Multi-Table Models for Connectivity Analysis

Parts One: DISTATIS

Hervé Abdi

The University of Texas at Dallas

With a lot of help … from my friends

- Part I:

Douglas Garret, **Cheryl Grady**, Natasha Kovacevic, Anjali Krishnan, Randy McIntosh, Dominique Valentin, Lynne Williams

Overview

- Part I
  - What to do with correlation based connectivity
  - MDS, CORSTATIS, R_V-PCA, etc.
  - Illustration: Age related changes in connectivity

References

Part I

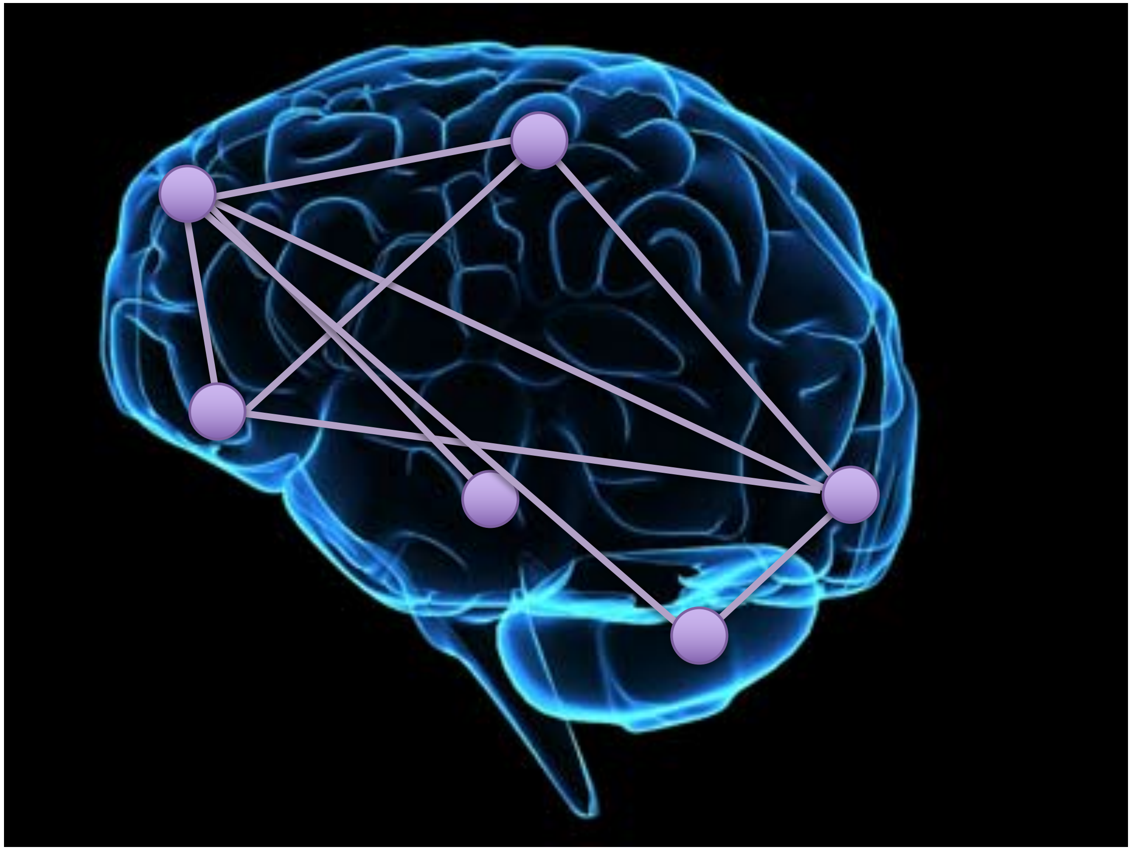
- Abdi, H., Williams, L.J., & Valentin, D. (2013). Multiple factor analysis: Principal component analysis for multi-table and multi-block data sets. *Wiley Interdisciplinary Reviews: Computational Statistics*, **5**, 149-179
- Abdi, H., Williams, L.J., Valentin, D., & Bennani-Dosse, M. (2012). STATIS and DISTATIS: Optimum multi-table principal component analysis and three way metric multidimensional scaling. *Wiley Interdisciplinary Reviews: Computational Statistics*, **4**, 124-167.
- Abdi, H., Dunlop, J.P., & Williams, L.J. (2009). How to compute reliability estimates and display confidence and tolerance intervals for pattern classifiers using the Bootstrap and 3-way multidimensional scaling (DISTATIS). *NeuroImage*, **45**, 89-95.

Part I

Part I:
What to do with correlation based connectivity matrices?

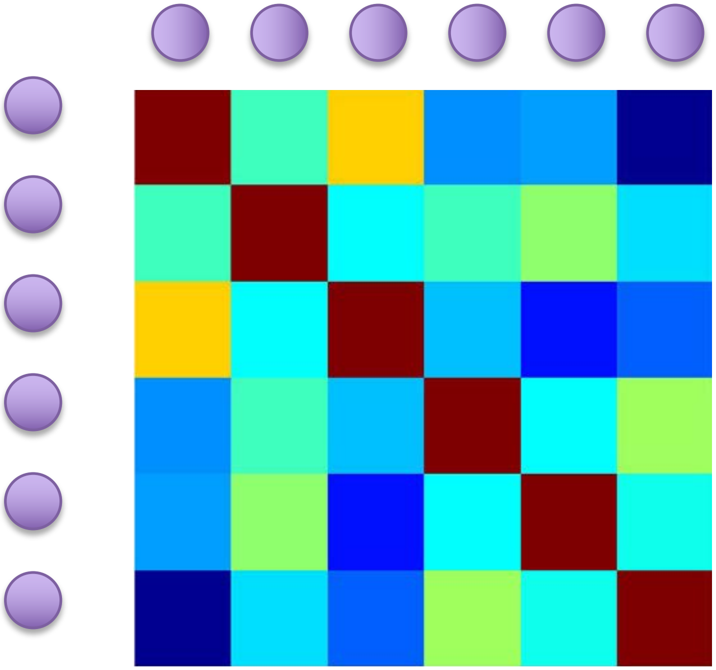
What to do with correlation based connectivity matrices? What we start with (ROIs)

What to do with correlation based connectivity matrices? What we want (connections).



What to do with correlation based connectivity matrices?

What do we have (today)? Correlations.



We want a nice picture of the ROIs and their relationships

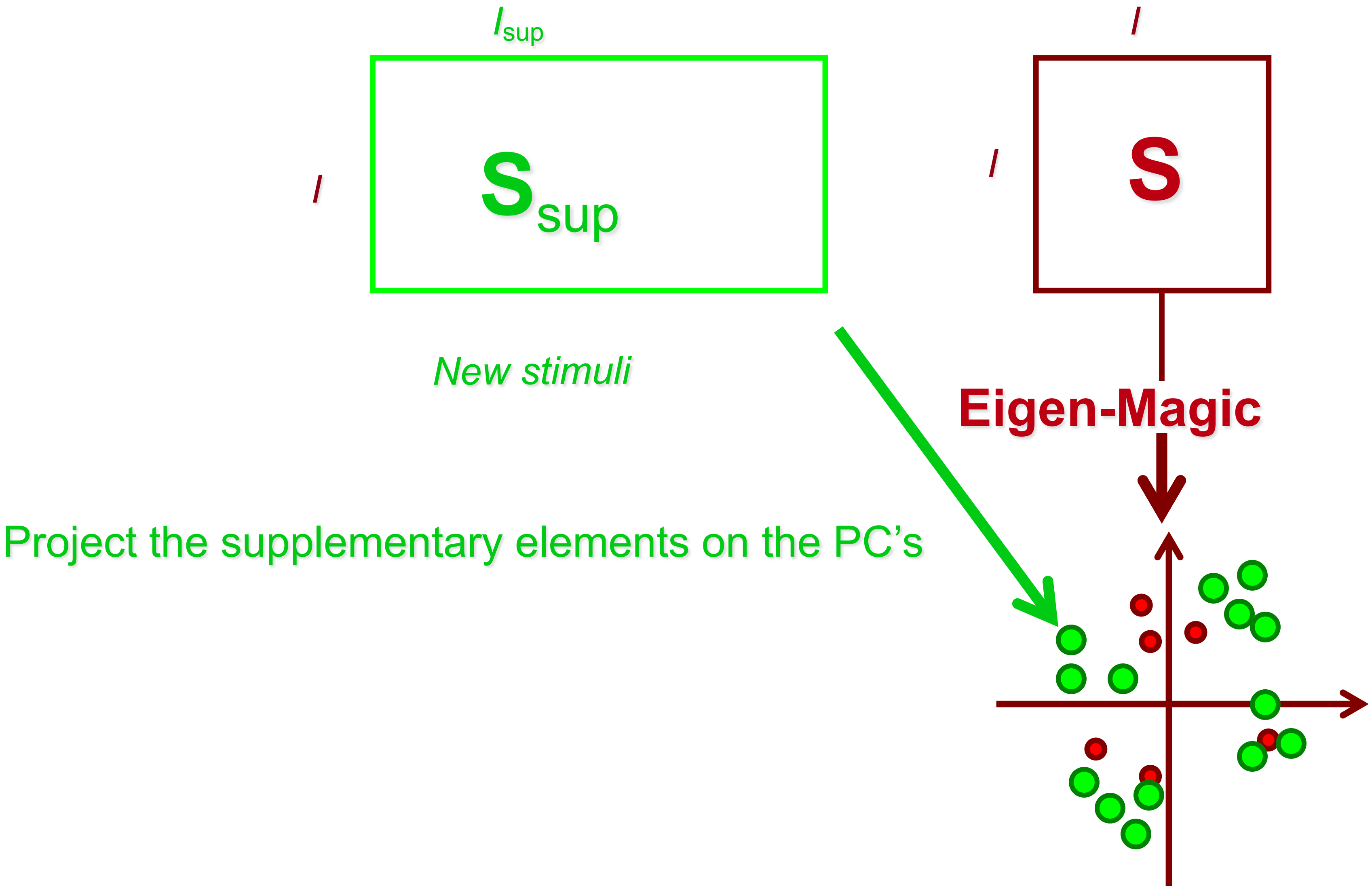
Looks like a problem for: Euclid's Multivariate Tool Kit

THE BEAUTY OF EUCLIDE

- It works with the Pythagorean theorem!
- Means, Variance, Inertia are *natural*.
- We love to minimize sum of squares
- (… and for that the eigenworld is *magic*).
- Good routines means very large data sets.
- Generalizes beautifully (masses, weights, …)
- And if it is multivariate normal we are *in Paradise!*

The beauty of Euclide … for correlation

- *I* by *I* (Similarity/Covariance) data sets: MDS/PCA
- Here: Coordinates are *Eigenvectors*.
- Variance of coordinates are *Eigenvalues*.



Refresher: Supplementary (illustrative) elements

Standard Idea: Metric (Classical) Multidimensional Scaling (MDS) a.k.a Principal Coordinate Analysis

- Simple Idea:
- Transform Correlation into a Distance

  $d = 1 - r$
- and go for MDS

MDS?
Refresher: Analysis of a Distance Table

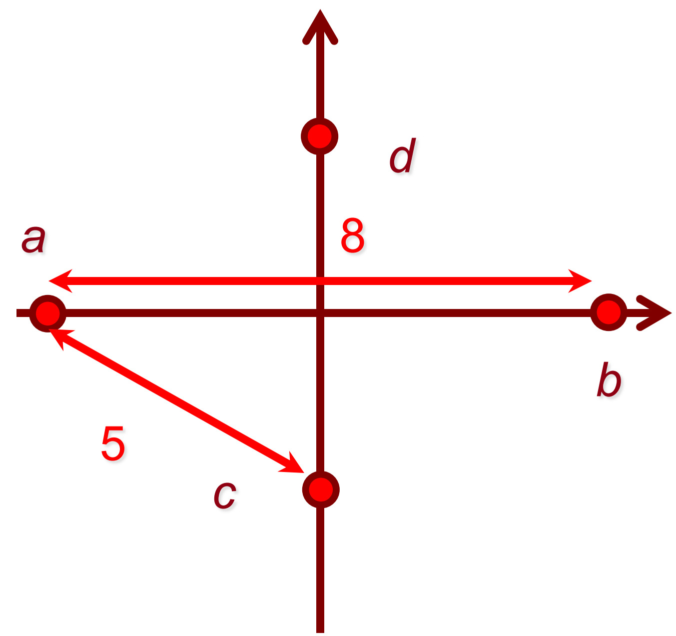
- Metric Multidimensional Scaling MDS
  (Young-Householder, 1938; Torgerson 1958)
- Principal Coordinate Analysis (Gower, 1962)

Representing a distance matrix as a map: MDS

	a	b	c	d
a	0	8	5	5
b	8	0	5	5
c	5	5	0	6
d	5	5	6	0

Some

Magic



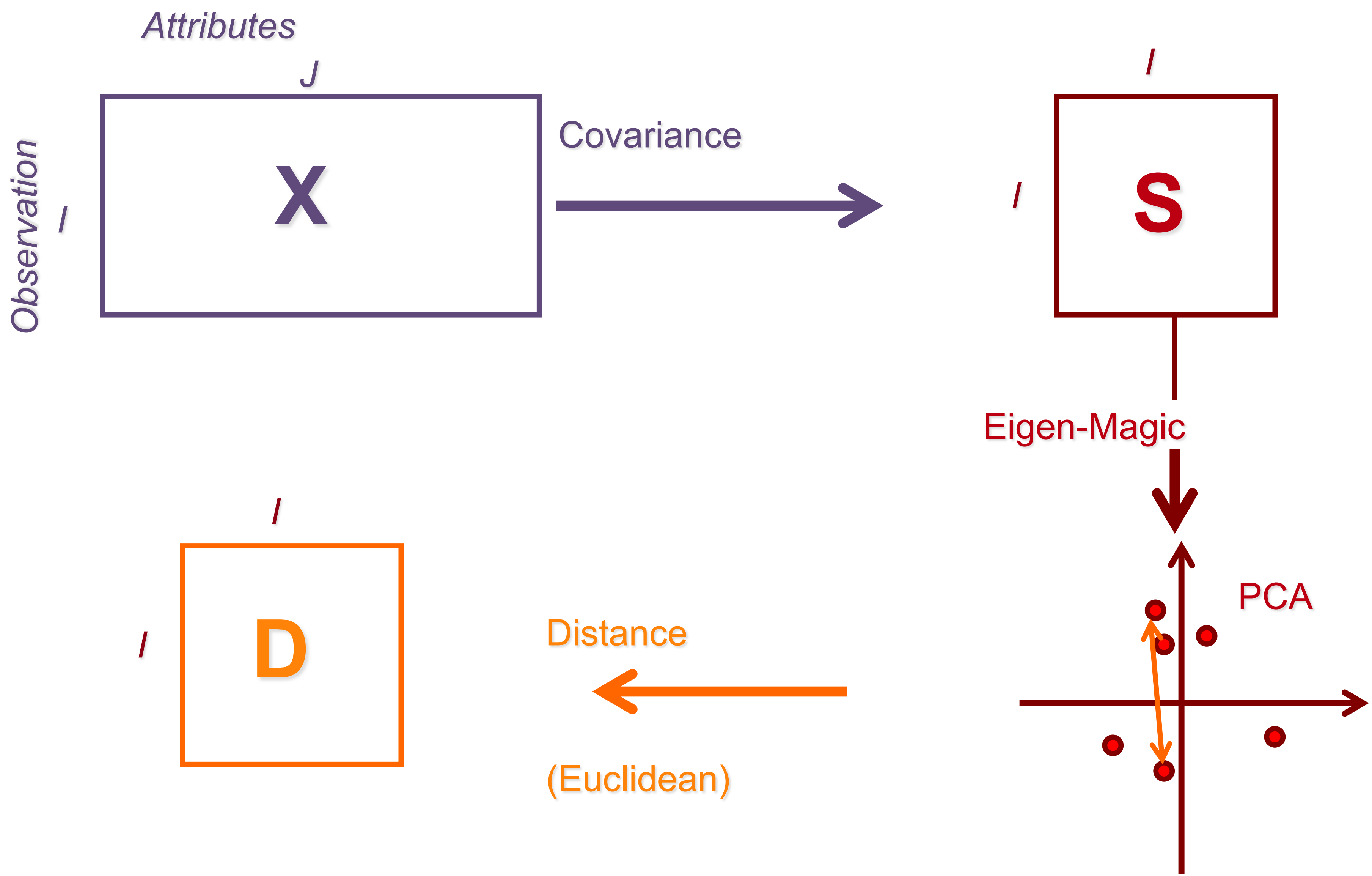
$d(a,a) = 0$

$d(a,b) > 0$

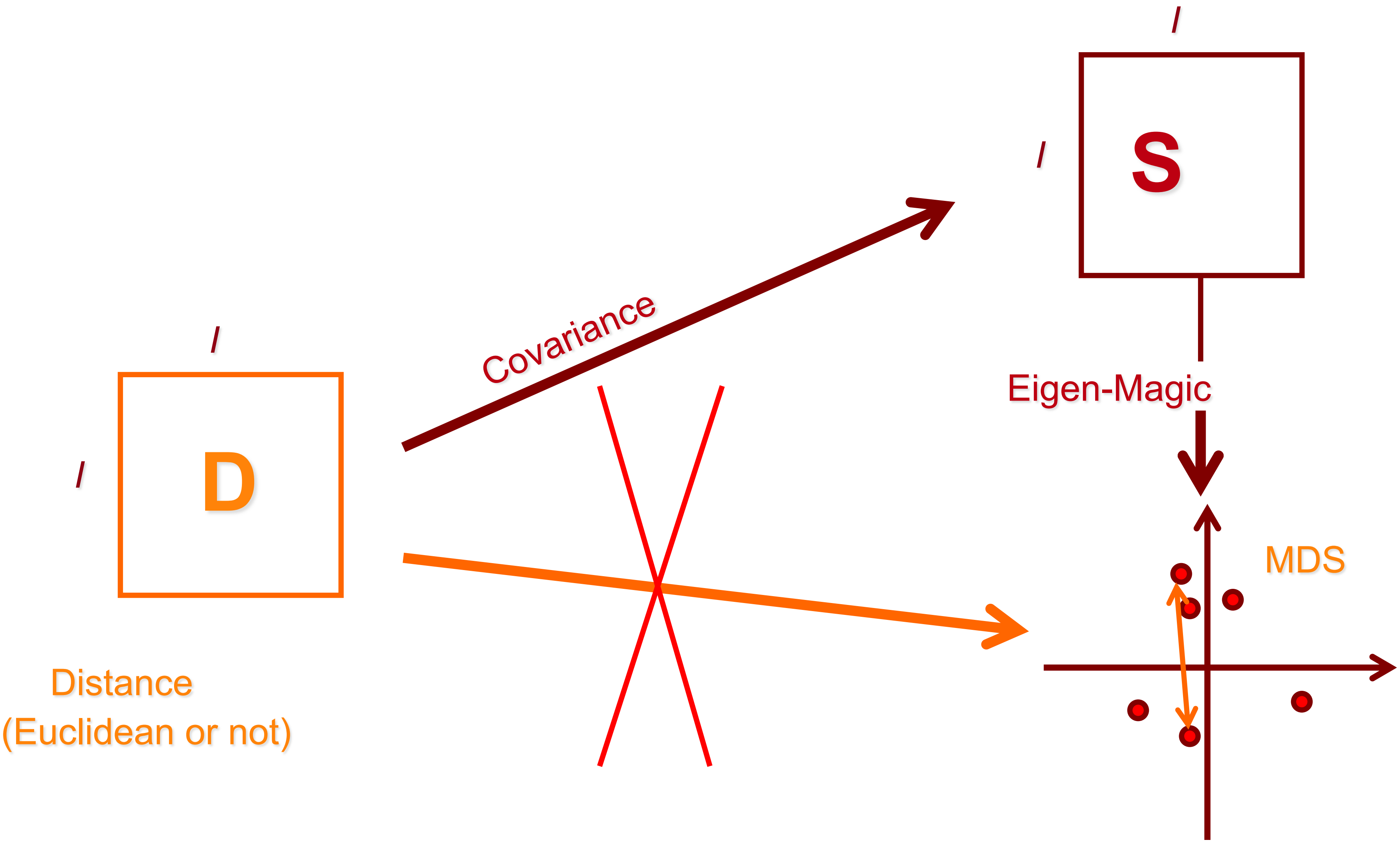
$d(a,c) \leq d(a,b) + d(b,c)$

How to do it?

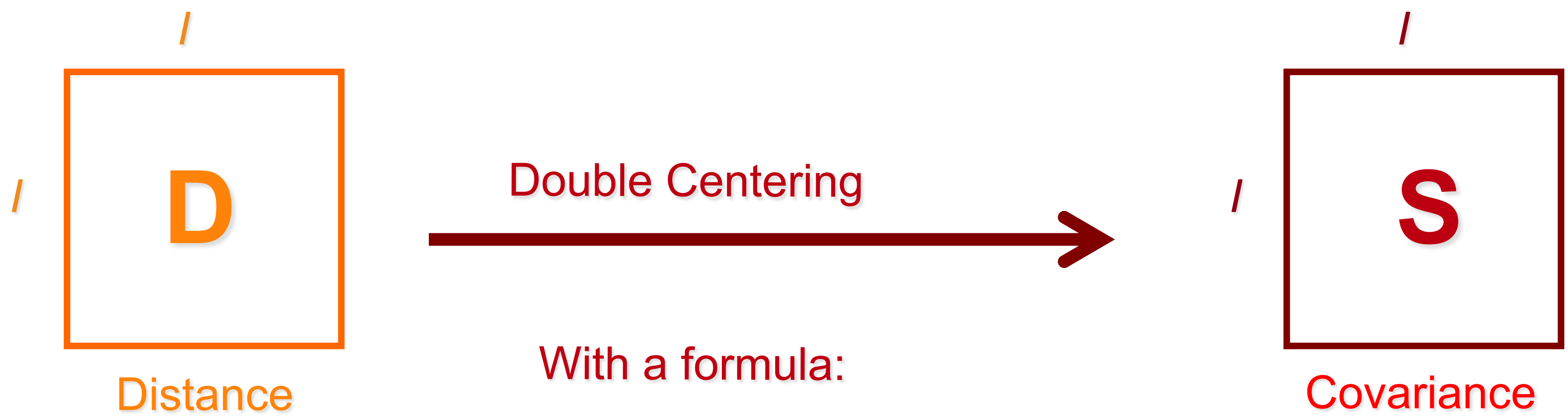
- A PCA to MDS refresher



From PCA to MDS: 1 PCA ...



. . . From (metric) MDS to PCA . . . to MDS



$$s_{i,j'} = -\tfrac{1}{2}[(d_{i,j'} - Md_{+,+}) - (Md_{i,+} - Md_{+,+}) - (Md_{+,j'} - Md_{+,+})]$$
$$= -\tfrac{1}{2}[d_{i,j'} - Md_{i,+} - Md_{+,j'} + Md_{+,+}]$$

With matrices:

$$\mathbf{S} = -.5\,\Xi\mathbf{D}\Xi^{T} \text{ with } \Xi = \mathbf{I} - \mathbf{1}\mathbf{m}^{T} \text{ and } \mathbf{m}^{T}\mathbf{1} = 1$$

Transform the Distances into Covariances

And then? Eigen-Magic

- Eigen-decomposition = Factor scores
- Problem solved!

Sounds Silly!

- No need to transform correlations into distances:

  simply

  go for the eigen-decomposition of the correlations

Right?

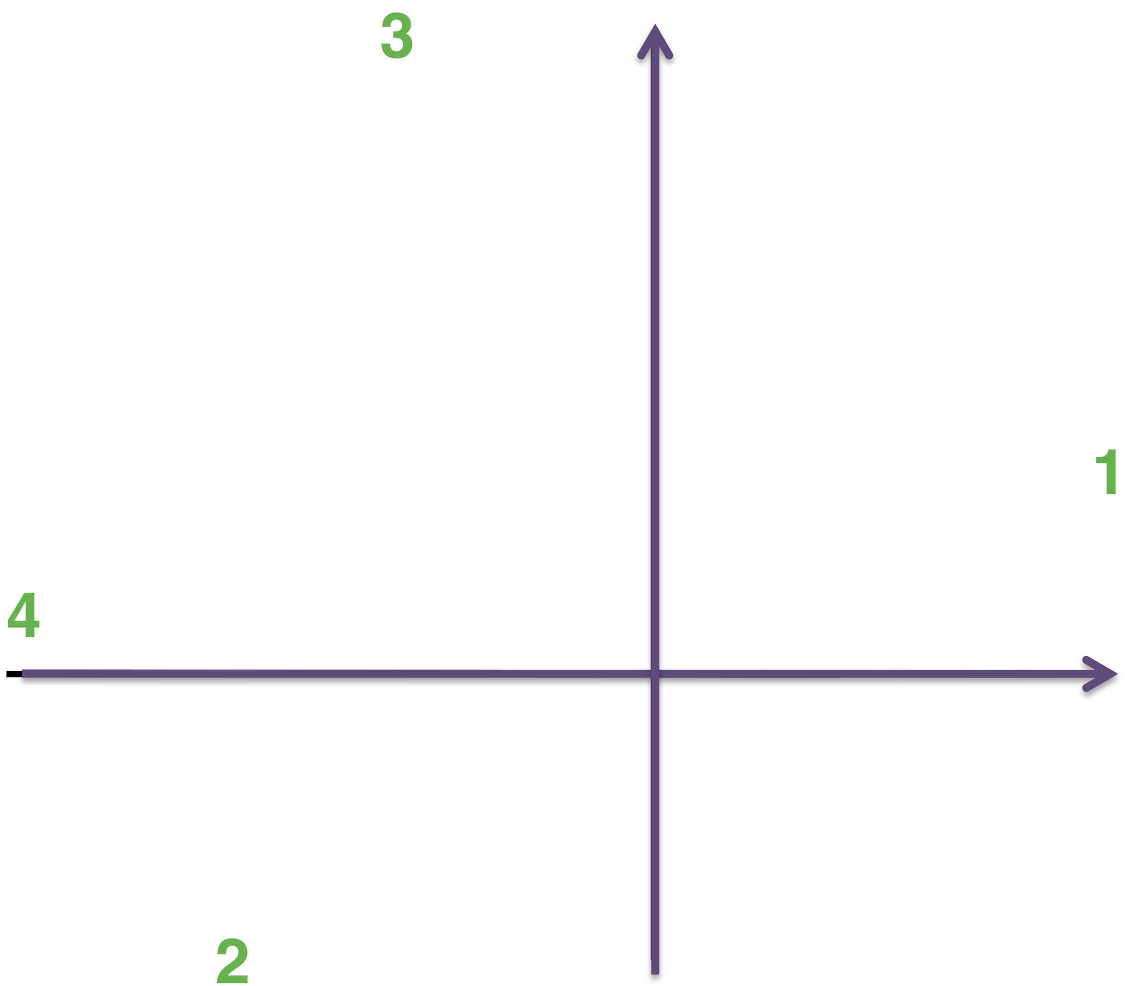
- Here is a connectivity correlation matrix

	1	2	3	4
1	1	-.10	.01	-.42
2	-.10	1	-.03	.43
3	.01	-.03	1	.33
4	-.42	.43	.33	1

Some eigen-magic and then a plot!

Right?

- The connectivity correlation matrix … and its map



Looks good for one Matrix. What about two matrices?

Integrating two correlation matrices

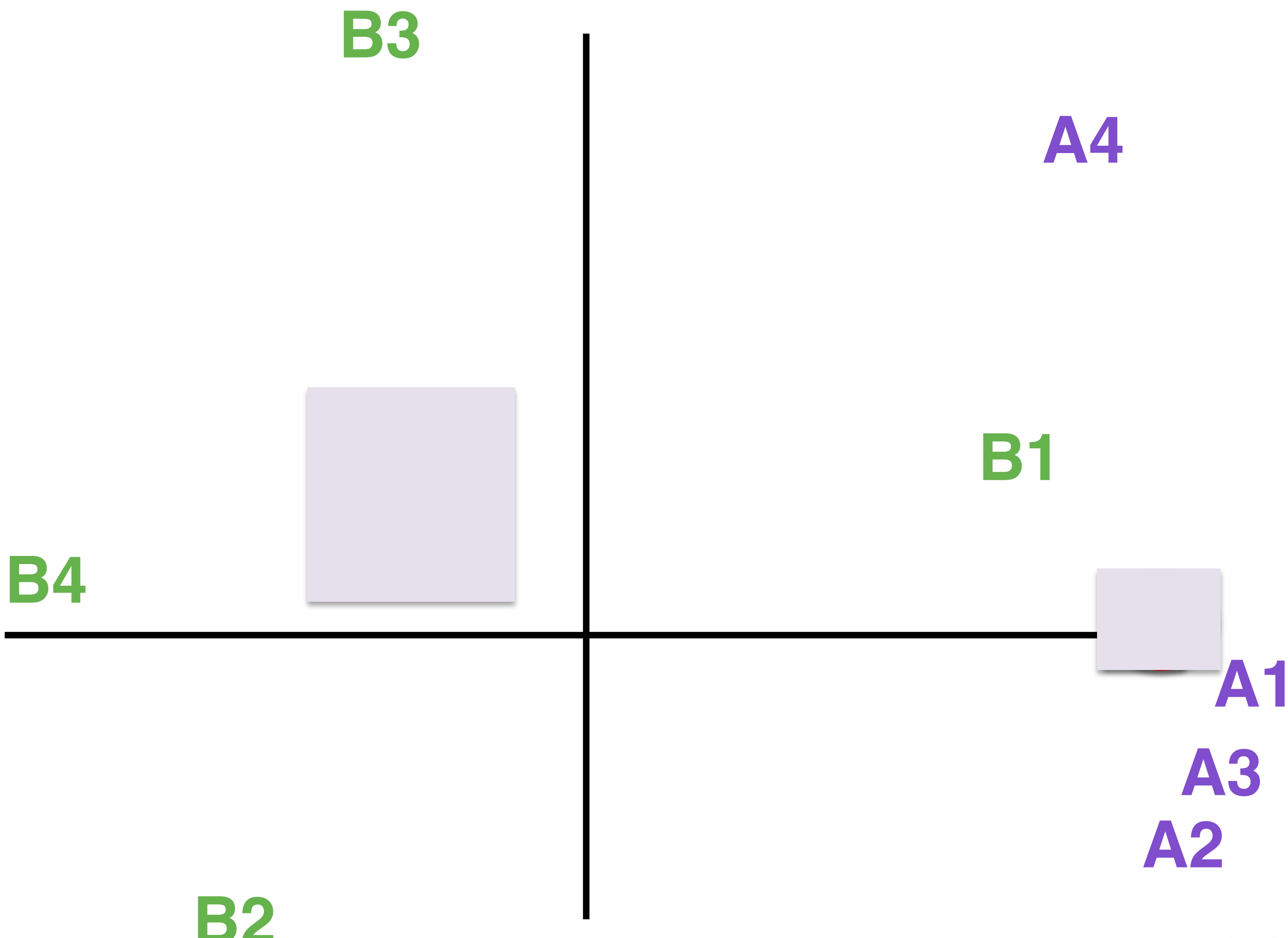
Two correlation matrices

	1	2	3	4
1	1	.69	.79	.54
2	.69	1	.62	.37
3	.79	.62	1	.43
4	.54	.37	.43	1

	1	2	3	4
1	1	-.10	.01	-.42
2	-.10	1	-.03	.43
3	.01	-.03	1	.33
4	-.42	.43	.33	1

Two eigen-decompositions plot them together

Where are the centers?



How to get rid of centers: (double) centering.

- BTW. Not a new idea:
- Horst P. (1965). *Factor Analysis of Data Matrices.* New York: Holt. ("Ipsative Measurements")

I

I **R** — Double Centering → I **S**

Correlation

With a formula:

$$s_{i,j'} = ½[(r_{i,j'} - Mr_{+,+}) - (Mr_{i,+} - Mr_{+,+}) - (Mr_{+,j'} - Mr_{+,+})]$$

$$= ½[r_{i,j'} - Mr_{i,+} - Mr_{+,j'} + Mr_{+,+}]$$

With matrices:

$$\mathbf{S} = .5\boldsymbol{\Xi}\mathbf{R}\boldsymbol{\Xi}^{T} \text{ with } \boldsymbol{\Xi} = \mathbf{I} - \mathbf{1}\mathbf{m}^{T} \text{ and } \mathbf{m}^{T}\mathbf{1} = 1$$

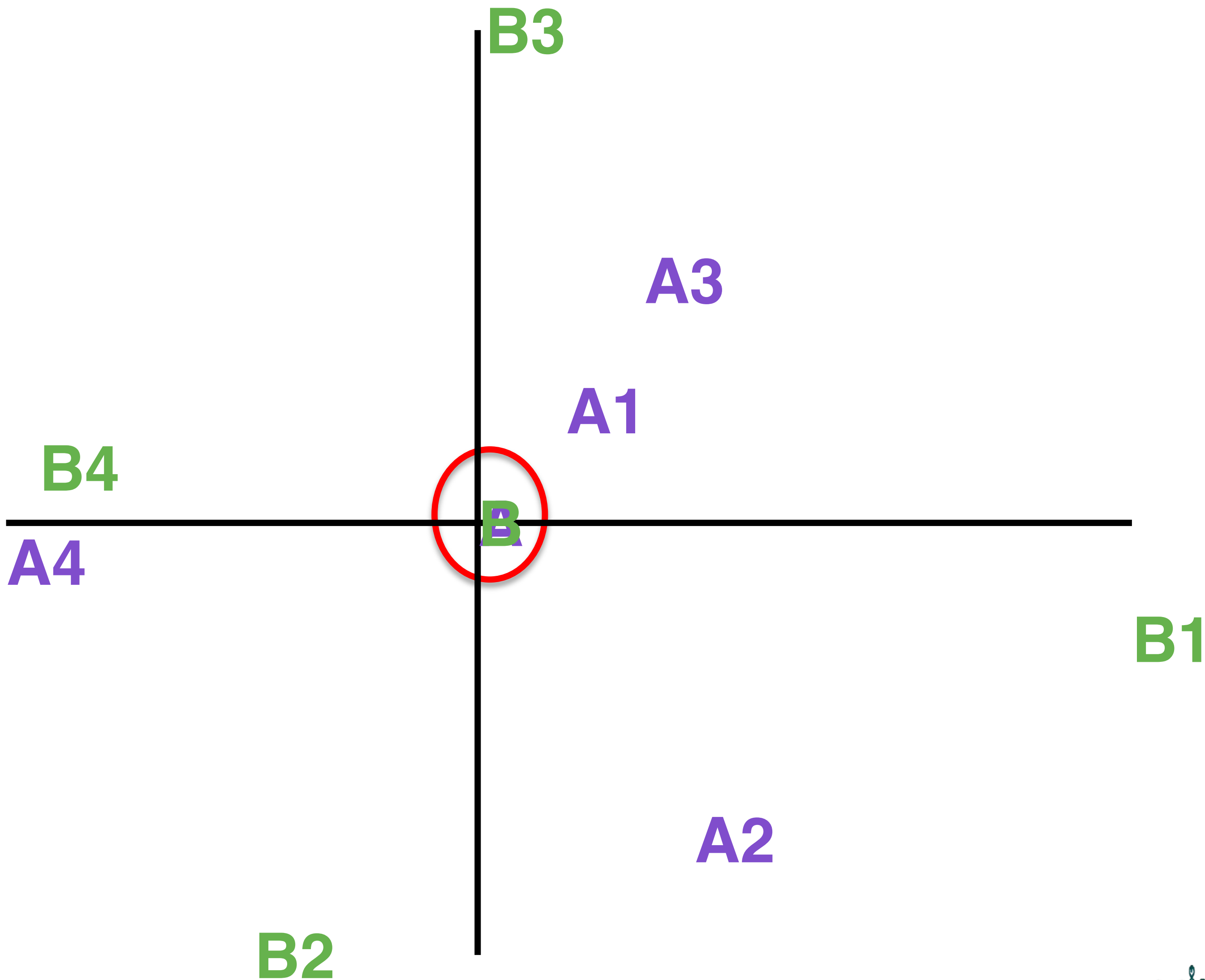
Center the correlation matrix (note: no "–")

The centered matrices

	1	2	3	4
1	.17	-.06	.01	-.12
2	-.06	.34	-.08	-.20
3	.01	-.08	.26	-.19
4	-.12	-.20	-.19	.51

	1	2	3	4
1	1.03	-.27	-.16	-.60
2	-.27	.63	-.41	.05
3	-.16	-.41	.62	-.05
4	-.60	.05	-.05	.60

… and their plot

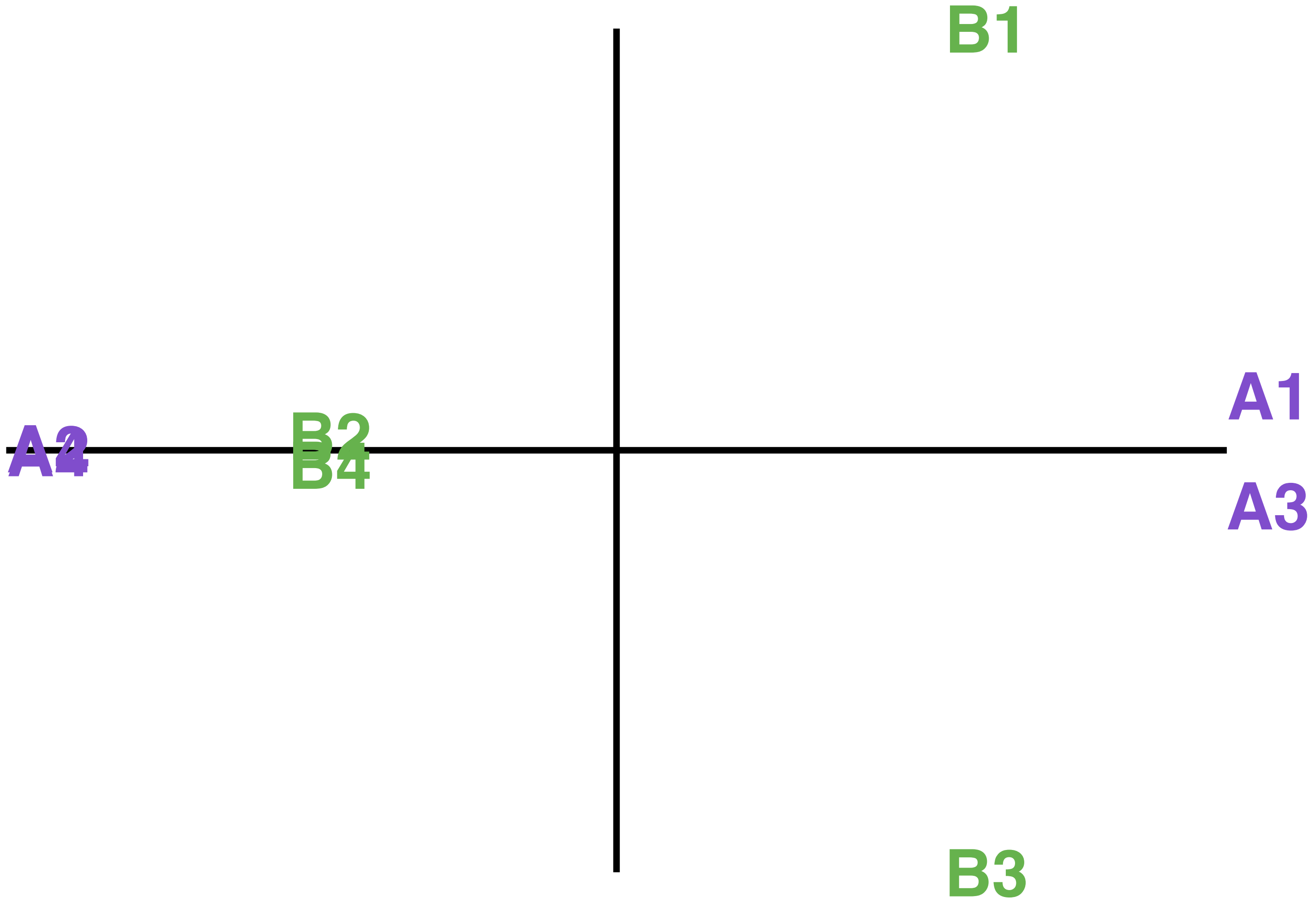


Integrating centered correlation matrices. Is it enough?

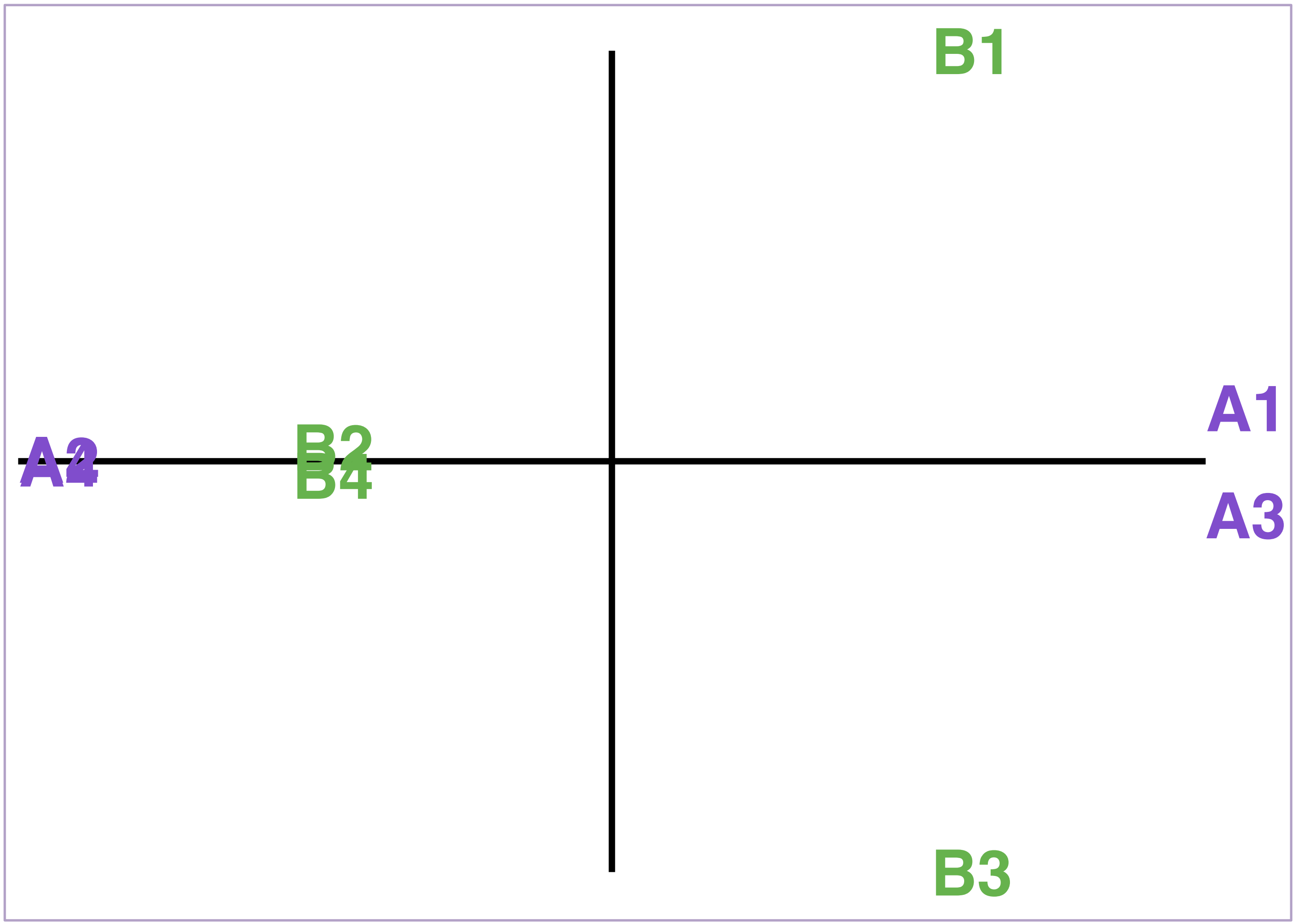
Two new centered matrices

	1	2	3	4
1	.50	-.49	.49	-.49
2	-.49	.50	-.49	.49
3	.49	-.49	.50	-.49
4	-.49	.49	-.49	50

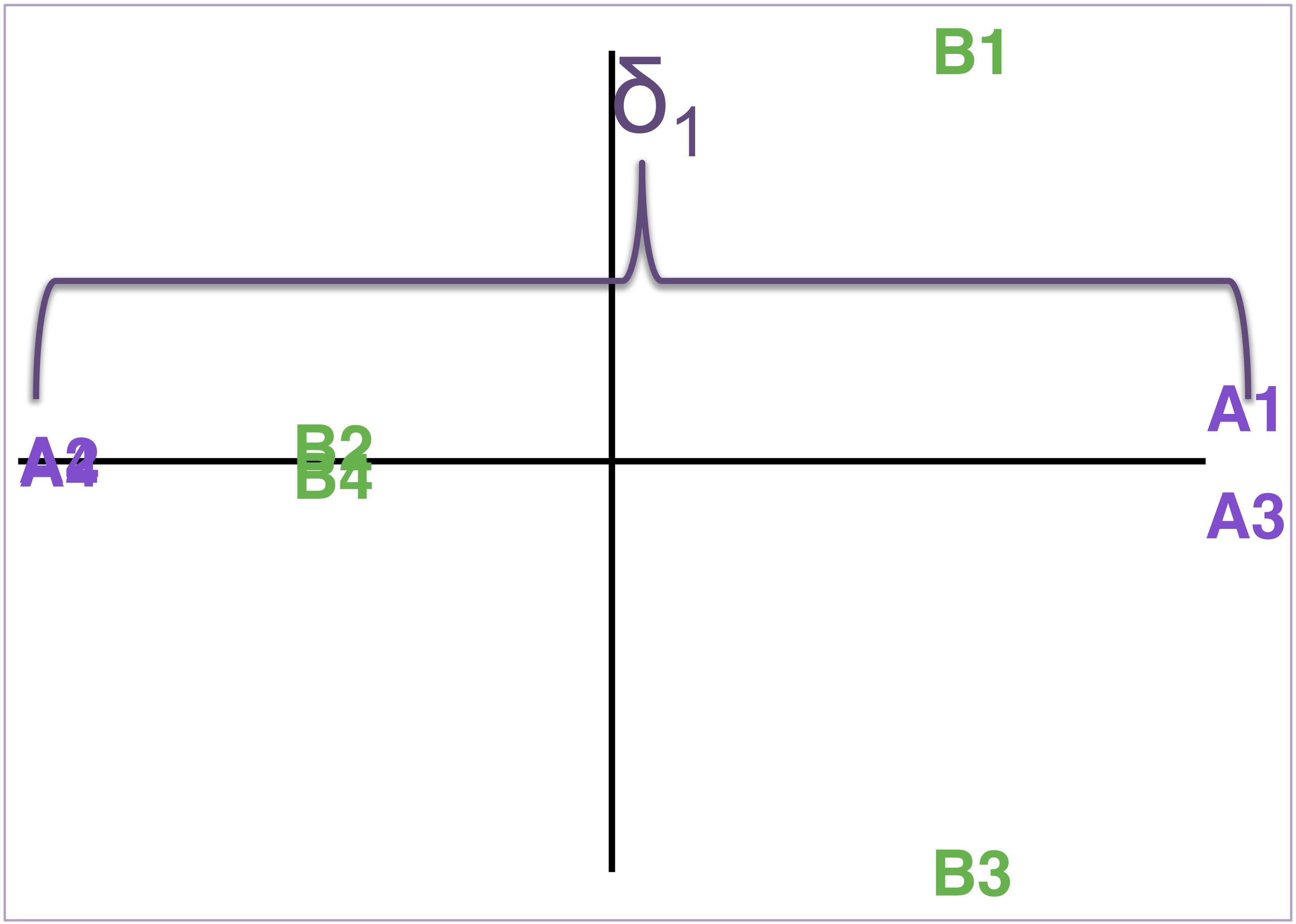
	1	2	3	4
1	.38	-.14	.09	-.14
2	-.14	.38	-.14	-.09
3	.09	-.14	.38	-.14
4	-.14	-.09	-.14	.38


B1
A1
B2
B4
A2
A4
A3
B3

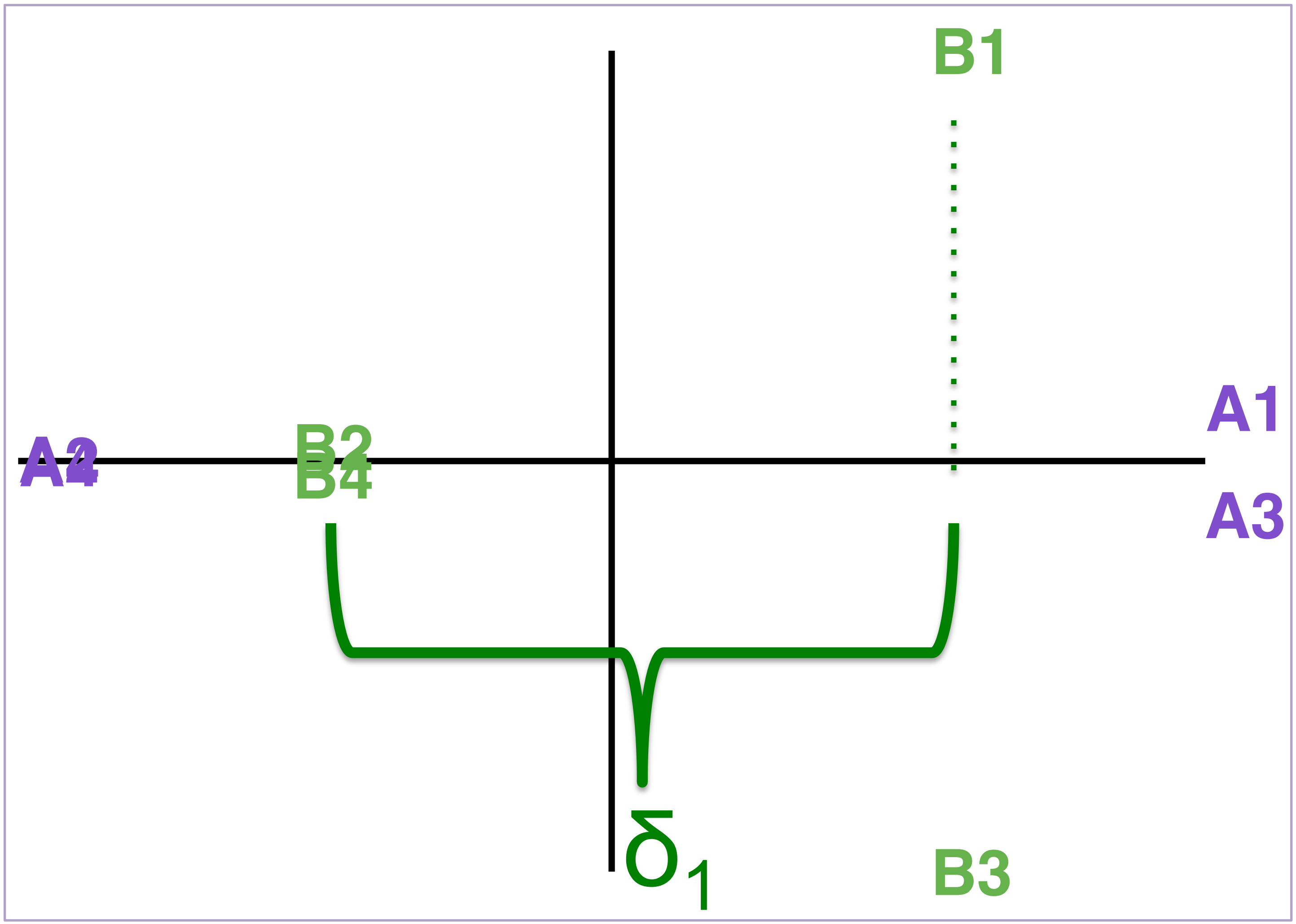
The “Length Problem”



Dimension 1 for A



Dimension 1 for B



Eliminating Length: The Need for Normalization

Normalizing by what?

Normalizing! Yes but how?

- *Z* score:

divide by standard deviation (square root of variance)

Normalizing! Yes but how?

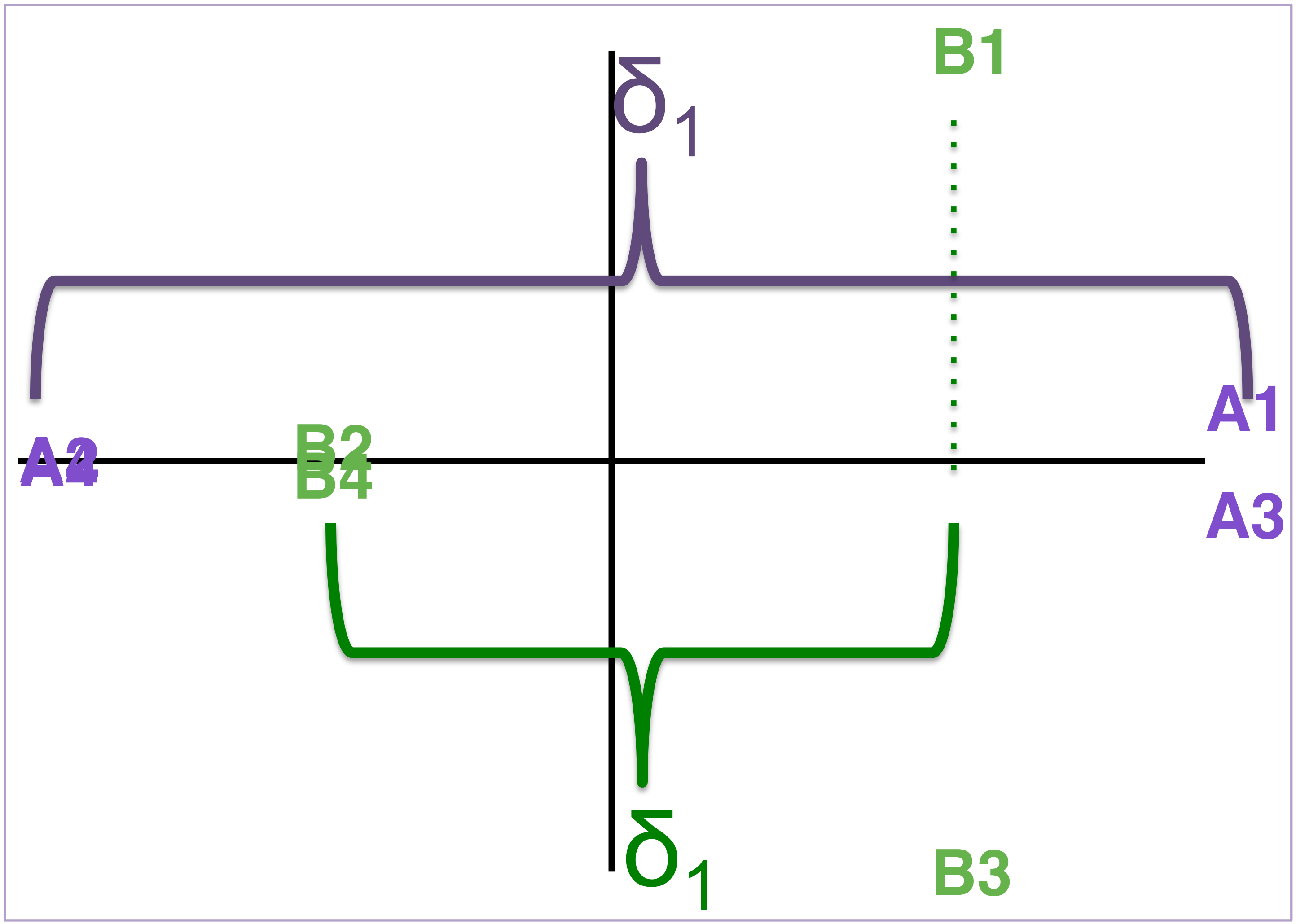
- *Z* score:
  divide by standard deviation (square root of variance)

- For a matrix:
  eigenvalues: variance, singular values: standard deviation.

- So normalize by the first singular value

- And now all first PCs are created equal

A and B have different Dimensions 1



The length of first component is δ_1

- Matrix 1: $\delta_1 = 1.41$

- Matrix 2: $\delta_1 = 0.76$

Need Normalization: Divide each matrix by its first singular value

- BTW. Not a new idea:
- Escofier, B. & Pagès, J. (1983). Méthode pour l'analyse de plusieurs groupes de variables: Application à la caractérisation des vins rouges du Val de Loire. *Revue de Statistique Appliquée*, **31**, 43–59.
- Escofier, B., & Pagès, J. (1990). Multiple factor analysis. *Computational Statistics and Data Analysis*, **18**, 121-140.

The length of Dimension 1 is δ_1. *So divide by* δ_1

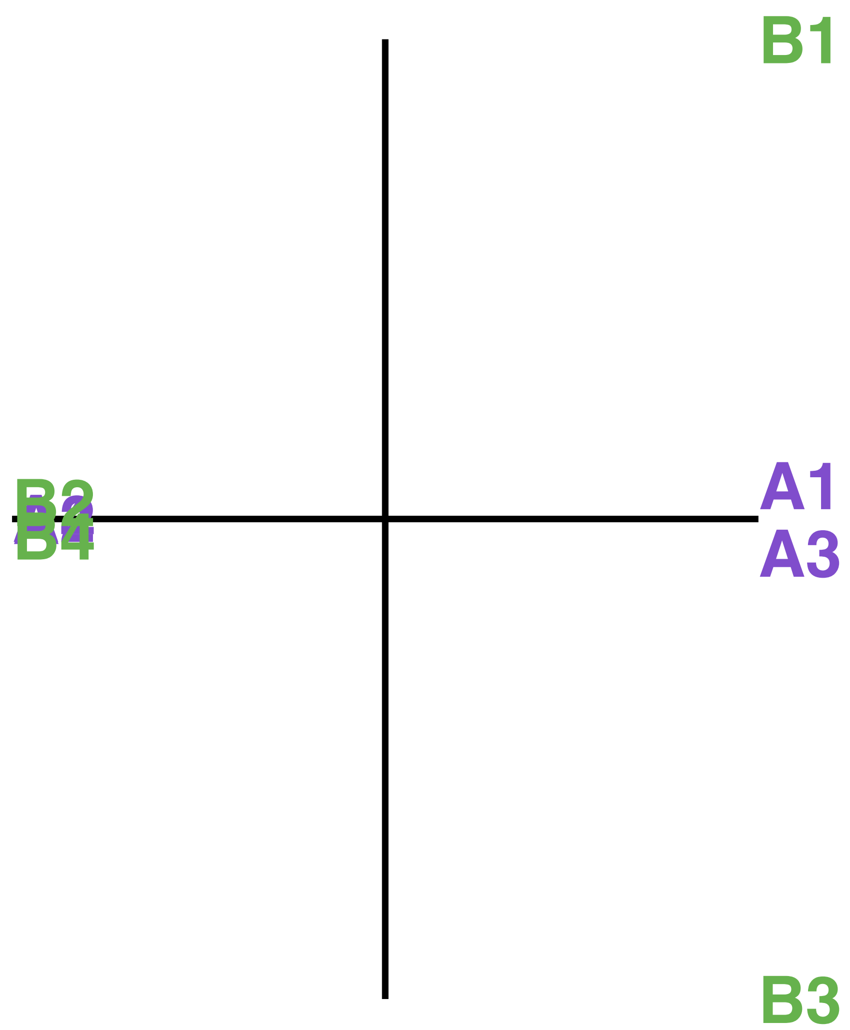
- Matrix 1: $\delta_1 = 1$.
- Matrix 2: $\delta_1 = 1$.

The normalized matrices

	1	2	3	4
1	.25	-.25	.24	-.25
2	-.25	.25	-.25	.24
3	.24	-.25	.25	-.25
4	-.25	.24	-.25	.25

	1	2	3	4
1	.66	-.25	-.16	-.25
2	-.25	.66	-.25	-.16
3	-.16	-.25	.66	-.25
4	-.25	-.16	-.25	.66

The plot: same length for Dimensions 1



Moral: "Z-scores for matrices"
or: correlation matrices are ipsative (Horst)

- To compare matrices we need first to
- 1. Center them
- 2. Normalize them

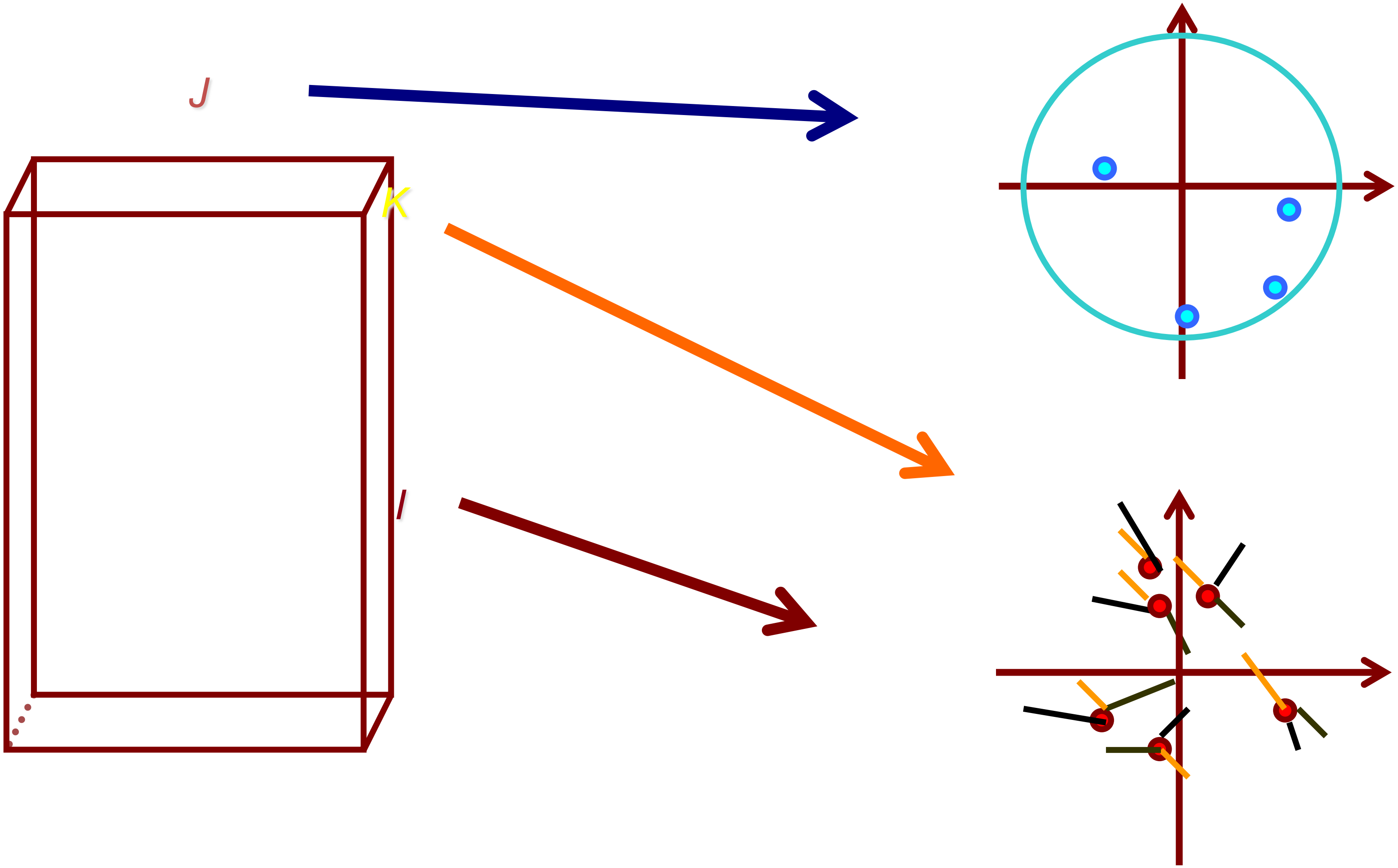
Part Ib

What about several connection matrices?

STATIS & DISTATIS etc.
and
the multi-table family

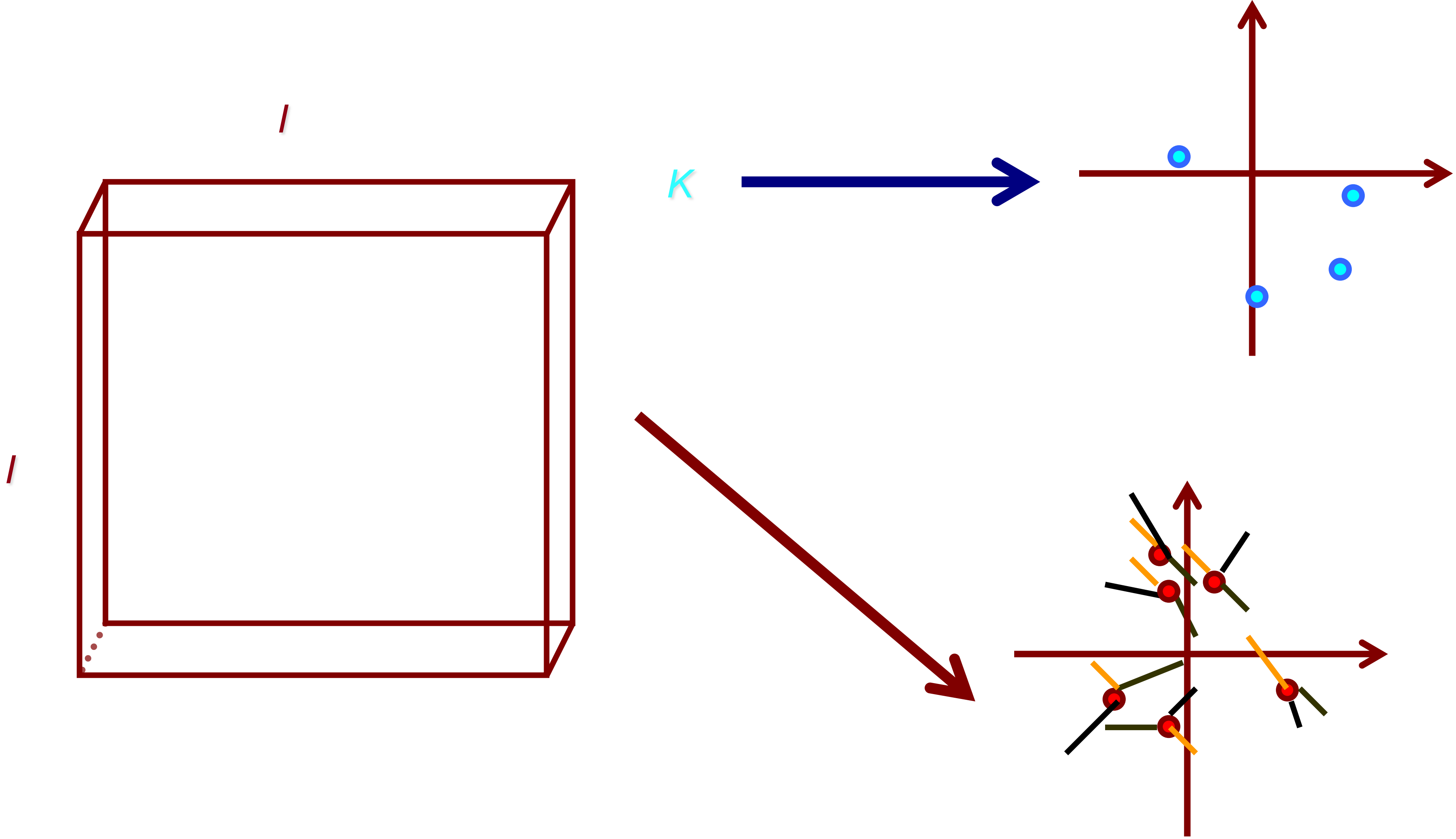
- If you wonder, "STATIS:"

- An (horrible) acronym for:
- *Structuration des Tableaux à Trois Indices de la Statistique*
- (three tables …)



- *I* by *J* by *K* data sets: MFA, STATIS Tucker-*n etc.*

Euclide … For the Three Table family



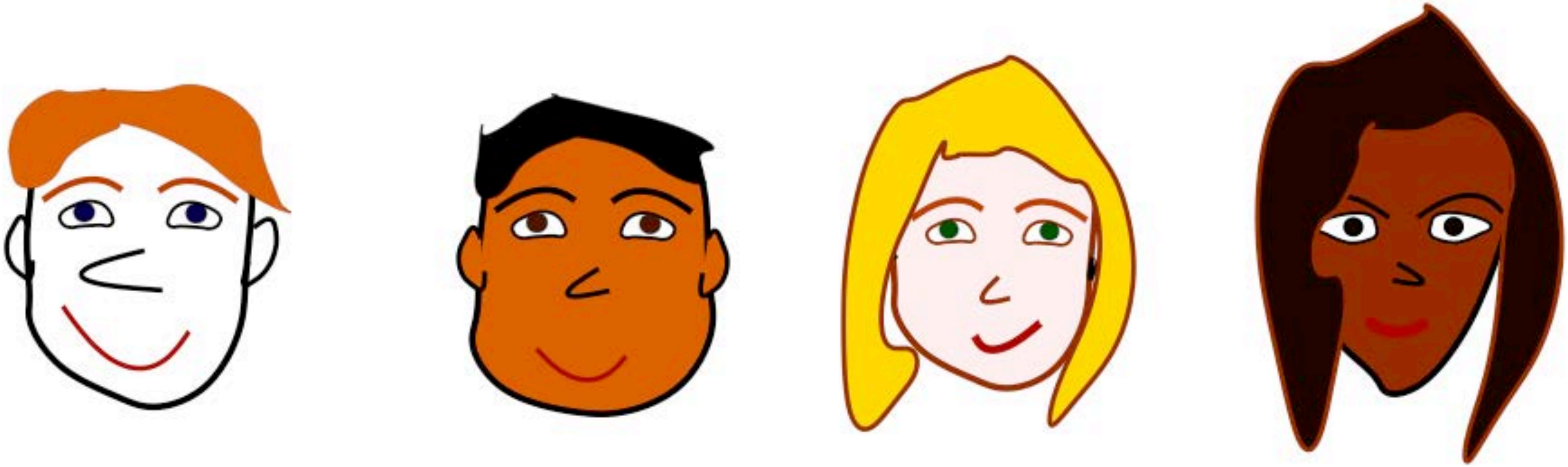
- *I* by *I* by *K* data sets: Procrustean family, DISTATIS**,** COVSTATIS, *etc.*

Euclide … Three (Squared) Table family

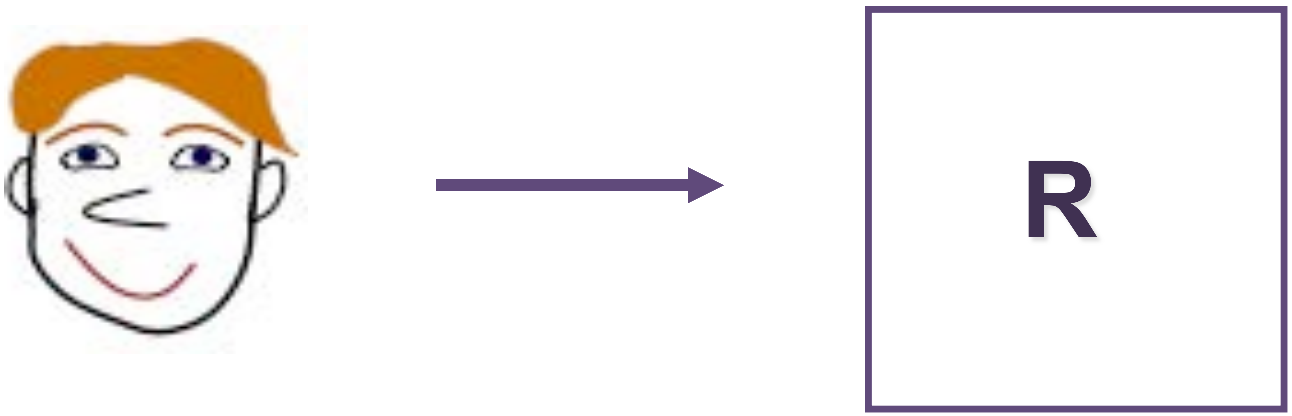
How to analyze a cube of distance

a picture tour of COVSTATIS (DISTATIS)

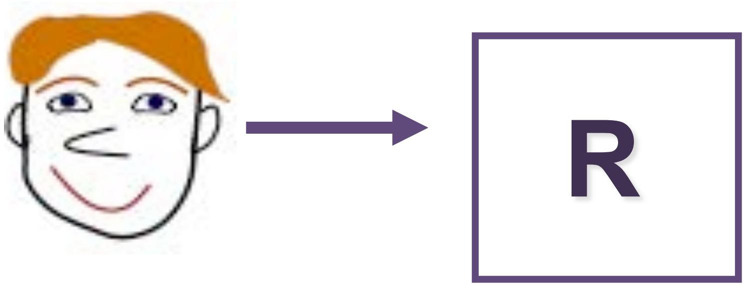
Fake example: 4 Participants and their connectivity matrices



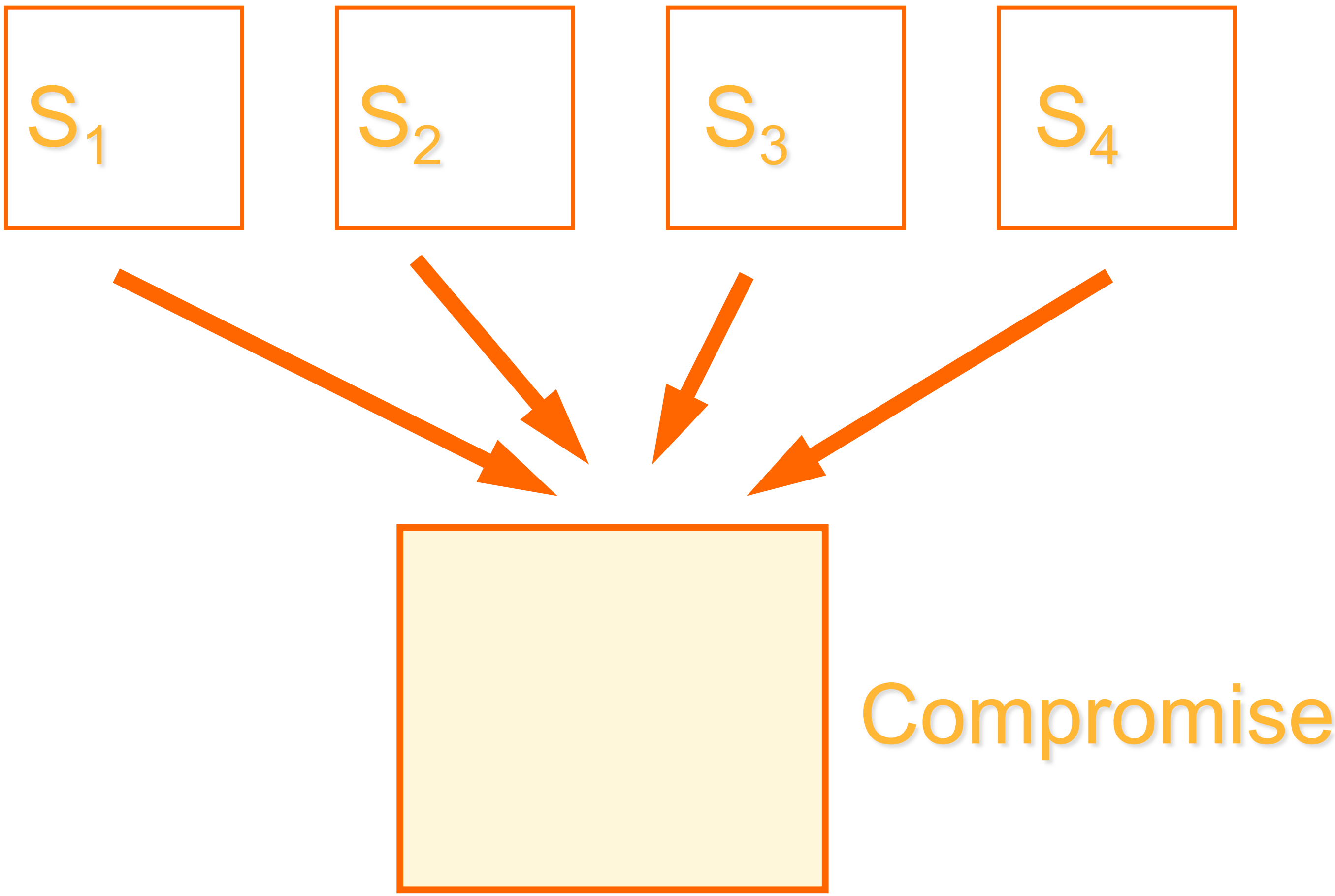
One participant gives a connectivity matrix



CORRELATION



One S matrix per participant What to do with these covariance matrices? Find the best compromise



The Compromise

- We have one **S** matrix per participant
  What to do with these covariance matrices?
  Find the best *compromise*

How to find a compromise?

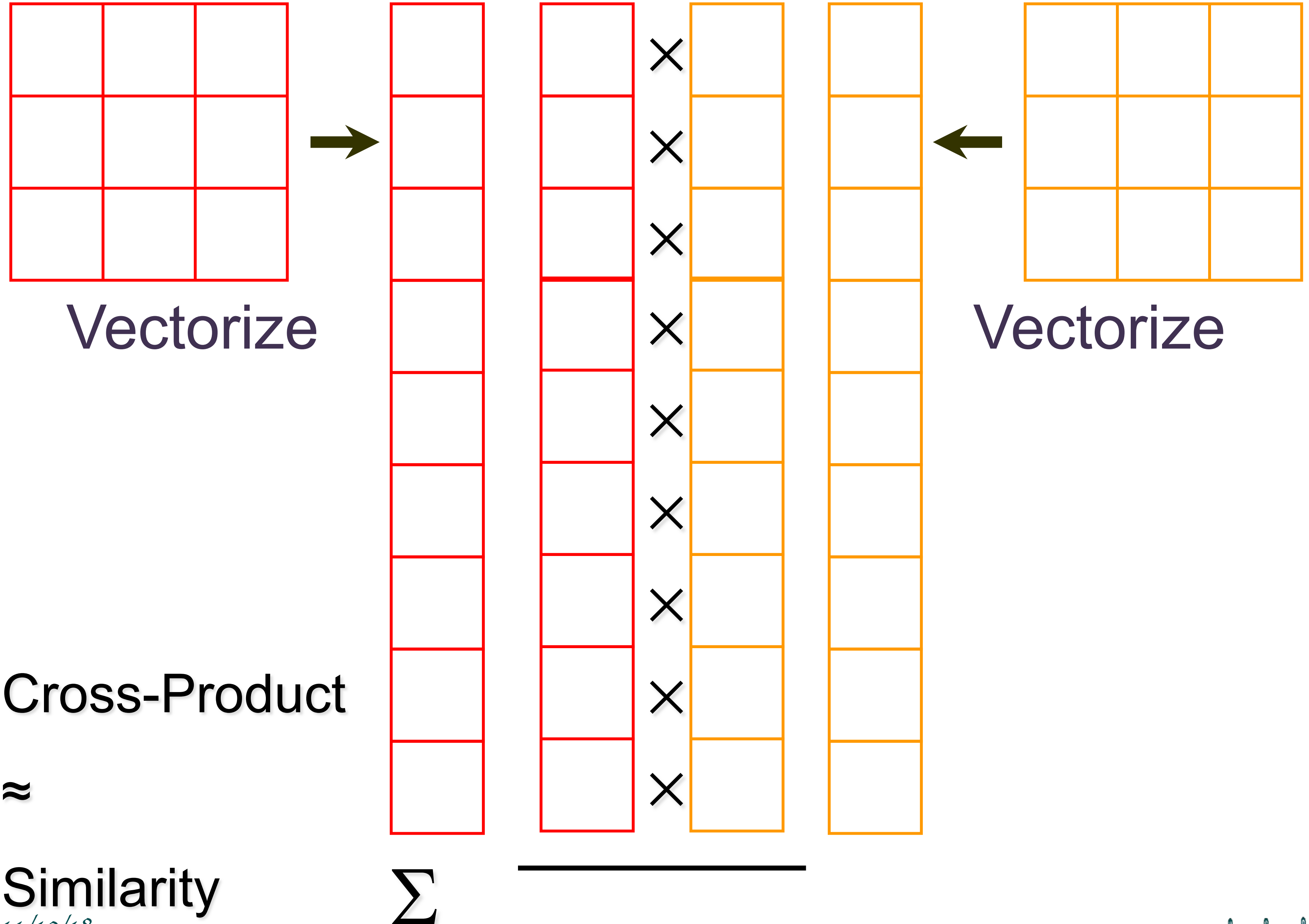
- To find the compromise:
- Mix them up! This means: find an optimal linear combination of the **S** matrices
- *Optimal*?
- Studies close to a common pattern should be weighted the most

How to find the compromise?

- Look at the between respondent structure:
- Need a coefficient of similarity between **S** matrices (for "*connaisseurs:*" semi-positive matrices)

- This is the R_V coefficient (Escoufier, 1973)

Similarity of Two Matrices with R_V



$$\mathit{Cross\text{-}Product} = \sum x_i \times y_i$$

- Problem: Cross-Product has "x by y" unit
- Solution: Normalize!
- $\Rightarrow$
- Unitless Number: This is R_V

$\mathcal{R}_V$ (Formally): *X* and *Y*: positive semi definite (psd) matrices

$$R_V = \frac{\mathrm{trace}\{\mathbf{X}^\mathsf{T}\mathbf{Y}\}}{\sqrt{[\,\mathrm{trace}\{\mathbf{X}^\mathsf{T}\mathbf{X}\}\,] \times [\,\mathrm{trace}\{\mathbf{Y}^\mathsf{T}\mathbf{Y}\}\,]}}$$

- R_V is a "cosine" between psd matrices.
- psd ⇨ $0 \leq R_V \leq 1$

$\mathcal{R}_V$ Formally (alternative version)
X and Y are 2 positive definite matrices

$$R_V = \frac{\langle \mathbf{X}^\mathsf{T}\mathbf{Y} \rangle}{|\mathbf{X}^\mathsf{T}\mathbf{X}| \times |\mathbf{Y}^\mathsf{T}\mathbf{Y}|}$$

$\langle , \rangle$: scalar product; $||$: matrix norm

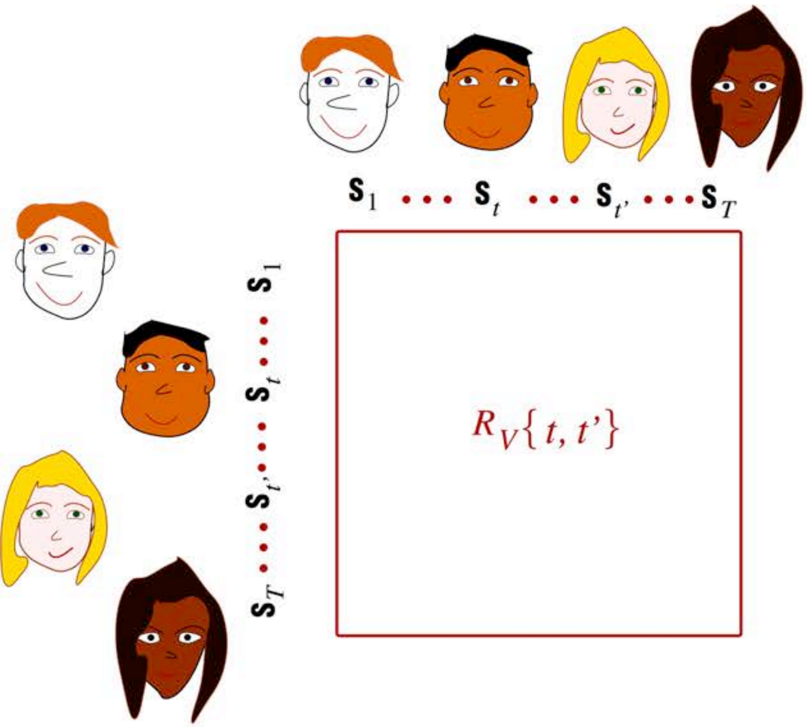
- R_V is a “cosine” between psd matrices.
- psd ⇨ $0 \leq R_V \leq 1$

$\mathcal{R}_V$ *Computational:* *X and Y are 2 matrices*

$$R_V = \frac{\mathrm{vec}(\mathbf{X})^T \mathrm{vec}(\mathbf{Y})}{\sqrt{[\,\mathrm{vec}(\mathbf{X})^T \mathrm{vec}(\mathbf{X})\,] \times [\,\mathrm{vec}(\mathbf{Y})^T \mathrm{vec}(\mathbf{Y})\,]}}$$

- R_V is a (squared)“cosine”
- between vectorized (psd) matrices

Question: Why is the first component all positive? Because R_V is a squared cosine



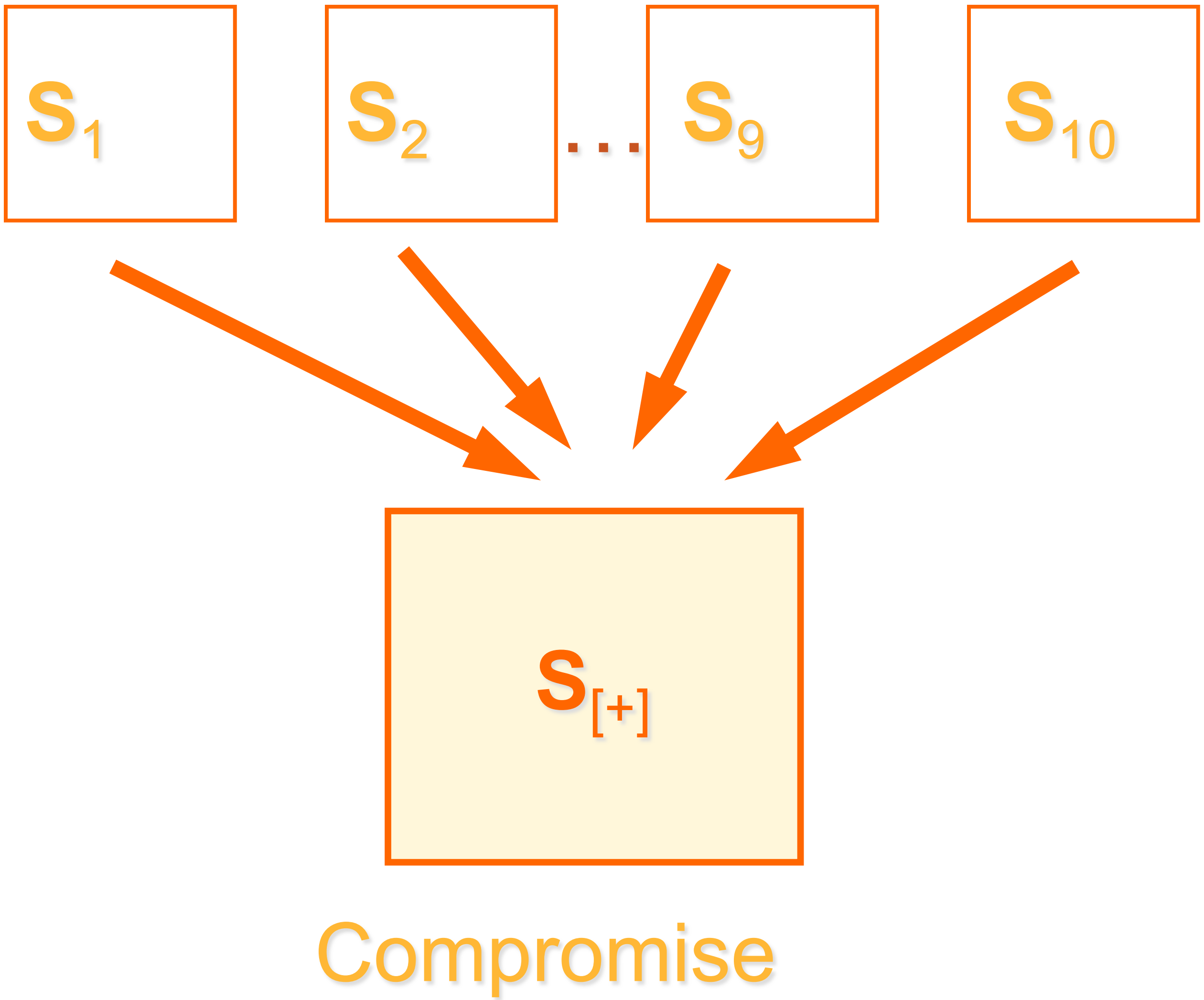
PCA of the participants (R_V matrix)

Transform the projections to get the α_t

- A good linear combination has weights summing to 1.

- So: rescale the projections
  into weights with a sum of 1

Combine the S matrices to get the compromise

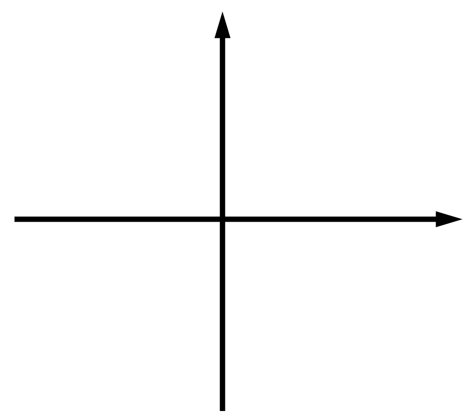


Combine the studies: Get the Compromise

$$\boxed{w_1\mathbf{S}_1} + \boxed{w_2\mathbf{S}_2} + \ldots + \boxed{w_k\mathbf{S}_k} + \ldots + \boxed{w_K\mathbf{S}_K}$$

$$=$$

$$\boxed{\mathbf{S}_{[+]}} \rightarrow$$



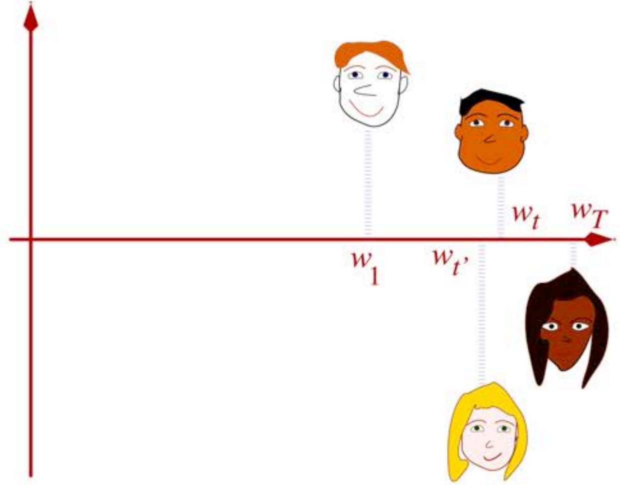
Compromise

So what does the compromise optimize?

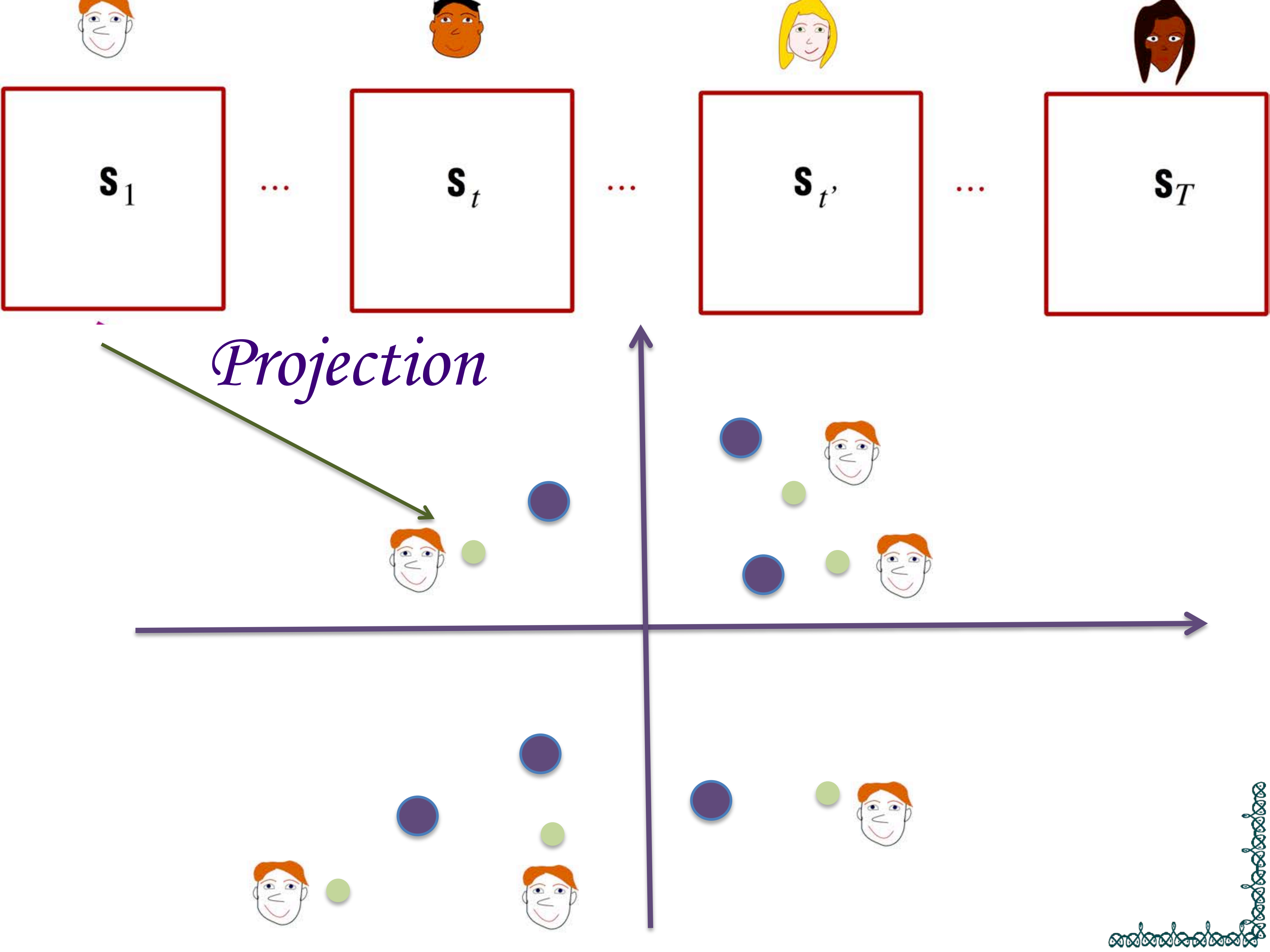
- The compromise satisfies:
- max $\sum R_V{}^2(\mathbf{S}_+, \mathbf{S}_i)$

How Good is the compromise?

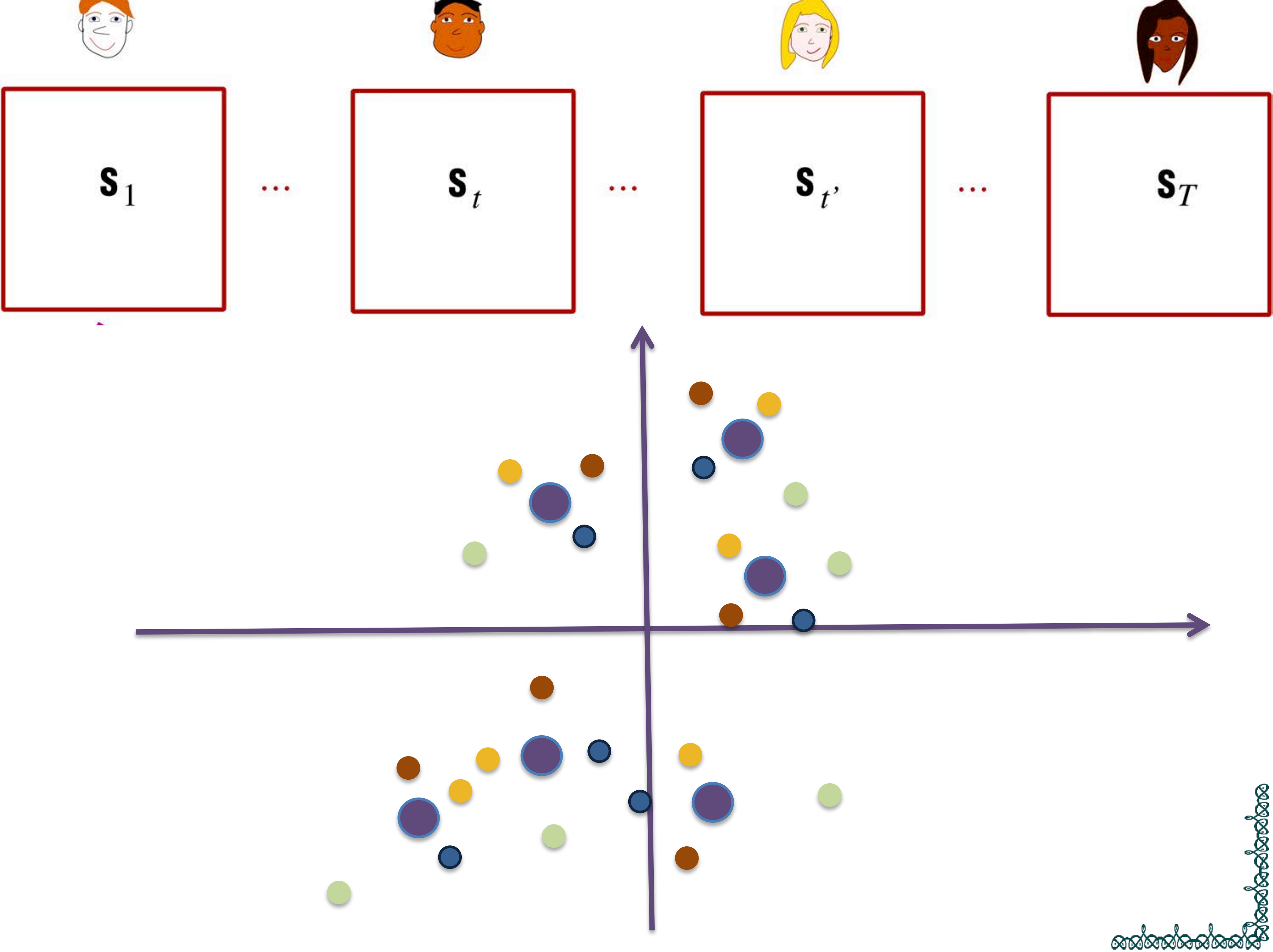
- Quality of the compromise:
- Percentage of explained variance
- Ratio of first eigenvalue to total



w_1
$w_{t'}$
w_t
w_T



$\mathbf{s}_1$
...
$\mathbf{s}_t$
...
$\mathbf{s}_{t'}$
...
$\mathbf{s}_T$
Projection


$\mathbf{S}_1$
$\mathbf{S}_t$
$\mathbf{S}_{t'}$
$\mathbf{S}_T$

How stable are the results?
Is the earth round enough for p <.05
(Merci Jacob Cohen!)

- Confidence intervals, Bootstrap ratios etc.

For the observations and their positions

Use Bootstrap resampling

What is the Bootstrap? (Efron, 1979)

or cross-validation

for a missing population

How Stable are the observations?
What are the important observations/variables

- *We have the (infinite) Population*:
  - big loadings
  - How to find them: resample

- Replace *Population* by *sample*
  - Sample repeatedly from *finite* sample
  - Make sure that probabilities do *not* change
    - This means

sample *with* replacement!

A B C D E F

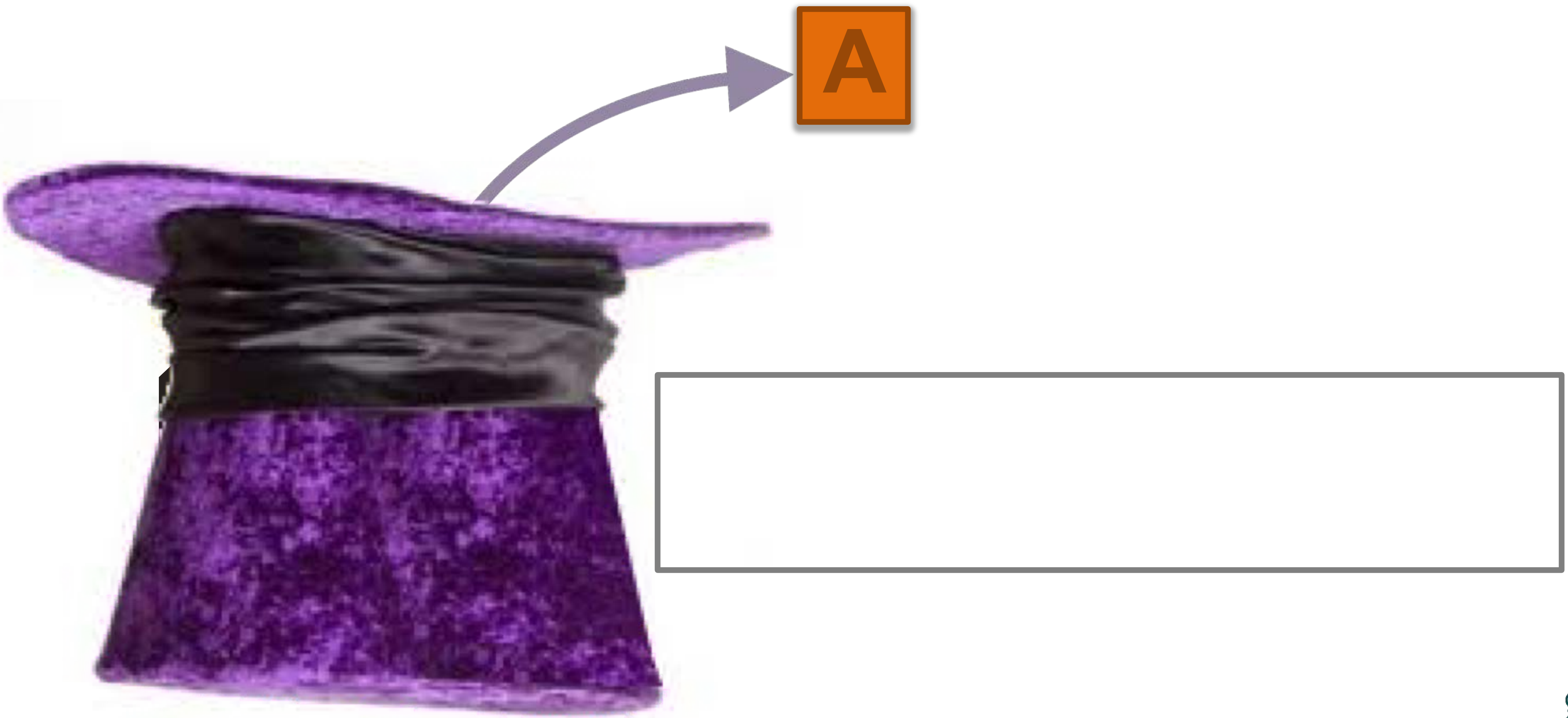
Sampling with Replacement:
p does not change

Sampling with Replacement:

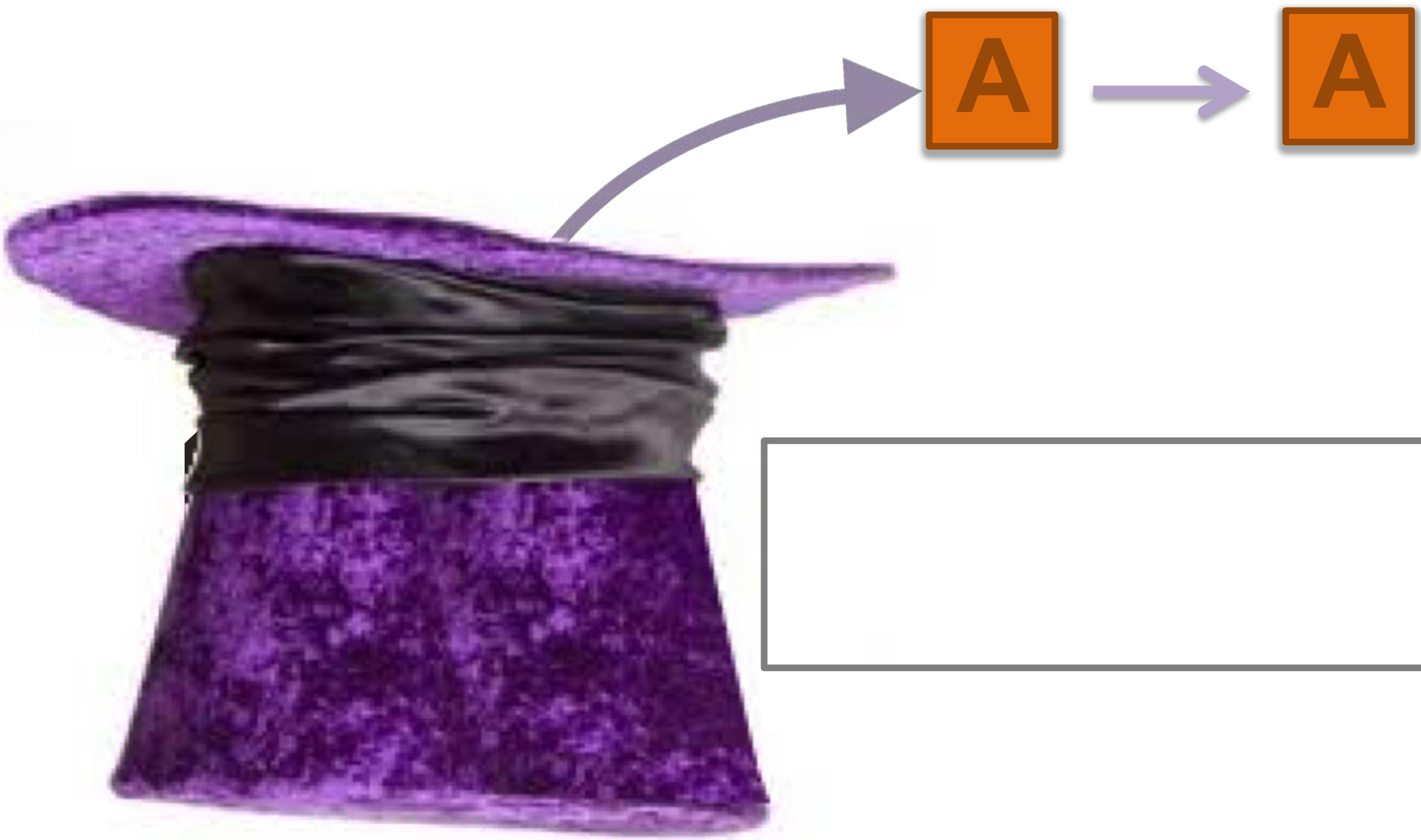
A Shake

Ready?

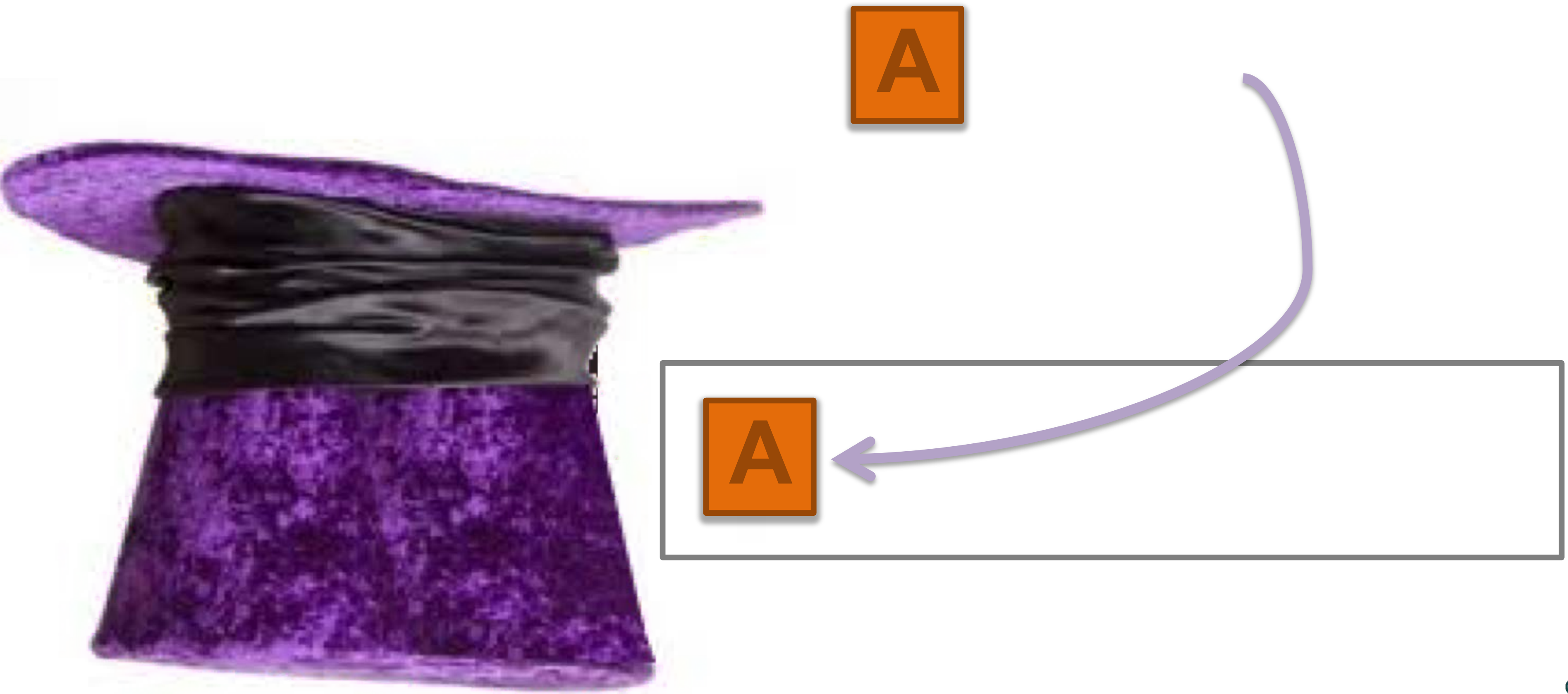
Get the first one



Copy it



Put the copy in the box



Sampling with Replacement:

Put the original back in the hat

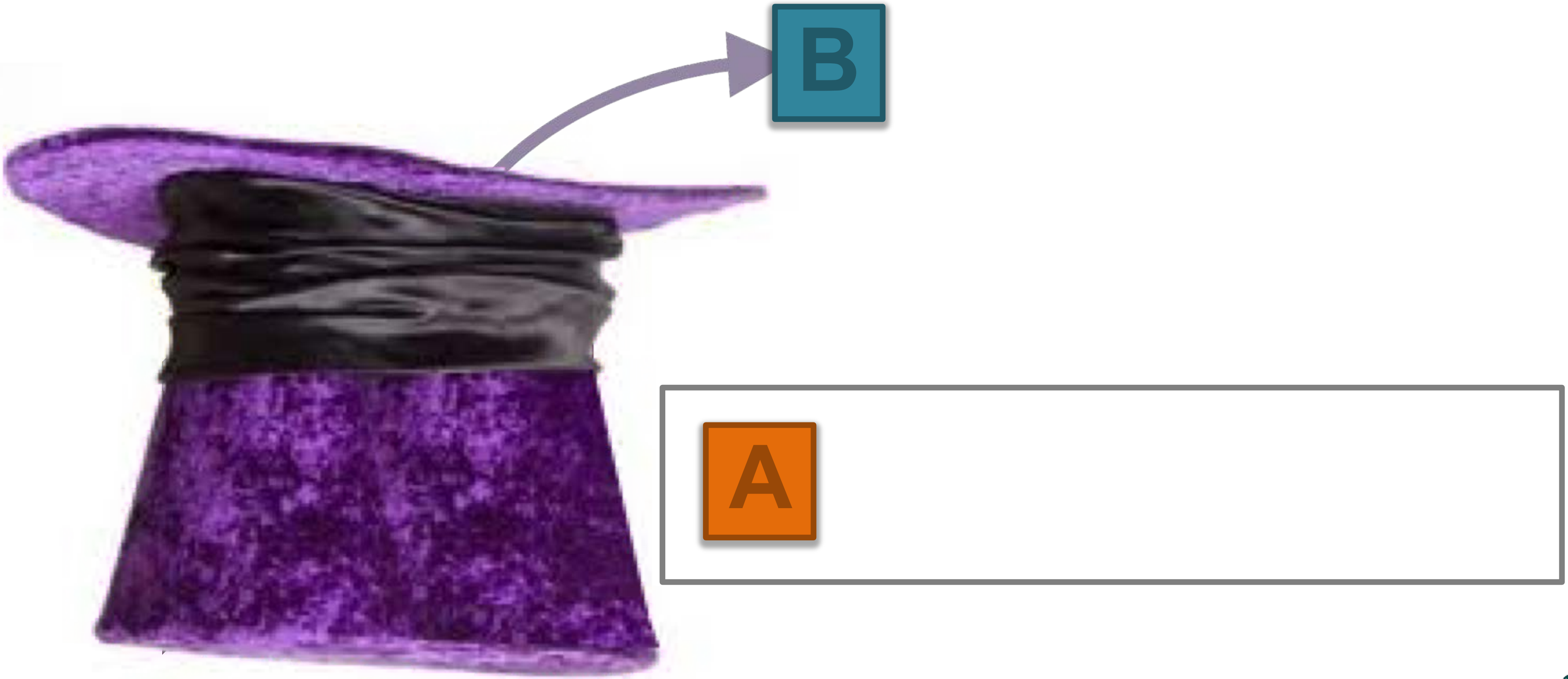
A

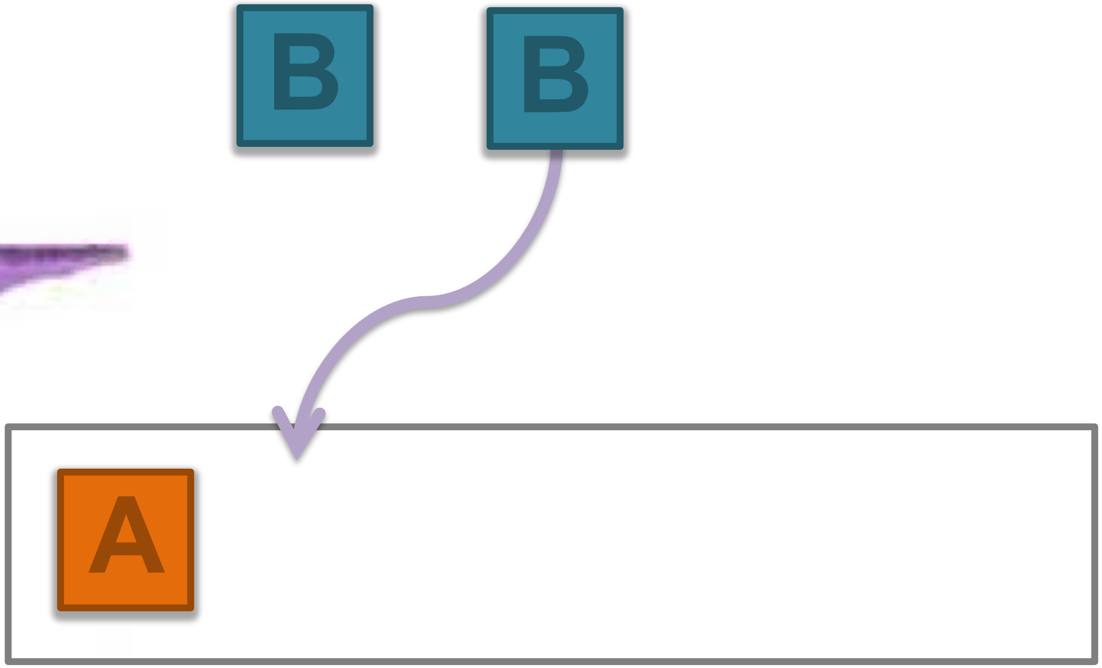
A

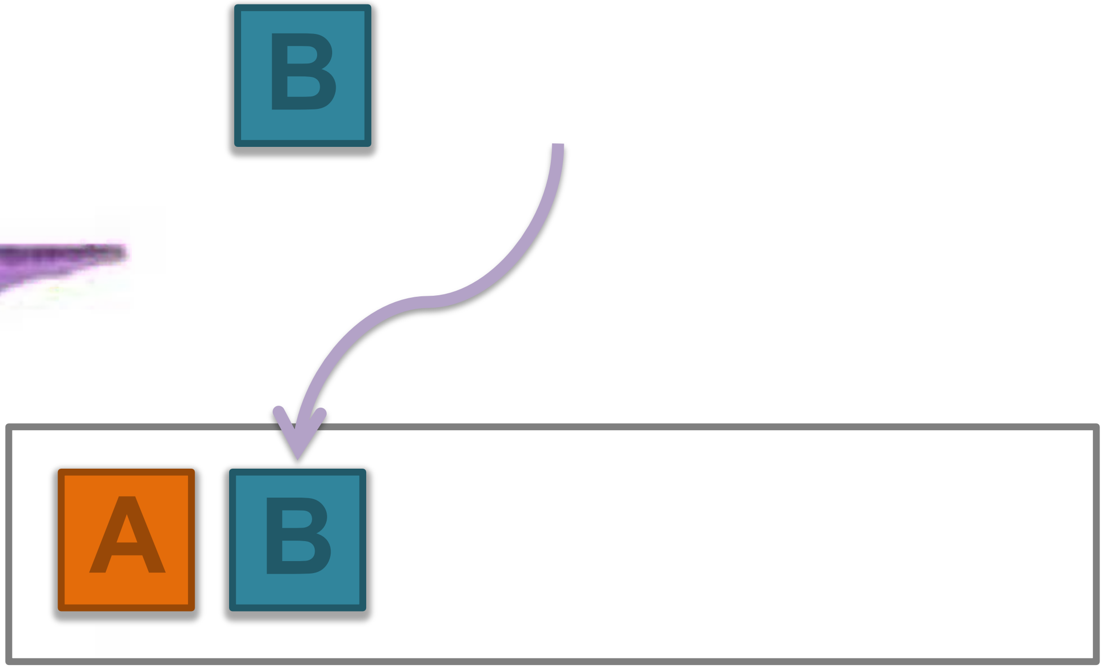
A

A Shake

A


B
A

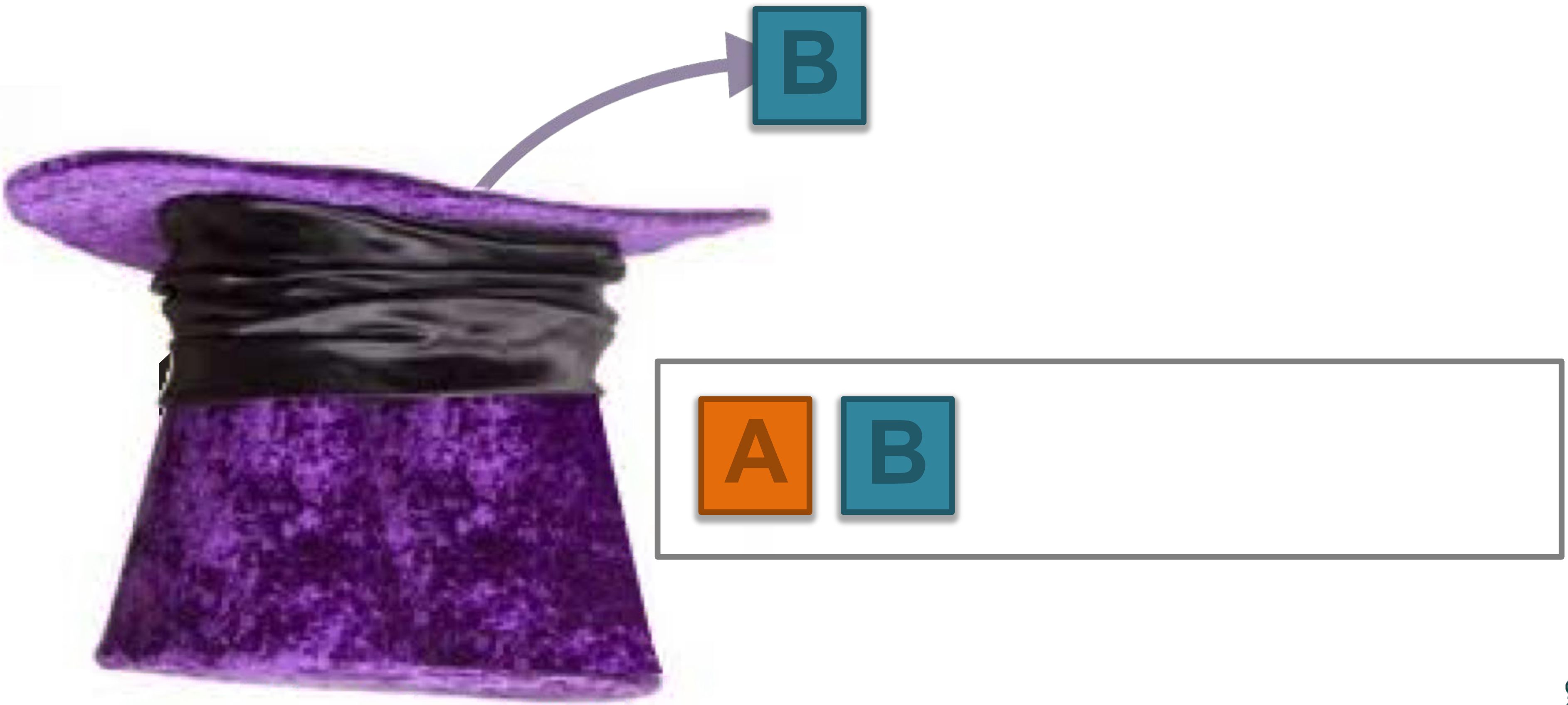

B
B
A

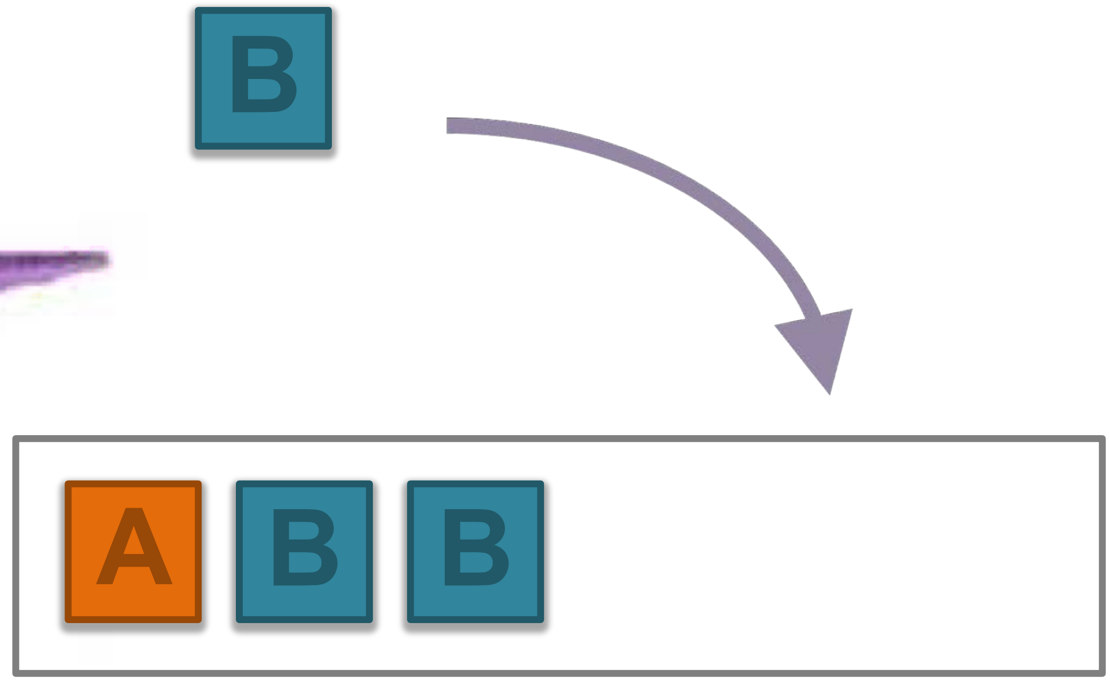

B
A
B

B

A B

A B

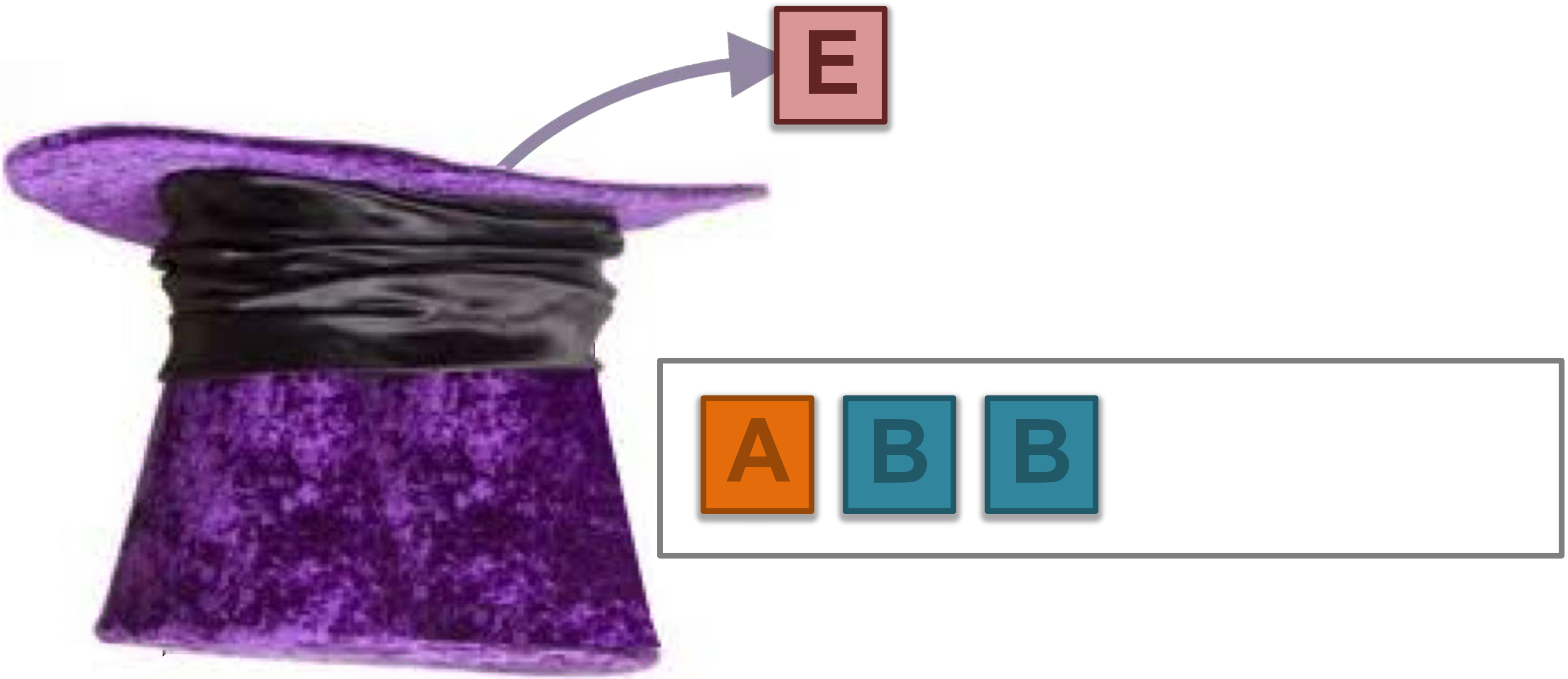

B
A
B

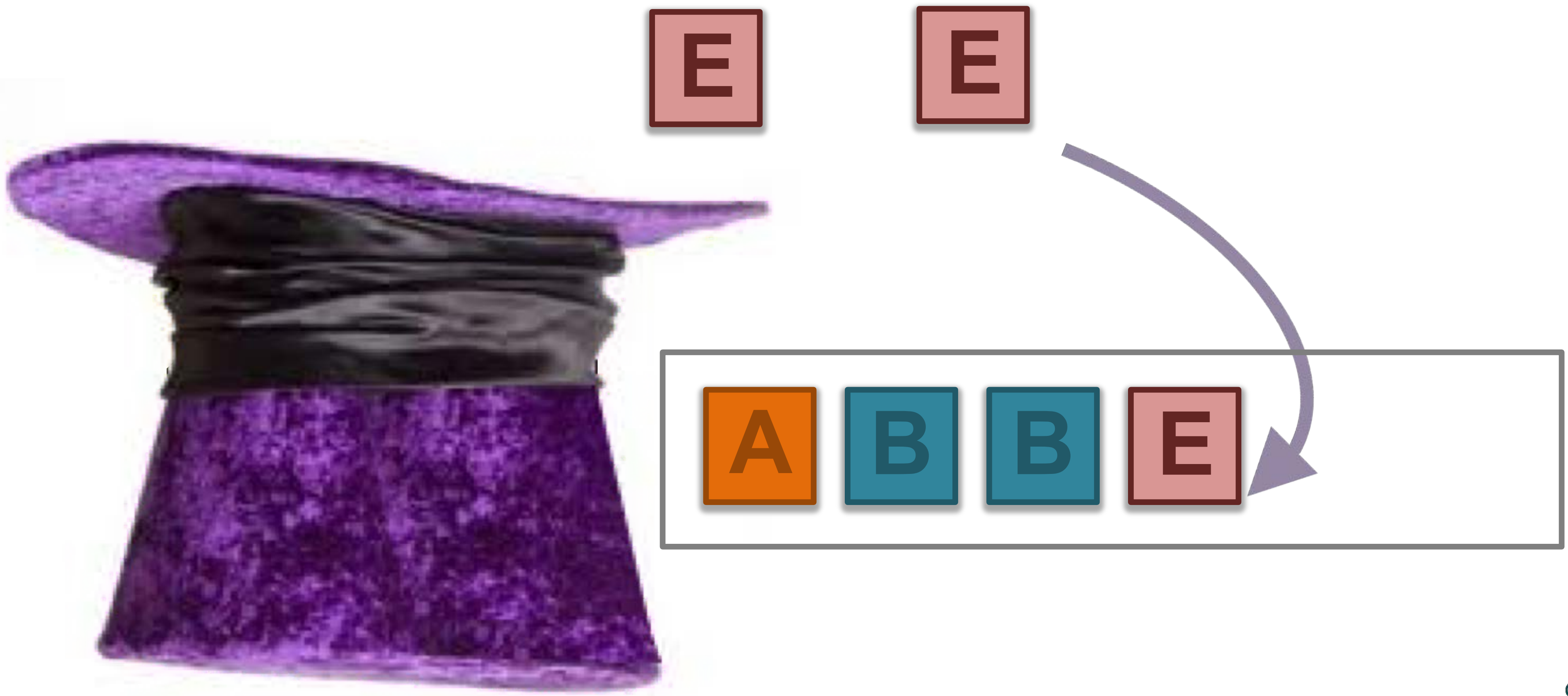

B
A
B
B

B

A B B

A B B

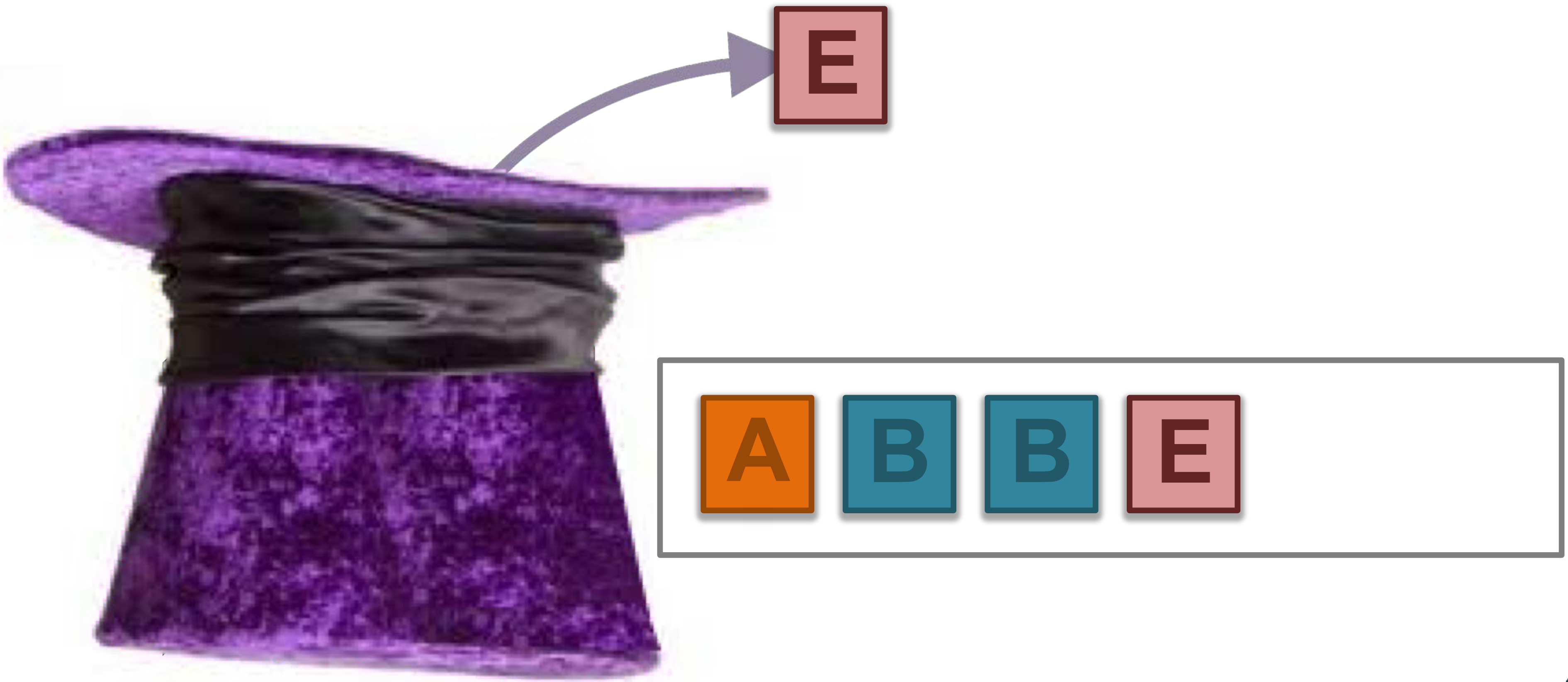

E
A
B
B

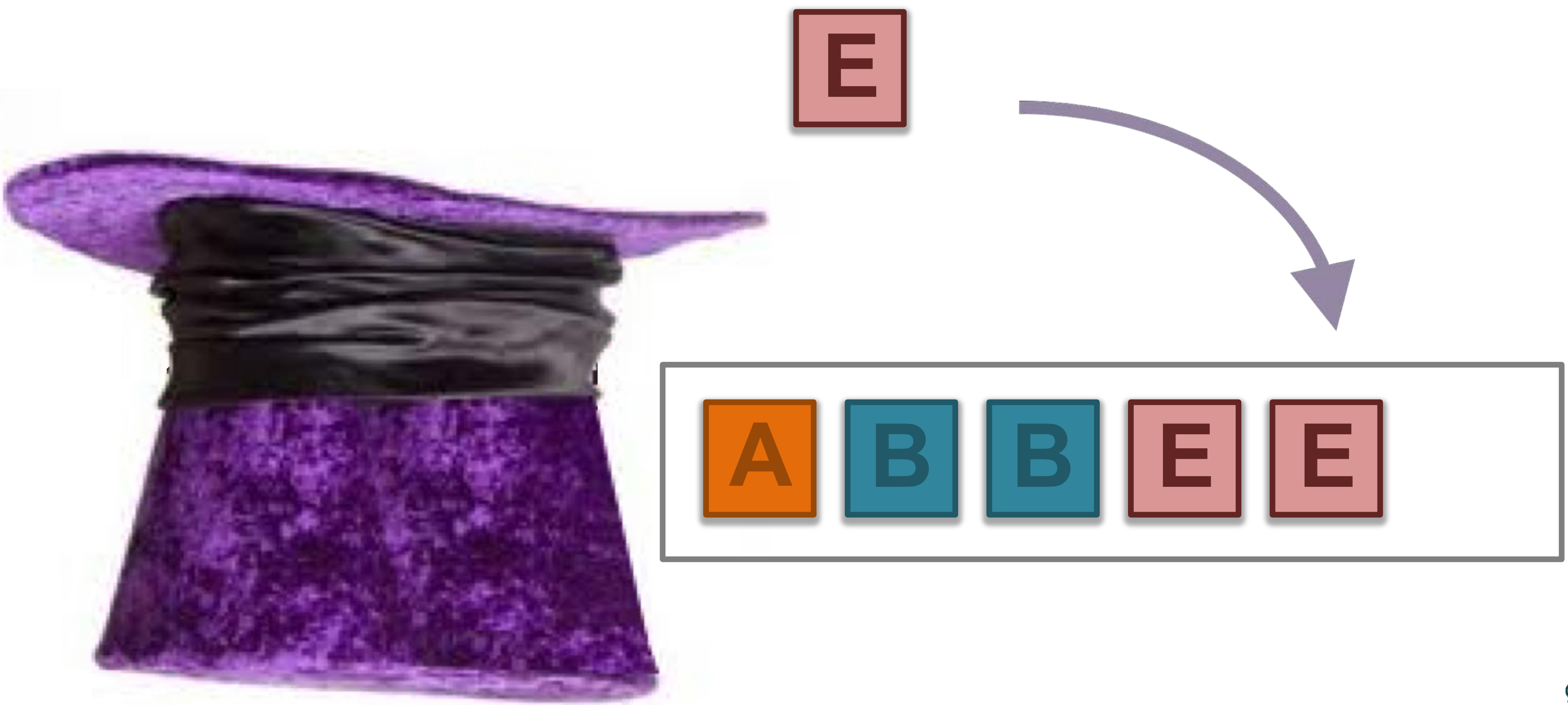

E
E
A
B
B
E

E

A B B E

A B B E

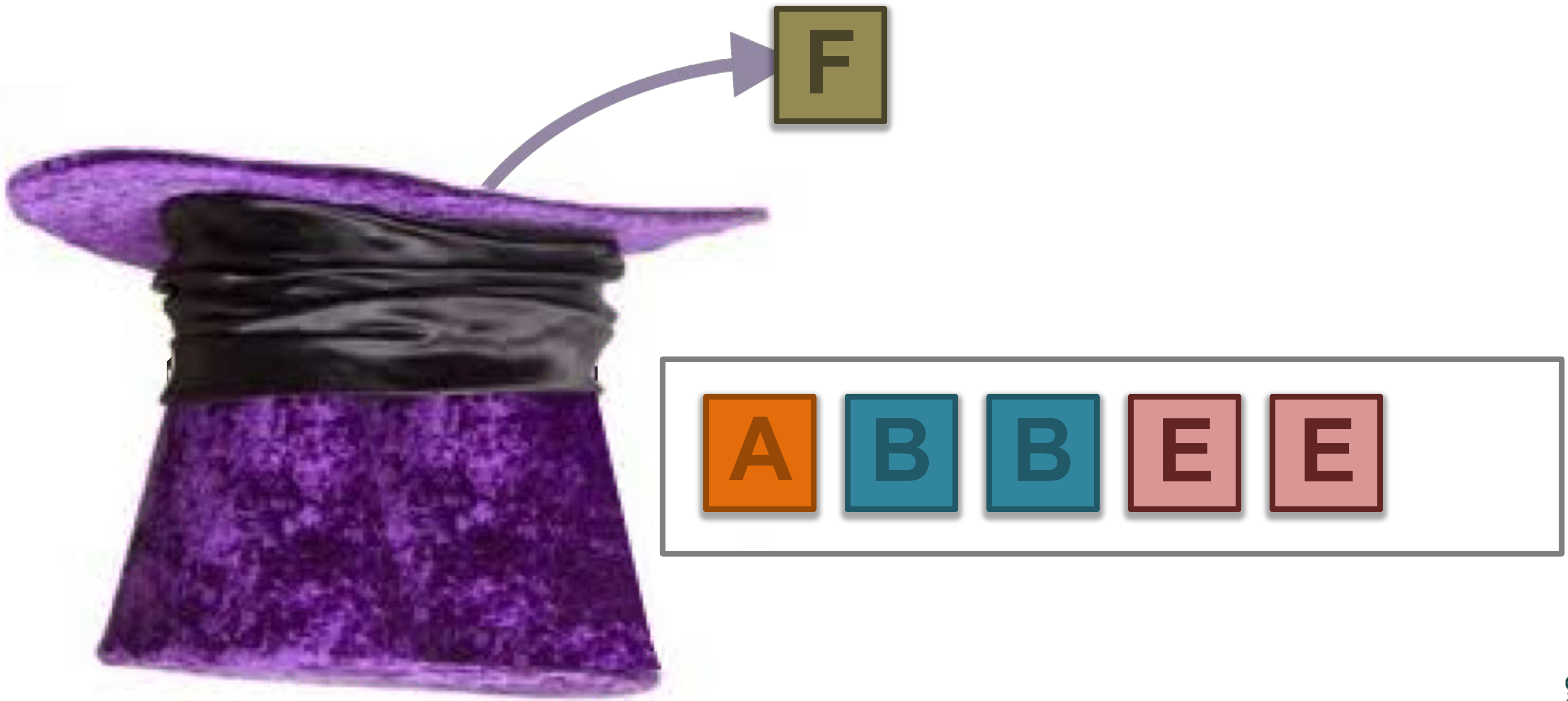

E
A B B E

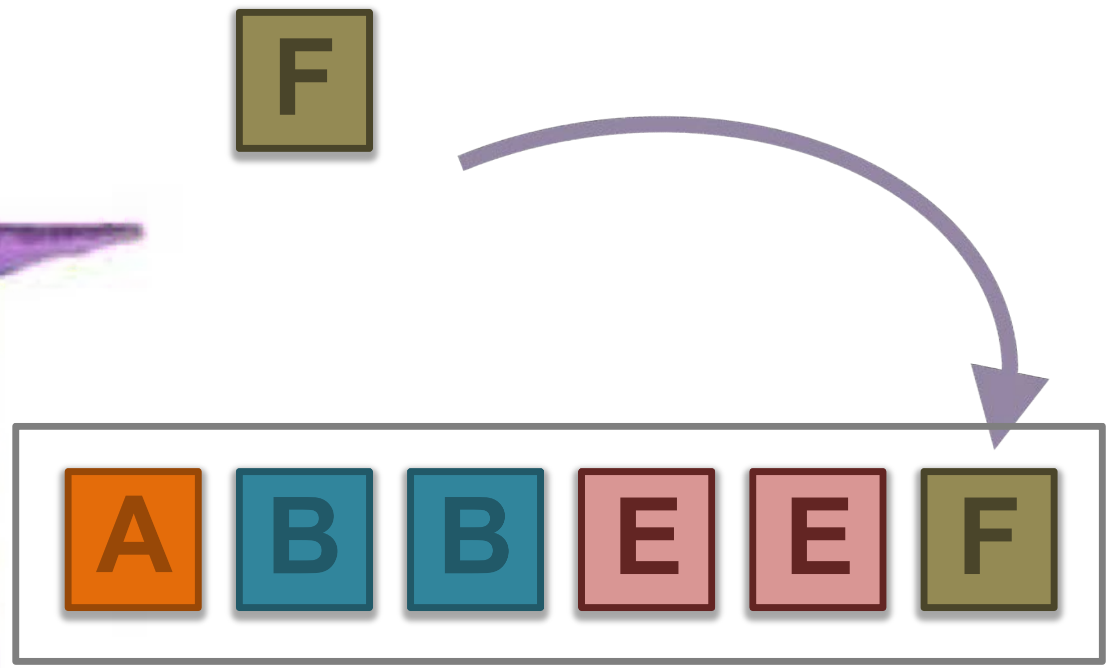

E
A
B
B
E
E

E

A B B E E

A B B E E


F
A
B
B
E
E


F
A
B
B
E
E
F

F

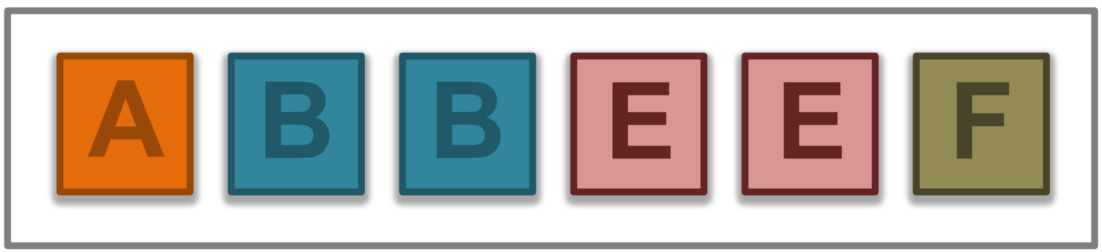
A B B E E F

A B B E E F

Moral: Sampling with Replacement

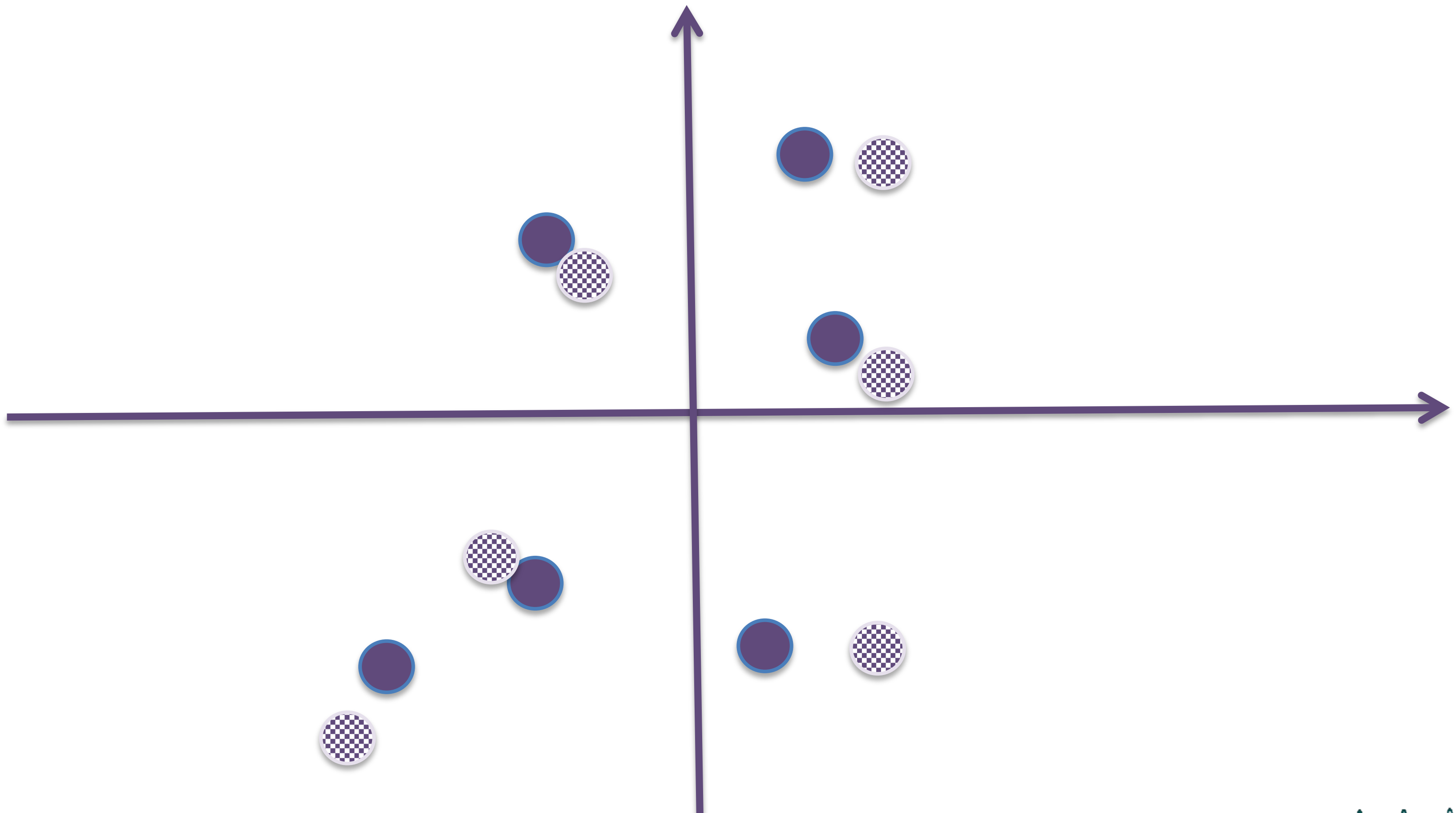
p does not change:

Some are repeated, some are lost

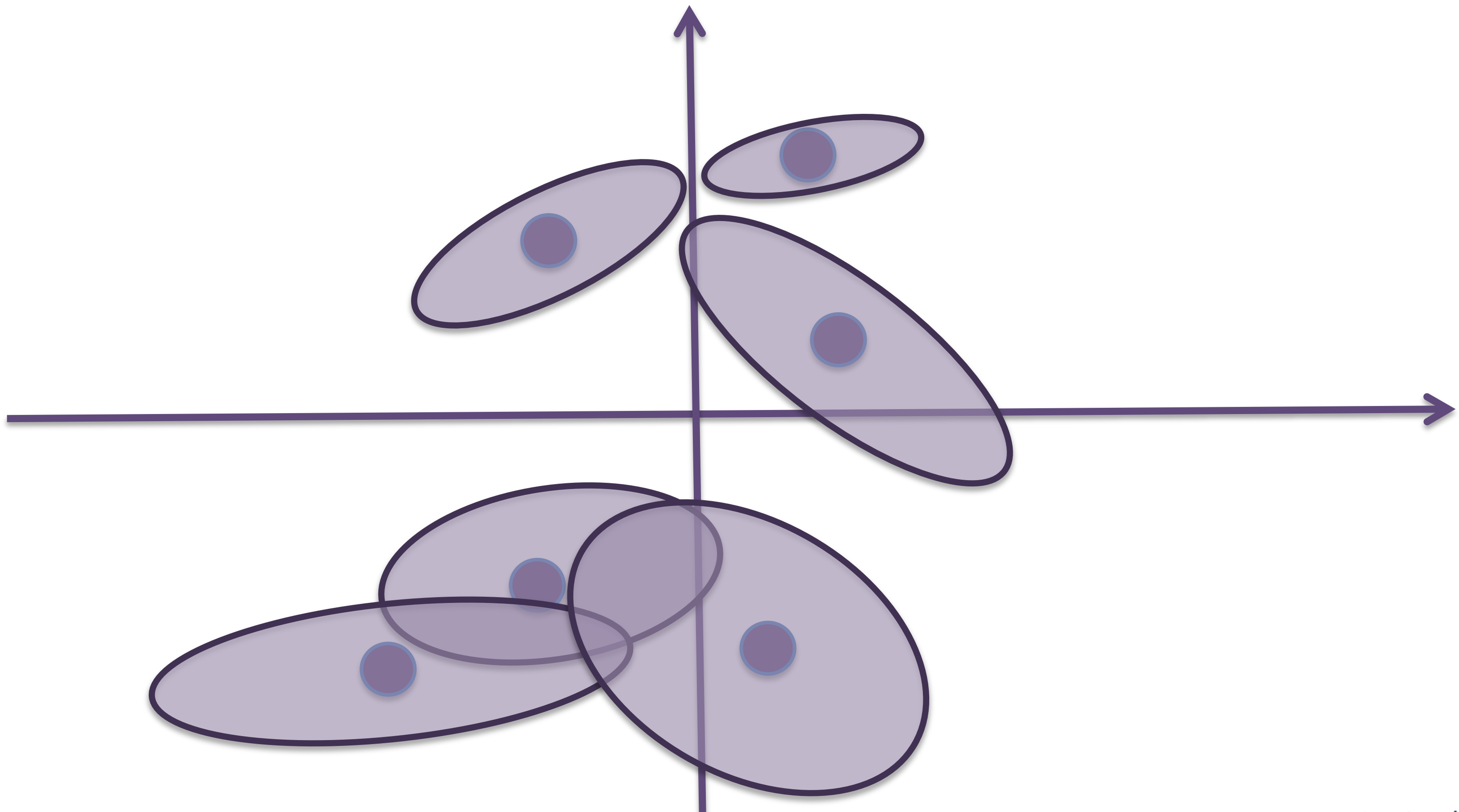


Confidence Ellipsoids from Bootstrap

Project the bootstrap sample onto the space



Confidence Ellipsoids from Bootstrap
Do that a lot of time and draw ellipsoids



Or compute t-like Bootstrap ratios

- $t = M_{\text{boot}} / \sigma_{\text{boot}}$

- So say $p < .05 \approx t_{\text{critical}} = 2.00$