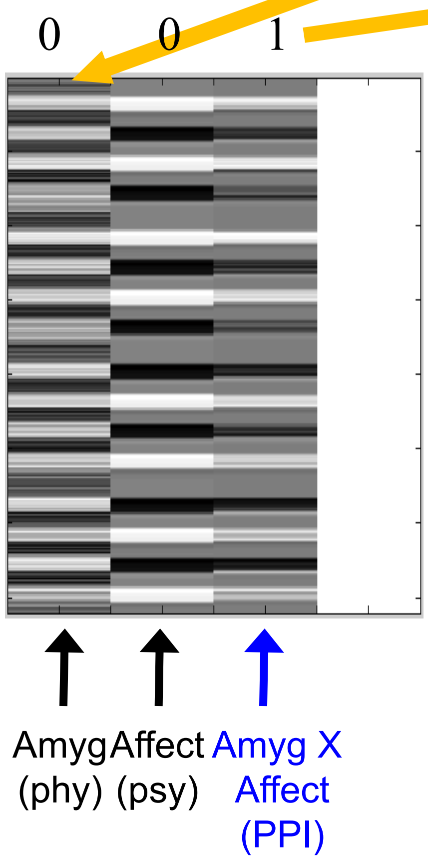
- LSS better when scan noise higher than trial variability (normally the case); LSA better when trial variability higher than scan noise

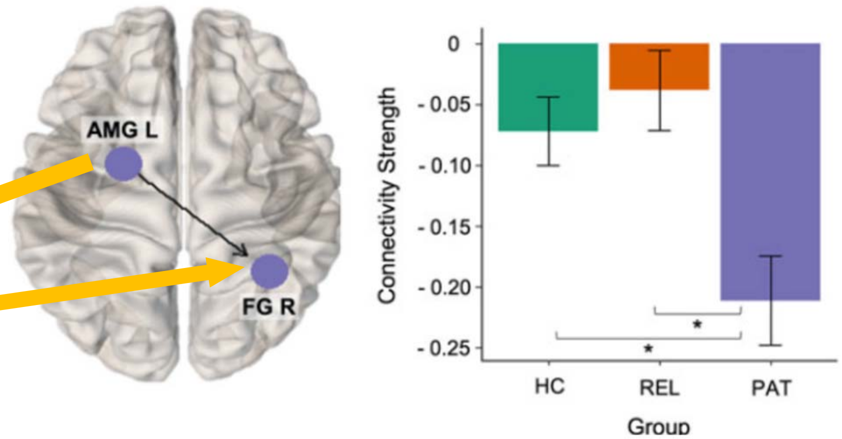
Abdulrahman & Henson (2016) Neuroimage

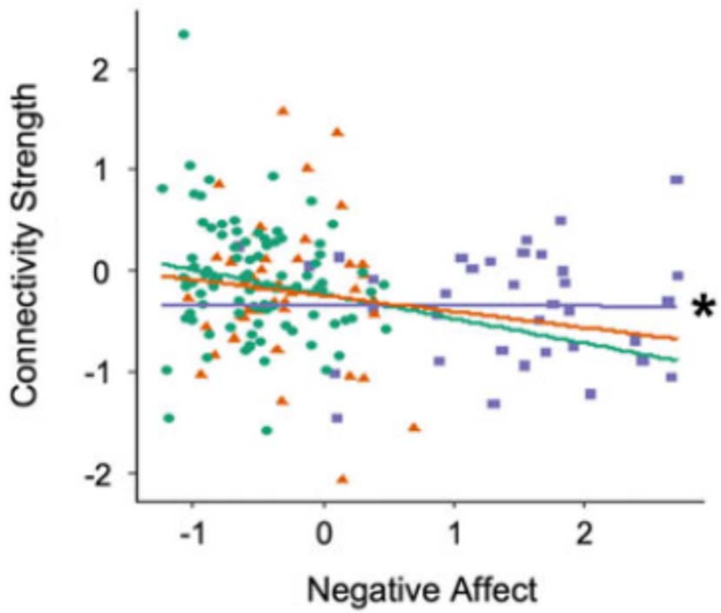
Psycho-Physiological Interactions (PPI)

Parametric, factorial design, in which one factor is psychological (eg affective state)

...and other is physiological *(viz. activity extracted from a brain region of interest)*



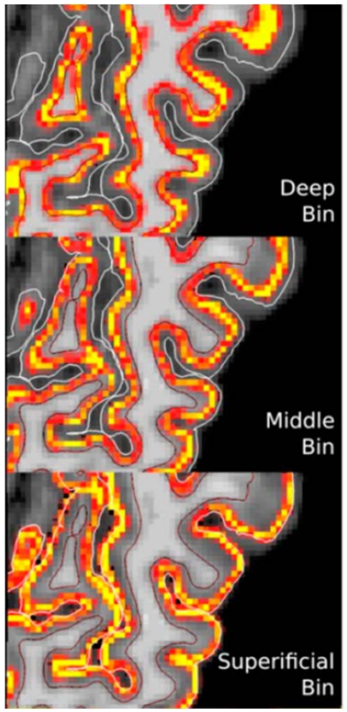




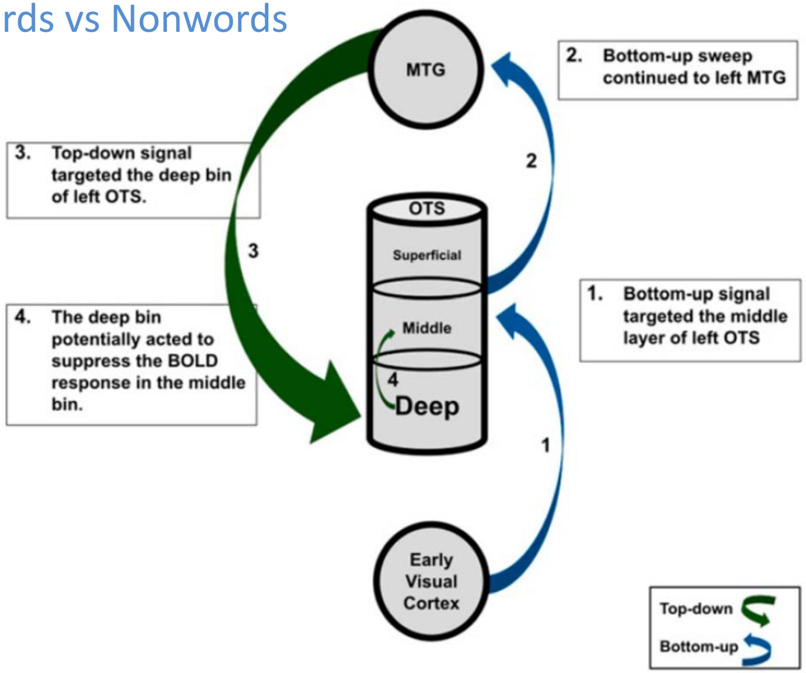
Sacu et al. (2023) Psychol. Med.

Laminar fMRI

- High-field (eg 7T) fMRI offers higher spatial resolution (<1mm), distinguishing layers (laminae) within cortex, eg deep vs middle vs superficial
- Top-down connections tend to target the deep and superficial layers, whereas bottom-up connections preferentially target the middle layer
- So can (indirectly) infer direction of information flow?



PPI for Words vs Nonwords



Sharoh et al (2019) PNAS

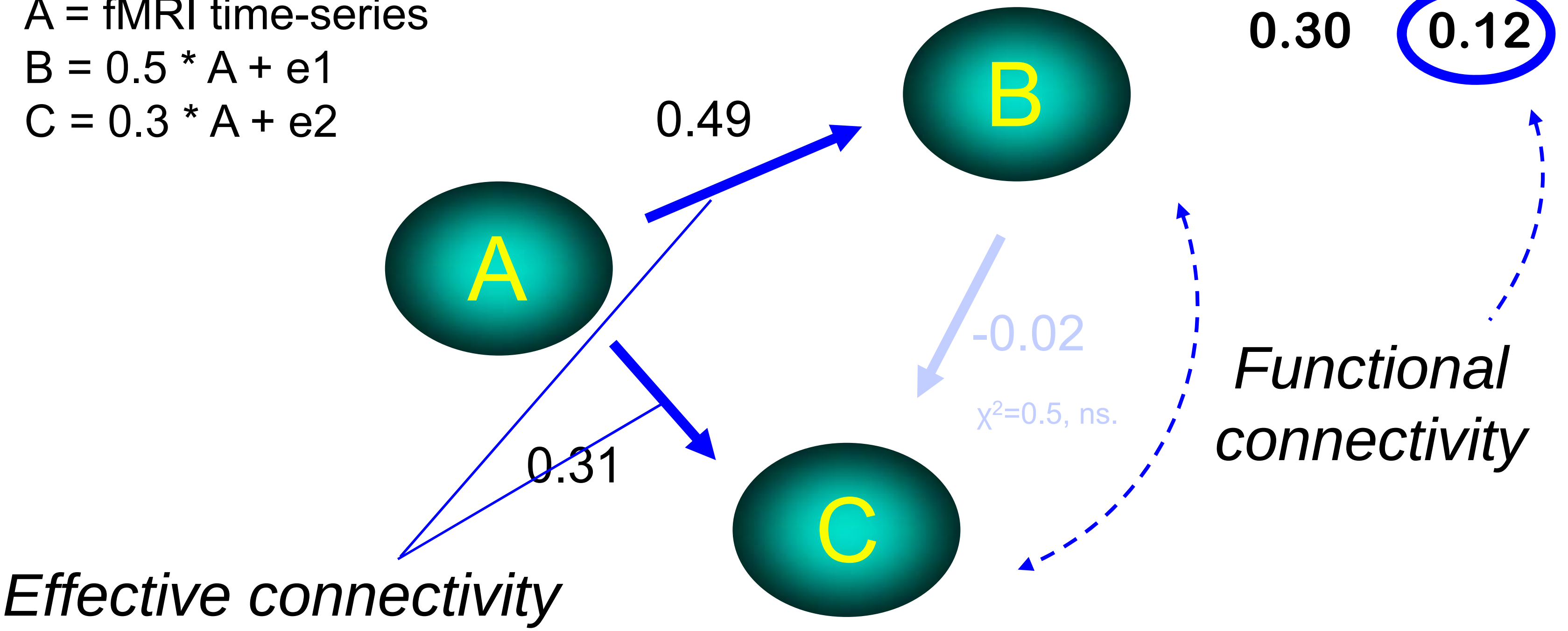
Functional vs Effective Connectivity

No connection between B and C,
yet B and C correlated because of
common input from A, eg:

A = fMRI time-series
B = 0.5 * A + e1
C = 0.3 * A + e2

Correlations:

A	B	C
1		
0.49	1	
0.30	0.12	1



(One can calculate *partial correlations* across whole connectomes,
but normally need to regularise their estimation…)

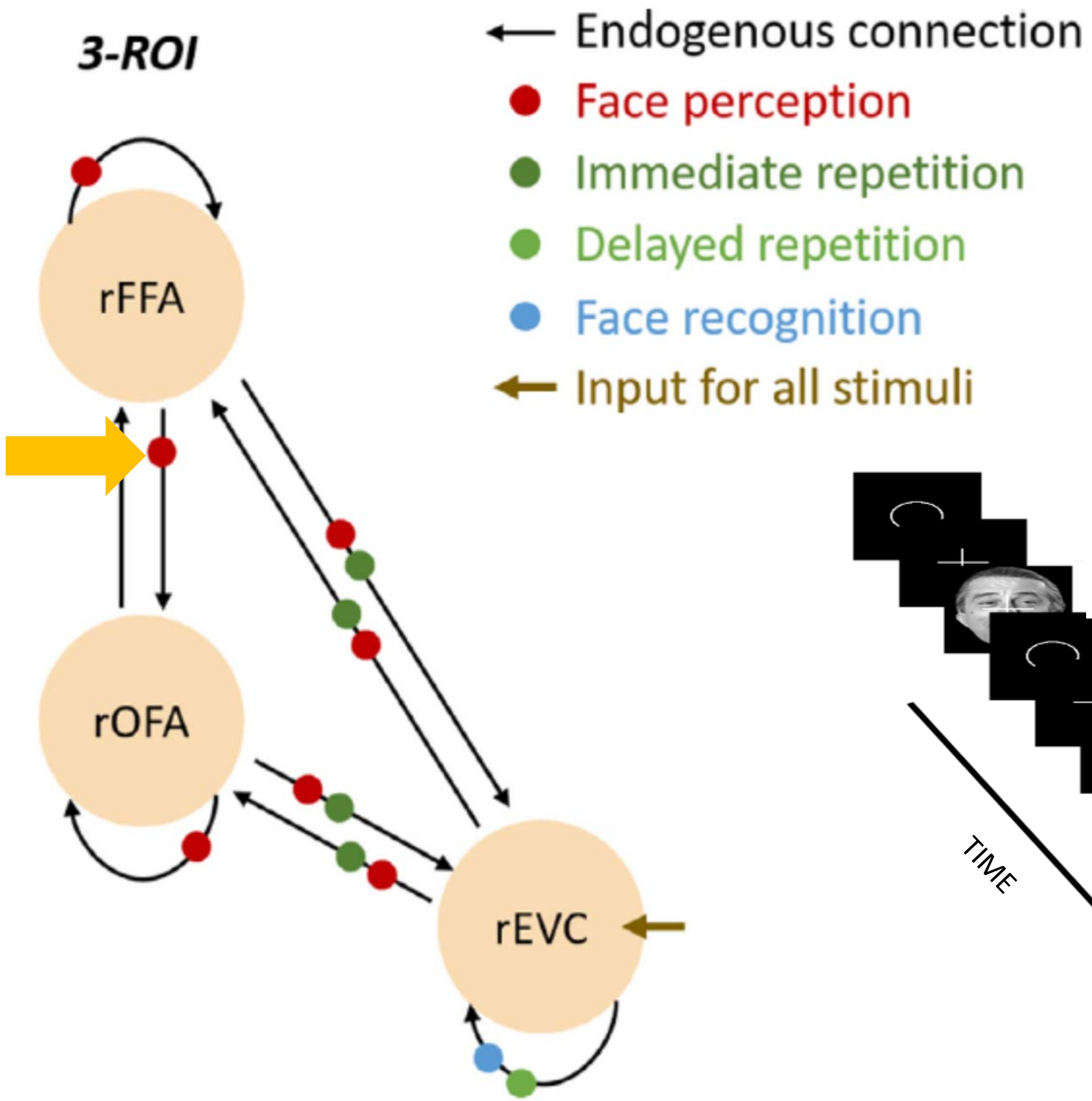
Effective Connectivity and Causality

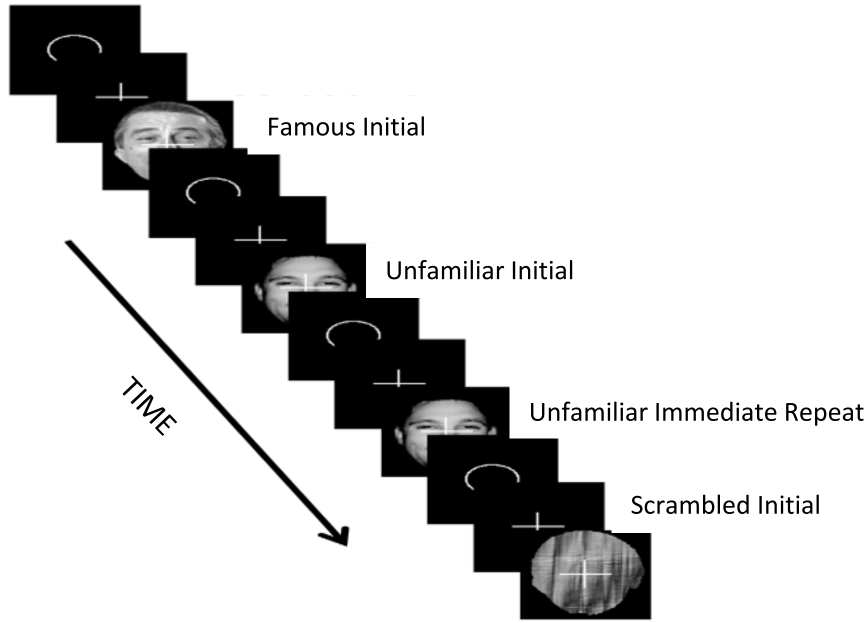
1. Direct experimental interventions (e.g, lesion, drugs)
2. Indirect experimental manipulations (e.g, PPI, DCM)
3. Network model inference (e.g, SEM, DCM)
4. Temporal precedence (e.g, Granger Causality, DCM)
5. …

Effective Connectivity and Causality

	Experimental modulation	Temporal/ Dynamical	Network model	Haemodynamic Model (for fMRI)
Correlation / ICA / PCA				
BSR / PPI	Y			
Granger		Y		
SEM			Y	
DCM	Y	Y	Y	Y

Dynamic Causal Modelling (DCM)





Lee et al. (2022) Neuroimage

Applications

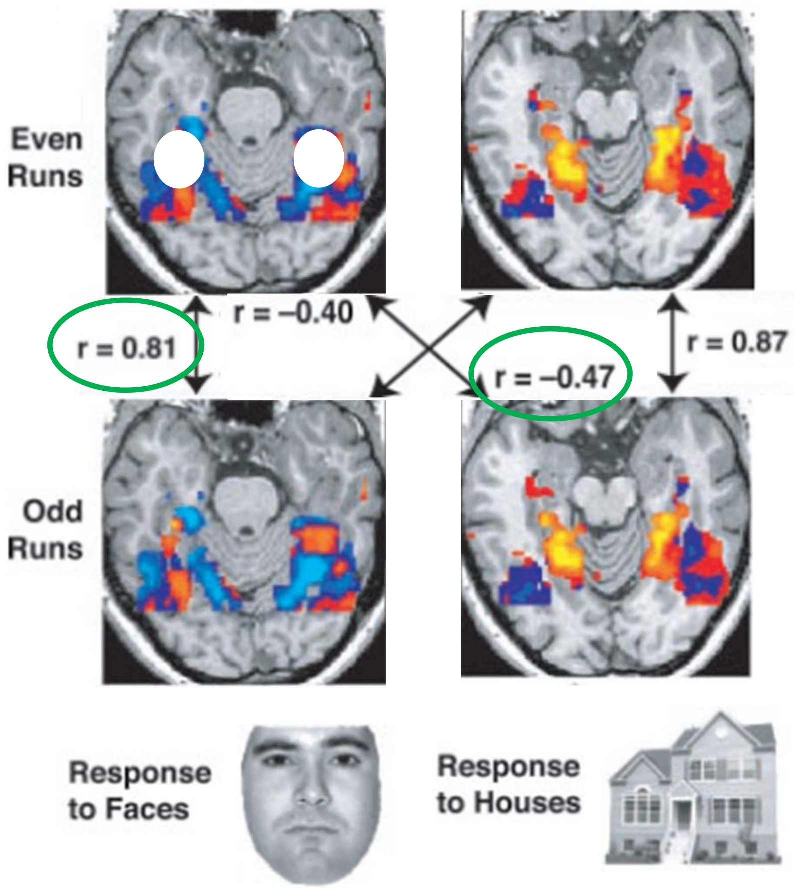
- Activation Analysis
  - Brain Mapping (mass univariate)
  - Subtraction logic (pure insertion)
  - Blocked versus Intermixed (Epoch vs Event-related)
  - Functional localisation; Forward and Reverse Inference

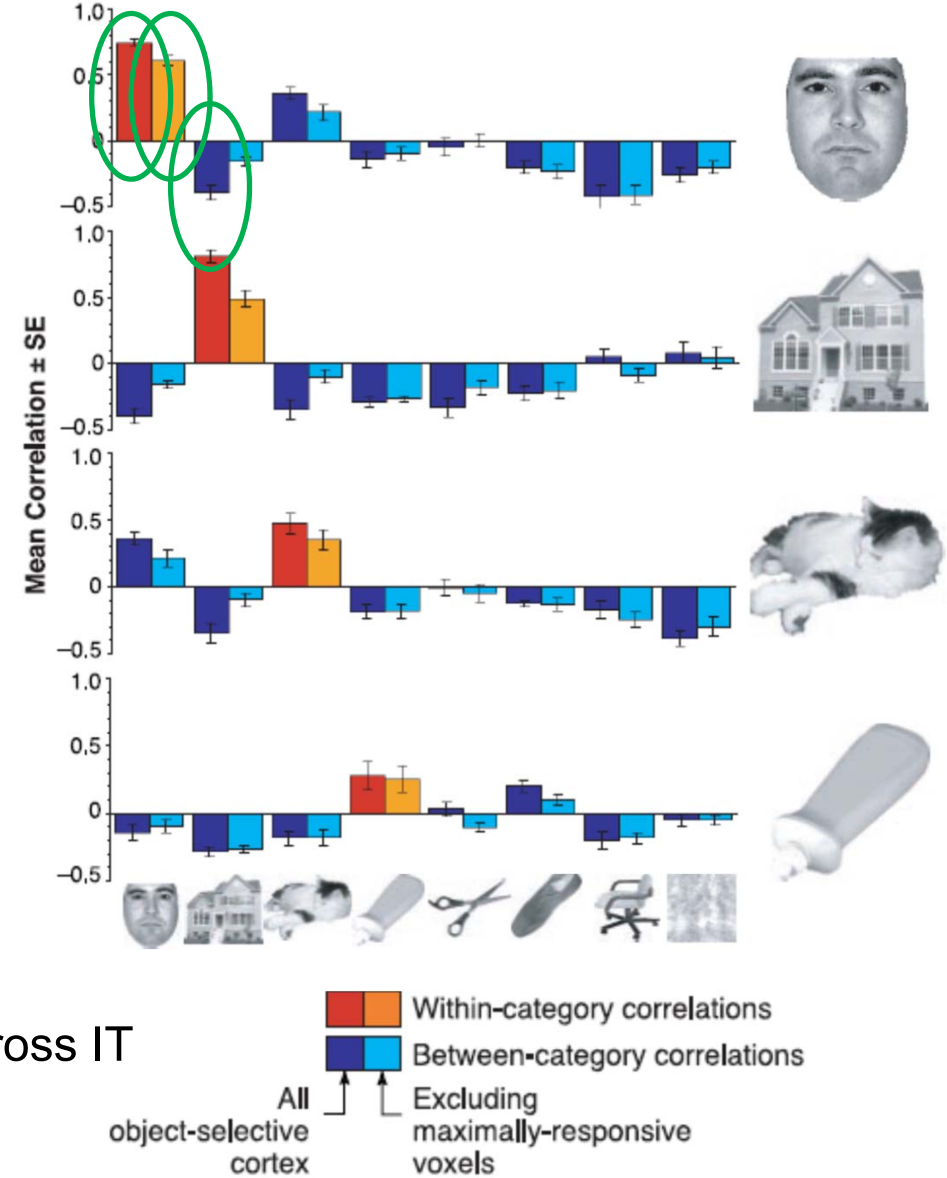
- Functional Connectivity
  - Matrix factorisation (Independent Component Analysis)
  - ROI-based connectomes, graph theory
  - Task-based (effective) connectivity

- Pattern (Information) Analysis
  - Multi-voxel pattern analysis (MVPA)
  - Representational Similarity Analysis (RSA)

Theme: Fusiform Face Area (FFA)

Information Outside Maximally-Responsive Regions

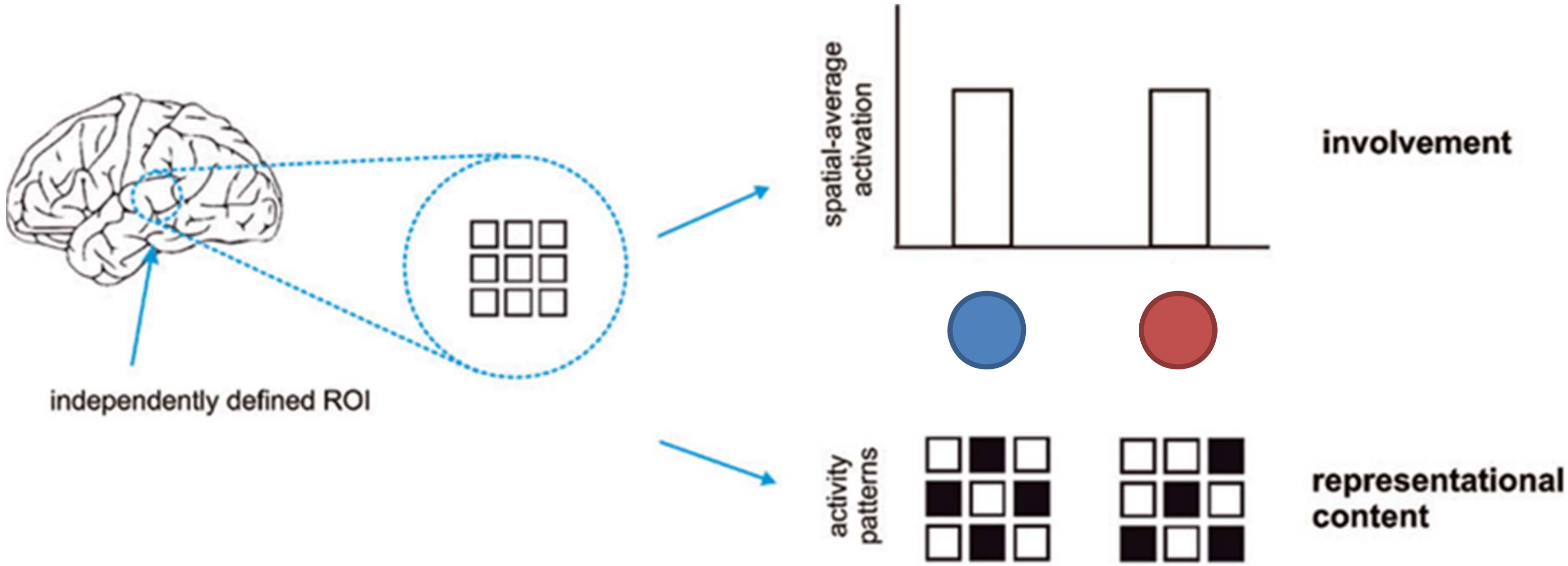




- Face representations distributed across IT (not just FFA)

Haxby et al (2001) Science

Multi-Voxel Pattern Analysis (MVPA)

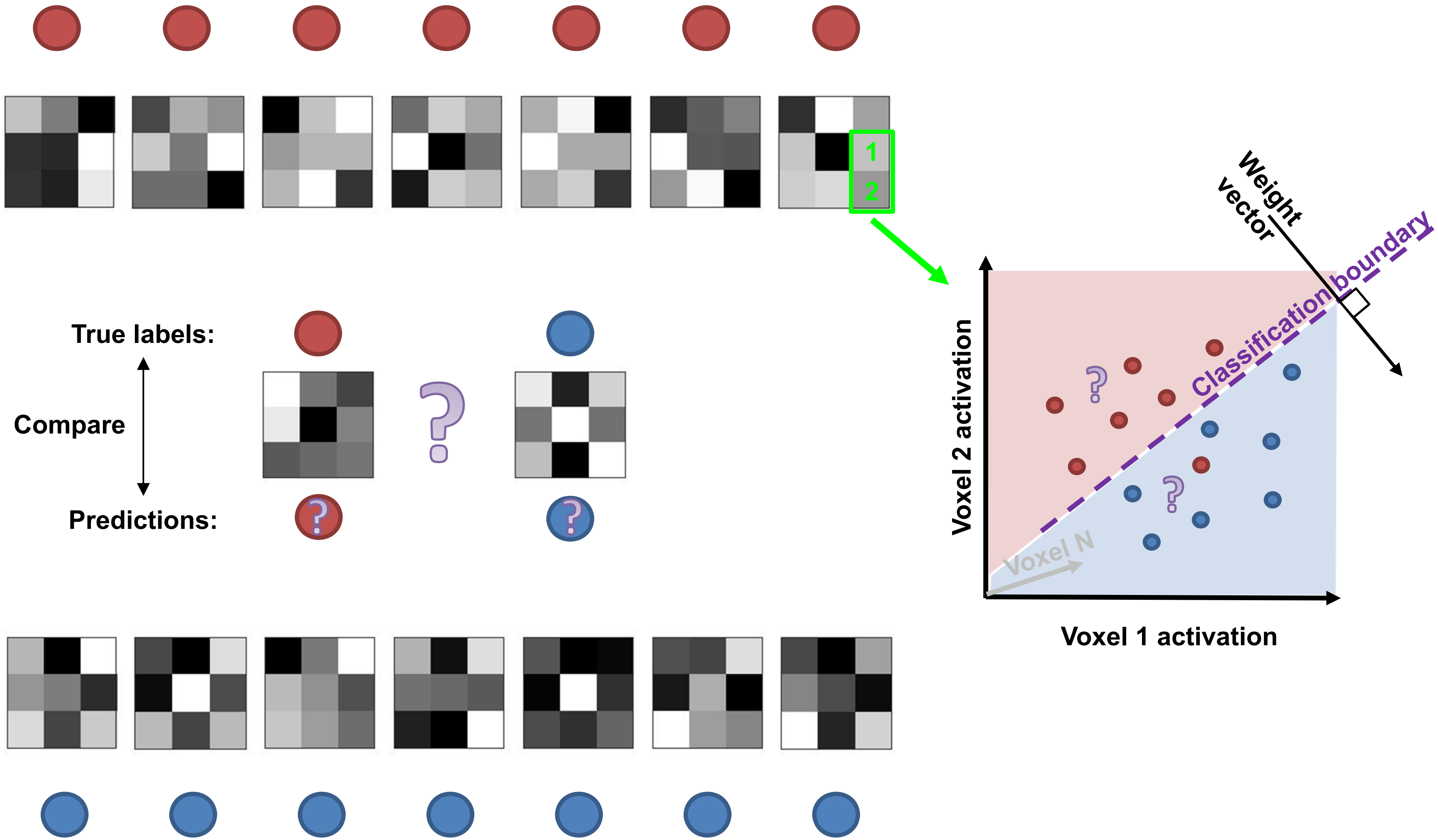


- If we can classify experimental conditions based on activity patterns (better than chance) …
- … then the activity pattern carries information about the experimental conditions

Mur et al. (2009) Soc. Cog. Aff. Neuro.

How does MVPA work?

classification example



Thanks to Danny Mitchell

Decoding Model

- Linear classifiers like a "reversed" (logistic) GLM:

$$X(t) = f(y(t,v) \times \beta(v)) + \varepsilon(t)$$

y (measured) = Activations for each trial *t* in voxel *v*

X (specified) = "Design matrix" coding condition label of each trial *t*

β (estimated) = "Betas" (weights) for each voxel *v*

ε (estimated) = residual error for each trial *t*

- Normally *v>t,* so regularisation needed, e.g. minimise:

$$\|\varepsilon\|_p + \lambda\|\beta\|_p$$

e.g., Ridge Regression (*p*=2), LASSO (*p*=1), Elastic Net, λ = regularisation parameter…

- (Decoding model actually more "causal", e.g., predict behaviour from all voxels / ROIs in brain…)

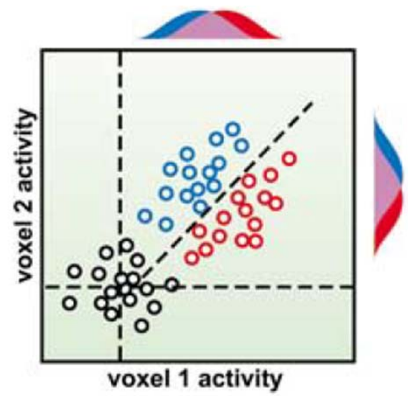
Why do MVPA?

- Scientific questions
  - What information does a brain region represent? (classification: are patterns reliably different?)
  - In what format does it represent the information? (RSA: what are distance relations between patterns?)

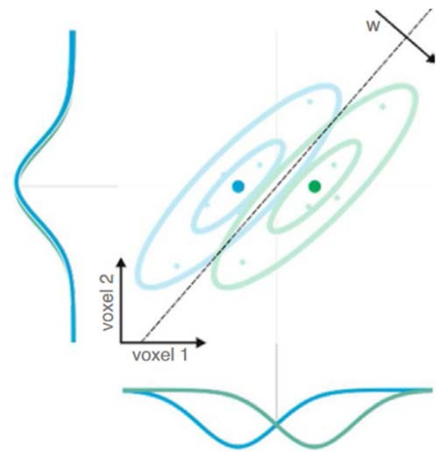
- Practical uses
  - Mind Reading? Disease status?
  - **More sensitive than univariate analysis…**

Why can MVPA be more sensitive?

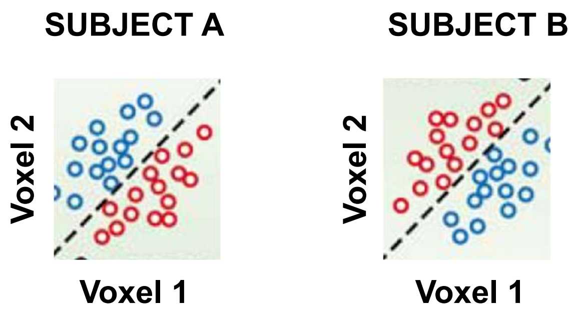
- Uses covariance between voxels



- Suppresses correlated noise

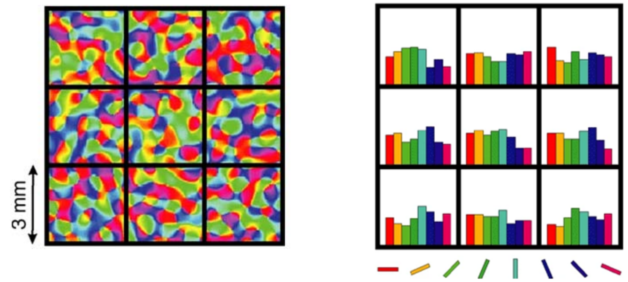


- Allows for spatial hetereogeniety across individuals



Physiology: fine or coarse patterns?

- Original view: biased sampling of sub-voxel patterns



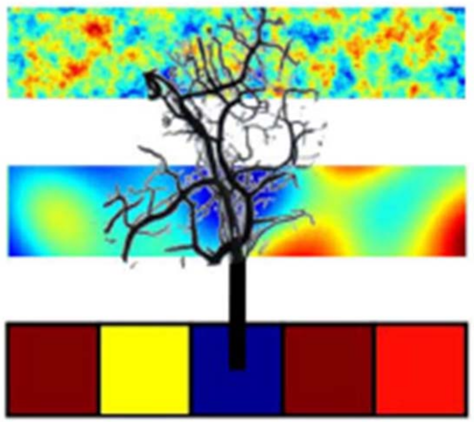
Kamitani & Tong (2005) Nat. Neuro.

- Later spatial filtering experiments:
  - Smoothing can actually improve classification (Op de Beeck, 2010, Neuroimage)
  - Therefore some patterns may be coarse

- Much evidence that information is represented at multiple spatial scales
  (e.g. FFA/PPA – see Haxby et al slide earlier)

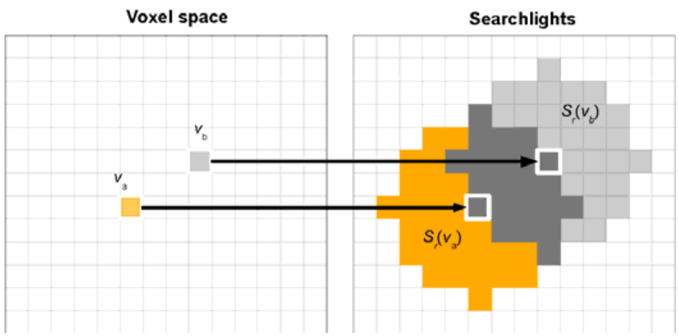
- Veins as multipronged sensors? (We sample from veins, which sample from neurons…)

- Practical consequences, e.g., smoothing, motion, search-light size



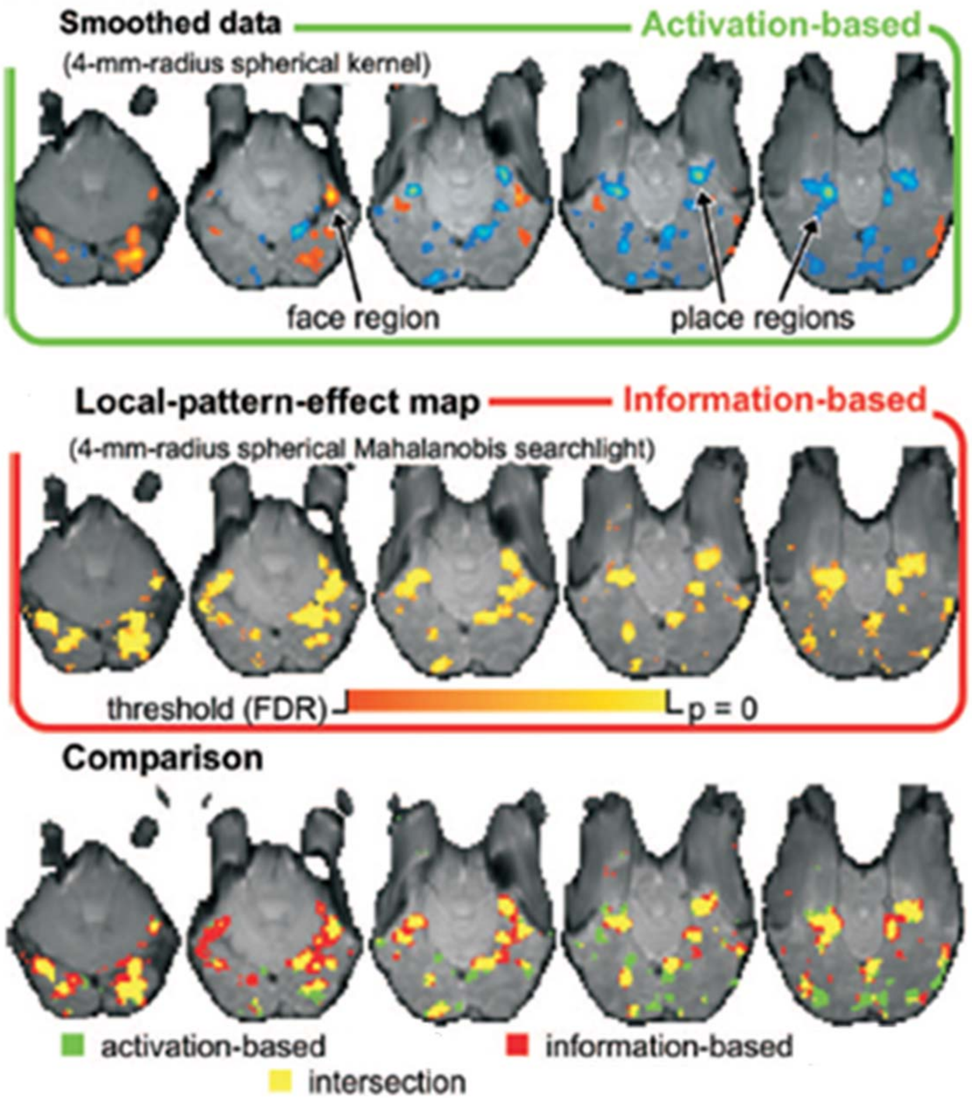
Kriegeskorte et al. (2010) Neuroimage

Localisation: Information-Mapping!



3 voxel radius
searchlight

- One can map out information!



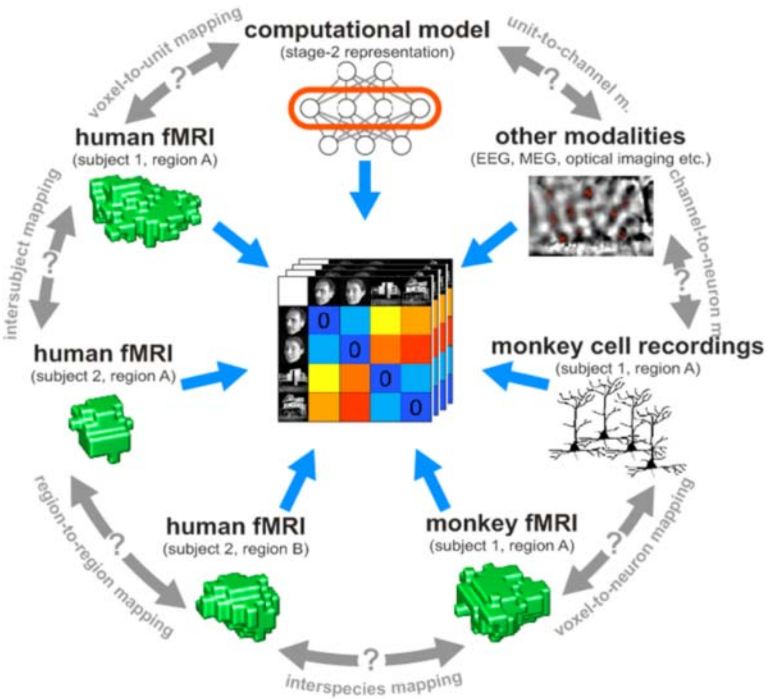
Kriegeskorte et al (2006) PNAS

Caveats

- Just because we can decode information in one brain region, does not mean that other brain regions also can (or use this information)!
- Eg can decode Faces vs Houses in V1 (EVC), but neurons there do not respond to these categories specifically

Representational Similarity Analysis (RSA)

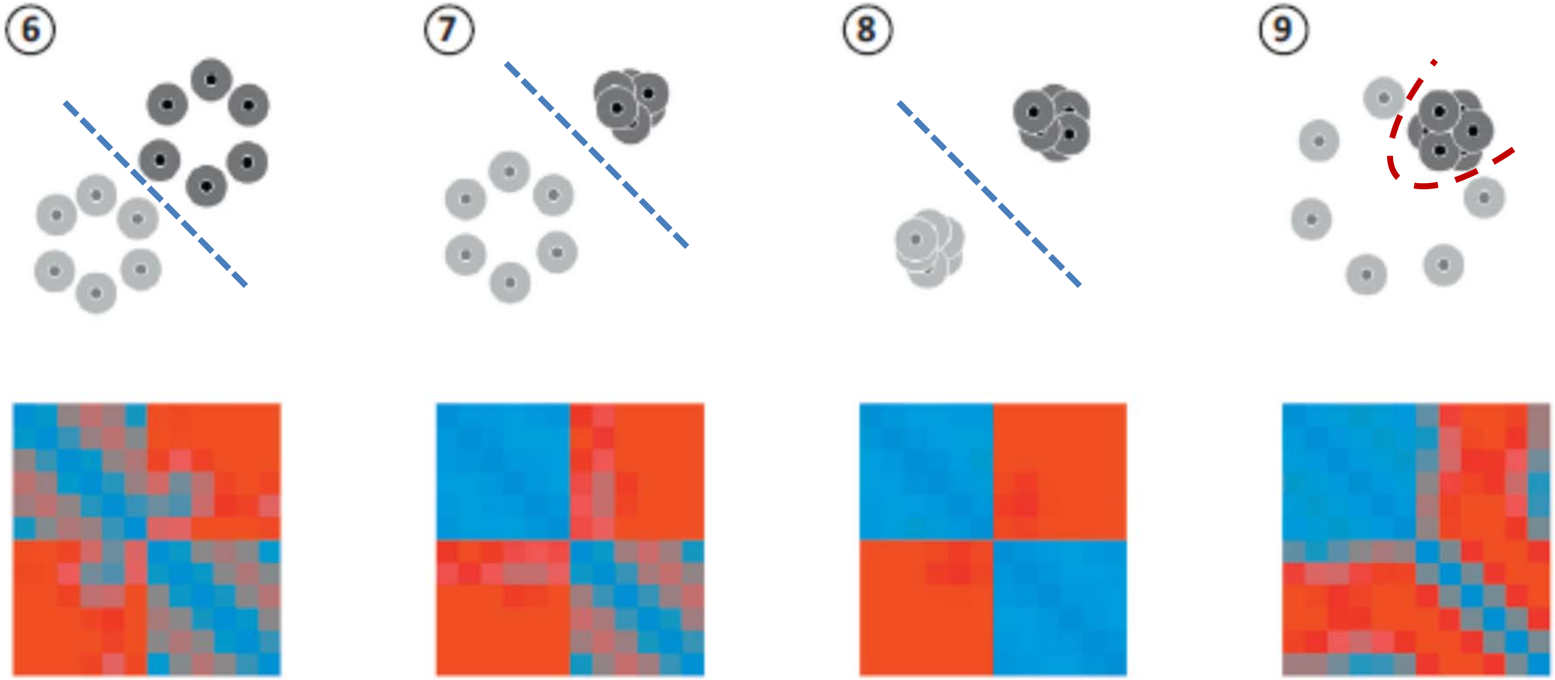
- Quantifies neural relationships between conditions in an abstract space (e.g, space of stimuli), derived from, but no longer dependent on, the measurement format (cf. kernel trick in machine learning)
- Allows different representational structures to be distinguished, and compared in common format



Kriegeskorte et al. 2008 Front. Syst. Neurosci.

Representational Similarity Analysis (RSA)

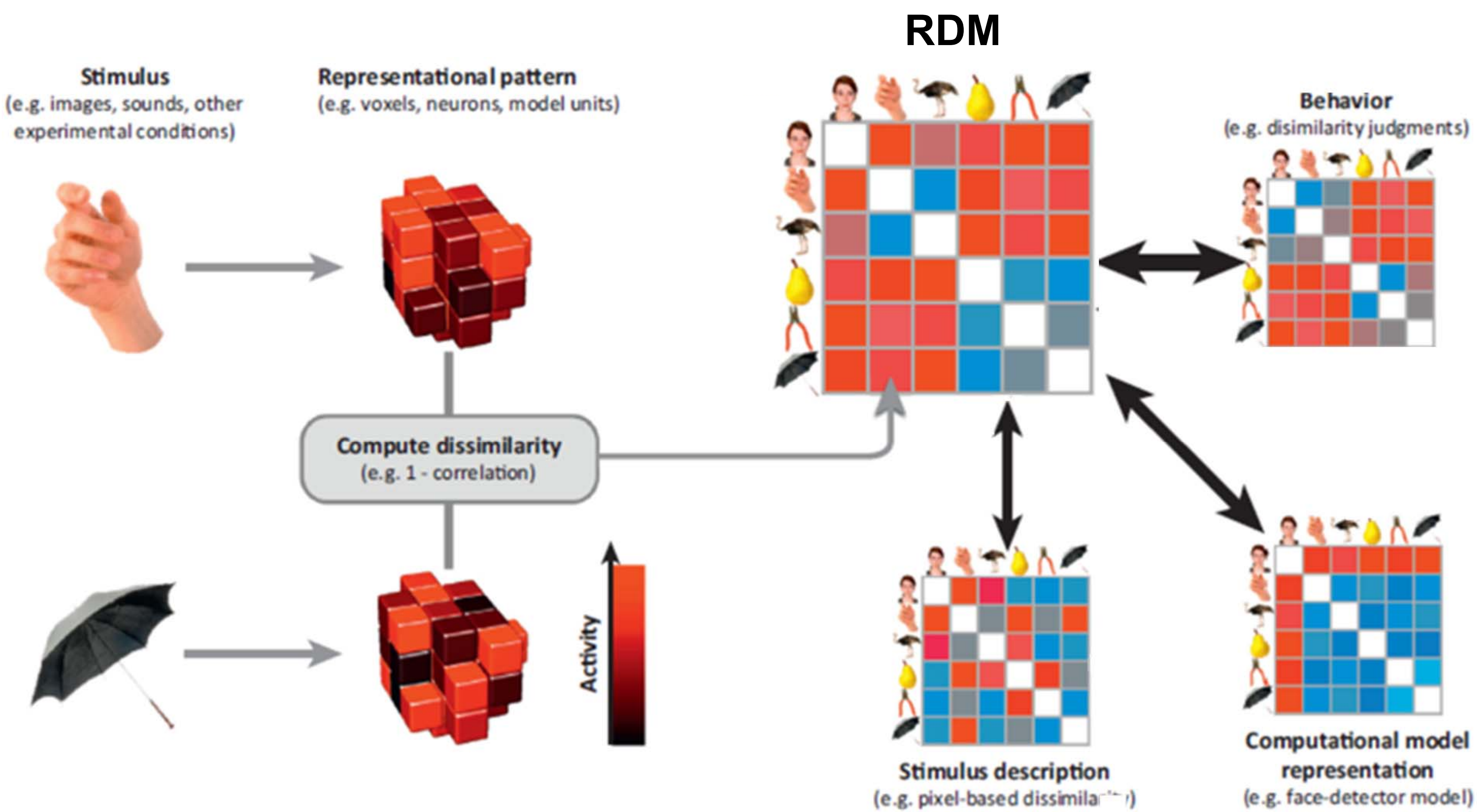
Equivalent classification, different representations:



Kriegeskorte & Kievit (2013) Trends. Cog.

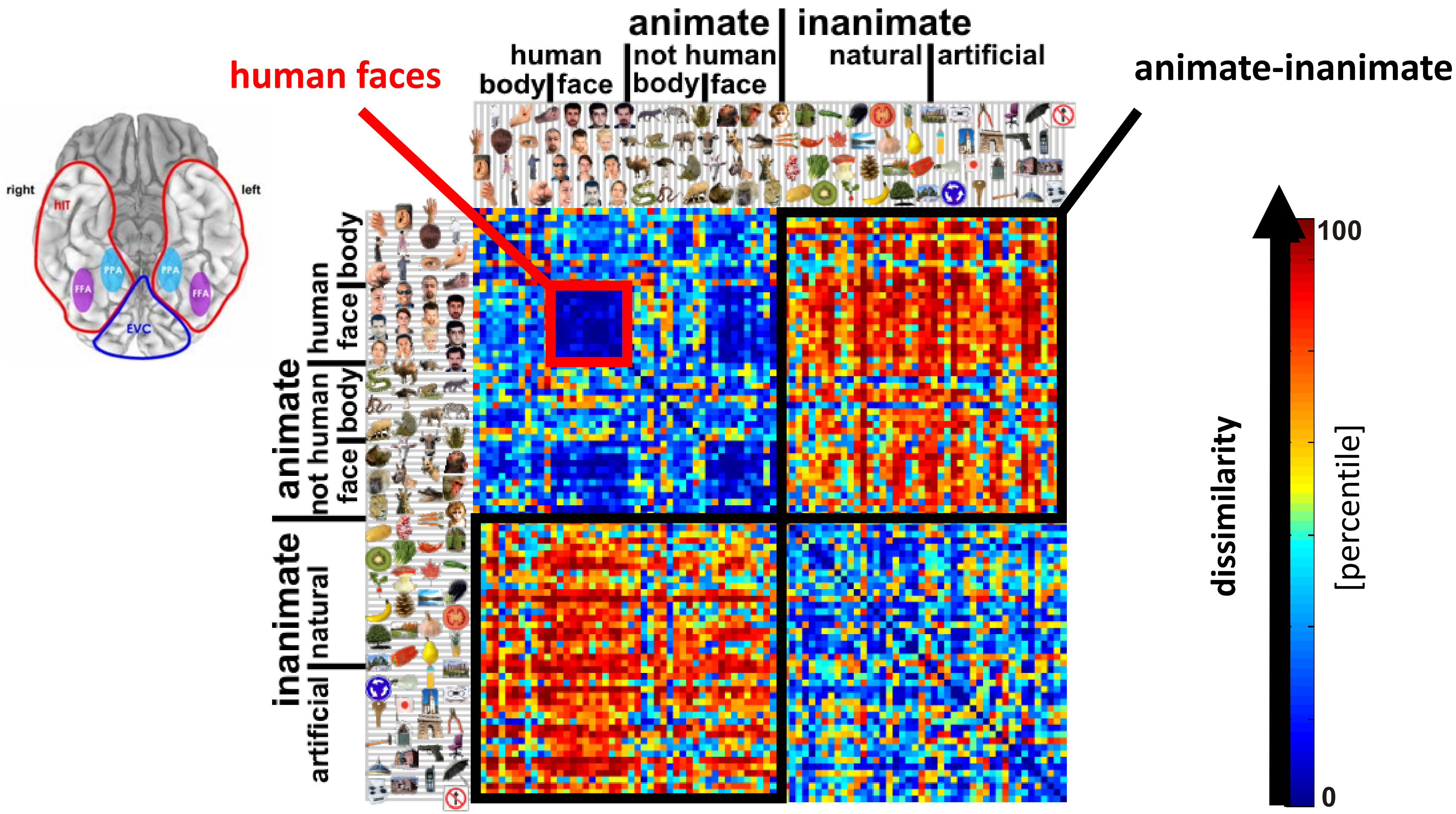
Representational Dis-similarity Matrix (RDM)

Assuming a space of N stimuli, the size/nature of the original data space is irrelevant, e.g. whether over P voxels, or Q participants, or R units of a Neural Network Model…



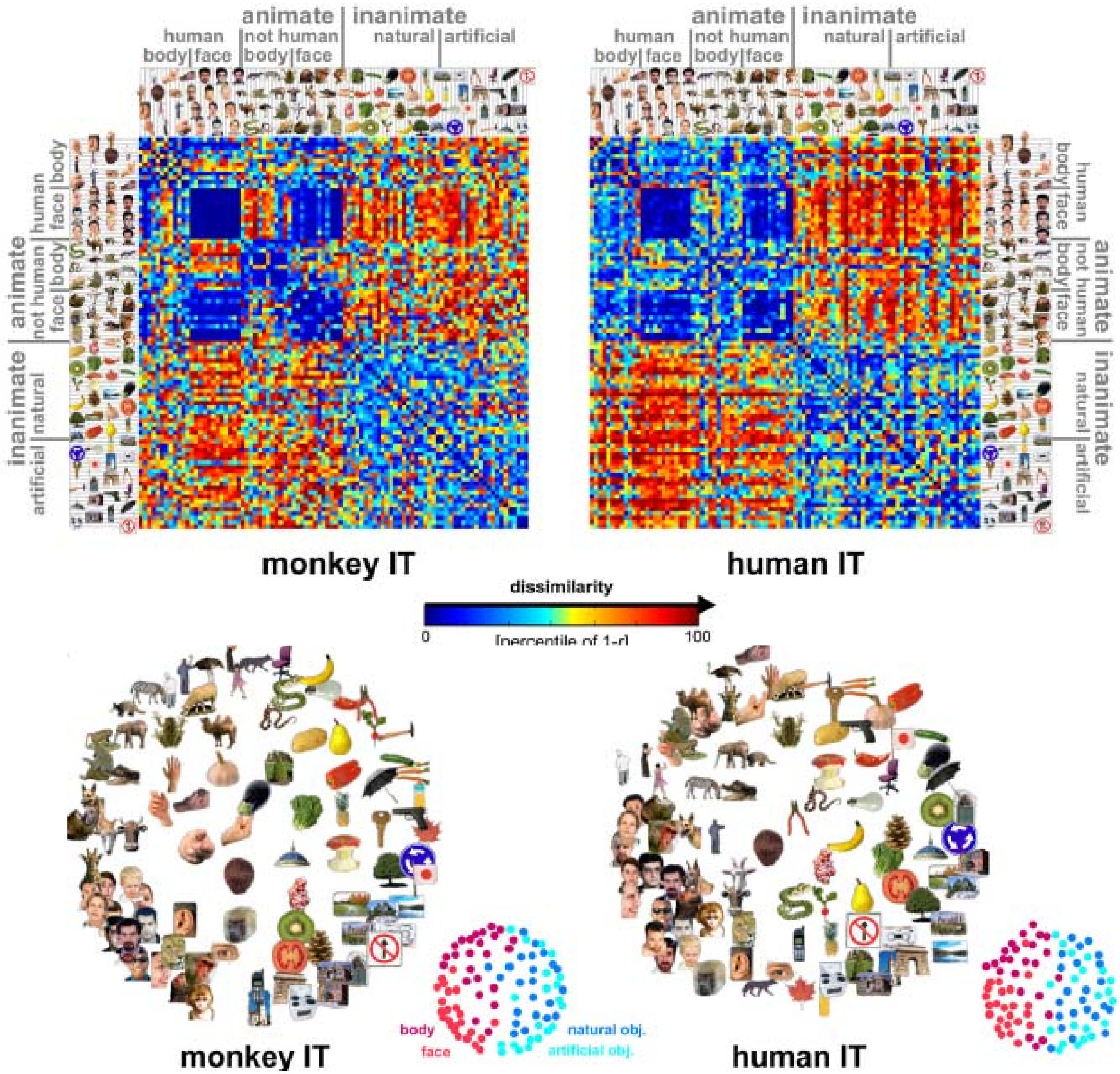
Kriegeskorte & Kievit (2013) Trends. Cog.

RDM for Human InferoTemporal (IT) cortex



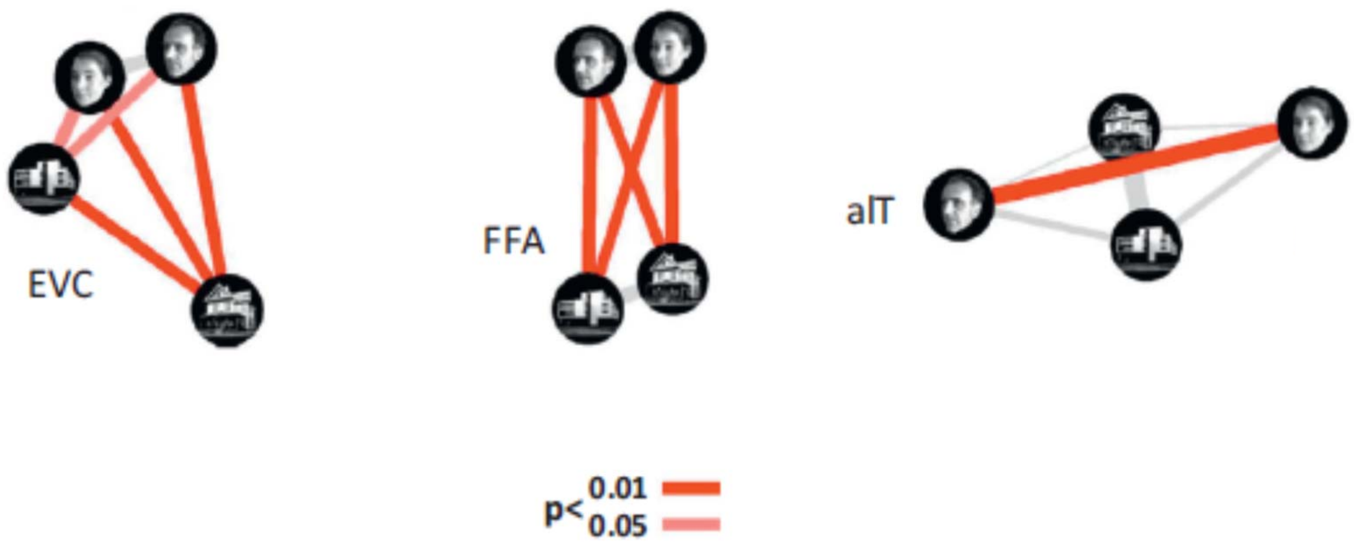
Kriegeskorte et al (2008) Neuron

RDMs across Species



Kriegeskorte et al (2008) Neuron

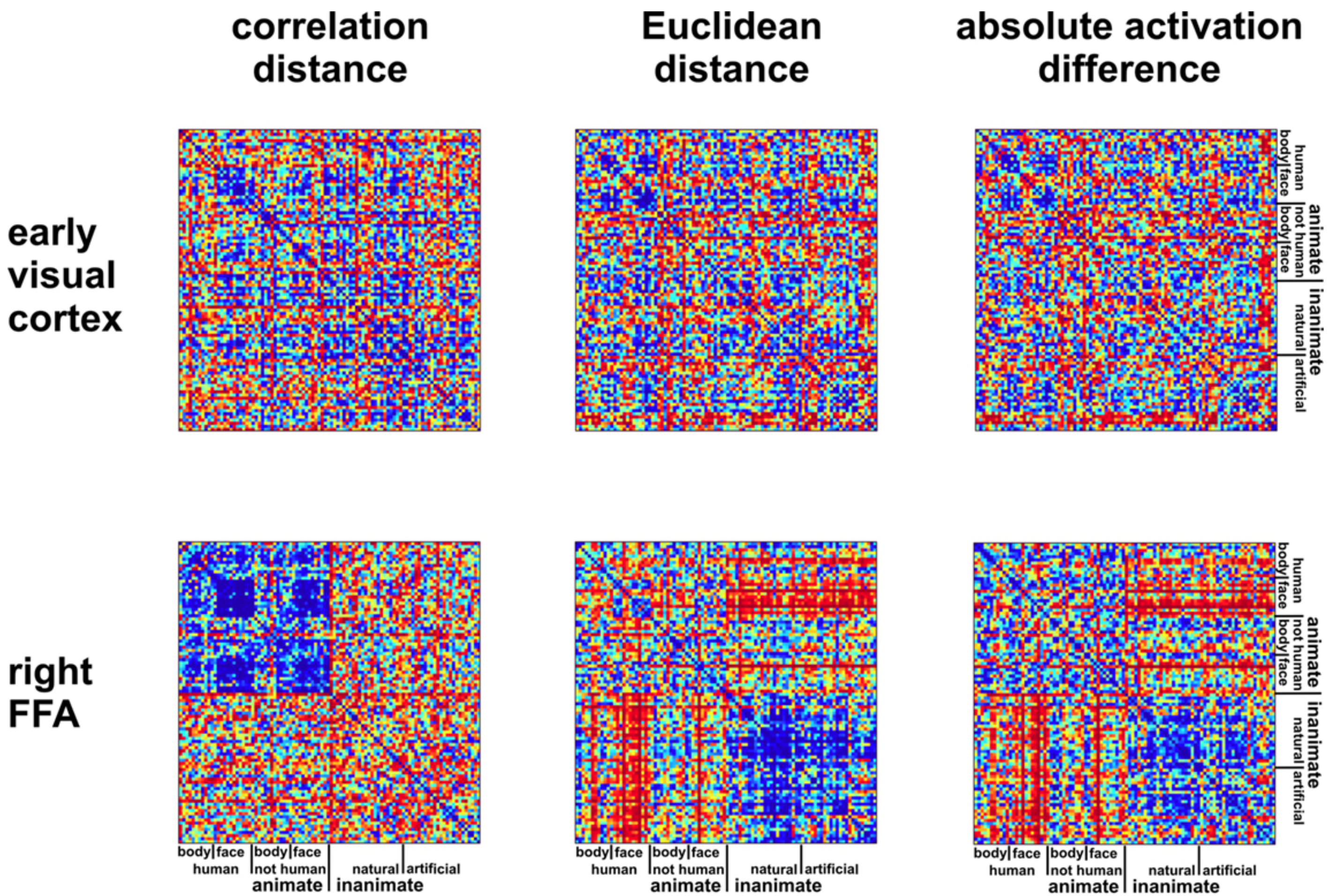
RDMs across Regions (human fMRI)



- Early visual cortex (EVC) distinguished all stimuli
- FFA distinguished Faces from Houses
- Anterior IT (aIT) distinguished facial identities

Kriegeskorte et al (2007) PNAS

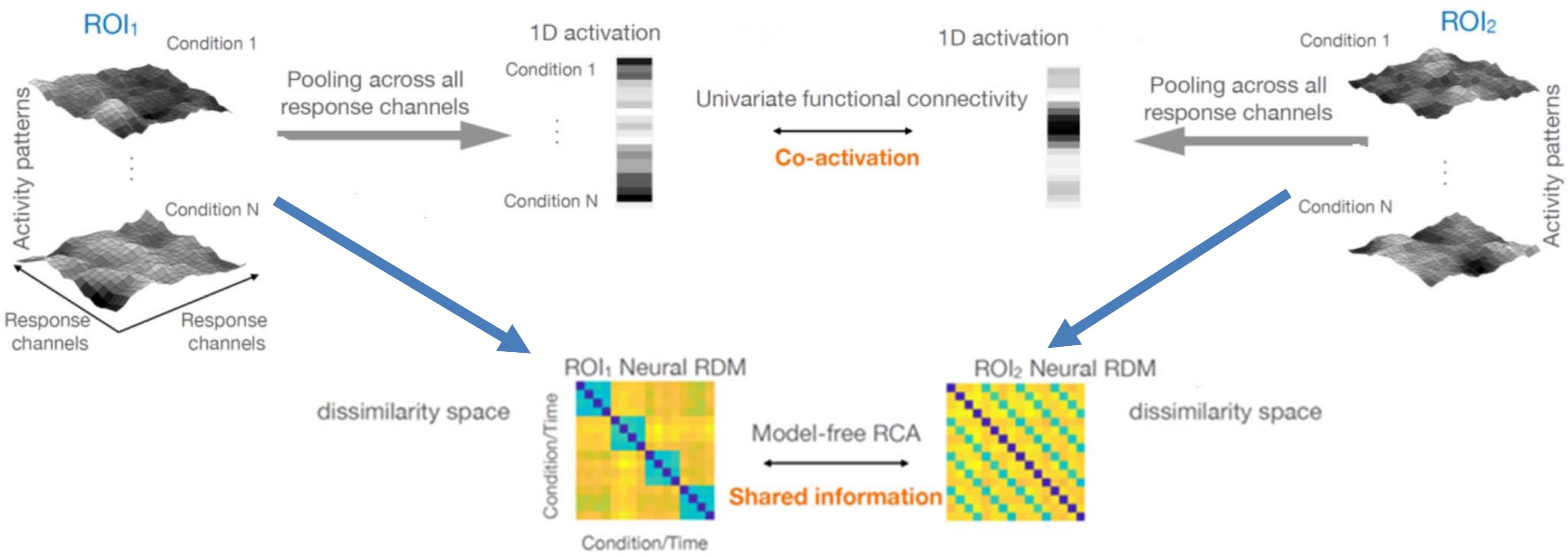
Which measure of Dis-Similarity?



- More sophisticated, eg cross-validated Mahalanobis distance

Kriegeskorte et al. 2008 Front. Syst. Neurosci.

Multi-dimensional (RSA) Connectivity



Karimini-Rouzbahani et al (2022) Front. Neuro.

THE END!

(Thanks to Danny Mitchell for some MVPA slides)