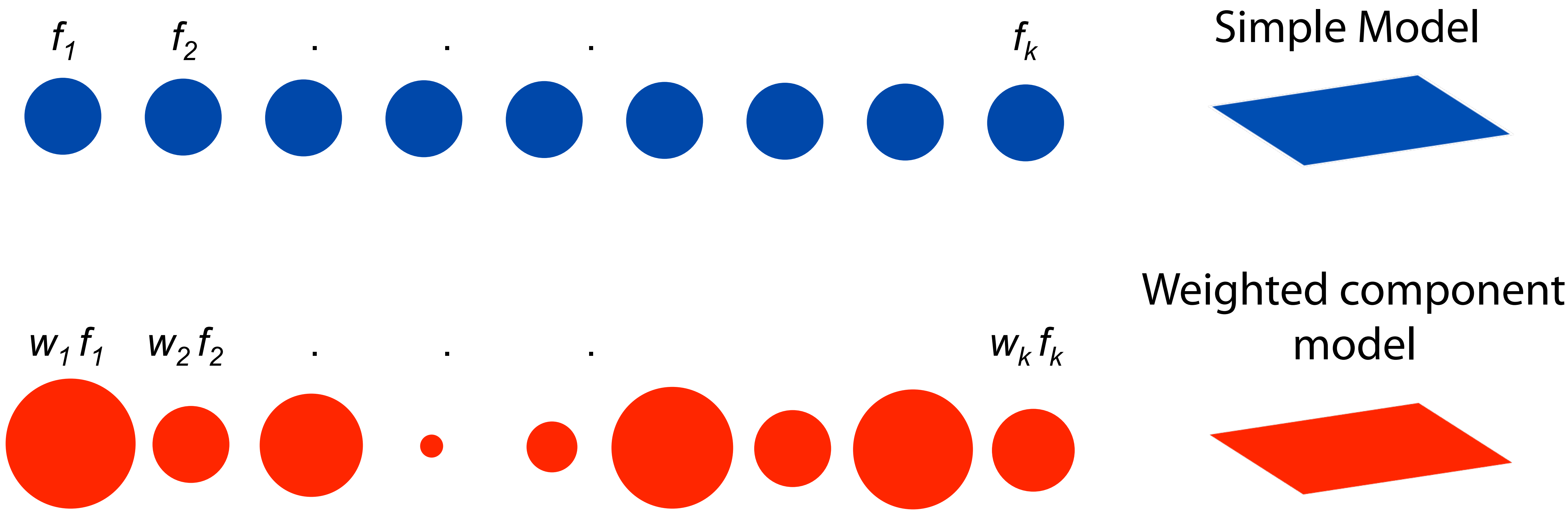
# Weighted Representational Component Models I

Jörn Diedrichsen

Motor Control Group
Institute of Cognitive Neuroscience
University College London

# Weighting of different features




Examples:

- Different aspects of objects: color, shape
- Different aspects of movements: force, sequence, timing
- Different layers of a computation vision model

## Overview

# Representational component modelling

- Features or groups of features (components) can be differently weighted
- Component weights can be estimated from the data
- Inferences can be made directly on component weights

    or

- Model fit can be assessed using cross-validation

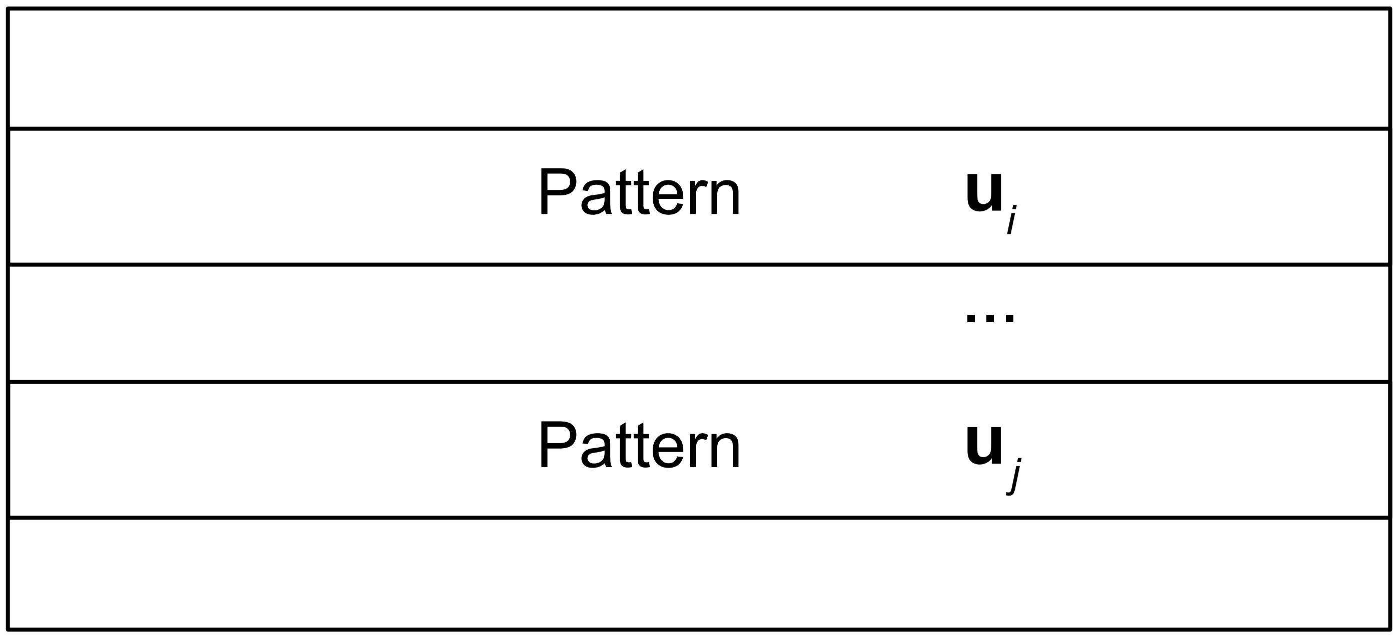
## *Overview*

- Covariances and Distances
- Features and representational components
- Factorial models (MANOVA)
- Linear representational models
- Nonlinear representational models
- Summary

# Covariances and Distances

P Voxels

K Conditions




Distances (LDC)

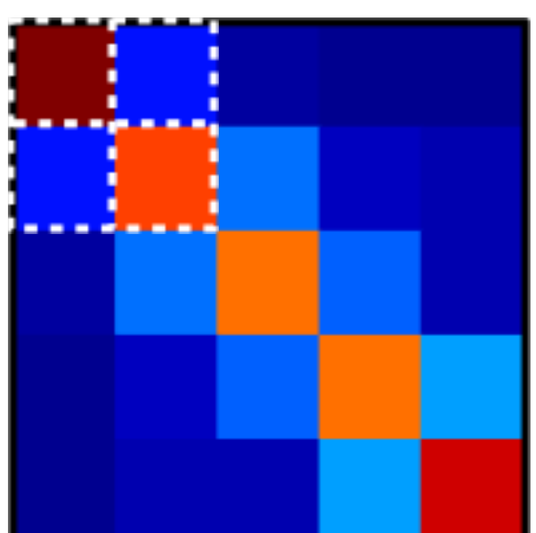
$$d_{i,j} = (\mathbf{u}_i - \mathbf{u}_j)(\mathbf{u}_i - \mathbf{u}_j)^T$$
$$= \mathbf{u}_i\mathbf{u}_i^T + \mathbf{u}_j\mathbf{u}_j^T - 2\mathbf{u}_i\mathbf{u}_j^T$$

Inner product of pattern

$$\mathbf{G}_{i,j} = \mathbf{u}_i\mathbf{u}_j^T$$

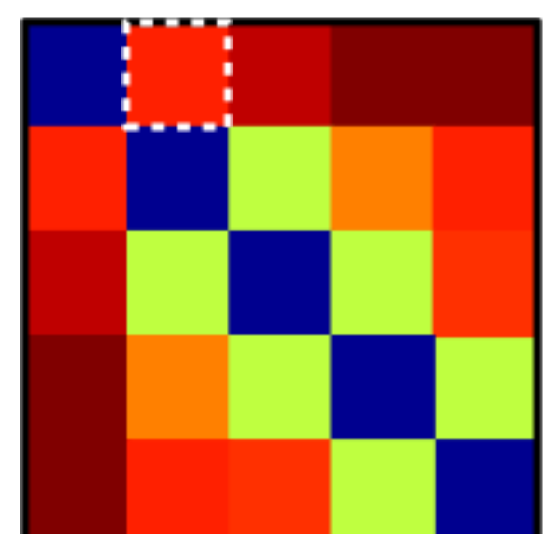
Pattern covariance

$$\mathbf{G} = \mathbf{U}\mathbf{U}^T$$



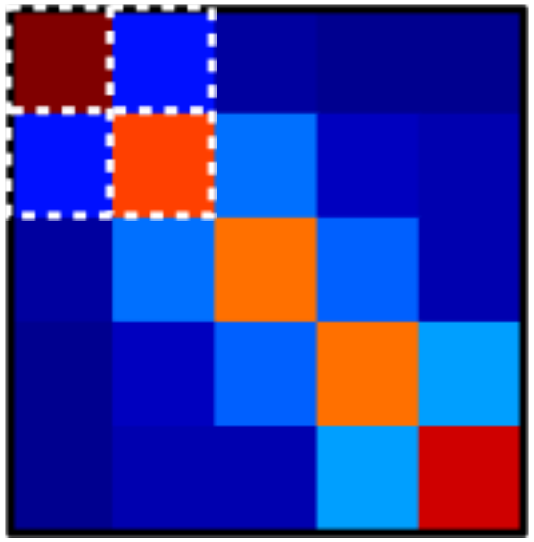
Patter distance

$$d_{ij} = \mathbf{G}_{ii} + \mathbf{G}_{jj} - \mathbf{G}_{ij} - \mathbf{G}_{ji}$$



# Covariances and Distances

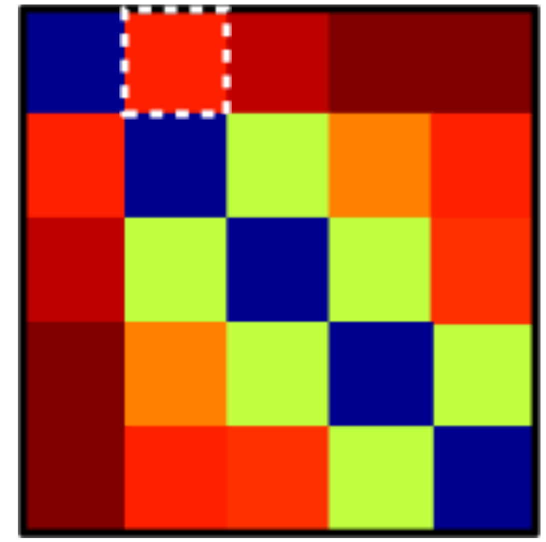
Pattern covariance matrix
(inner product matrix)



$$\mathbf{G} = \mathbf{U}\mathbf{U}^T$$

Contains baseline information

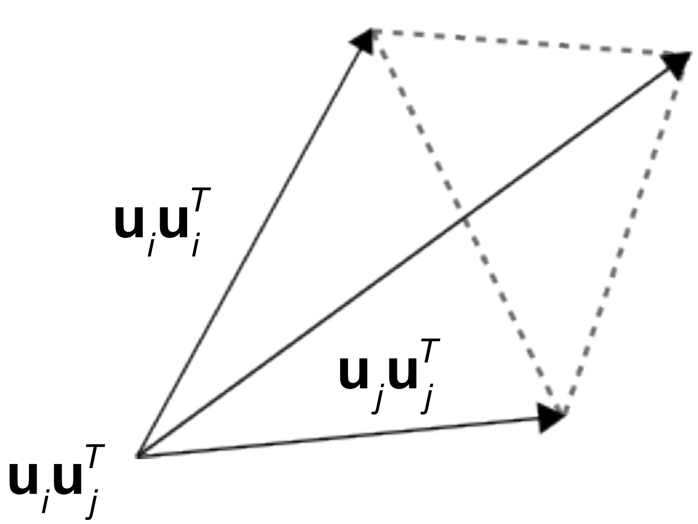
Mahalanobis-distance
matrix



$$\mathbf{D}_{i,j} = \mathbf{G}_{i,i} + \mathbf{G}_{j,j} - \mathbf{G}_{i,j} - \mathbf{G}_{j,i}$$

Contains no baseline information

$$\mathbf{u}_i\mathbf{u}_i^T + \mathbf{u}_j\mathbf{u}_j^T - 2\mathbf{u}_i\mathbf{u}_j^T$$




Assuming baseline in the
mean of all patterns,
columns and row means of
G are zero

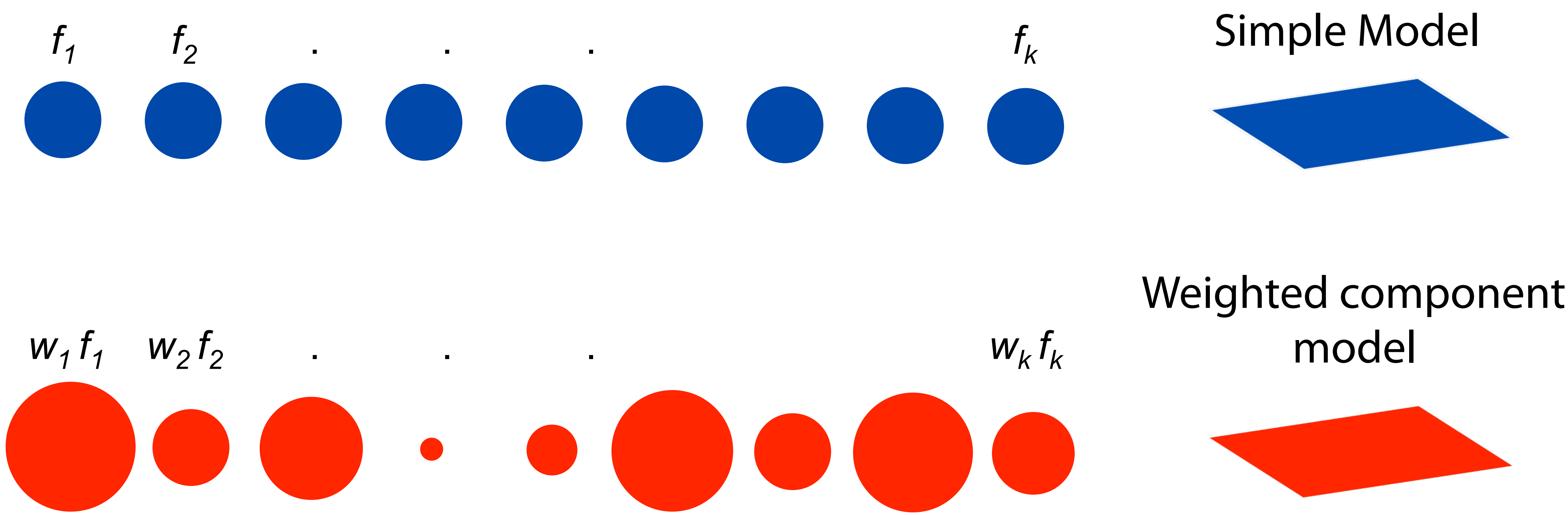
$$\mathbf{G} = -\frac{1}{2}\mathbf{H}\mathbf{D}\mathbf{H}^T$$

$$\mathbf{H} = \mathbf{I}_k - \mathbf{1}_K\mathbf{1}_k^T$$

# Overview

- Covariances and Distances
- Features and representational components
- Factorial models (MANOVA)
- Linear representational models
- Nonlinear representational models
- Summary

# *Weighting of different features*

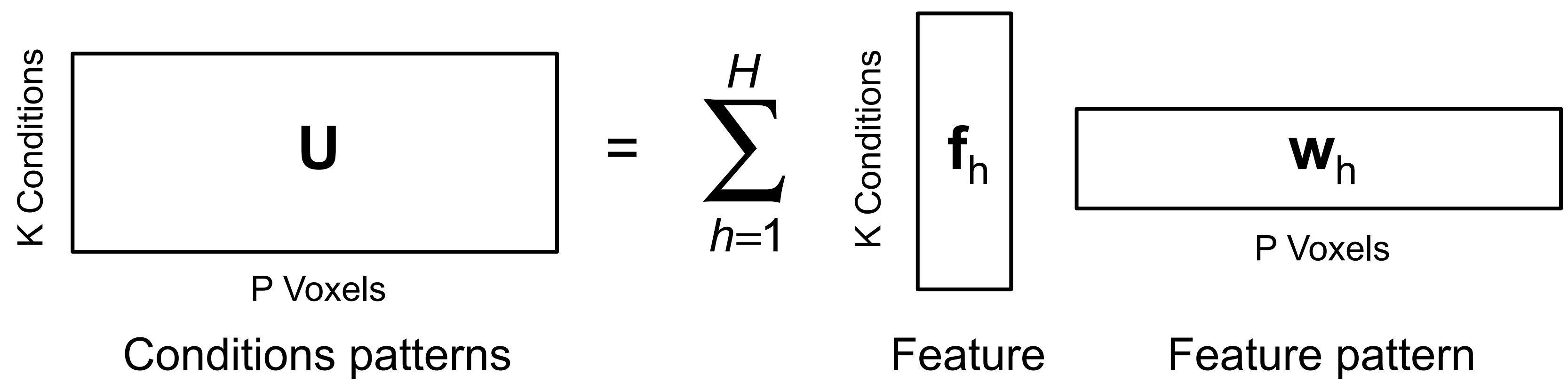


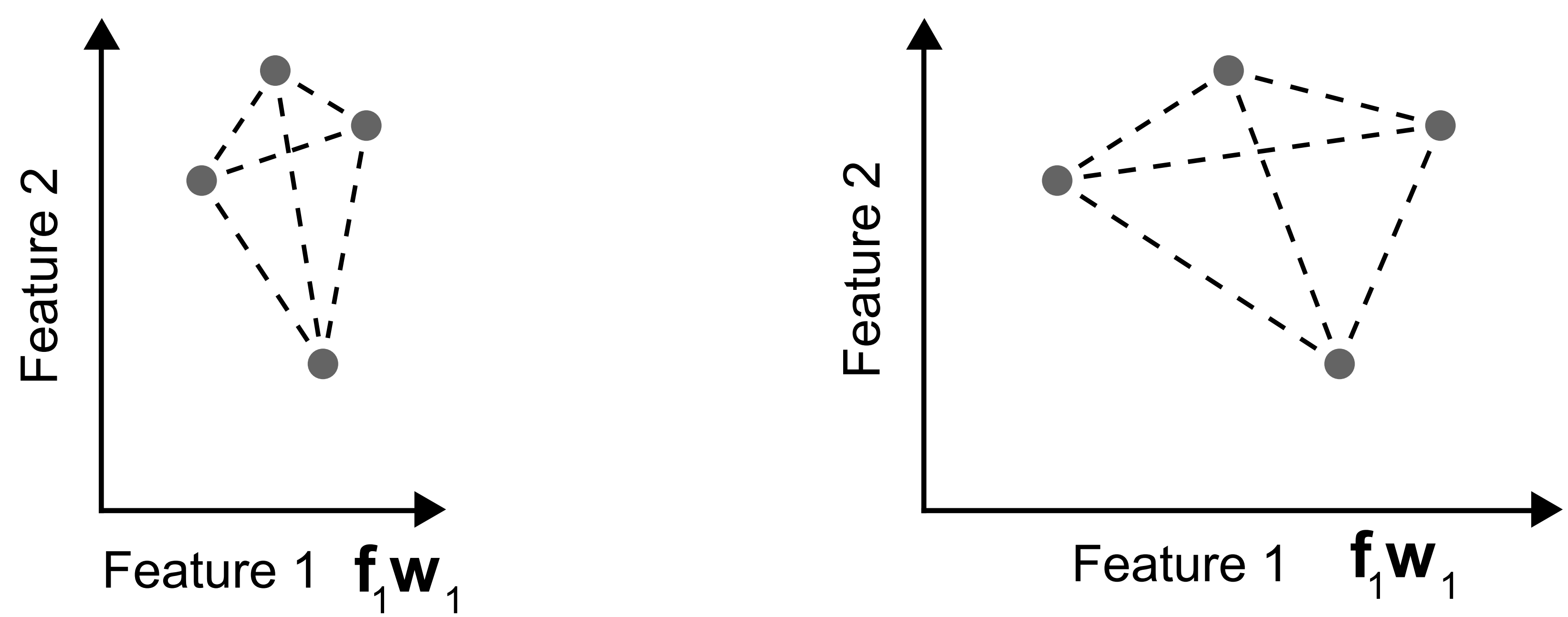

Examples:

- Different aspects of objects: color, shape
- Different aspects of movements: force, sequence, timing
- Different layers of a computation vision model

# *Features and representational components*

Pattern for each condition is caused by different features, each associated with a feature pattern.



# Features and representational components

Pattern for each condition is caused by different features, each associated with a feature pattern.

$$\underset{\text{K Conditions} \times \text{P Voxels}}{\mathbf{U}} = \sum_{h=1}^{H} \underset{\text{K Conditions}}{\mathbf{f}_h} \; \underset{\text{P Voxels}}{\mathbf{w}_h}$$

Conditions patterns — Feature — Feature pattern

Covariance matrix

$$\mathbf{G} = \mathbf{U}\mathbf{U}^T$$

$$= \sum_{h=1}^{H} \mathbf{f}_h \mathbf{w}_h \sum_{h=1}^{H} \mathbf{w}_h^T \mathbf{f}_h^T$$

*Assuming independence of feature patterns:*

$$\mathbf{w}_i \mathbf{w}_j^T = 0$$

$$= \sum_{h=1}^{H} \mathbf{f}_h \mathbf{w}_h \mathbf{w}_h^T \mathbf{f}_h^T$$

$$= \sum_{h=1}^{H} \underbrace{\left(\mathbf{w}_h \mathbf{w}_h^T\right)}_{\omega_h} \underbrace{\left(\mathbf{f}_h \mathbf{f}_h^T\right)}_{\mathbf{G}_h}$$

Covariance matrix

$$\mathbf{G} = \sum_{h=1}^{H} \omega_h \mathbf{G}_h$$

Component weight — Component matrix (G)

Distance matrix

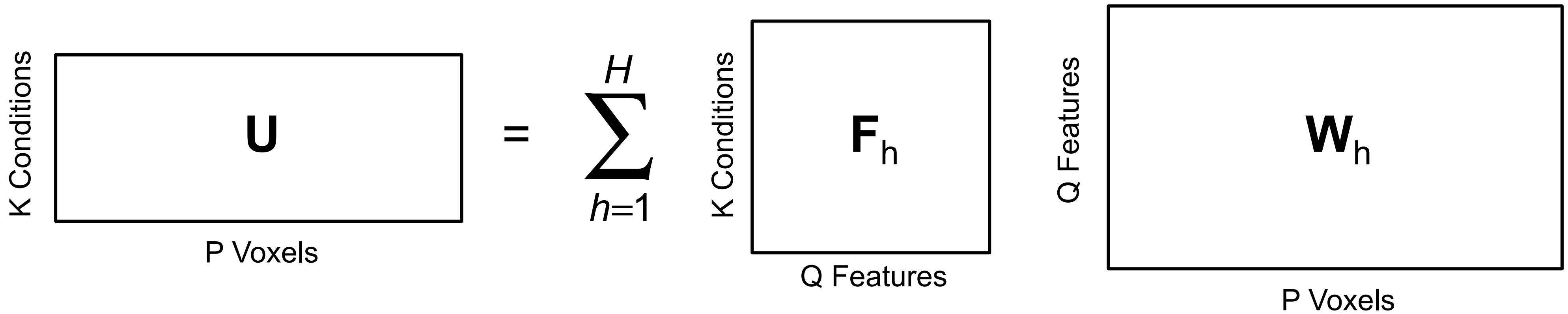
$$\mathbf{D} = \sum_{h=1}^{H} \omega_h \mathbf{D}_h$$

Component weight — Component matrix (G)

$\omega_h$ The component weight is the variance or power of the feature pattern.

# Features and representational components

Most often we do not weight single features, but groups of features: representation components




Conditions patterns | Feature | Feature pattern

Covariance matrix

$$\mathbf{G} = \mathbf{U}\mathbf{U}^T$$

$$= \sum_{h=1}^{H} \mathbf{F}_h \mathbf{W}_h \sum_{h=1}^{H} \mathbf{W}_h^T \mathbf{F}_h^T$$

*Assuming independence of feature patterns across components*

$$\mathbf{W}_i \mathbf{W}_j^T = \mathbf{0}$$

$$= \sum_{h=1}^{H} \mathbf{F}_h \mathbf{W}_h \mathbf{W}_h^T \mathbf{F}_h^T$$

*Assuming covariance of feature patterns*

$$\mathbf{W}_h \mathbf{W}_h^T = \mathbf{V}_h \omega_h$$

$$= \sum_{h=1}^{H} \omega_h \underbrace{\left(\mathbf{F}_h \mathbf{V}_h \mathbf{F}_h^T\right)}_{\mathbf{G}_h}$$

Covariance matrix

$$\mathbf{G} = \sum_{h=1}^{H} \omega_h \mathbf{G}_h$$

Component weight

Component matrix (G)

Distance matrix

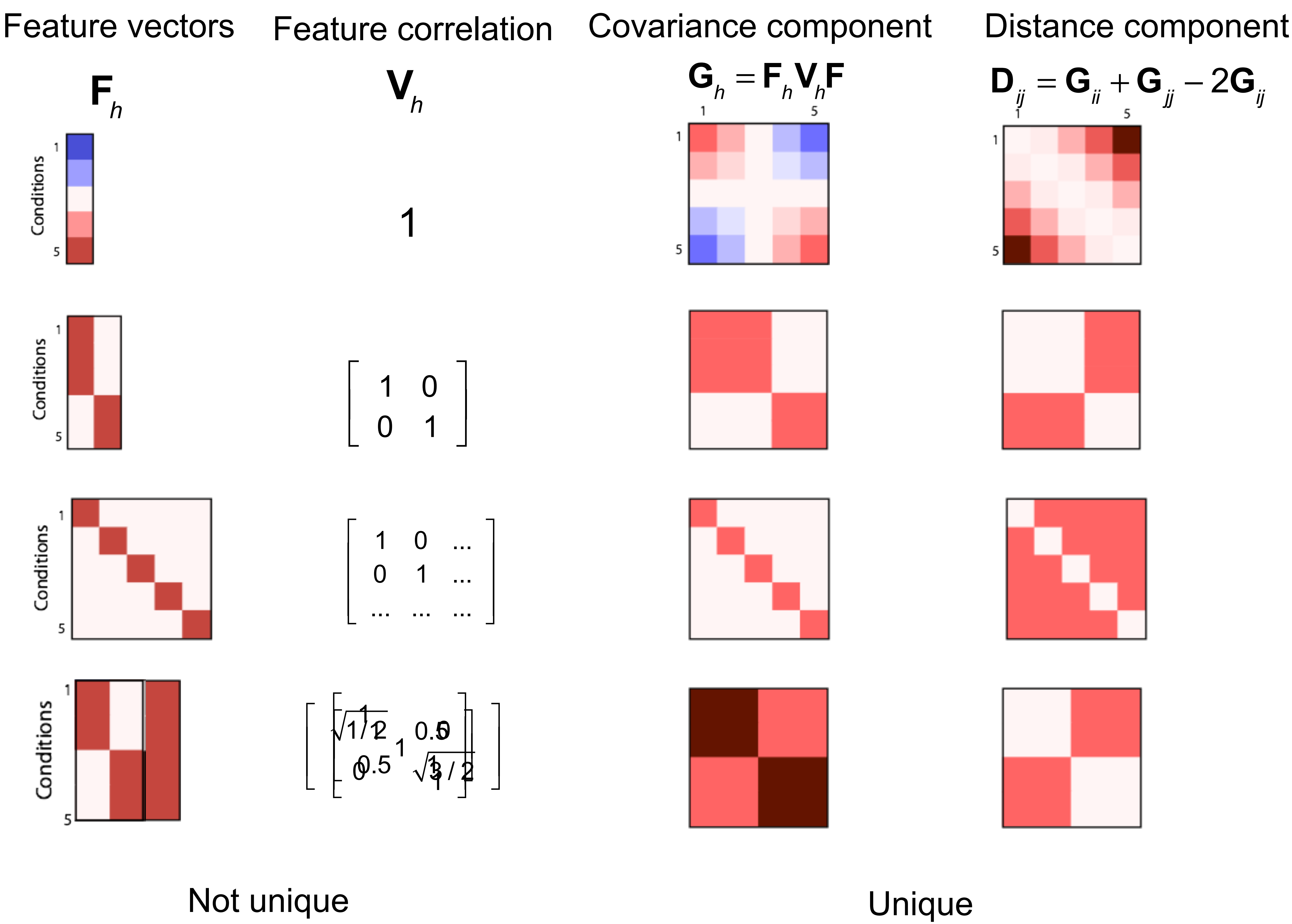
$$\mathbf{D} = \sum_{h=1}^{H} \omega_h \mathbf{D}_h$$

Component weight

Component matrix (G)

$\omega_h$ The component weight is the variance or power of the feature pattern.

# *Features and representational components*



# *Features and representational components: Estimation*

How do we estimate component weights?

$$\mathbf{D} = \sum_{h=1}^{H} \omega_h \mathbf{D}_h$$

*Vectorise*

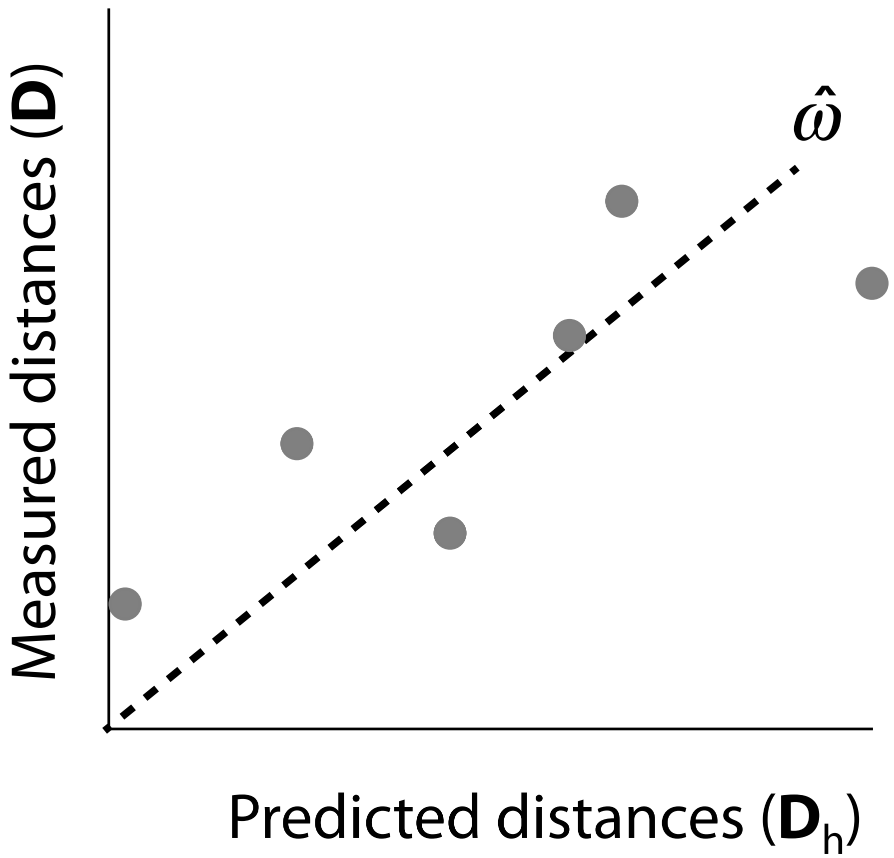
$$\mathbf{D} \longrightarrow \mathbf{d}$$

*Build component matrix*

$$\mathbf{X} = \begin{bmatrix} \mathbf{d}_1 & \mathbf{d}_2 & \ldots \end{bmatrix}$$

*Ordinary least-squares*

$$\hat{\omega} = \underbrace{\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T}_{\mathbf{X}^+}\hat{\mathbf{d}}$$



# *Features and representational components*

- *Features* are variables encoded in neuronal elements
- Groups of features with similar encoding strength form a *representational component*
- Features of different components are assumed to be mutually independent
- Many feature sets can lead to the same representational component
- Models are uniquely specified via their component matrices (representational similarity trick)
- Component weights estimate variance (or power) of representations

# Overview

- Covariances and Distances
- Features and representational components
- Factorial models (MANOVA)
- Linear representational models
- Nonlinear representational models
- Summary

## *Integrated vs. independent encoding*

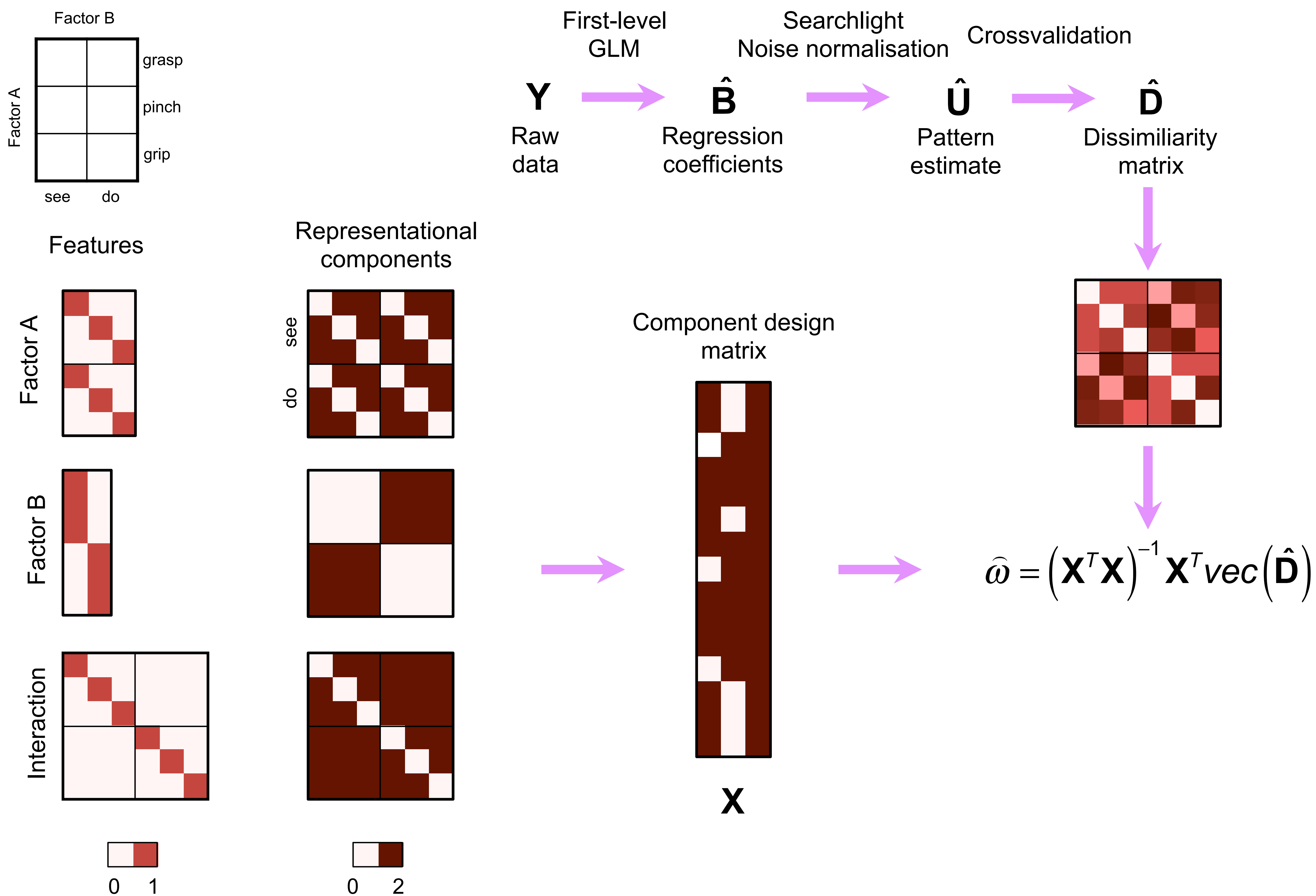
- Now, we assumed that different components are encoded independently in the brain
- This does not mean that they are encoded in different regions / voxel: only that their patterns are unrelated to each other

- BUT: Can we test this?

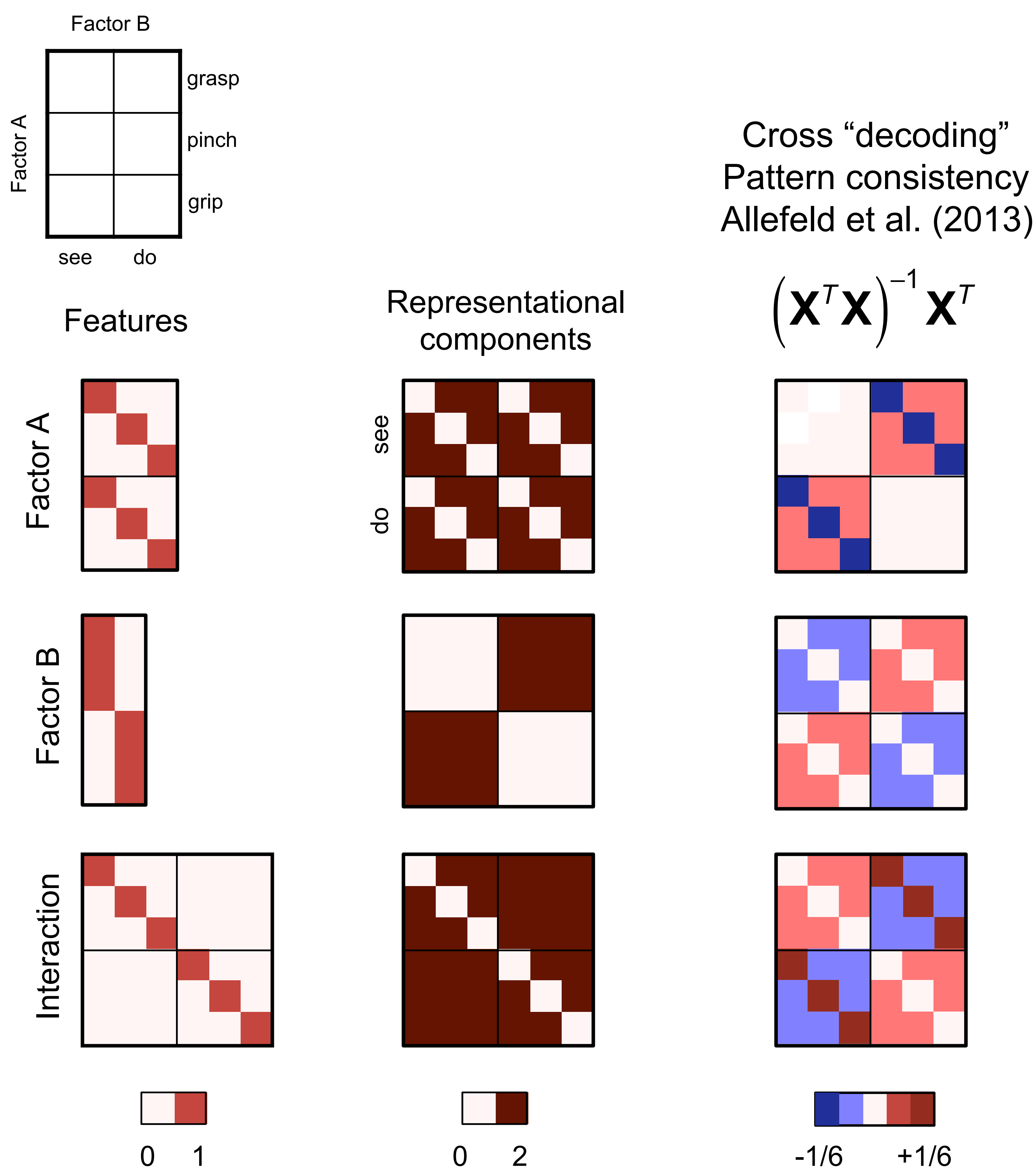
## *Integrated vs. independent encoding: factorial models*

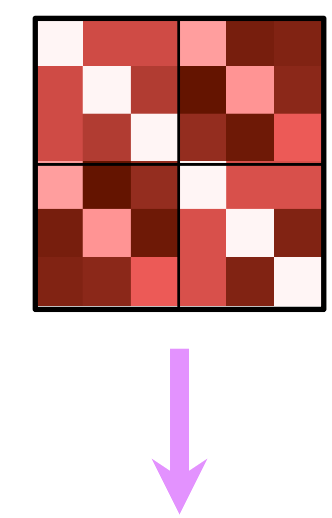
- Are two groups of features (variables) encoded independently or dependently?
- Vary the 2 factors in a fully crossed design
  - Condition (see / do) x Action (3 gestures)
  - Rhythm x Spatial sequence
  - Reach directions (3) x Grasps (3) ....
- Where is factor A encoded, where is B encoded?
- Are A and B encoded in an integrated or independent fashion?
- Is Factor B consistently encoded across levels of factor A ("cross-decoding")?

# Factorial representational models (MANOVA)



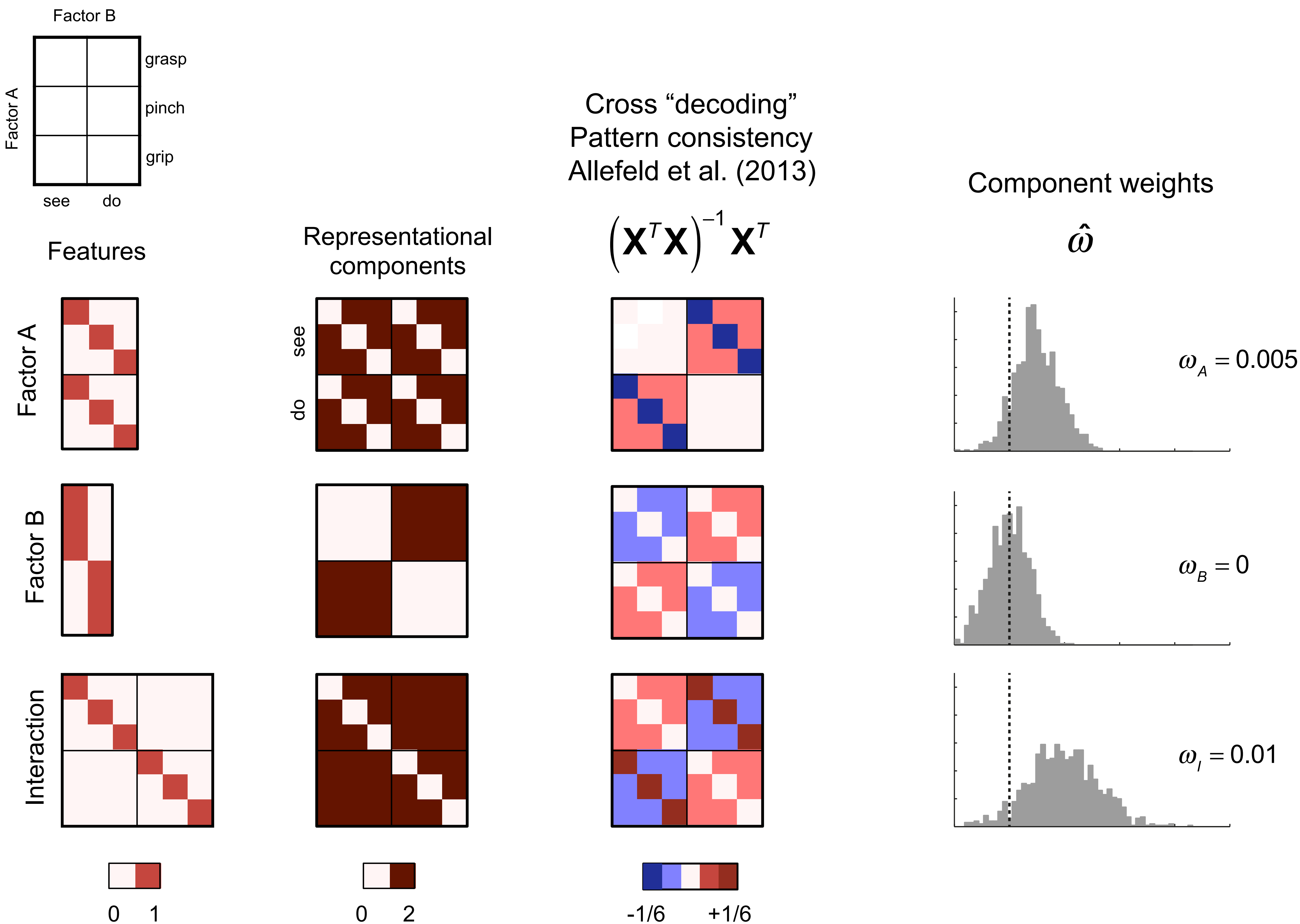
# Factorial representational models (MANOVA)



$$\hat{\omega} = \underbrace{\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T}\, vec\left(\hat{\mathbf{D}}\right)$$

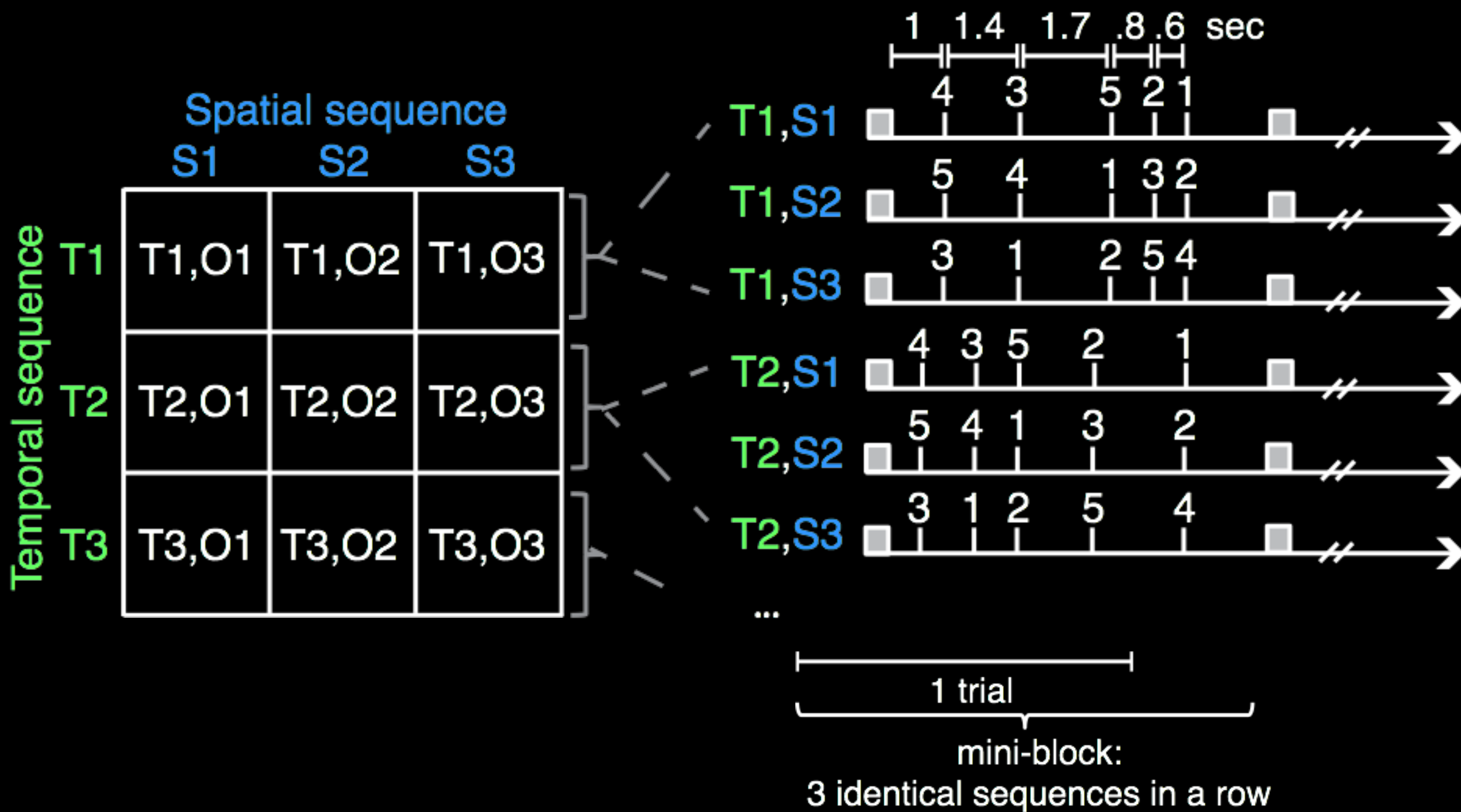
# Factorial representational models (MANOVA)



# *Factorial representational models (MANOVA)*

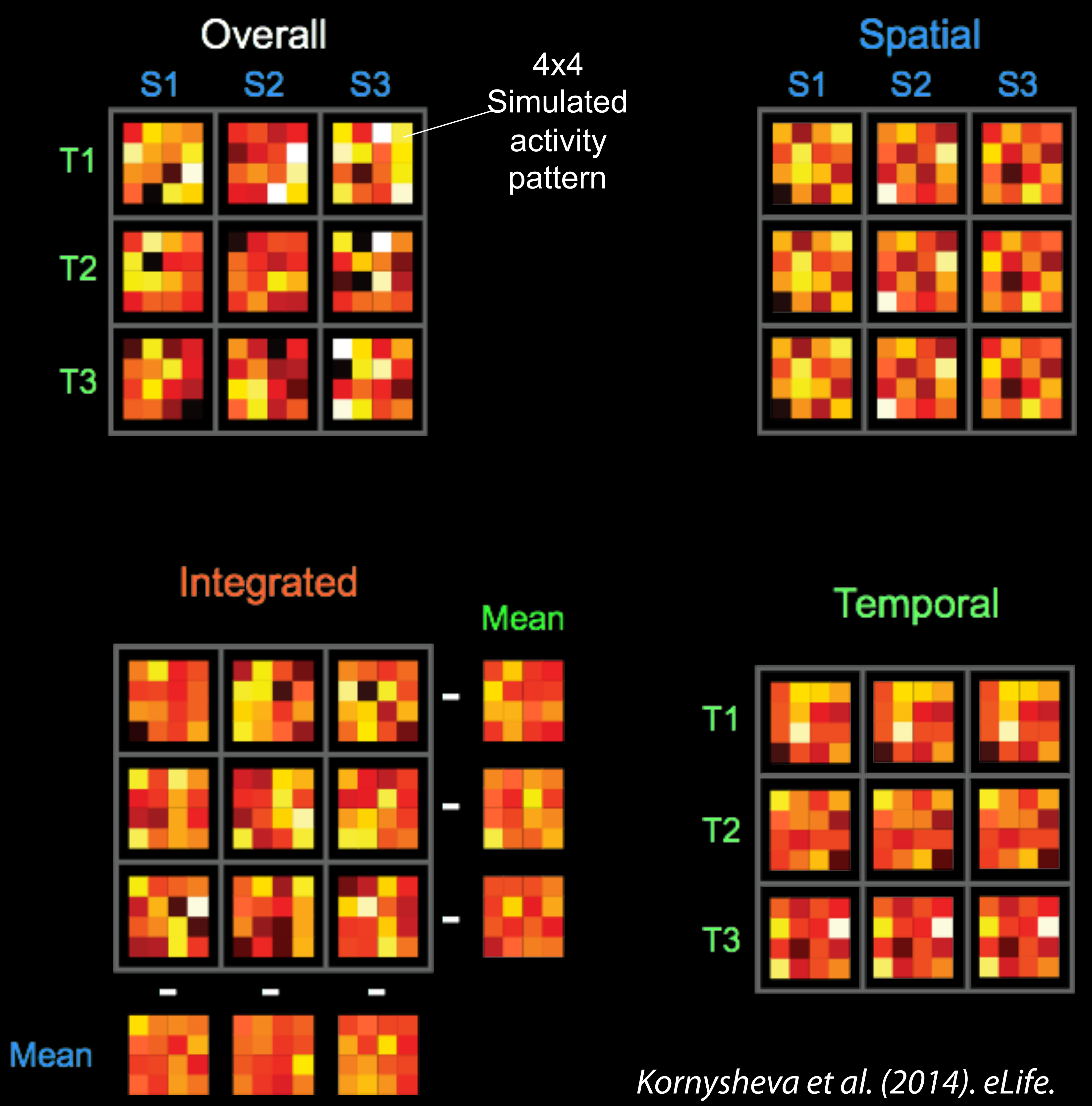
- Factorial models can reveal mean encoding effect and interactions
- Component weight estimates are unbiased and can be directly tested in group analysis
- Main effects are assess by pattern consistency across levels of the other variable (replaces cross-classification)
- Mathematically identical to approach suggested by Allefeld et al. (2013)

# *Factorial MANOVA designs (example)*



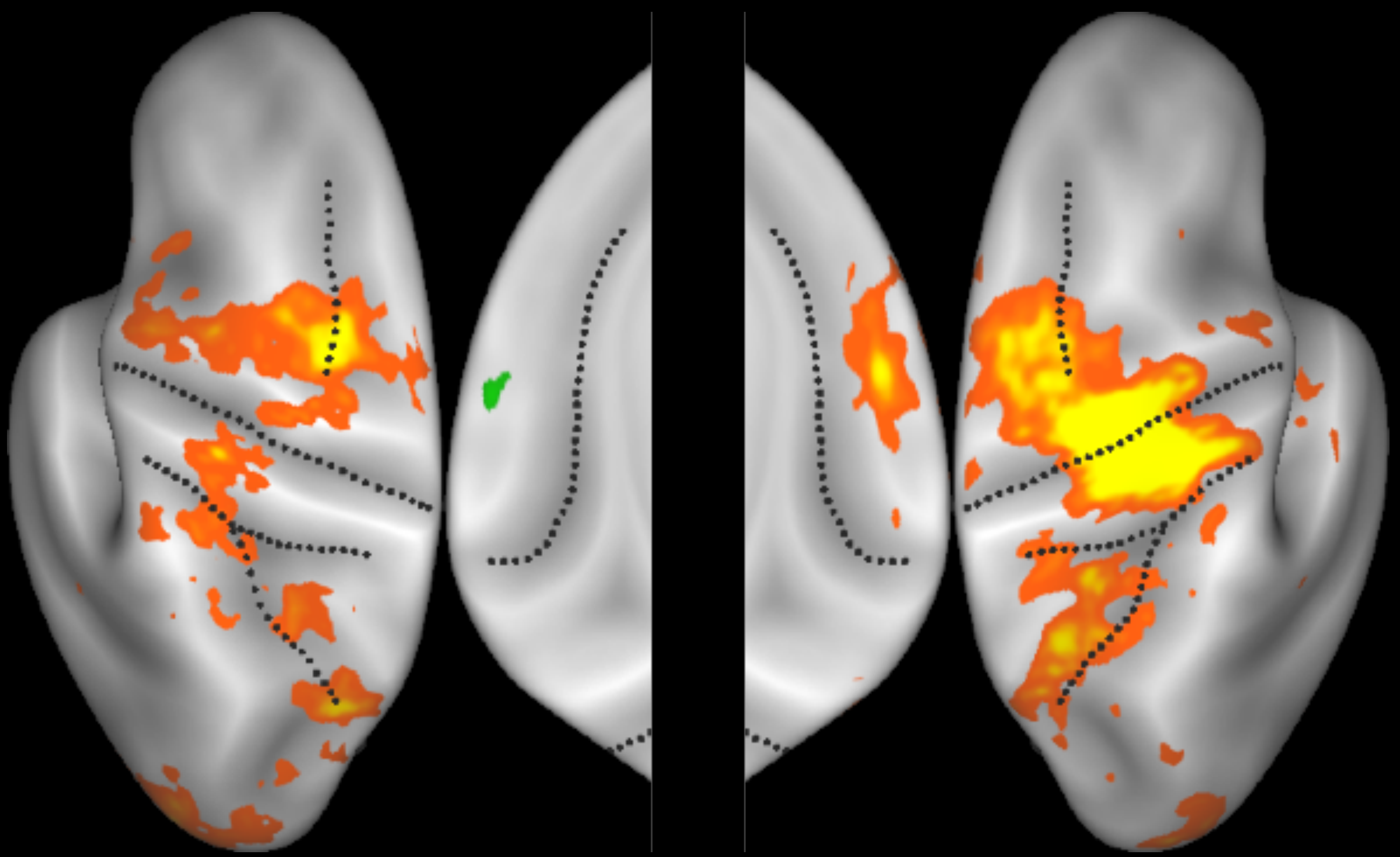

Kornysheva et al. (2014). eLife.

# Factorial MANOVA designs (example)




*Kornysheva et al. (2014). eLife.*

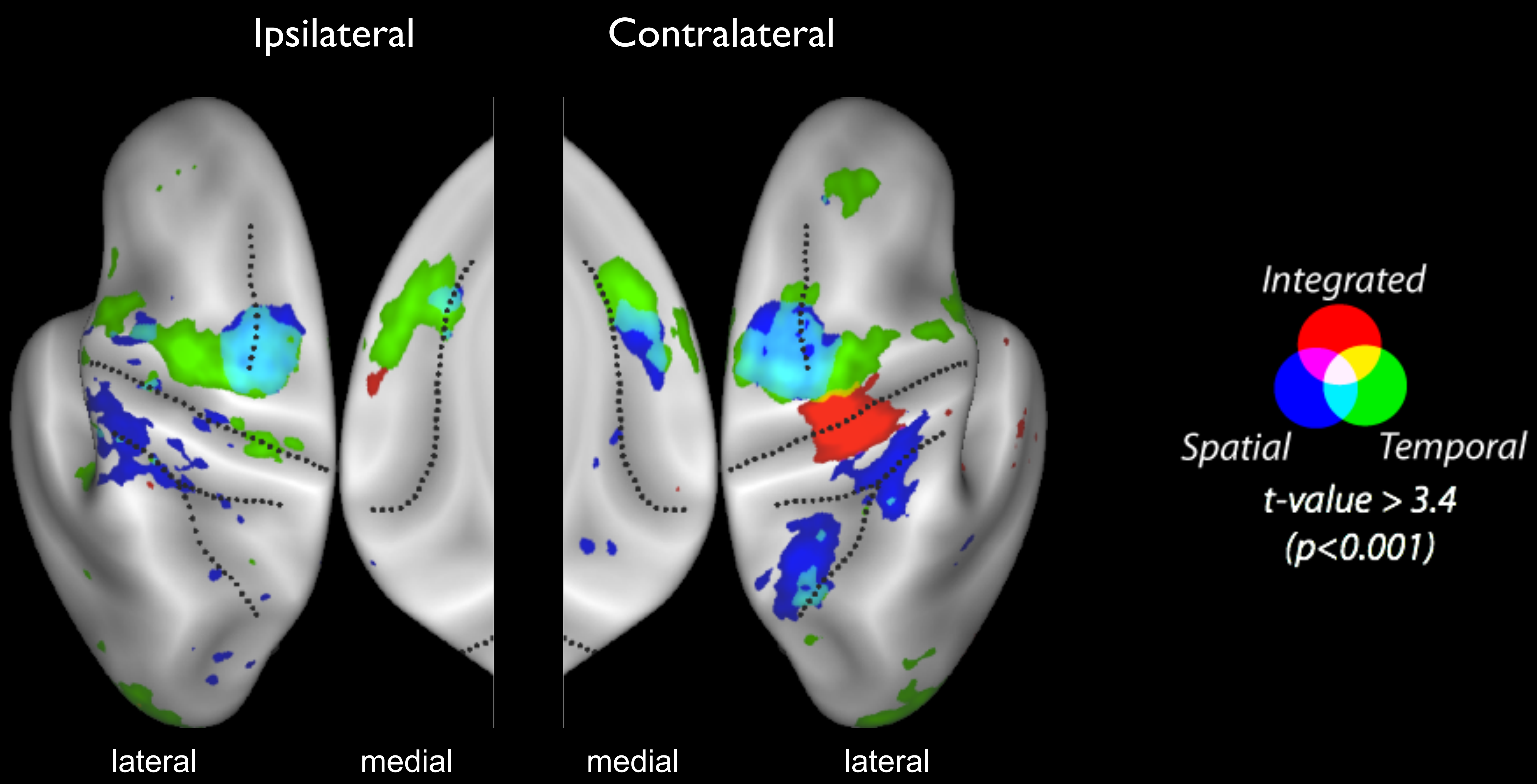
# *Factorial MANOVA designs (example)*




Overall encoding

Kornysheva et al. (2014). eLife.

# *Factorial MANOVA designs (example)*




Kornysheva et al. (2014). eLife.

# *Factorial representational models (MANOVA)*

## *Linearity Assumption*

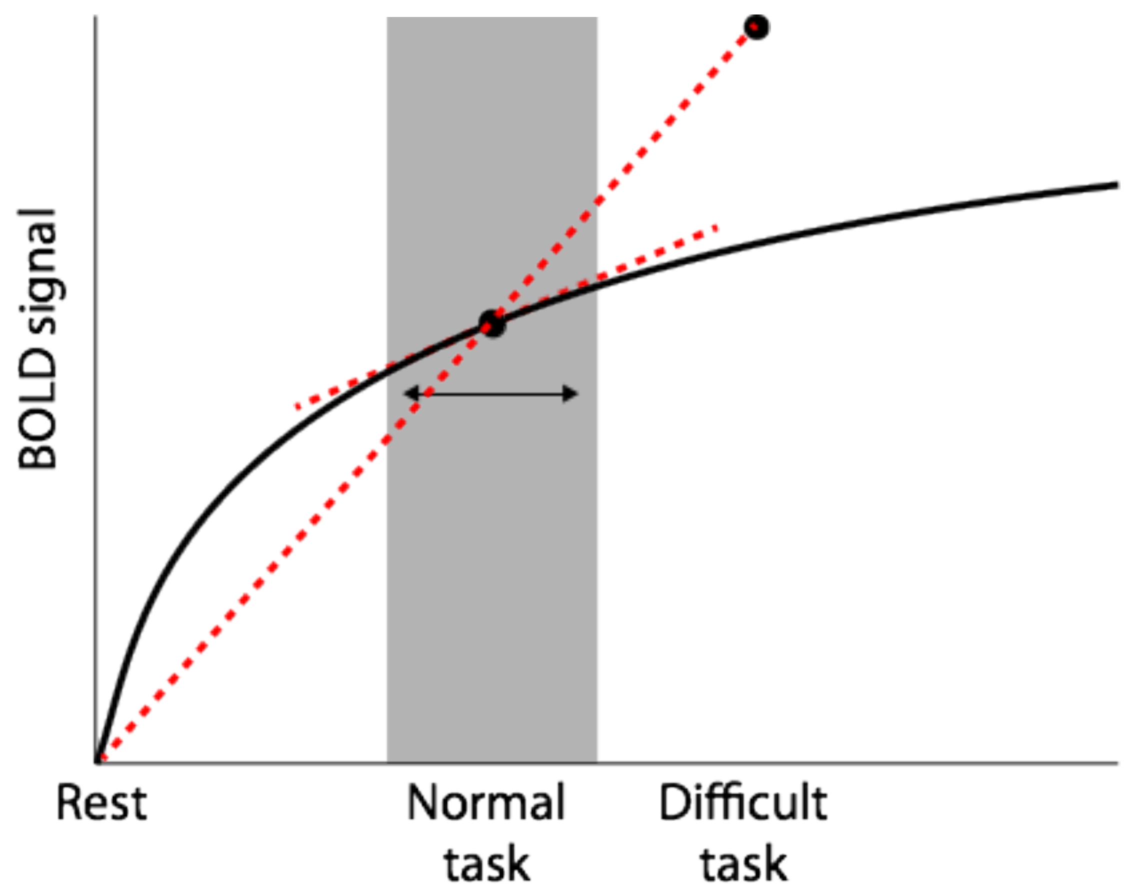
Patterns for different components overlap linearly

if they engage independent neuronal subpopulations

if they combine linearly to determine firing rate

***AND***

if the relationship between neural activity and BOLD is approximately linear




Experimental conditions should be similar in overall activity

Note: mean value subtraction in analysis does not fix this!

## *Key insights I*

**D1.** Pattern covariance matrices and squared Euclidean distance matrices capture the same information, but the former retain the baseline

**D2.** A representational component (RC) is a group of representational features.

**D3.** A representation can be modelled as weighted combination of RCs (one weight per RC).

**D4.** Weighted combinations of RCs correspond to weighted combinations of representational distance matrices.

**D5.** Component weights can be estimated using regression and tested directly (against zero) in group analyses.

# Weighted Representational Component Models II

Jörn Diedrichsen

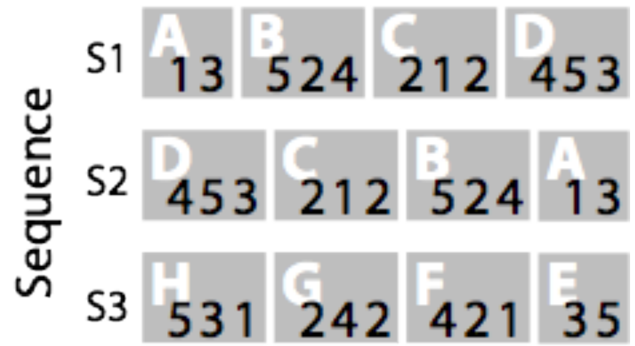
Motor Control Group
Institute of Cognitive Neuroscience
University College London

## Overview

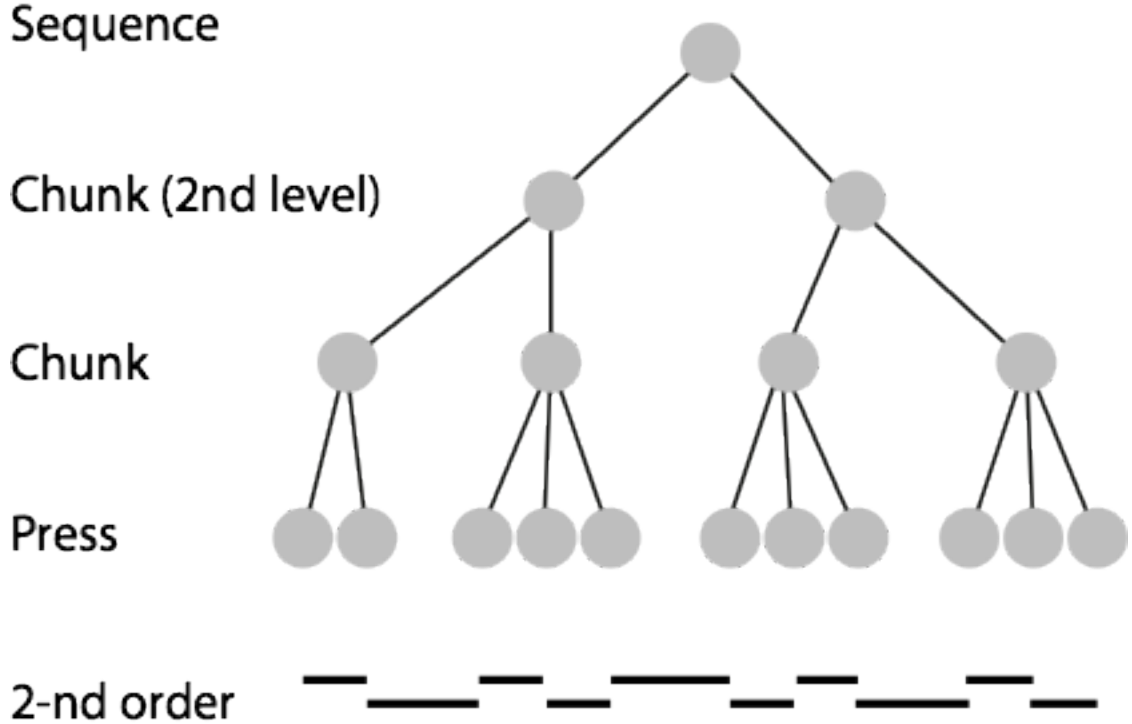
- Covariances and Distances
- Features and representational components
- Factorial models (MANOVA)
- Linear representational models
- Nonlinear representational models
- Summary

# Linear representational models

sequence consisting of chunks

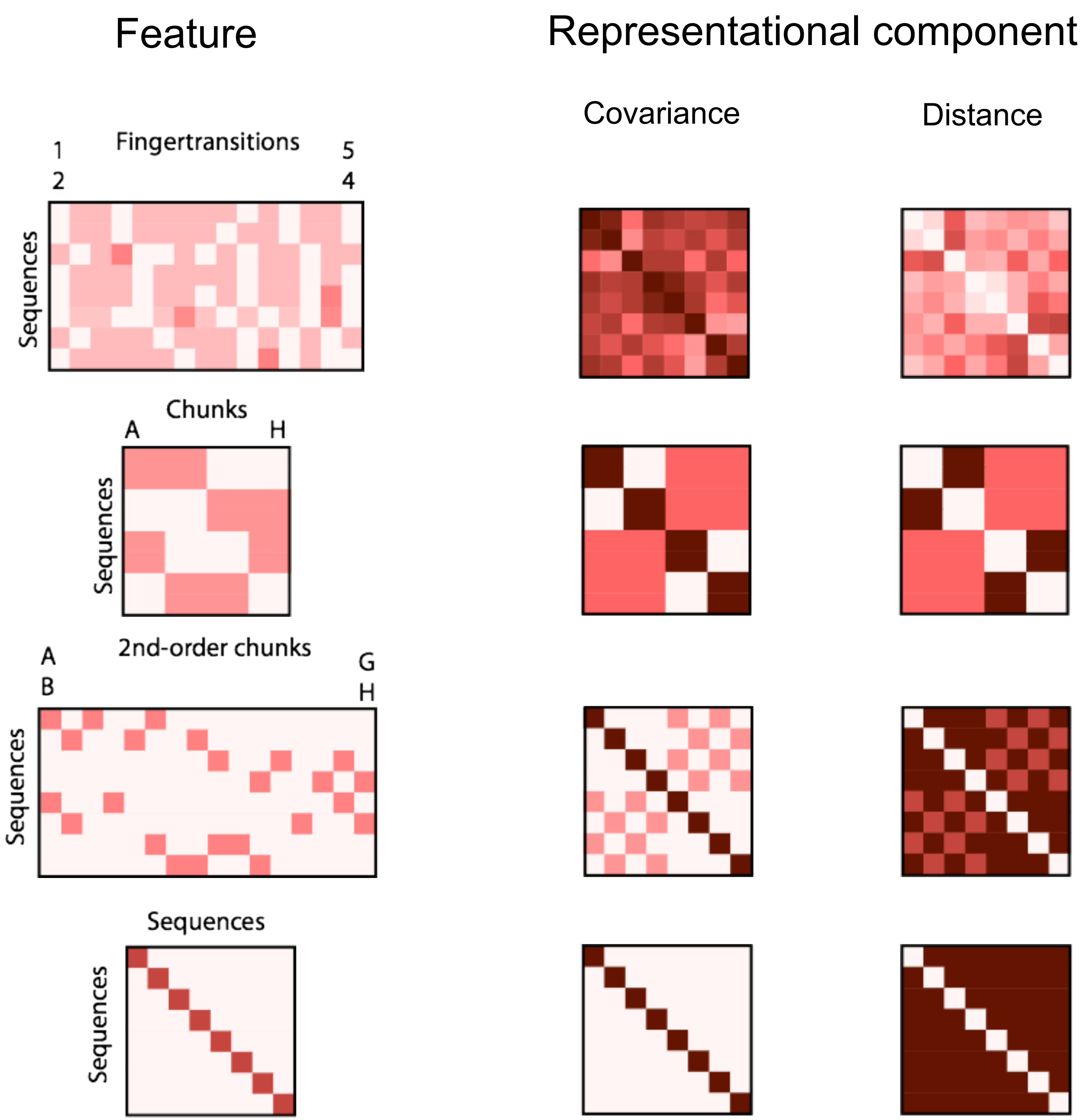



At what level are sequences represented?




*Yokoi et al. (in prep)*

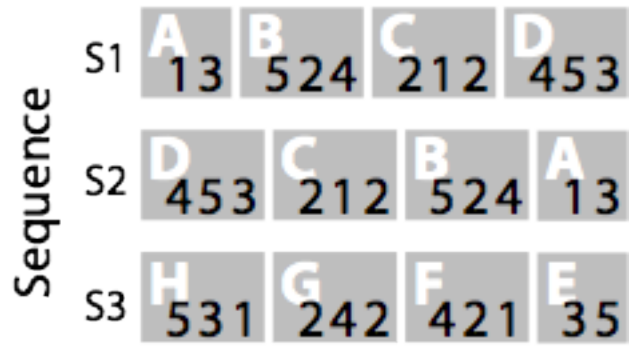
# Linear representational models

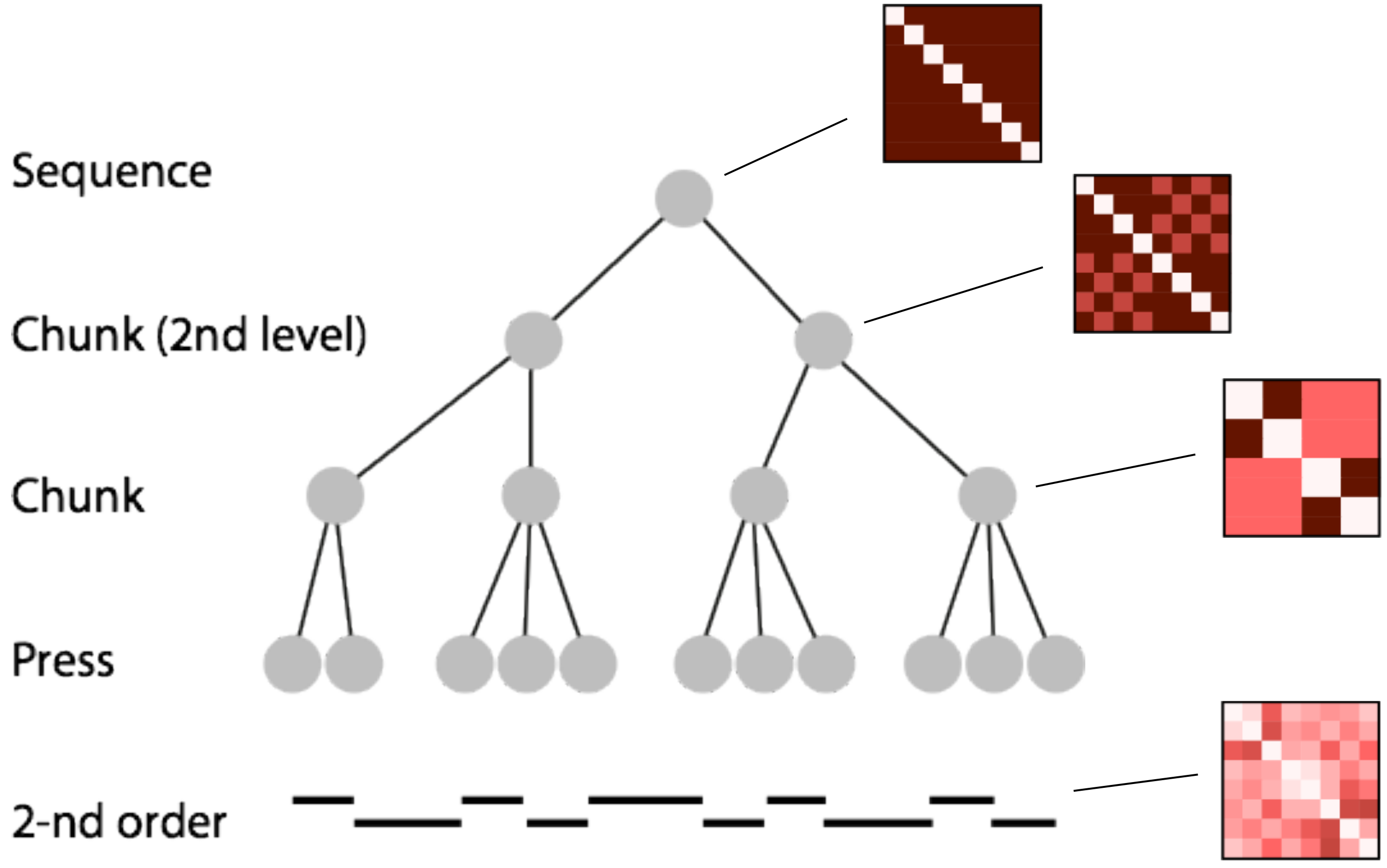



*Yokoi et al. (in prep)*

# Linear representational models

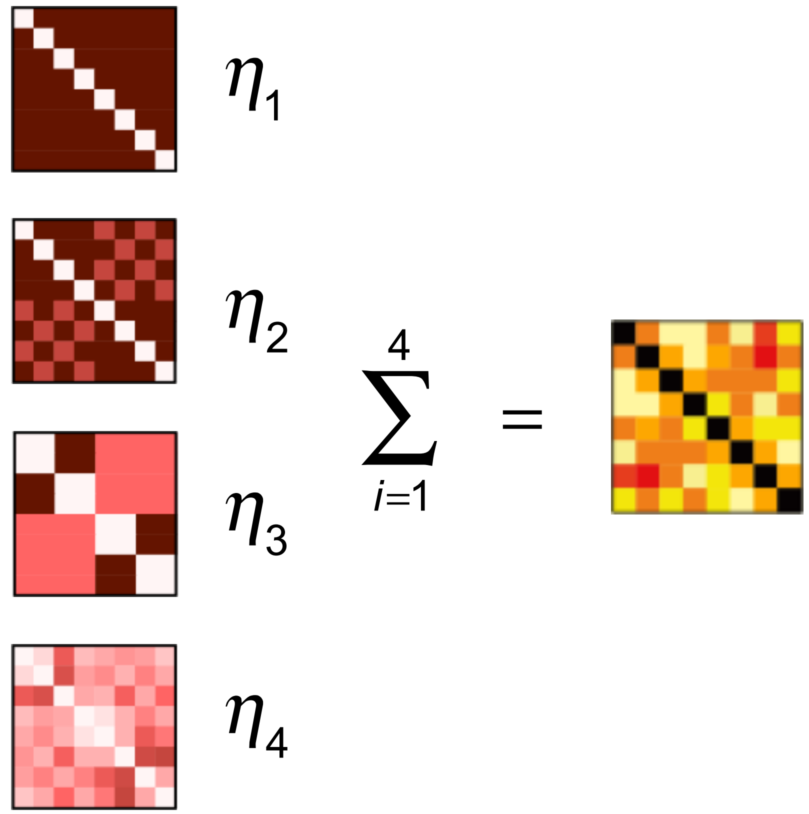
sequence consisting of chunks




Yokoi et al. (in prep)

# *Linear representational models*

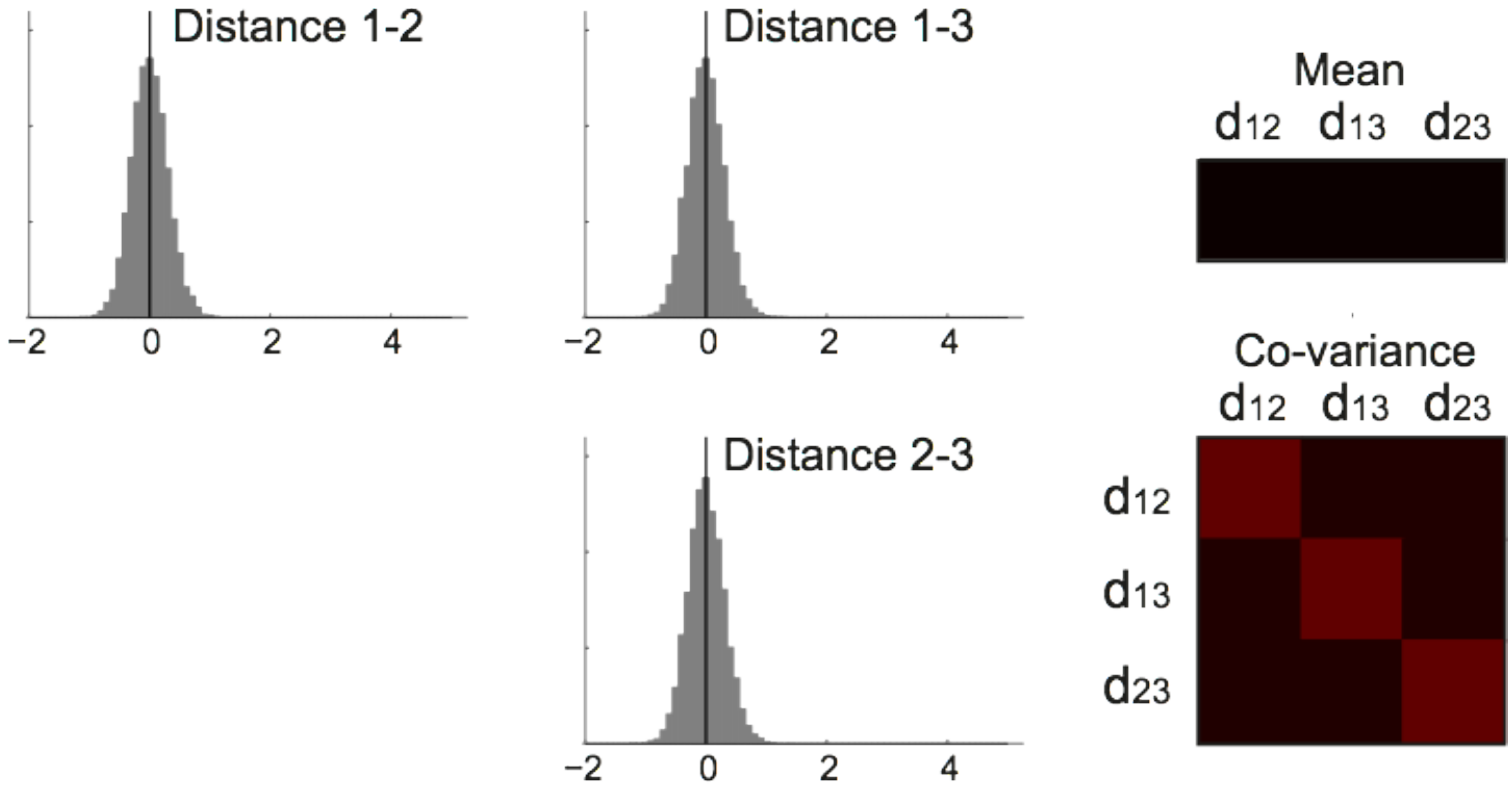



The use of simple regression (OLS) assumes that distances:

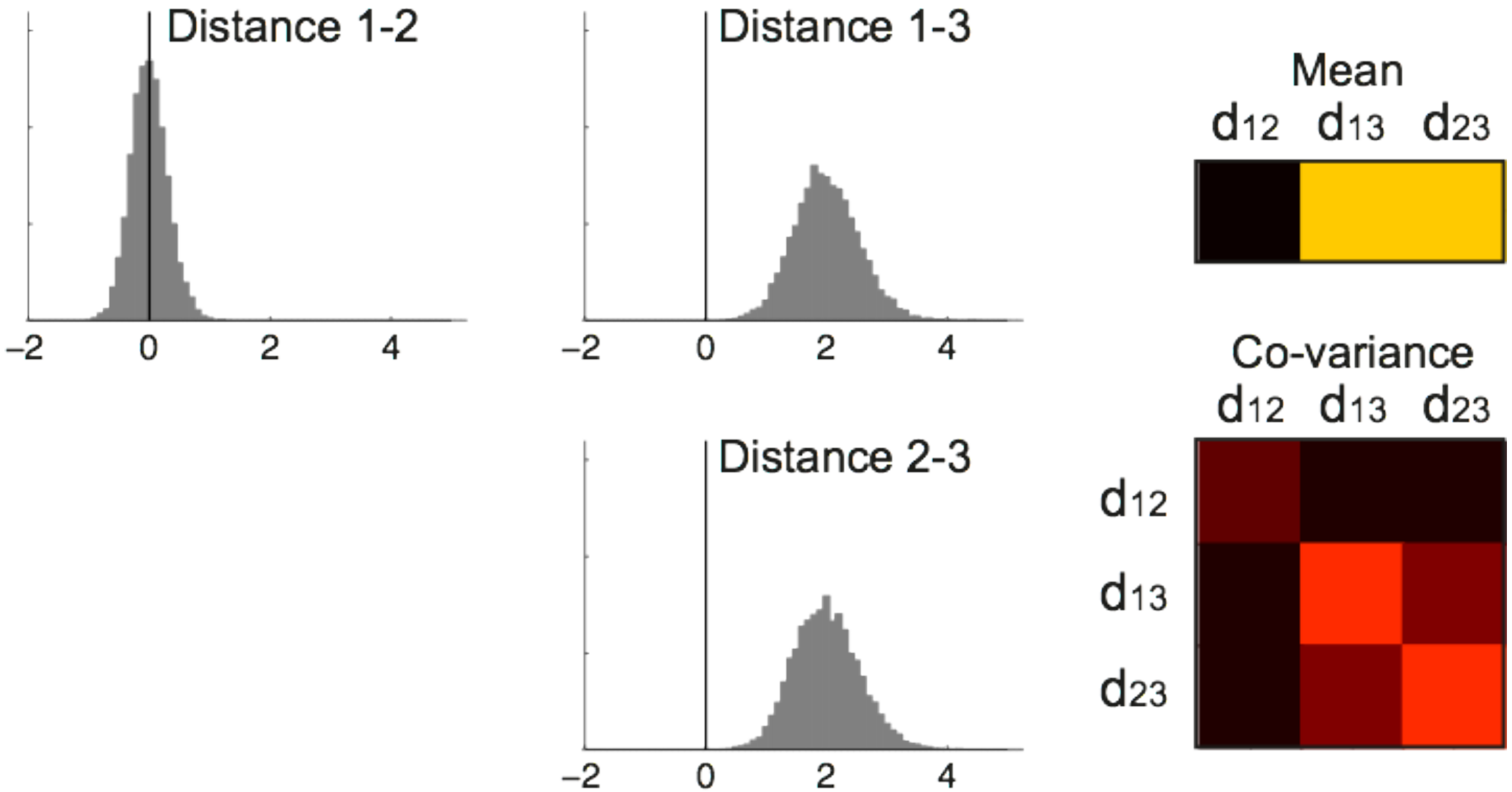
- are independent — i.
- have equal variance — i.
- are ~ normally distributed — d.

If this is violated we have an *unbiased* estimator, but not the *best linear unbiased estimator (BLUE)*

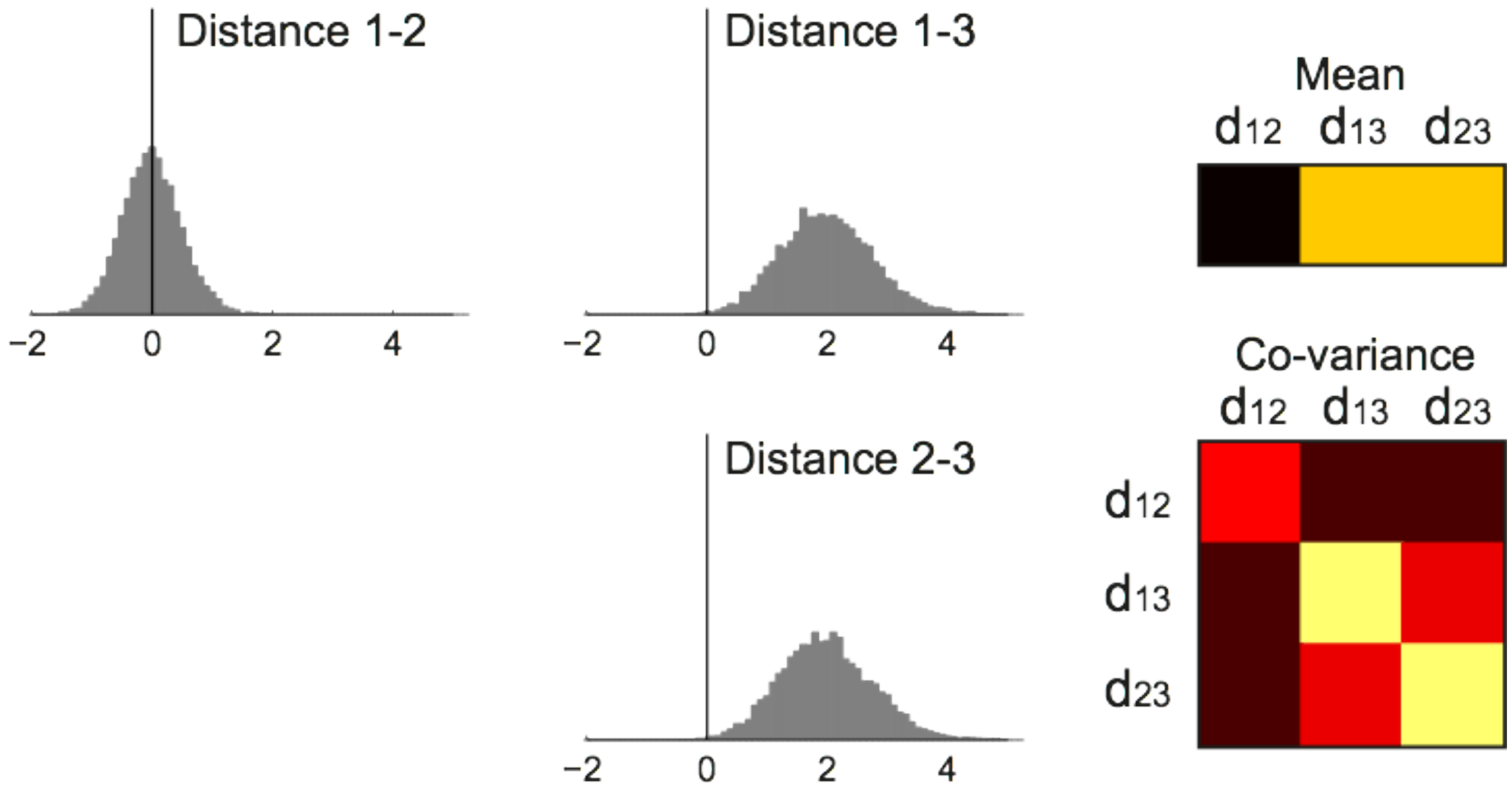
# Linear representational models



# *Linear representational models*



# Linear representational models



# Linear representational models: variance of distances

$$\hat{\delta}_{ij} = \mathbf{u}_i - \mathbf{u}_j \qquad \hat{\delta}_{ij} = \delta_{ij} + \varepsilon$$

Differences between patterns are measured with noise

$$\hat{d} = \hat{\delta}^{(m)}\hat{\delta}^{(n)T} = \left(\delta + \varepsilon^{(m)}\right)\left(\delta + \varepsilon^{(n)}\right)^T$$

$$\hat{d} = \delta\,\delta^T + \varepsilon^{(m)}\delta^T + \delta\varepsilon^{(n)T} + \varepsilon^{(m)}\varepsilon^{(n)T}$$

Squared distances are a sum of inner products,
Signal with signal, signal with noise,
and noise with noise

$$E\left(\hat{d}\right) = \delta\delta^T + \underbrace{\varepsilon^{(m)}\delta^T + \delta\varepsilon^{(n)T} + \varepsilon^{(m)}\varepsilon^{(n)T}}_{0}$$

In the expected value, the inner products containing noise drop out

$$\operatorname{var}\left(\hat{d}\right) = \underbrace{\operatorname{var}\left(\delta\delta^T\right)}_{0} + \sigma_\varepsilon^2\delta\delta^T + \sigma_\varepsilon^2\delta\delta^T + \sigma_\varepsilon^2\sigma_\varepsilon^2 P$$

Distance dependent Constant

For the variance, we obtain a part that depends on the distances, and one part that only depends on the noise.

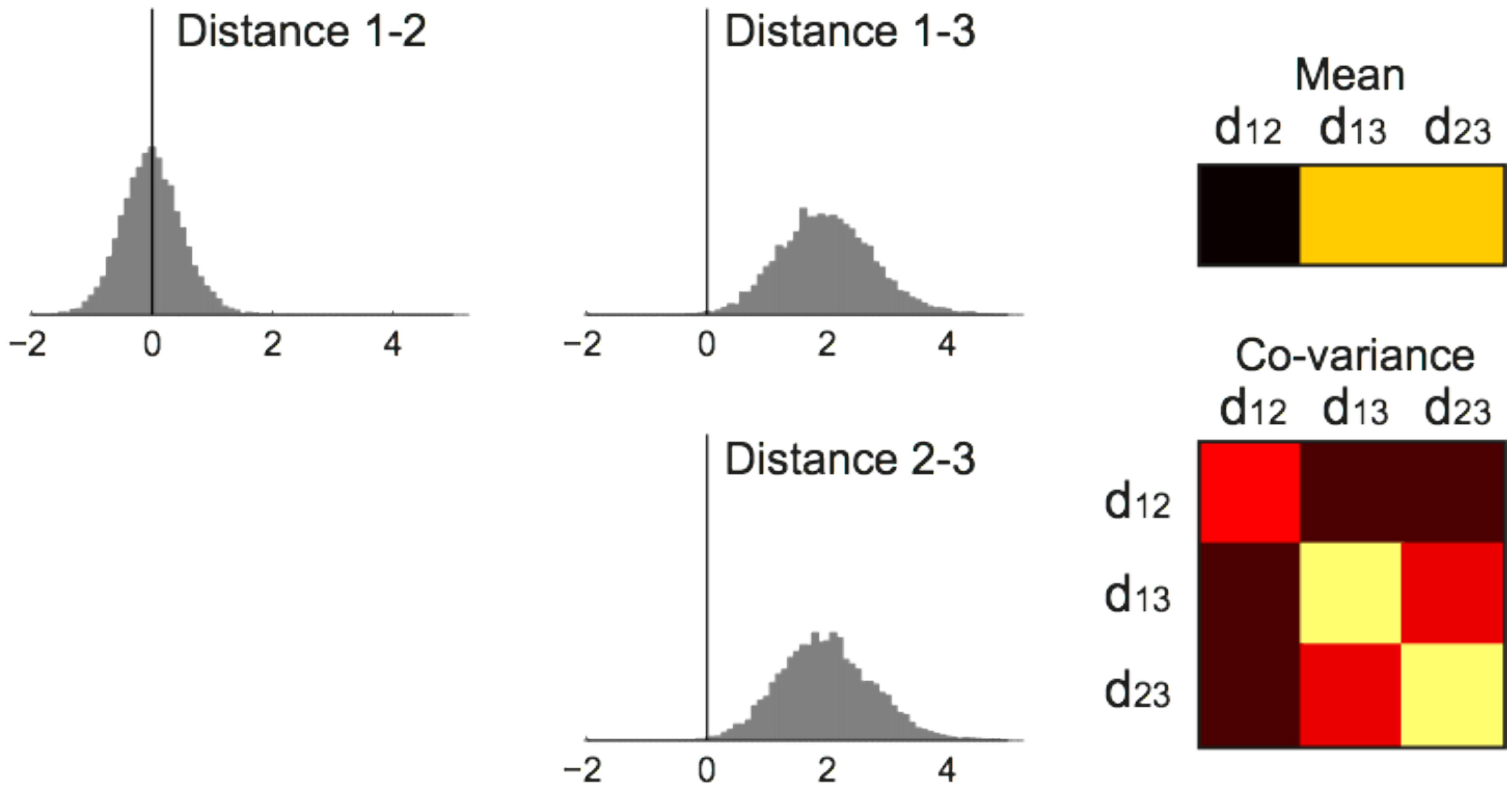
$$\operatorname{var}\left(\hat{\mathbf{d}}\right) = \frac{4}{R}\Delta\circ\Sigma + \frac{2P}{R(R-1)}\Sigma\circ\Sigma$$

Distance dependent Constant

The covariance of p distances when doing exhaustive crossvalidation over *R* partitions:

- $\Sigma$ Within-run covariances of $\hat{\delta}$
- $\Delta$ True inner products $\langle\delta,\delta\rangle$
- $\circ$ Element-by-element multiplication

# *Linear representational models*



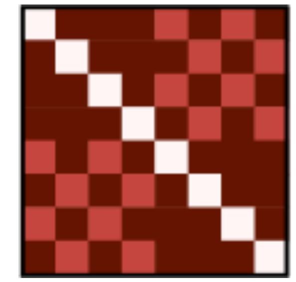

Co-variance of distances

$$\mathrm{var}\left(\hat{\mathbf{d}}\right) = \frac{4}{R}\Delta \circ \Sigma + \frac{2P}{R\left(R-1\right)}\Sigma \circ \Sigma$$

Distance dependent Constant

Taking this into account, we should do better than OLS

# *Linear representational models: IRLS Estimation*



Model

2.Predict distances

$$\hat{\mathbf{d}} = \mathbf{X}\omega$$

3. Calculate variance-covariance of d

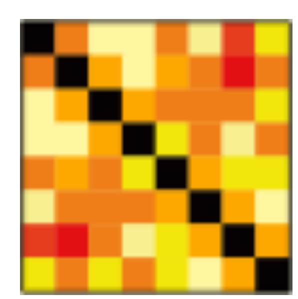
$$\mathbf{V} = \frac{4}{R}\Delta \circ \Sigma + \frac{2P}{R(R-1)}\Sigma \circ \Sigma$$

1.Start with initial guess of $\omega$

Until convergence

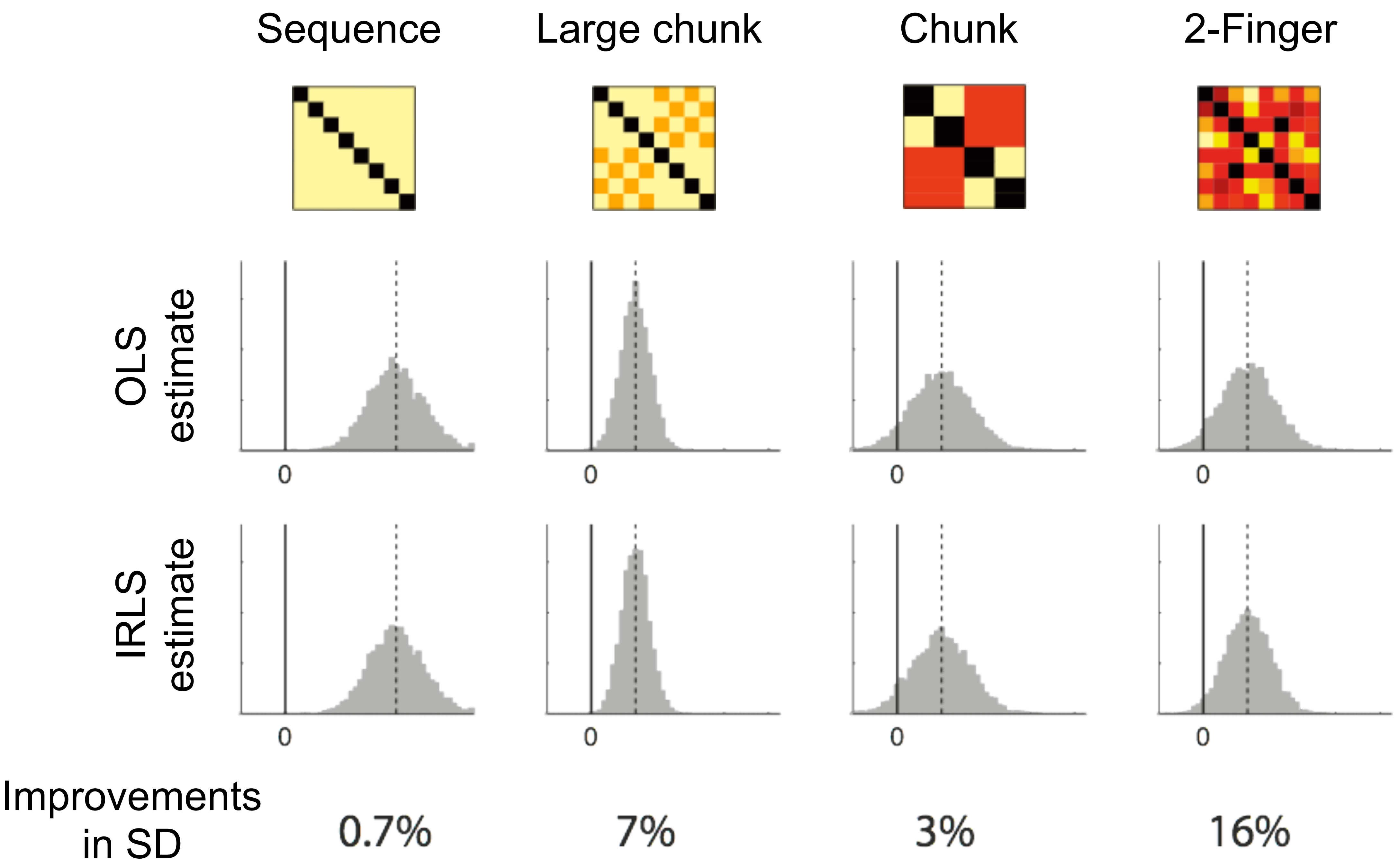
4. Use in estimation

$$\hat{\omega} = \left(\mathbf{X}^T\mathbf{V}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{d}$$



Data

# *Linear representational models*




*OLS is unbiased, but suboptimal*

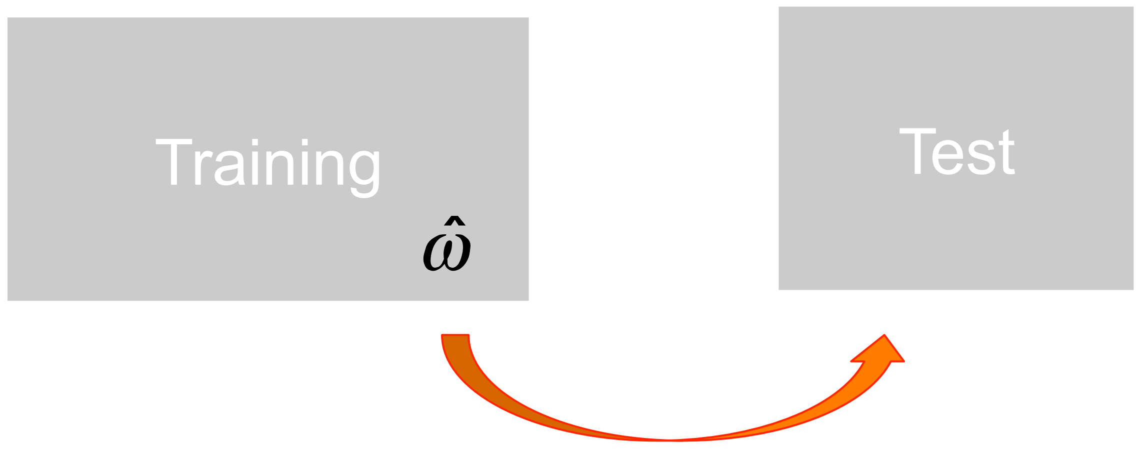
*IRLS can do better, but by how much depends on model structure*

*-> for factorial designs it does not matter*

# *Linear representational models: Estimation*

How do we best estimate and component weights?

| | | |
|---|---|---|
| Ordinary least-squares (OLS) | | • Unbiased estimates<br>• Can become negative<br>• Allows direct testing of parameters |
| Iteratively reweighted least-squares (IRLS) | | |
| Non-negative least-squares | *Khaligh-Razavi & Kriegeskorte (2014)* | • Positive estimates<br>• Biased<br>• Model testing by crossvalidation |
| Maximum likelihood | *Diedrichsen et al. (2011)* | |



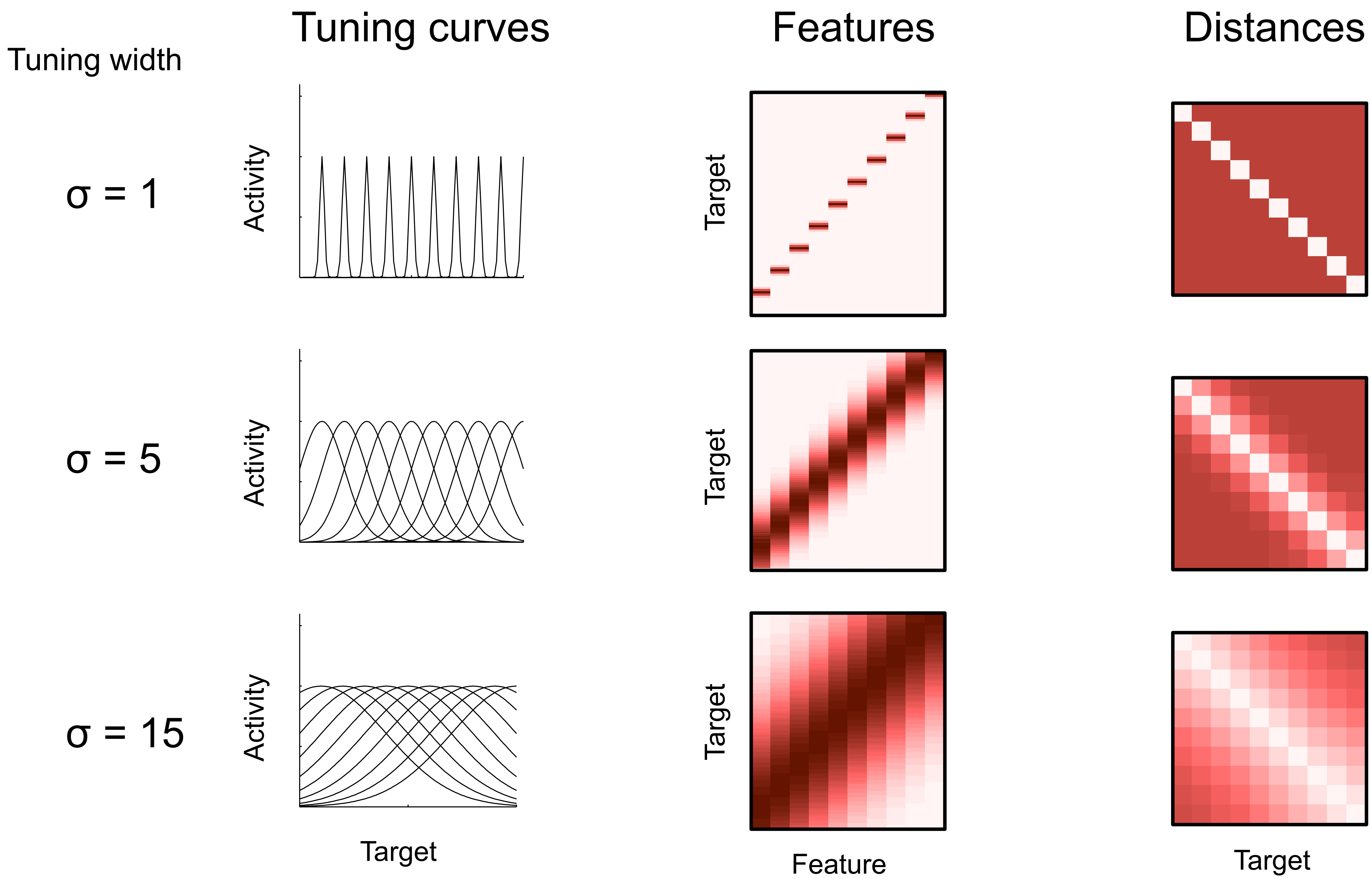

Tight link to encoding models

# Overview

- Covariances and Distances
- Features and representational components
- Factorial models (MANOVA)
- Linear representational models
- Nonlinear representational models
- Summary

# *Non-linear representational models*

Models in which component matrices are non-linear functions of parameters



# *Non-linear representational models*

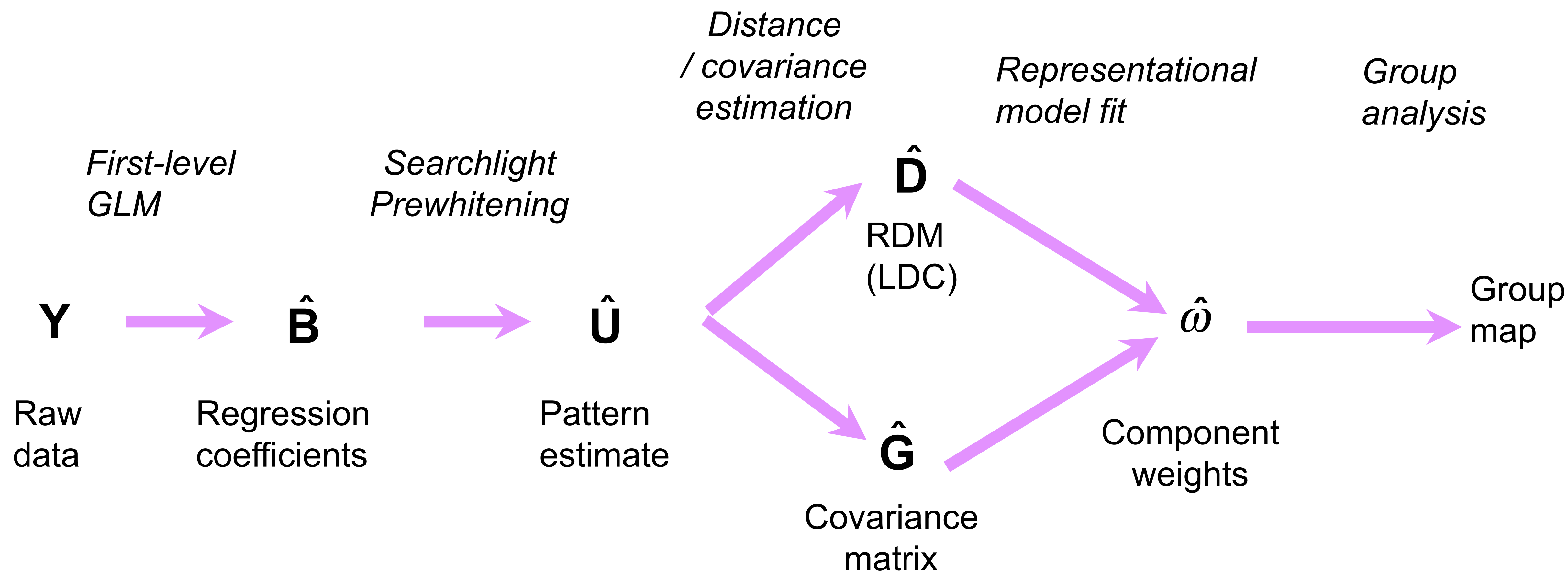
- Sometimes good linear approximations can be found (Example: AR-estimation in first-level SPMs)
- Otherwise, estimate nonlinear parameters to optimize the log-likelihood:

$$\log p\left(\hat{\mathbf{d}} \mid \theta\right) \propto -\frac{1}{2}\left(\hat{\mathbf{d}} - \mathbf{d}\right)^{T} V\left(\mathbf{d}\right)^{-1}\left(\hat{\mathbf{d}} - \mathbf{d}\right)$$

# *Overview*

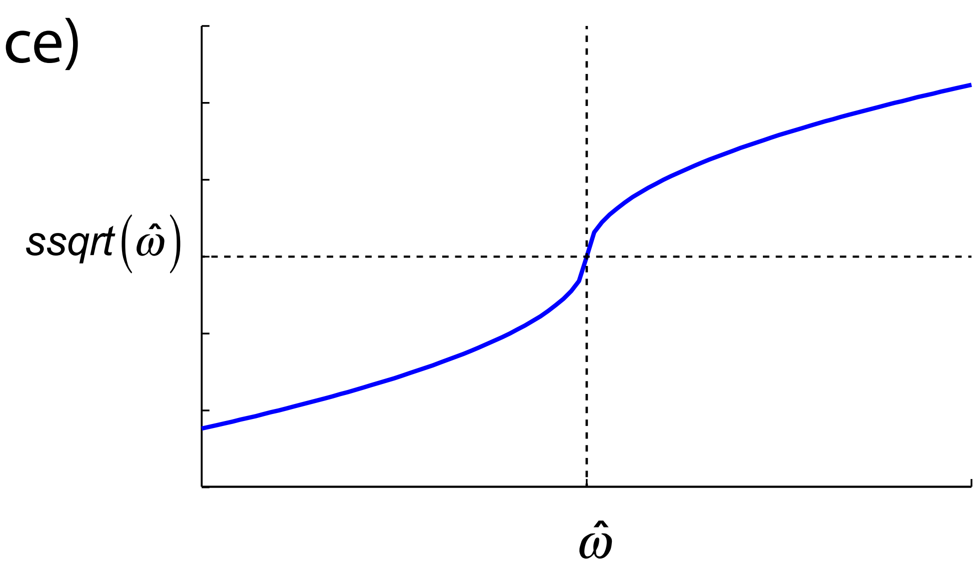
- Covariances and Distances
- Features and representational components
- Factorial models (MANOVA)
- Linear representational models
- Nonlinear representational models
- Summary

# The whole process




Group analysis

- Model coefficients can be directly tested (unbiased)
- Sometimes it is more sensible to use $\sqrt{\hat{\omega}}$ (SD vs. variance)



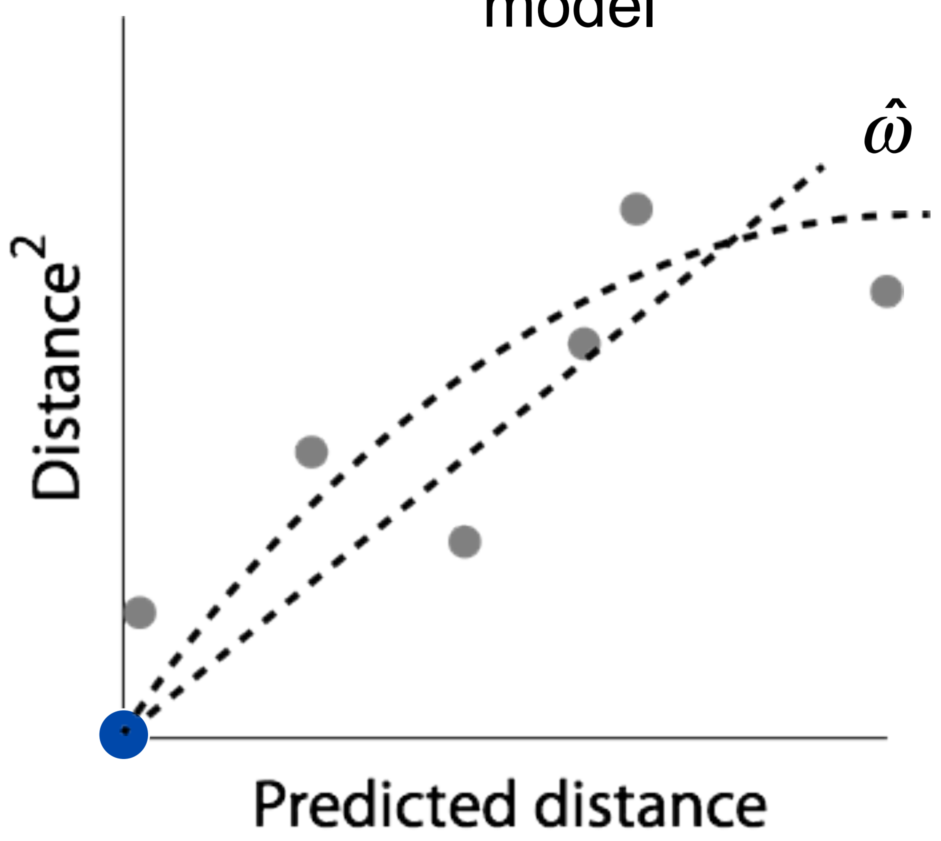
# *Representational component models*

- Representational component models assume
  - independence of data across partitions
    -> (zero-distance is meaningful)
  - independence of feature patterns across components
  - linear overlap of patterns (within small range of variations)

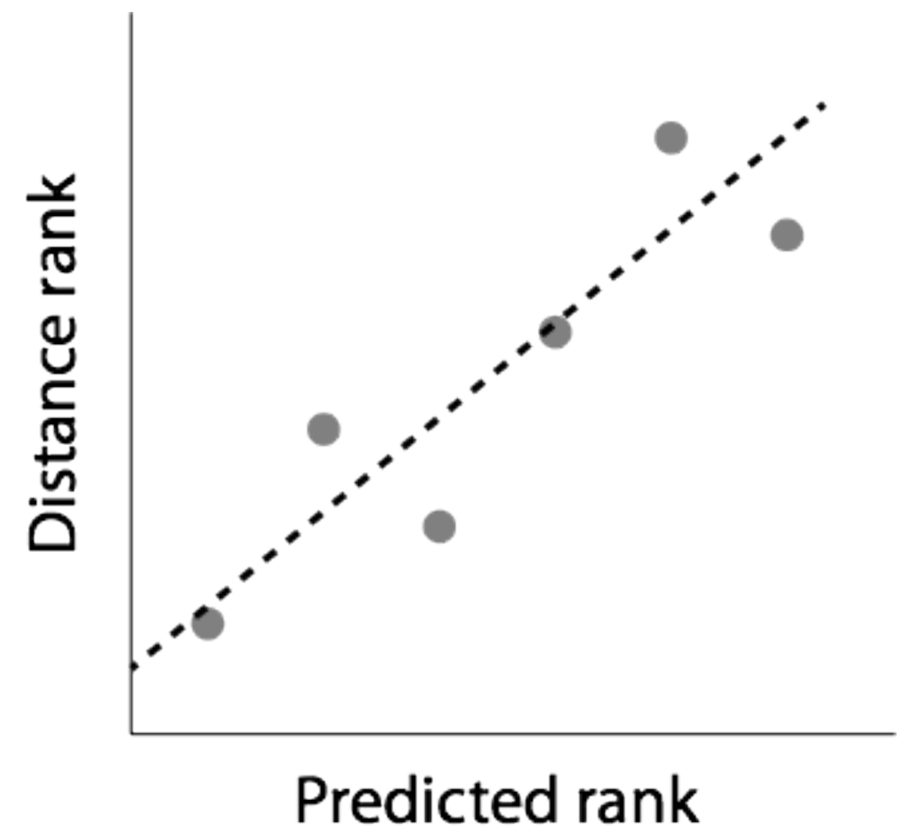
- Representational component models do *NOT* assume
  - normality of the data
  - independence of distance estimates
  - linear relationship between psychological variables and BOLD

# Representational component models

Representational component model




Rank-based RSA




Intercept is not included in fitting
Needs to be explicitly modeled

Predictions on ratio-scale

Flexible factorial and combined models

Linearity assumption (narrow range)

Non-linearity can be modeled

Intercept is implicitly removed
Does not contribute to model comparision

Predictions on ordinal scale

Single models

Non-linearity of distances removed

## *Representational component models: Key insights II*

**E1.** Large (squared Euclidean) distances are estimated with larger variability than smaller distances.

**E2.** Distance estimates are statistically dependent in a way that is determined by the true distance structure.

**E3.** Component weights can be estimated using iteratively reweighted least squares (IRLS), which yields better estimates than ordinary least squares (OLS) in some cases.

# *The end*

Thanks!