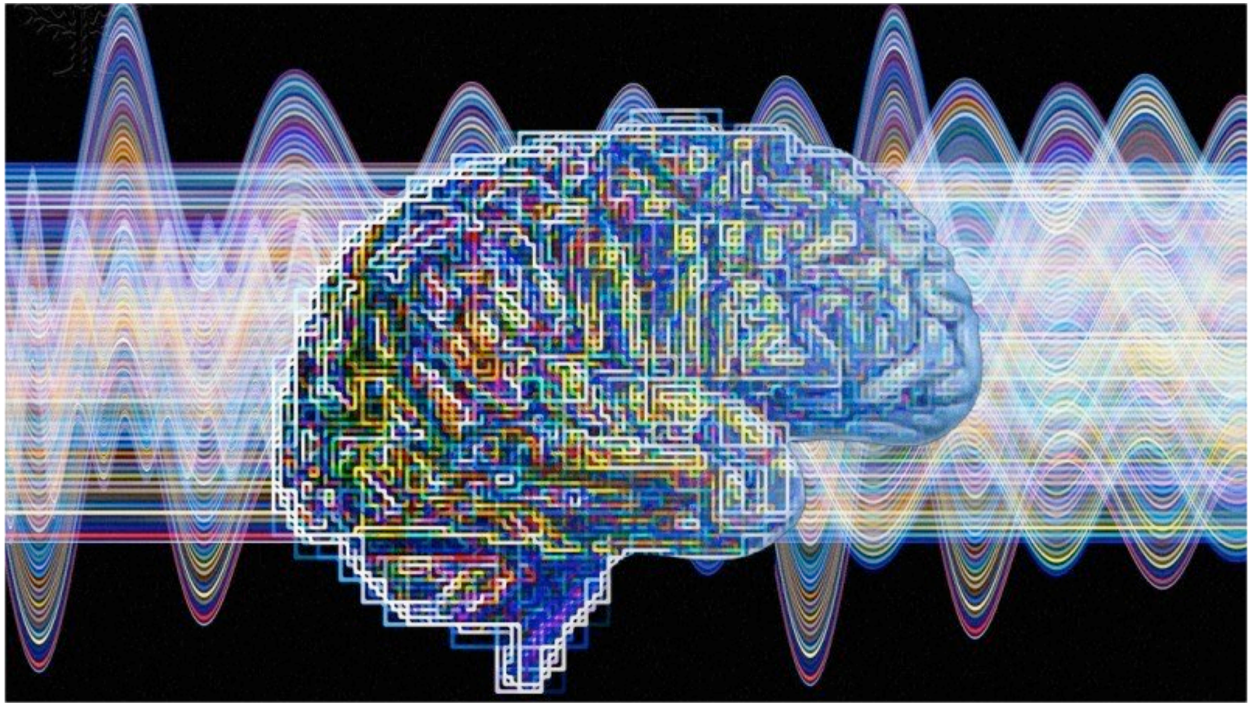
The multivariate partial least squares (PLS) framework for neuroimaging

FU Berlin, Sept 1, 2015



Max-Planck-Institut für Bildungsforschung
Max Planck Institute for Human Development

MAX PLANCK UCL CENTRE
for Computational Psychiatry and Ageing Research

UCL

In search of multivariate data "patterns" in neuroimaging

- Multivariate techniques make using complex imaging data simpler.
  - Leverage vast information to maximize our understanding of phenomena of interest.

- <u>MVPA</u>:
  - Often purported in neuroimaging as a multivariate method, but is often utilized as a <u>univariate</u> technique (e.g., linear discriminant and logistic models are typically *dimensionally* univariate).

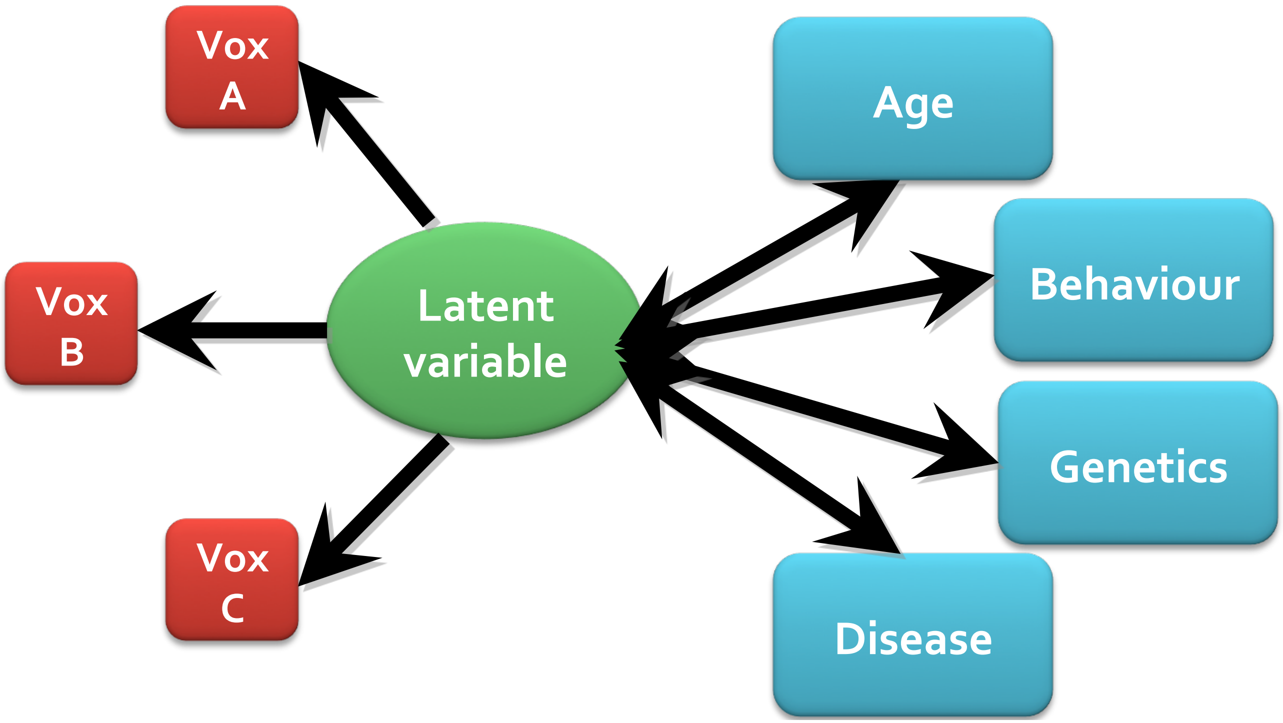
- What about truly multivariate (multidimensional) models?

Partial least squares (PLS)

- Is a general multivariate (multidimensional) statistical method:
  - Whereas MVPA often leverages multiple sources of brain activity to discriminate between discrete classes/groups/states;
  - PLS (McIntosh et al., 1996) is more general in form, allowing researchers to find multivariate, latent-level "patterns" linking brain data to *any other* variables of interest (classes, continuous variables, etc.) in one mathematical step.
  - Can be utilized in the context of EEG, fMRI (block design, event-related), structural MR, PET, network indices, etc.

PLS: What's to gain?

- PLS is an effective way of **reducing dimensionality**
  - Moves us to **latent space:** allows us to capture very complex phenomenon in fewer dimensions than univariate = PARSIMONY



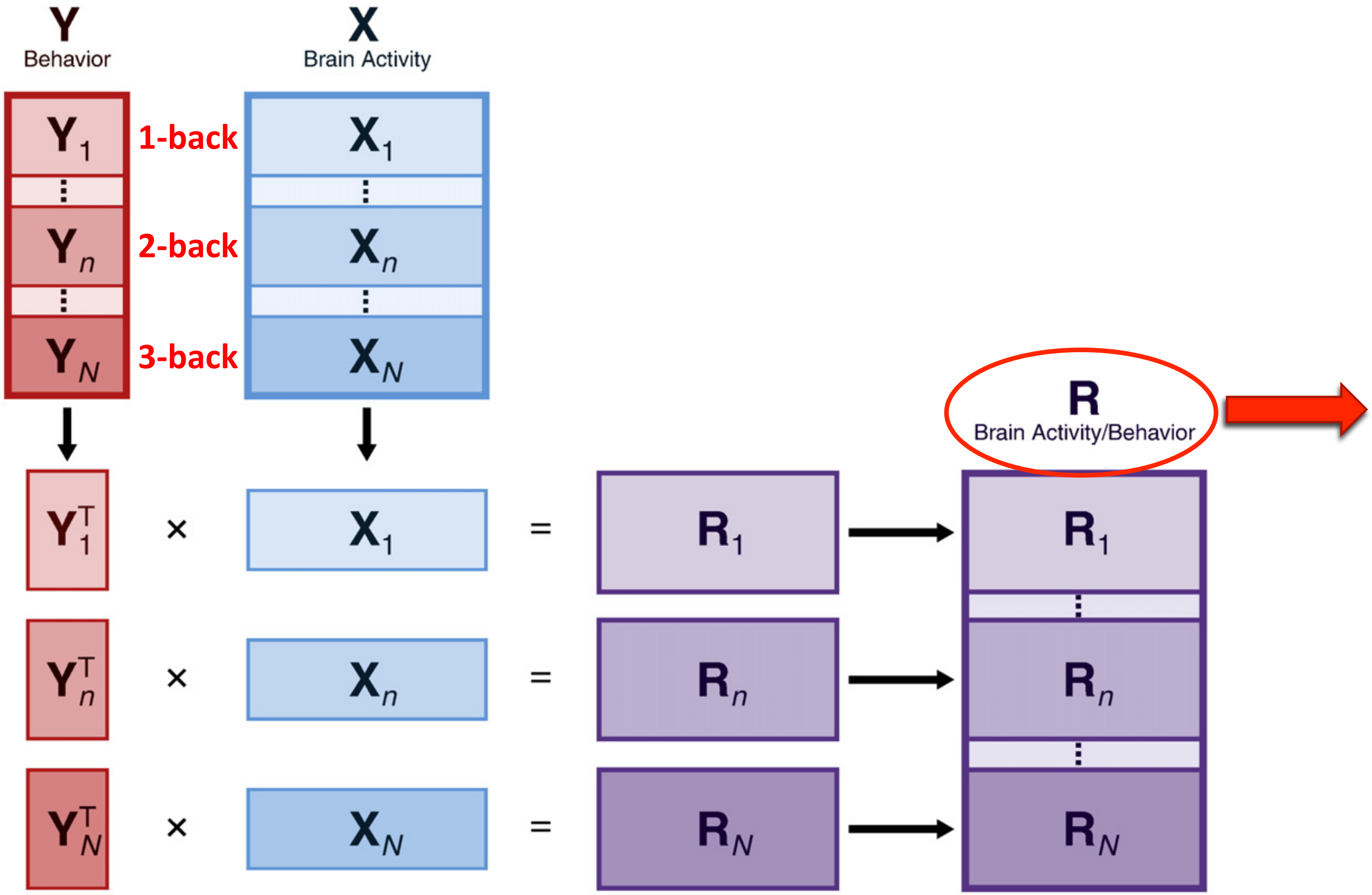
PLS: Details…

- PLS uses **singular value decomposition (SVD)** to decompose large matrices into orthogonal dimensions
  - E.g., linking **$\mathrm{ISD}_{\mathrm{RT}}$ and $\mathrm{mean}_{\mathrm{RT}}$ to fMRI voxel means** in three different task conditions (1-, 2-, 3-back); n=3.

$$
\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \mathbf{X}_3 \end{bmatrix} =
\begin{array}{c} \\ s1 \\ s2 \\ s3 \\ s1 \\ s2 \\ s3 \\ s1 \\ s2 \\ s3 \end{array}
\begin{array}{c} \text{Voxels} \\ \begin{bmatrix}
2 & 5 & 6 & 1 & 9 & 1 & 7 & 6 & 2 & 1 & 7 & 3 \\
4 & 1 & 5 & 8 & 8 & 7 & 2 & 8 & 6 & 4 & 8 & 2 \\
5 & 8 & 7 & 3 & 7 & 1 & 7 & 4 & 5 & 1 & 4 & 3 \\
3 & 3 & 7 & 6 & 1 & 1 & 10 & 2 & 2 & 1 & 7 & 4 \\
2 & 3 & 8 & 7 & 1 & 6 & 9 & 1 & 8 & 8 & 1 & 6 \\
1 & 7 & 3 & 1 & 1 & 3 & 1 & 8 & 1 & 3 & 9 & 5 \\
9 & 0 & 7 & 1 & 8 & 7 & 4 & 2 & 3 & 6 & 2 & 7 \\
8 & 0 & 6 & 5 & 9 & 7 & 4 & 4 & 2 & 10 & 3 & 8 \\
7 & 7 & 4 & 5 & 7 & 6 & 7 & 6 & 5 & 4 & 8 & 8
\end{bmatrix} \end{array}
\begin{array}{c} \\ \text{1-back} \\ \\ \\ \text{2-back} \\ \\ \\ \text{3-back} \\ \\ \end{array}
$$

$$
\mathbf{Y}_{\text{behavior}} = \begin{bmatrix} \mathbf{Y}_{\text{behavior},1} \\ \mathbf{Y}_{\text{behavior},2} \\ \mathbf{Y}_{\text{behavior},3} \end{bmatrix} =
\begin{array}{c} \\ s1 \\ s2 \\ s3 \\ s1 \\ s2 \\ s3 \\ s1 \\ s2 \\ s3 \end{array}
\begin{array}{c} \mathrm{ISD}_{\mathrm{RT}} \quad \mathrm{Mean}_{\mathrm{RT}} \\ \begin{bmatrix}
15 & 600 \\
19 & 520 \\
18 & 545 \\
22 & 426 \\
21 & 404 \\
23 & 411 \\
29 & 326 \\
30 & 309 \\
30 & 303
\end{bmatrix} \end{array}
\begin{array}{c} \\ \text{1-back} \\ \\ \\ \text{2-back} \\ \\ \\ \text{3-back} \\ \\ \end{array}
$$

PLS: Data set up…



Single, stacked R matrix submitted to SVD!

The SVD

- SVD (like PCA, but for rectangular matrices) then produces orthogonal latent variables that optimally express relations between X and Y.

- Singular values rank-ordered by strength.

$$\textbf{SVD of } R_{XY} = U\ S\ V'$$

Condition/group/class/behaviour weights

Singular values

Voxel weights

SVD: How many dimensions are mathematically possible?

- # is always equal to the **smaller** rank of X (e.g., voxel measures) or Y (e.g., behavioural measures/conditions)
  - Ask the audience! How many dimensions possible here (ignoring subjects)?
  - <u>**6**</u> (2 behav measures*3 conditions; brain=12 vox*3 conditions)

Voxels

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \mathbf{X}_3 \end{bmatrix} = \begin{array}{c} \text{s1} \\ \text{s2} \\ \text{s3} \\ \text{s1} \\ \text{s2} \\ \text{s3} \\ \text{s1} \\ \text{s2} \\ \text{s3} \end{array} \begin{bmatrix} 2 & 5 & 6 & 1 & 9 & 1 & 7 & 6 & 2 & 1 & 7 & 3 \\ 4 & 1 & 5 & 8 & 8 & 7 & 2 & 8 & 6 & 4 & 8 & 2 \\ 5 & 8 & 7 & 3 & 7 & 1 & 7 & 4 & 5 & 1 & 4 & 3 \\ 3 & 3 & 7 & 6 & 1 & 1 & 10 & 2 & 2 & 1 & 7 & 4 \\ 2 & 3 & 8 & 7 & 1 & 6 & 9 & 1 & 8 & 8 & 1 & 6 \\ 1 & 7 & 3 & 1 & 1 & 3 & 1 & 8 & 1 & 3 & 9 & 5 \\ 9 & 0 & 7 & 1 & 8 & 7 & 4 & 2 & 3 & 6 & 2 & 7 \\ 8 & 0 & 6 & 5 & 9 & 7 & 4 & 4 & 2 & 10 & 3 & 8 \\ 7 & 7 & 4 & 5 & 7 & 6 & 7 & 6 & 5 & 4 & 8 & 8 \end{bmatrix} \begin{array}{c} \\ \text{1-back} \\ \\ \\ \text{2-back} \\ \\ \\ \text{3-back} \\ \\ \end{array}$$

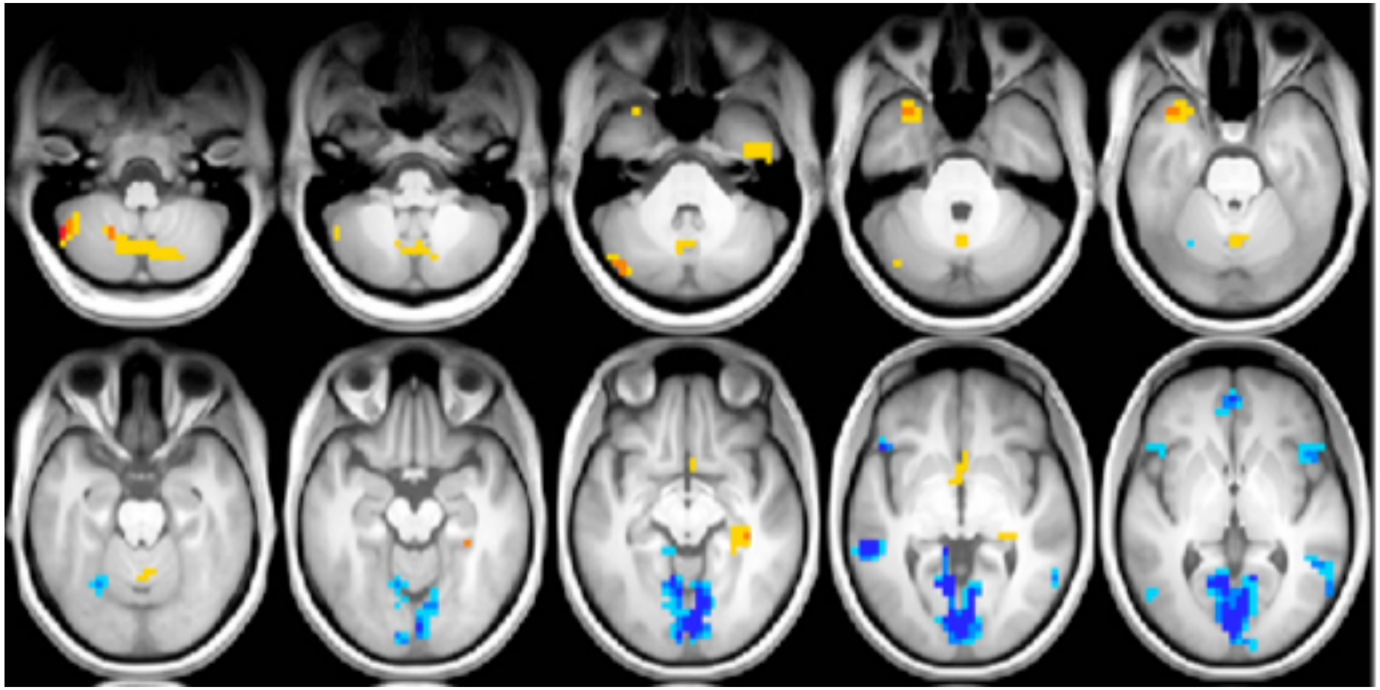
$$\mathbf{Y}_{\text{behavior}} = \begin{bmatrix} \mathbf{Y}_{\text{behavior},1} \\ \mathbf{Y}_{\text{behavior},2} \\ \mathbf{Y}_{\text{behavior},3} \end{bmatrix} = \begin{array}{c|cc} & \text{ISD}_{RT} & \text{Mean}_{RT} \\ \hline \text{s1} & 15 & 600 \\ \text{s2} & 19 & 520 \\ \text{s3} & 18 & 545 \\ \text{s1} & 22 & 426 \\ \text{s2} & 21 & 404 \\ \text{s3} & 23 & 411 \\ \text{s1} & 29 & 326 \\ \text{s2} & 30 & 309 \\ \text{s3} & 30 & 303 \end{array} \begin{array}{c} \\ \\ \text{1-back} \\ \\ \\ \text{2-back} \\ \\ \\ \text{3-back} \\ \\ \end{array}$$

We've gained parsimony, but what else?

- Multiple comparisons problem is minimized.
- "Brute force" of univariate approach not required.
- As in MVPA, brain data are leveraged together, but PLS simply extends this logic to an arguably more flexible model class than typically employed in the MVPA world.

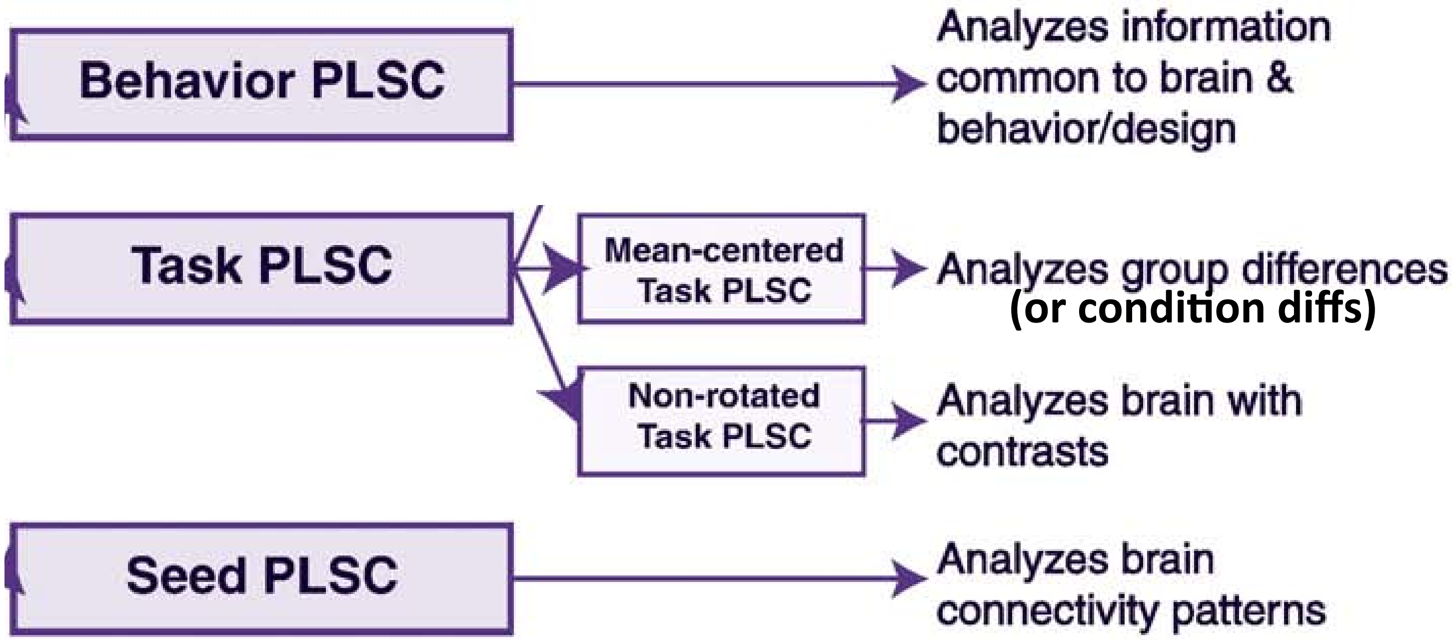
PLS: Another bonus…

- Brain weights can be either positive or negative across voxels within a single dimension
  - Unlike many univariate designs/packages, can visualize both positive and negative voxel effects simultaneously.



PLS: Practical details…

Typical forms of PLS



e.g., brain and RTs

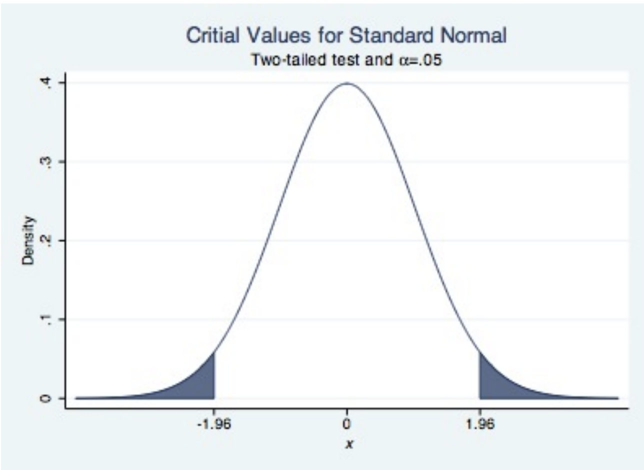
e.g., brain and young/middle/old

e.g., fixation vs. task

e.g., brain and DLPFC seed

Model evaluation

- SVD as described above = **fixed effect model**; so how can we generalize to the population (like random effects models)?
  - Inferential analytical approaches are available, but arguably make too many parametric assumptions to be used routinely across broad types of data. Why assume shape when we can evaluate using **<u>nonparametric</u>** methods?



Two-stage nonparametric model evaluation

1. Dimensions to keep? **Permutation on singular values**
   - A singular value represents the "strength" (like variance accounted for in the data) of a particular dimension of XY relations; how strong is strong enough?
   - We can test the robust strength of each singular value by randomly shuffling rows of X without changing Y, and test whether singular value would be just as high.
   - Run new SVD on this reshuffled dataset, get singular value; do that 1000 times to get distribution of singular values
     - If get singular value as strong as in original data < .05, then keep that dimension!

At voxel level…

- Now that we have chosen which latent variables to keep, we need to "threshold" which brain data (e.g., voxels) to reliably report within those latent variables.

2. Stage 2 then = **<u>Bootstrapping</u>** (with replacement)
   - Reach into sample hat, pull out subject x; put back in, draw again, etc., until have a new "resample" of same size as original.
   - Do that 1000 times, running SVD on each resample
     - Derive bootstrap standard errors for each voxel.
     - Original voxel weight/bootstrap standard error = **<u>BOOTSTRAP RATIO (BSR).</u>** Tells us how probable it would be that this voxel is "really" active, across multiple resamples.
     - Typical choice: BSR=3.00 (akin to z-score).

PLS: Individual differences

- Latent variables chosen, voxels thresholded... but what about the subject level?
  - <u>**“Brain score”**</u>

$$\textbf{SVD of } R_{XY} = U\ S\ V'$$

Condition/ group weights

Singular values

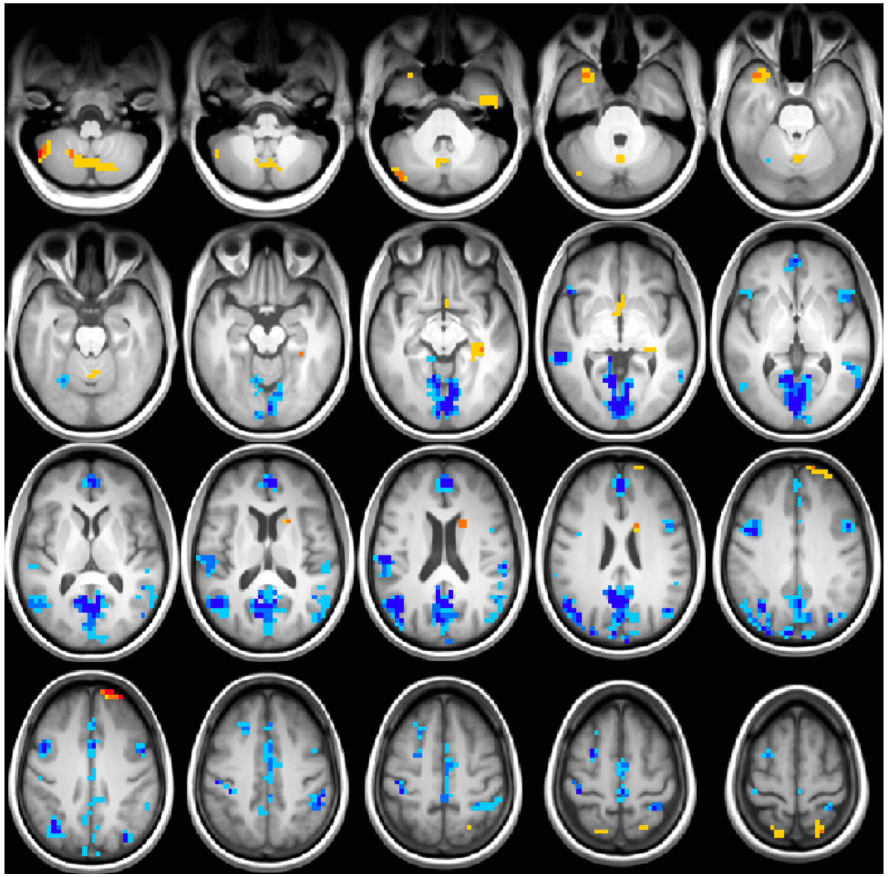
Voxel weights

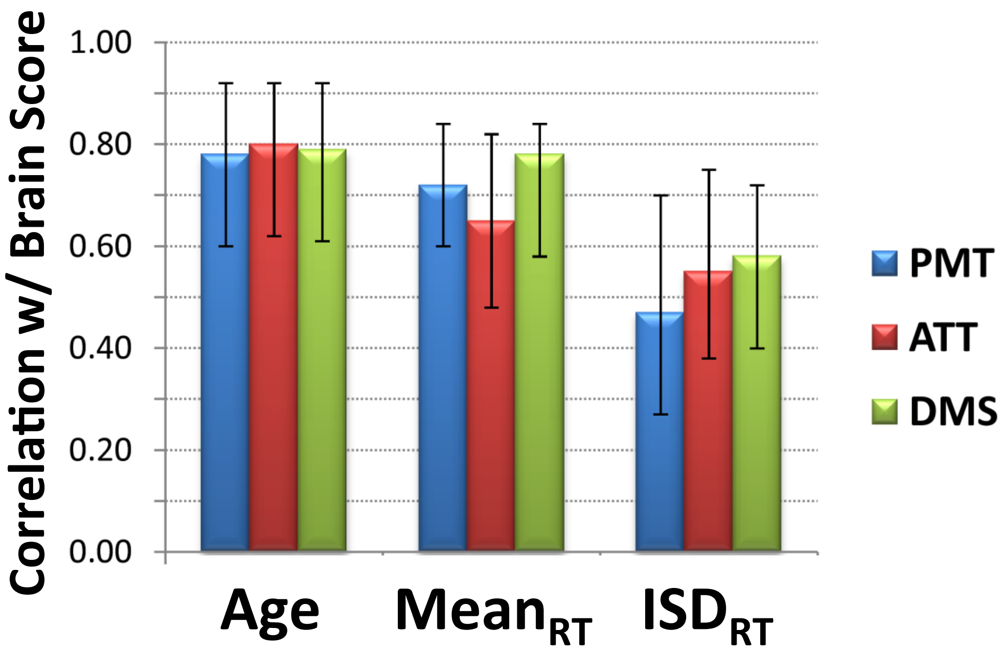
$$\textbf{Brain Score}_i = V_j * \textbf{Voxel Data}'_{ij}$$

PLS: Published examples

Example #1: Linking BOLD signal variability to age, RT_{mean}, RT_{sd} on three cognitive tasks

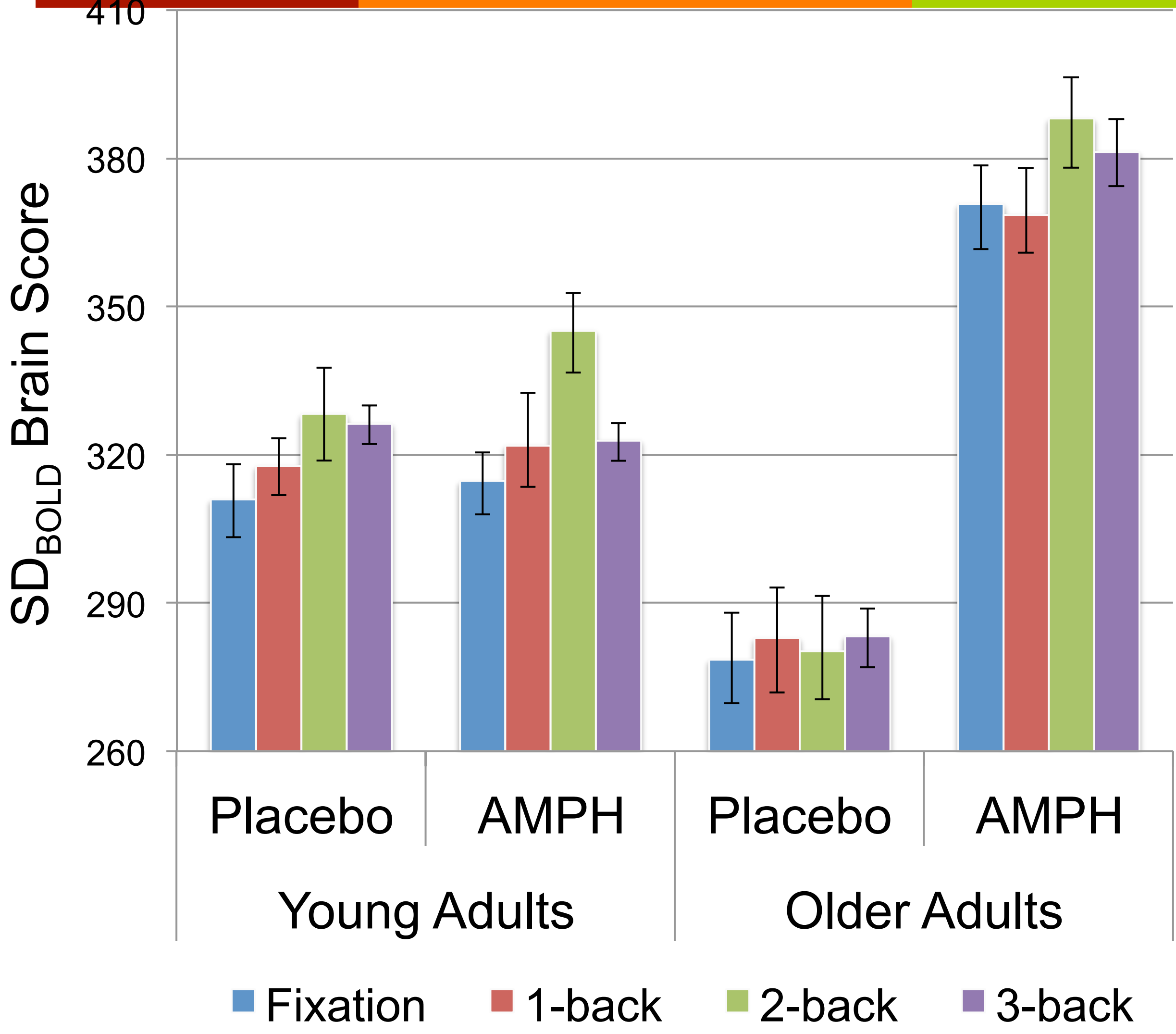
- Young, fast, stable adult performers = higher SD_{BOLD} (single robust LV).

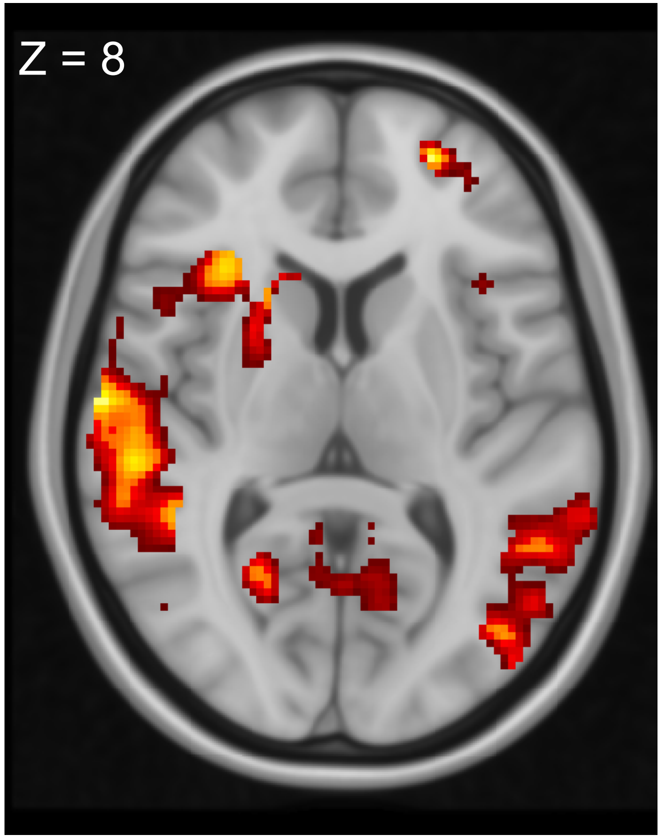




Garrett et al. (2011), *JNeurosci*

Example #2: Can AMPH boost SD_{BOLD} in older adults? (Garrett et al., 2015, *PNAS*)

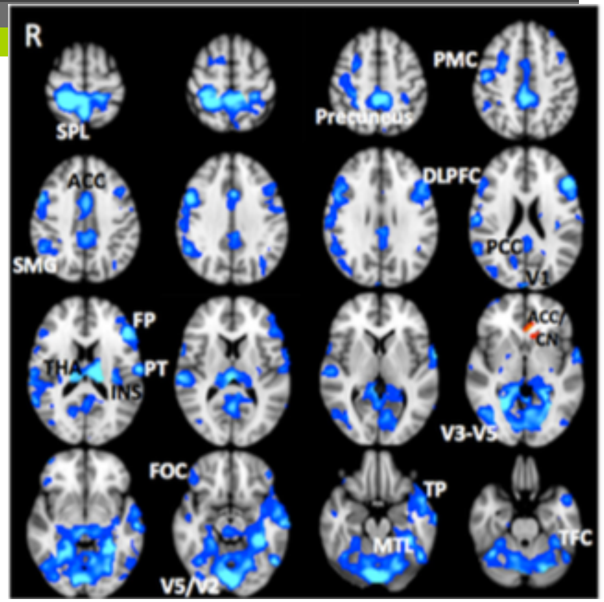


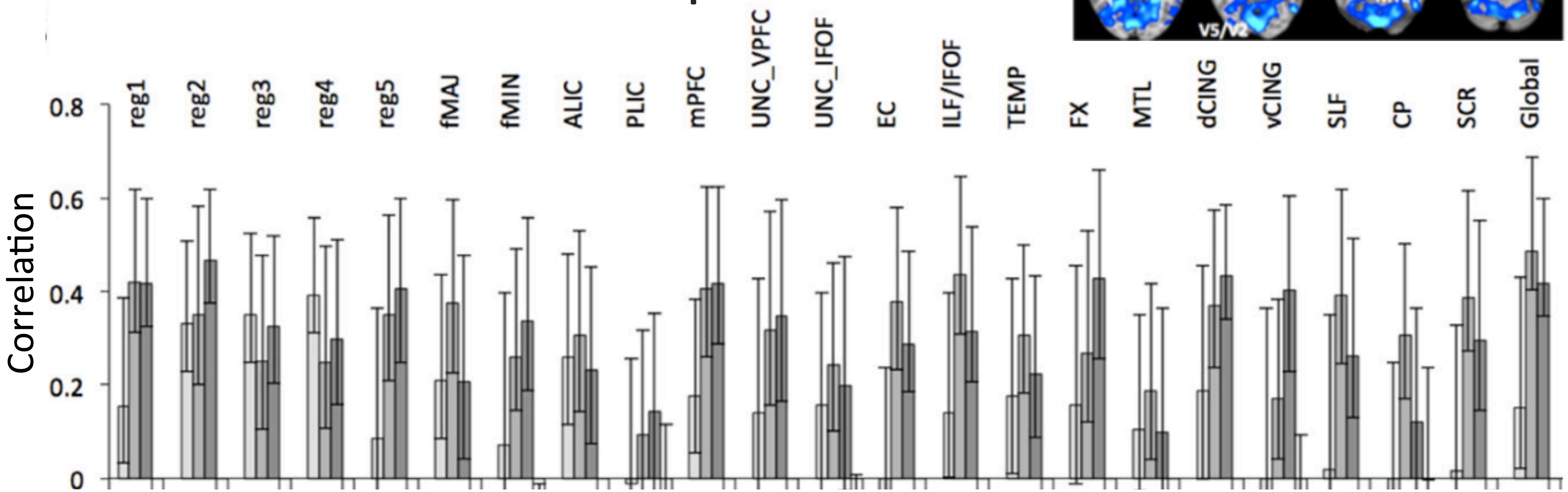


Example #3: Linking FA and mean$_{BOLD}$

(*Burzynska, *Garrett, et al., 2013, *JNeurosci*)

- Single robust LV; Higher FA correlated with lower BOLD during n-back (1-, 2-, 3-back)
  - Higher FA, lower n-back BOLD, better behavioral performance.





Summary

- PLS is a useful technique for reducing data dimensionality and finding “patterns,” and can capture broad scale relations between any brain data and any other variables of interest.
- Given non-parametric assumptions, there are no real bounds on the types of variables that can be examined.
- Can be either exploratory or hypothesis driven; it does what you tell it to do!

Many thanks to all!

- McIntosh PLS software: Google "PLS Rotman…" and the latest release will come up.

- Primary intro references:

1. Krishnan, A., Williams, L. J., McIntosh, A. R., & Abdi, H. (2011). Partial Least Squares (PLS) methods for neuroimaging: a tutorial and review. *NeuroImage*, *56*(2), 455–475.
2. McIntosh, A. R., & Lobaugh, N. J. (2004). Partial least squares analysis of neuroimaging data: applications and advances. *NeuroImage*, *23 Suppl 1*, S250–63.
3. McIntosh, A. R., Bookstein, F. L., Haxby, J. V., & Grady, C. L. (1996). Spatial pattern analysis of functional brain images using partial least squares. *NeuroImage*, *3*(3 Pt 1), 143–157.
4. McIntosh, A. R., Chau, W. K., & Protzner, A. B. (2004). Spatiotemporal analysis of event-related fMRI data using partial least squares. *NeuroImage*, *23*(2), 764–775.