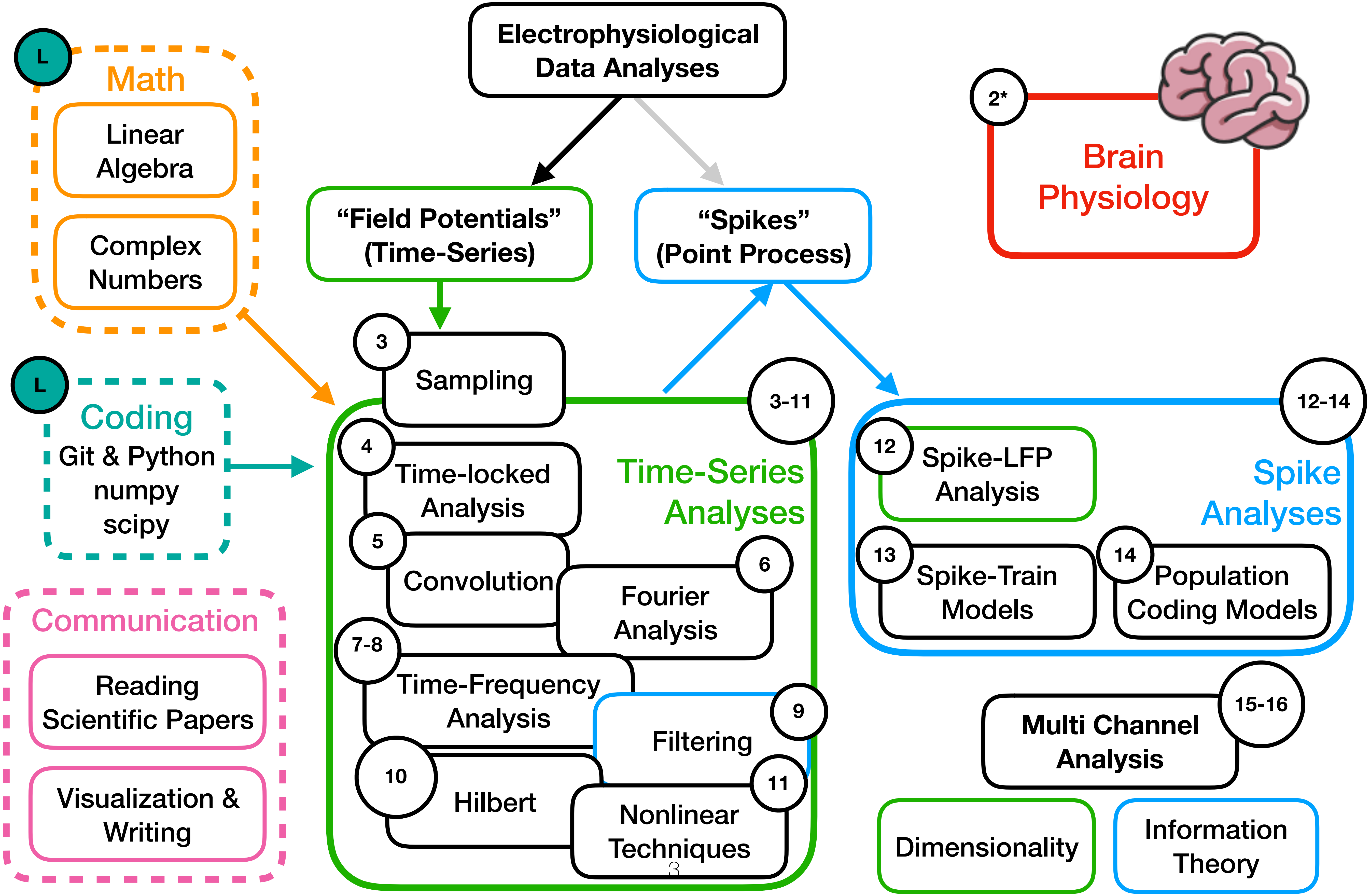
Question of the Day

How much “information” does the sun rising give you tomorrow? What about if it didn’t rise?

Decomposition & Information Theory

Lecture 16
July 31, 2019

Course Outline: Road Map



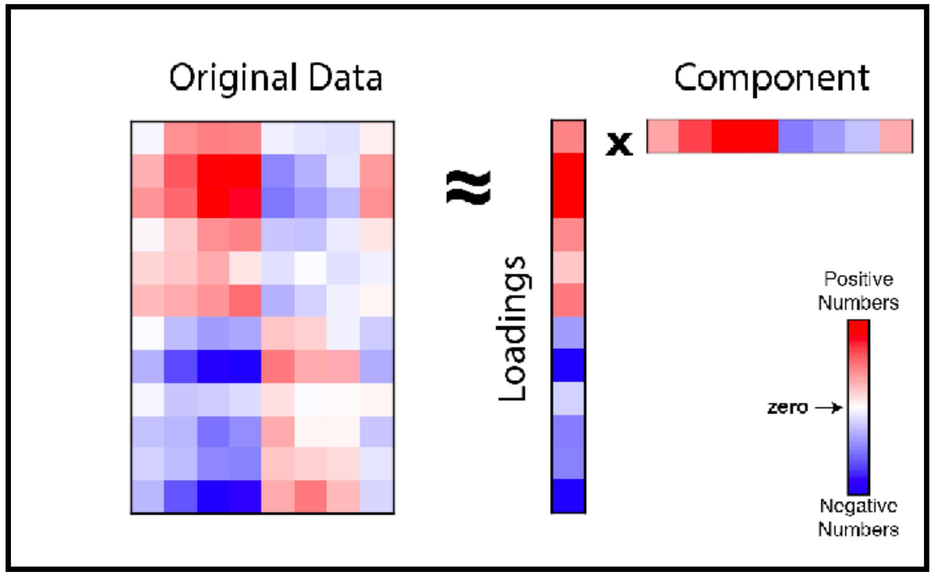
Goals for Today

1. More PCA intuition & examples

2. Define information & common quantities

3. Examples in neuroscience

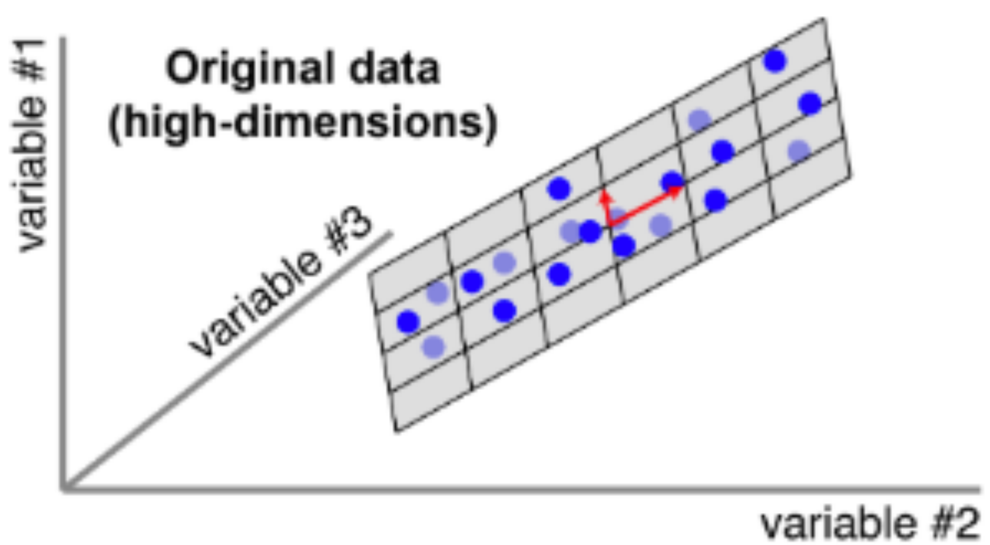
Principle Component Analysis

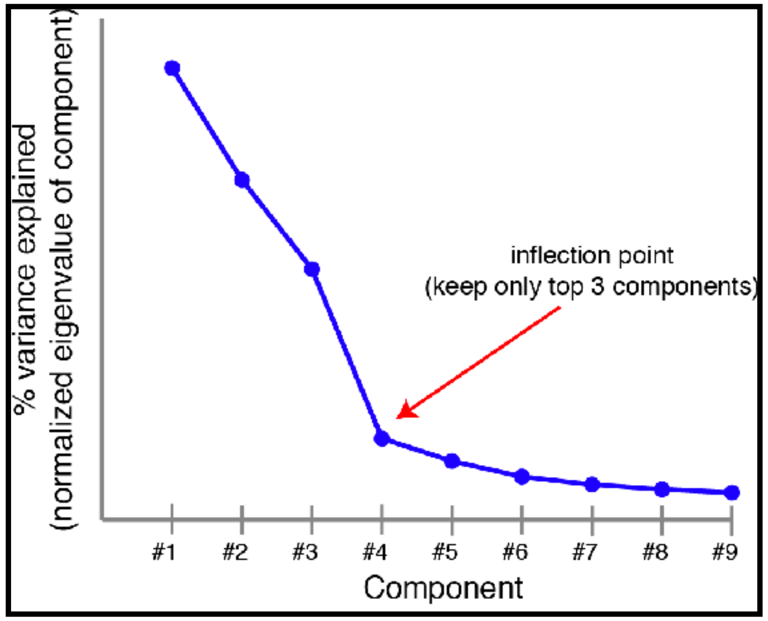


PCA decomposes correlated brain activity into a “smaller” set of orthogonal bases.

Bases are the eigenvectors of the correlation matrix

Eigenvalues represent how much variance is explained by each basis

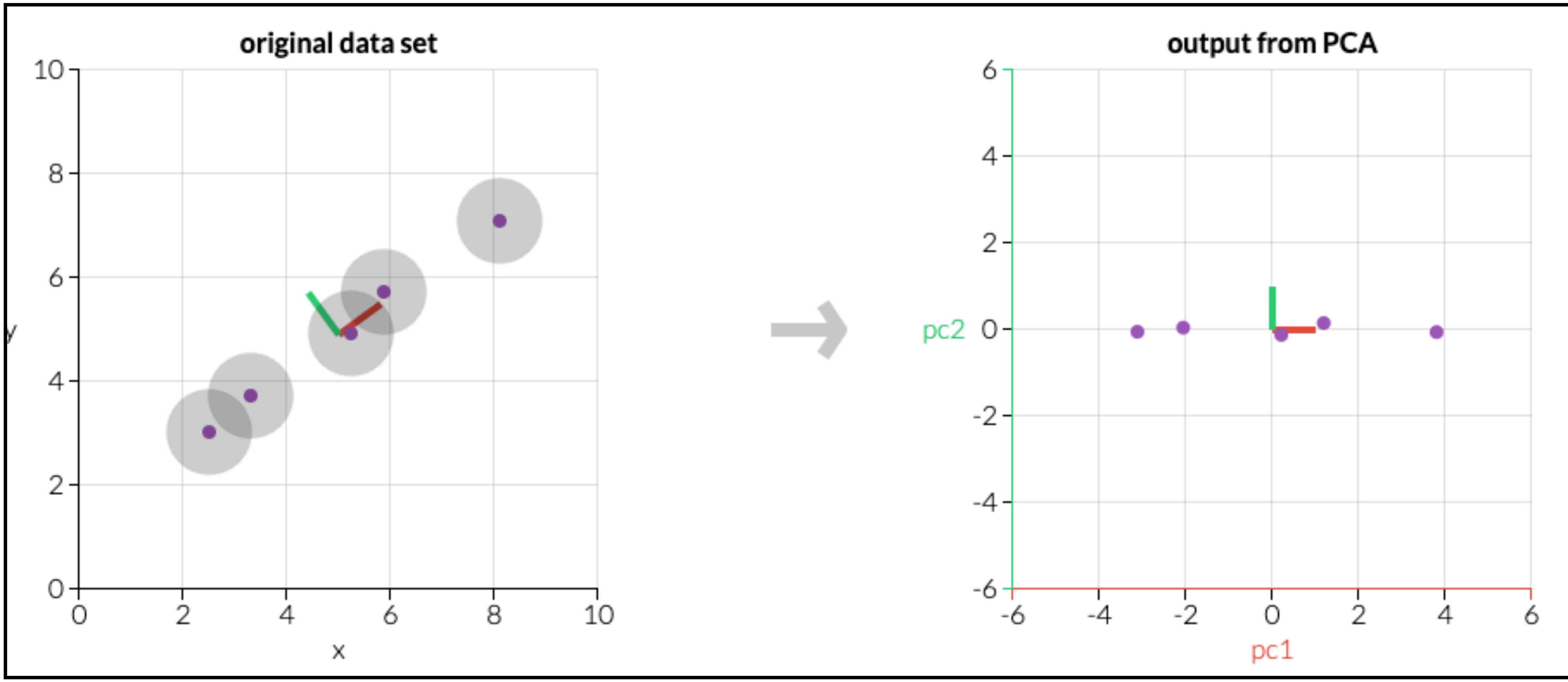




http://alexhwilliams.info/itsneuronalblog/2016/03/27/pca/

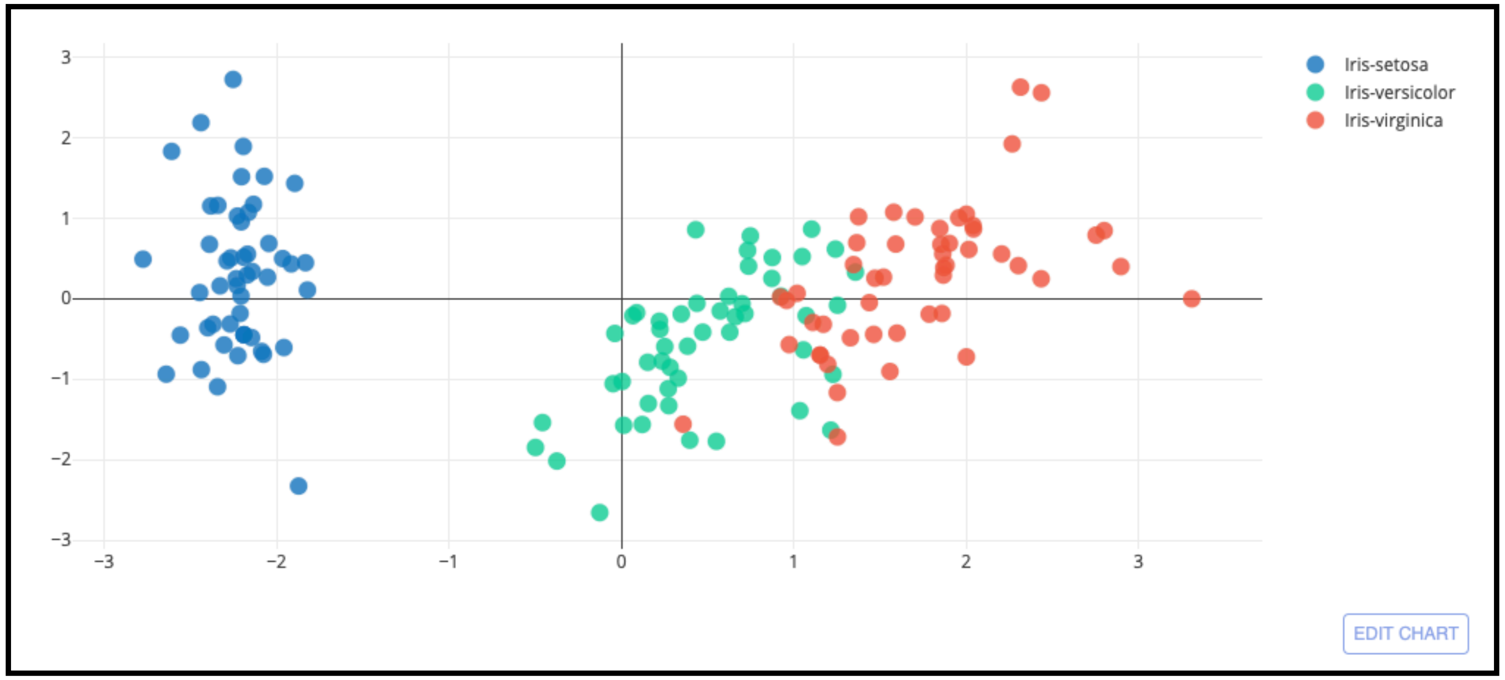
Principle Component Analysis

Rotation of basis vectors: from Cartesian to its Linear Combination (Empirical)



Sometimes refered to as “Latent Factors”

Principle Component Analysis



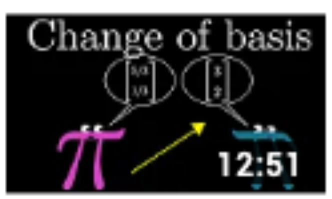
1. Mean-center data (subtract average of every feature)
2. Compute covariance/correlation matrix
3. Eigendecomposition of correlation matrix

Principle Component Analysis

sklearn.decomposition.PCA

class sklearn.decomposition. **PCA** (*n_components=None, copy=True, whiten=False, svd_solver='auto', tol=0.0, iterated_power='auto', random_state=None*) [source]

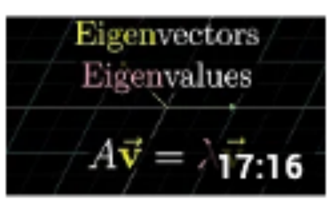
```python
from sklearn.decomposition import PCA as sklearnPCA
sklearn_pca = sklearnPCA(n_components=2)
Y_sklearn = sklearn_pca.fit_transform(X_std)
```

13 

3BLUE1BROWN SERIES S1 • E13

Change of basis | Essence of linear algebra, chapter 13

3Blue1Brown

14 

3BLUE1BROWN SERIES S1 • E14

Eigenvectors and eigenvalues | Essence of linear algebra, chapter 14

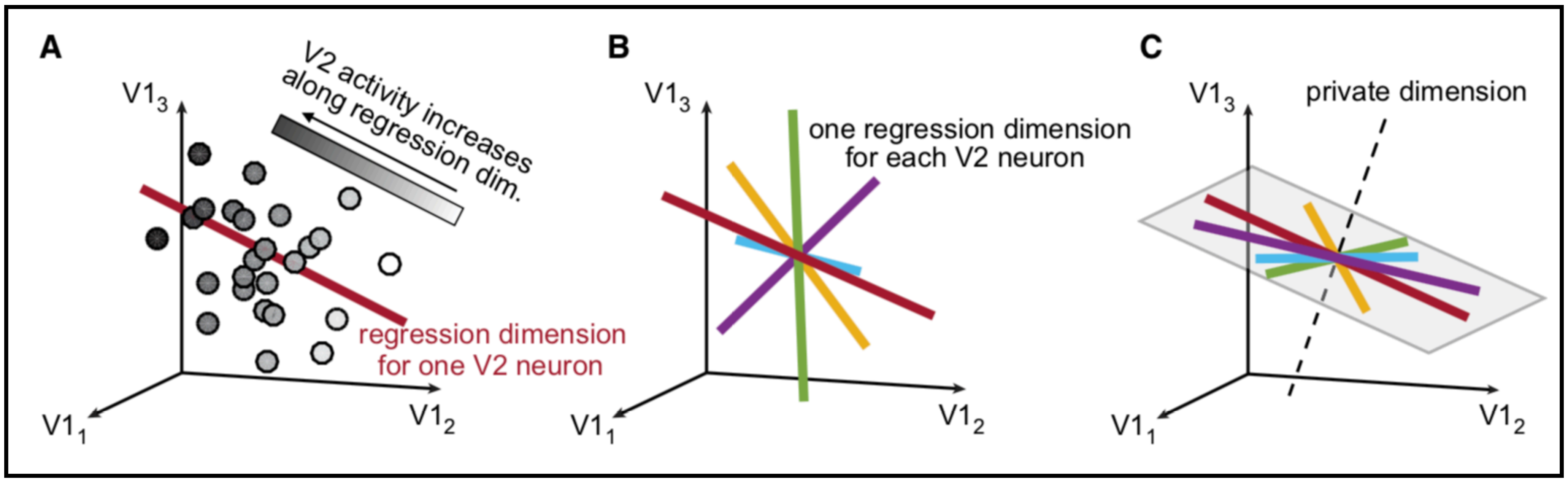
3Blue1Brown

Example in Neuroscience

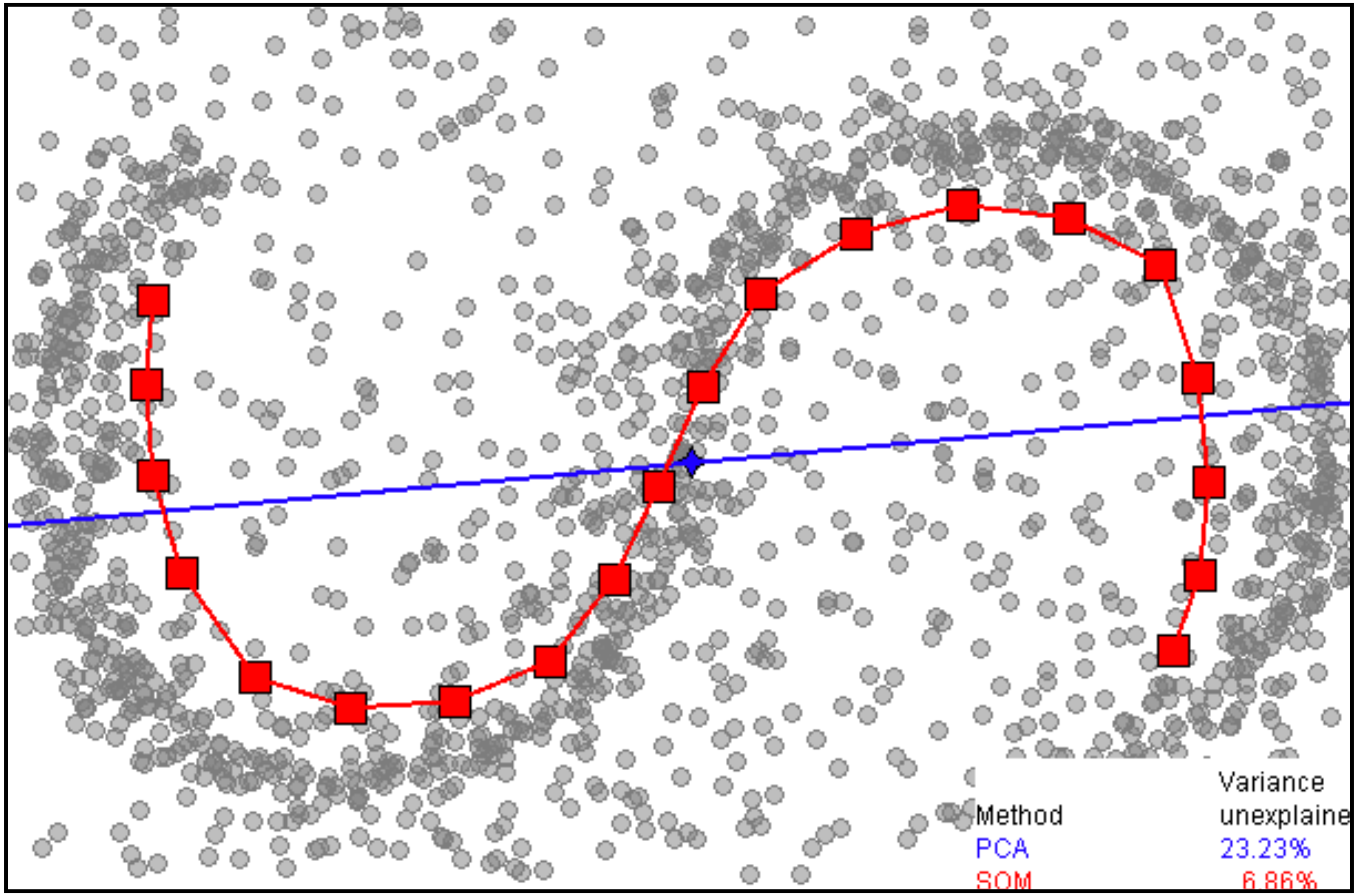
Neuron

Article

Cortical Areas Interact through a Communication Subspace



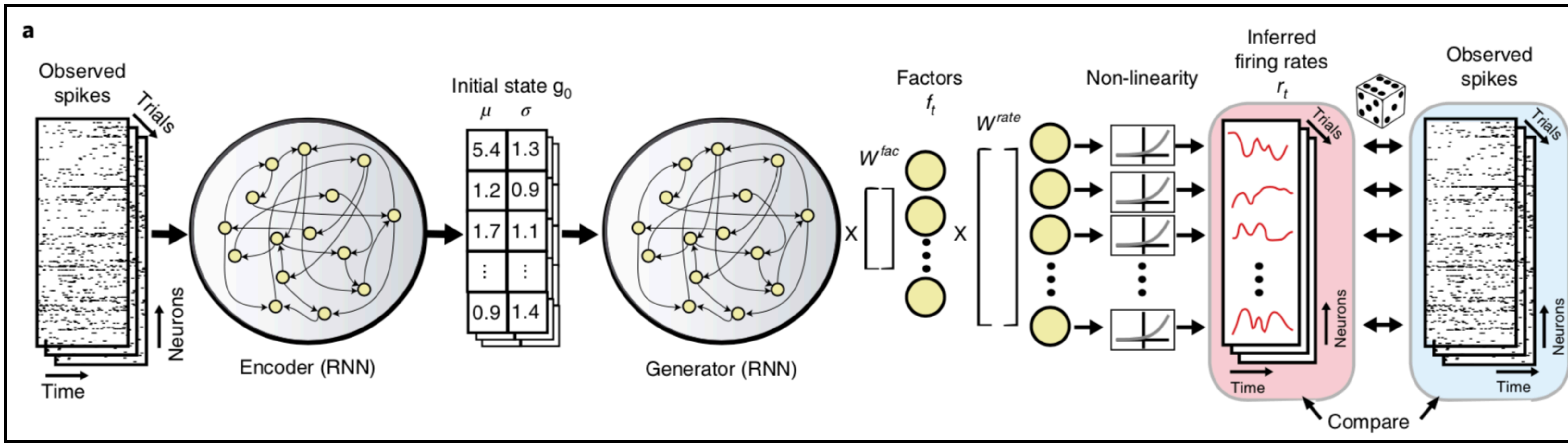
Nonlinear Dimensionality Reduction/Embedding

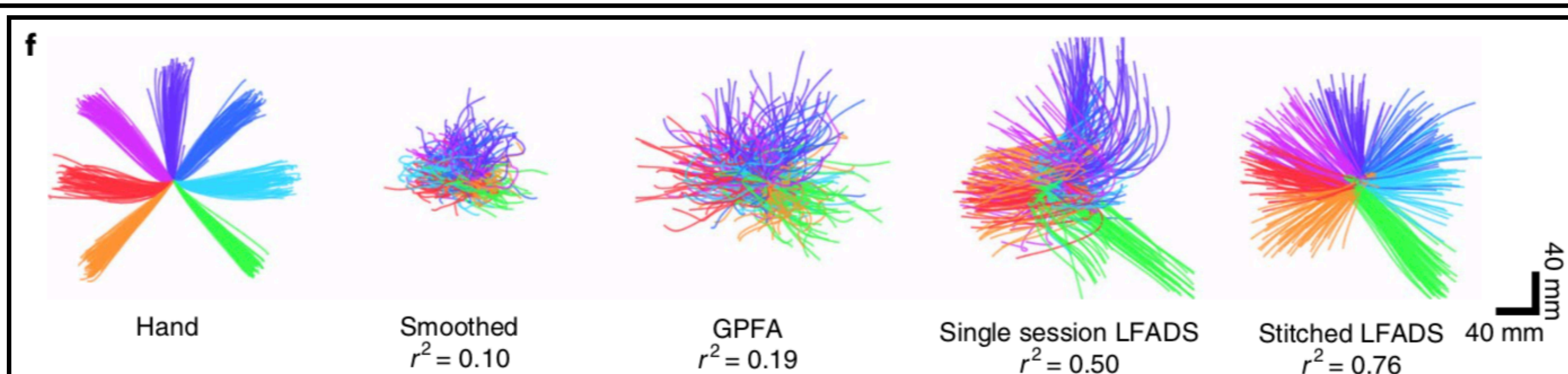


Nonlinear Dimensionality Reduction/Embedding

nature methods | ARTICLES
https://doi.org/10.1038/s41592-018-0109-9

Inferring single-trial neural population dynamics using sequential auto-encoders





Goals for Today

1. More PCA intuition & examples
2. Define information & common quantities
3. Examples in neuroscience

(Shannon) Information

Informal Definition:
Information reduces uncertainty of outcome, given some expectation

- observing an unlikely event is very surprising
- observing an likely event is not (does not convey a lot of information)

How many questions do you need to ask to guess a random number (with equal likelihood)?

Between 1-2?

Between 1-4?

Between 1-8?

Conversely, if given the outcome, how many questions does it “save” you?

(Shannon) Information

Formal Definition: "surprisal" of a message, m

$$\mathrm{I}(m) = \log\left(\frac{1}{p(m)}\right) = -\log(p(m))$$

Surprisal of observing a number:

Between 1-2?

Between 1-4?

Between 1-8?

Observing a single outcome gives you $-\log_2 P$ bits of information.

Formal Definition: "surprisal" of a message, m

$$\mathrm{I}(m) = \log\left(\frac{1}{p(m)}\right) = -\log(p(m))$$

Example	Possible Events	Probabilities	Surprisal	
Coin flip	H, T	1/2, 1/2	1,1	
Lottery	winning jackpot, not winning	1/(10mil), (10mil-1)/ 10mil	log10mil, ~0	
babies	B,G,BB,BG,GG, 3 or more	45.5%, 44.5%, 3%, 3%, 3%, 1%		
semantic incongruity	cream, sugar, dog	1/3, 1/5, 1/1000, rest		

Formal Definition: "surprisal" of an outcome/message, m

$$\mathrm{I}(m) = \log\left(\frac{1}{p(m)}\right) = -\log(p(m))$$

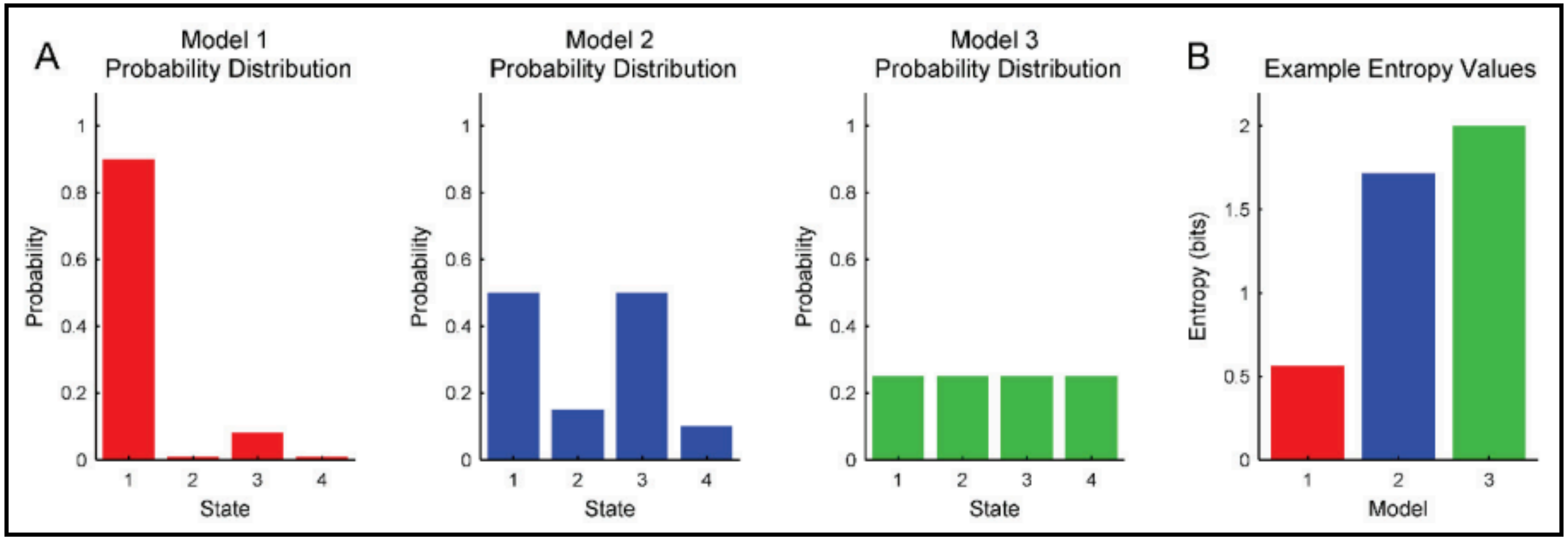
Property of a single outcome/message

Formal Definition: entropy (property of a variable's probability distribution)

$$H(X) = \mathbb{E}_X[I(x)] = -\sum_{x\in\mathbb{X}} p(x)\log p(x).$$

The amount of uncertainty about a variable X when its distribution is known.

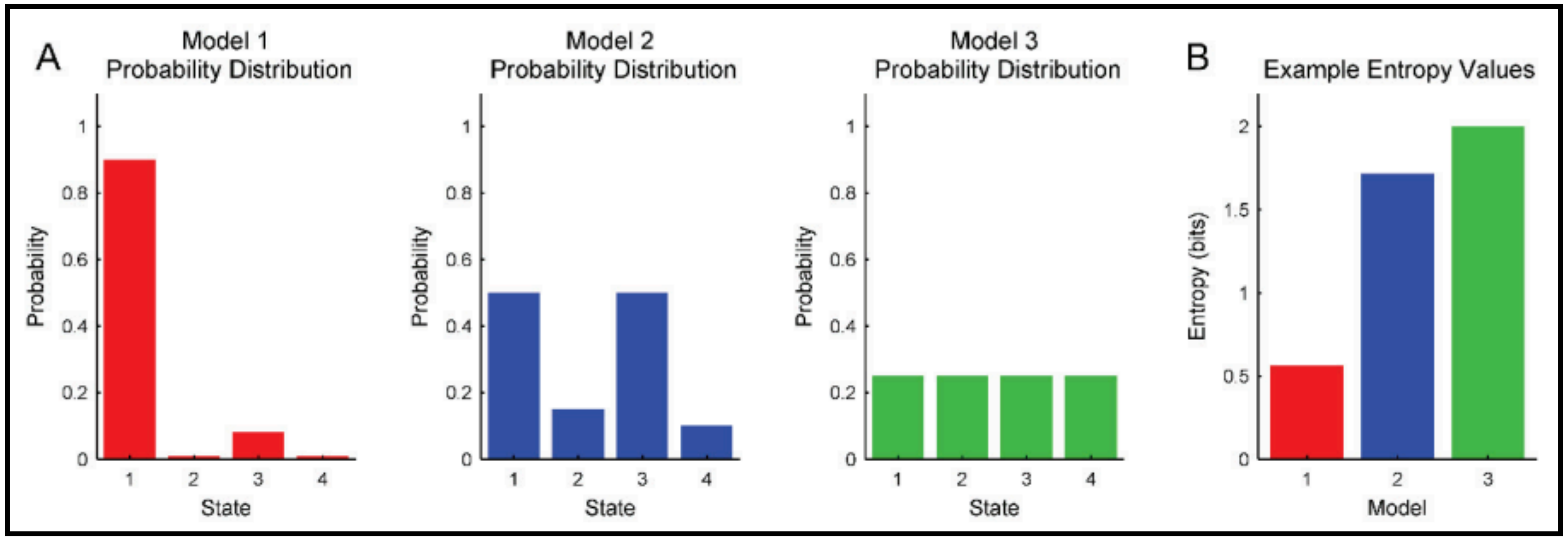
$$H(X) = \mathbb{E}_X[I(x)] = -\sum_{x \in \mathbb{X}} p(x) \log p(x).$$



Expected value of "surprisal" of individual outcomes (weighted average)

Practical consideration: for a continuous variable (voltage), this usually depends on histogram bin size.

Structure/Correlation Reduces Entropy



- Suppose English had no structure: P(a)=P(b)=P(c)=...=P(z)=1/26

$$H_{\text{independent letters}} = -\sum_{w=1}^{26} \frac{1}{26} \log_2 \frac{1}{26}$$
$$= \log_2 26 = 4.7 \text{ bits}$$

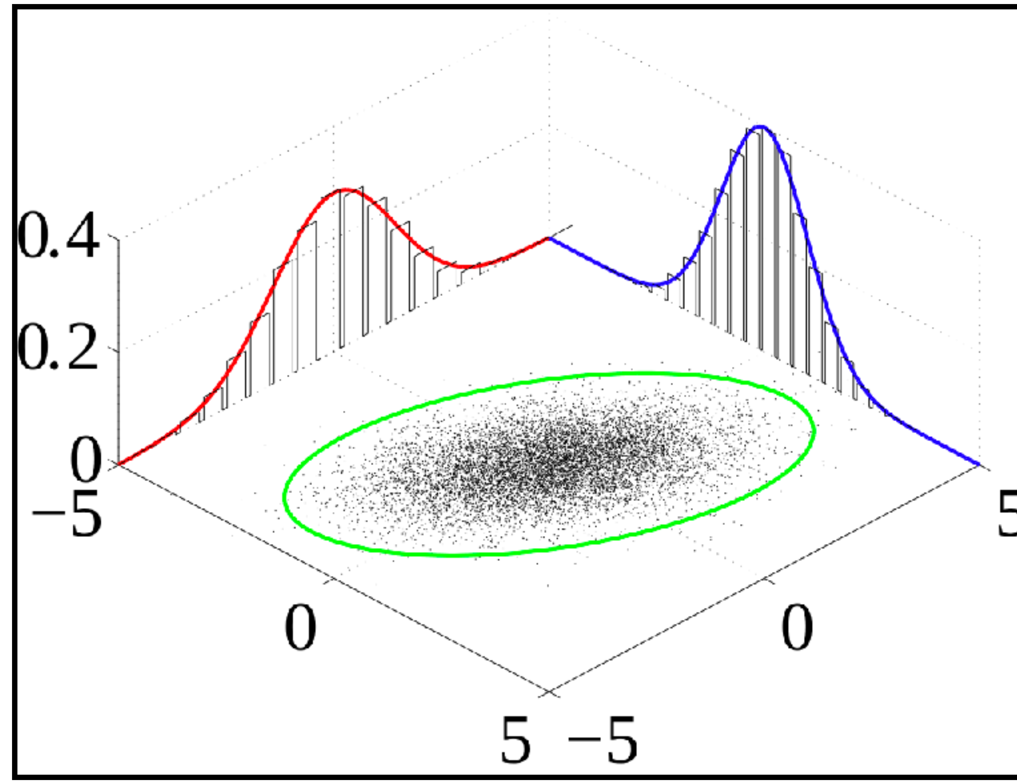
English text has between 0.6 and 1.3 bits of entropy per character of the message.

Joint Entropy

Joint Entropy (for multivariate distributions)

$$H(X,Y) = \mathbb{E}_{X,Y}[-\log p(x,y)] = -\sum_{x,y} p(x,y) \log p(x,y)$$

		BLACK (Y)				
	$f(x,y)$	1	2	3	4	$f_X(x)$
RED (X)	1	1/16	1/16	1/16	1/16	4/16
	2	1/16	1/16	1/16	1/16	4/16
	3	1/16	1/16	1/16	1/16	4/16
	4	1/16	1/16	1/16	1/16	4/16
	$f_Y(y)$	4/16	4/16	4/16	4/16	1

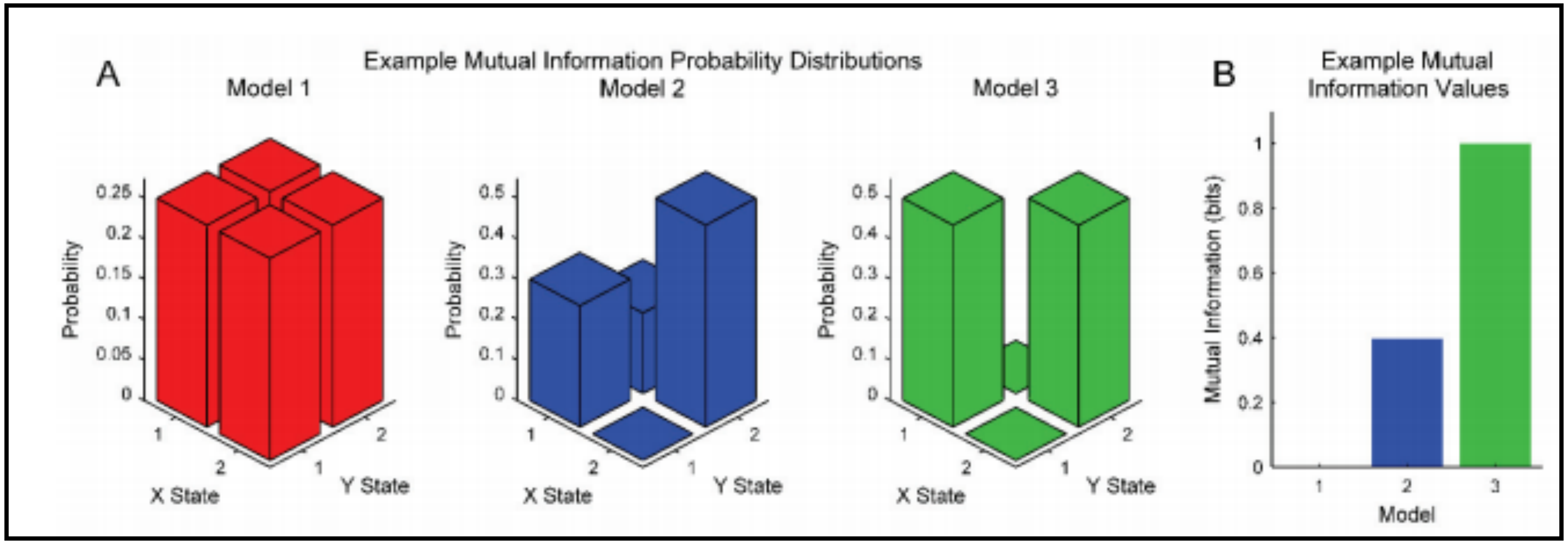


Property: if X and Y are **independent**, H(X,Y) = H(X) + H(Y)

Mutual Information

$$I(X;Y) = \mathbb{E}_{X,Y}[SI(x,y)] = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)\,p(y)}$$

How much information can be obtained, or how much uncertainty can be reduced, about one variable X when the other variable Y is observed.



$$I(X;Y) = I(Y;X) = H(X) + H(Y) - H(X,Y).$$

Kullback-Leibler (KL) Divergence

$$D_{\mathrm{KL}}(p(X)||q(X)) = \sum_{x \in X} -p(x)\log q(x) \; - \; \sum_{x \in X} -p(x)\log p(x) = \sum_{x \in X} p(x)\log\frac{p(x)}{q(x)}.$$

Measures the difference between two distributions:

if p(X) is the true distribution and q(X) is our guess, KL divergence measures how much more we are surprised.

Fair coin q(X=H) = 0.5 vs. p(X=H) = 0.9

1. More PCA intuition & examples
2. Define information & common quantities
3. Examples in neuroscience

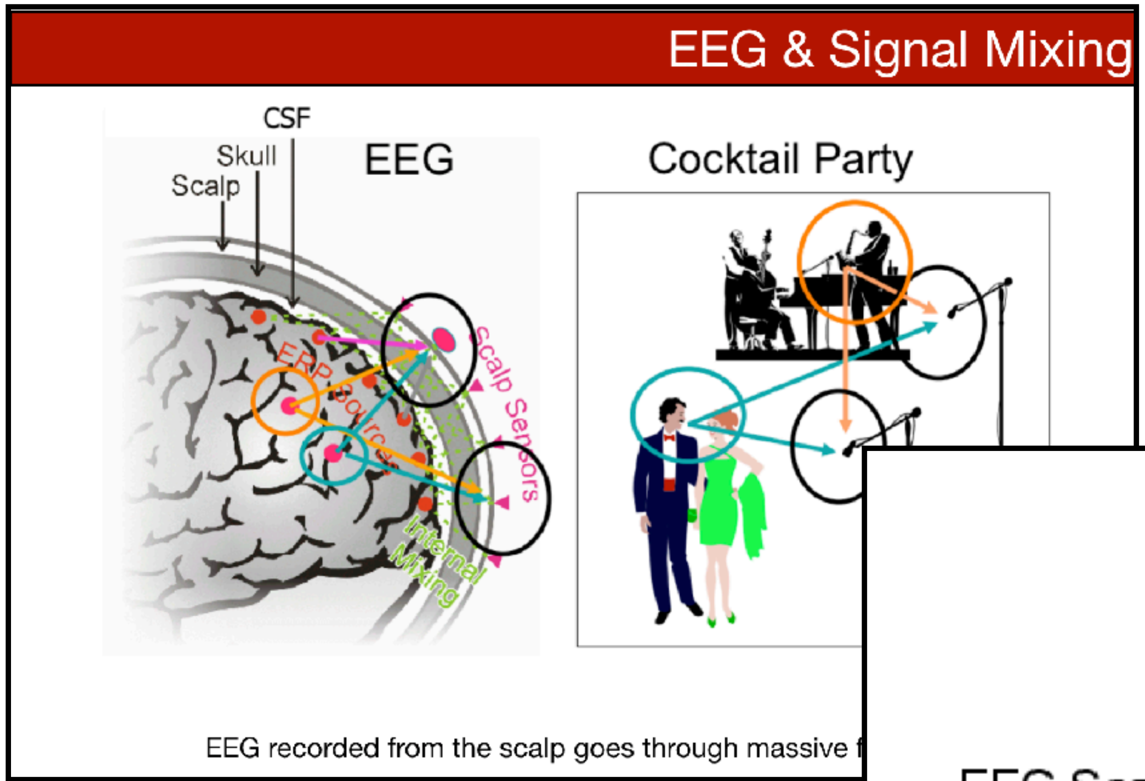
eNeuro

Reviews | Novel Tools and Methods

A Tutorial for Information Theory in Neuroscience

Nicholas M. Timme[1] and Christopher Lapish[1]

Independent Component Analysis



Blind Source Separation
(Cocktail Party Problem)

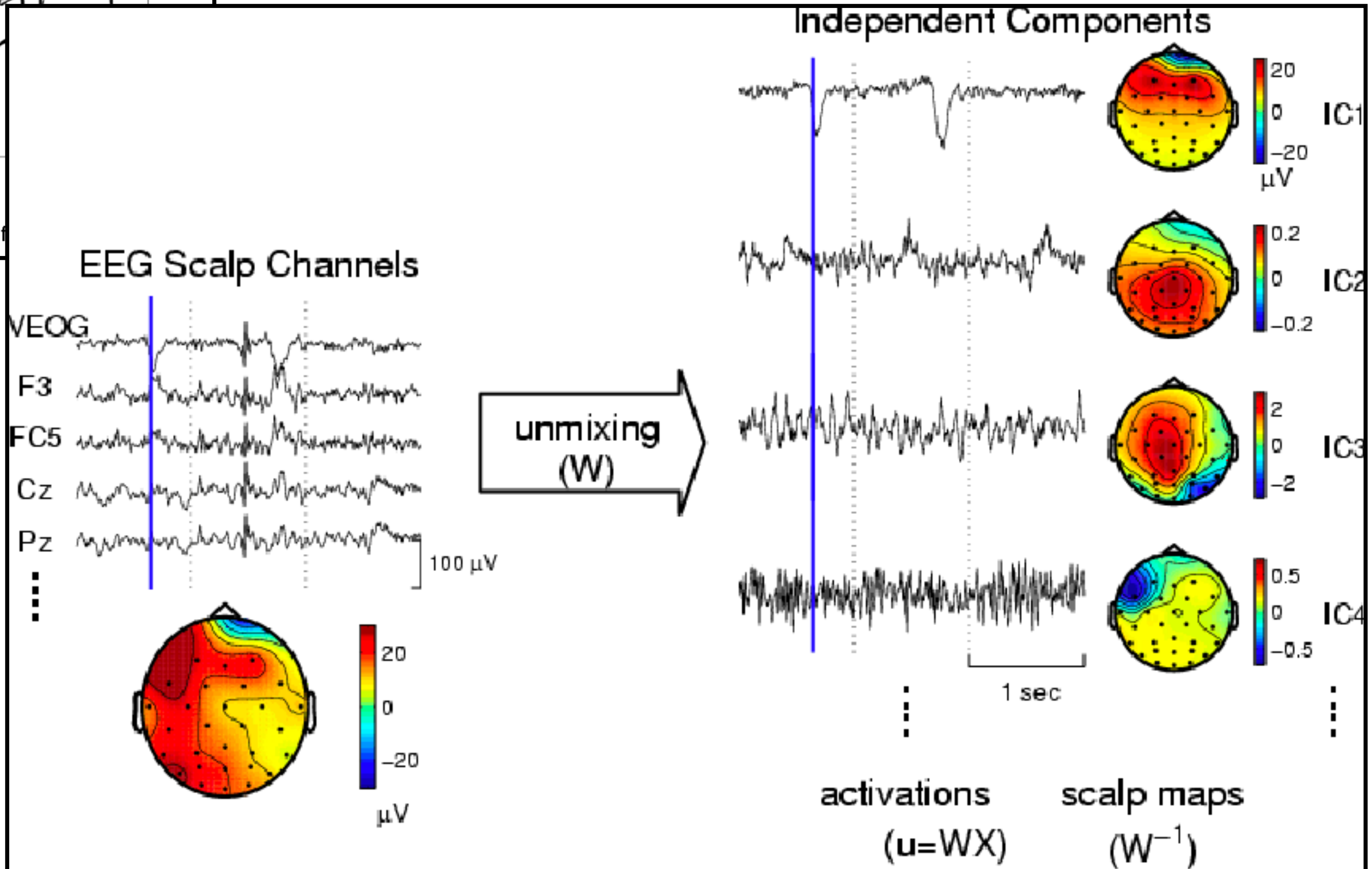


Fig. 1 EEG Signals being broken into ICs using ICA

Independent Component Analysis

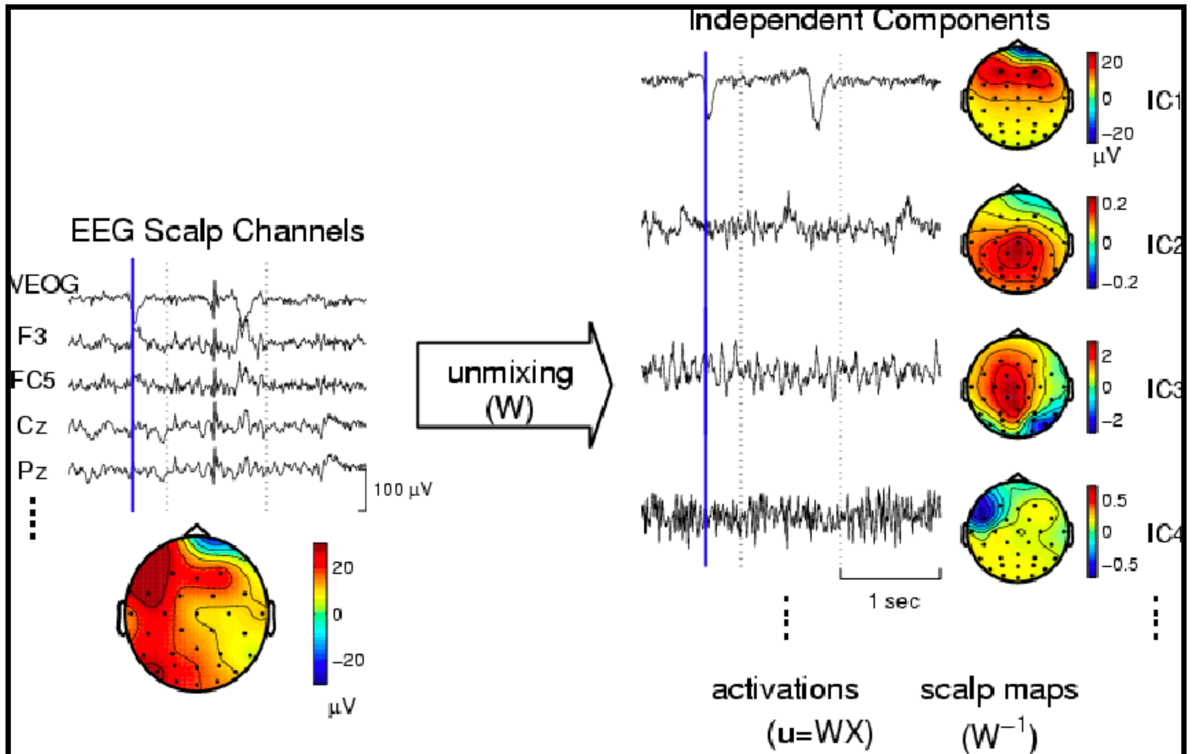


Fig. 1 EEG Signals being broken into ICs using ICA

An Introduction to Independent Component Analysis: InfoMax and FastICA algorithms

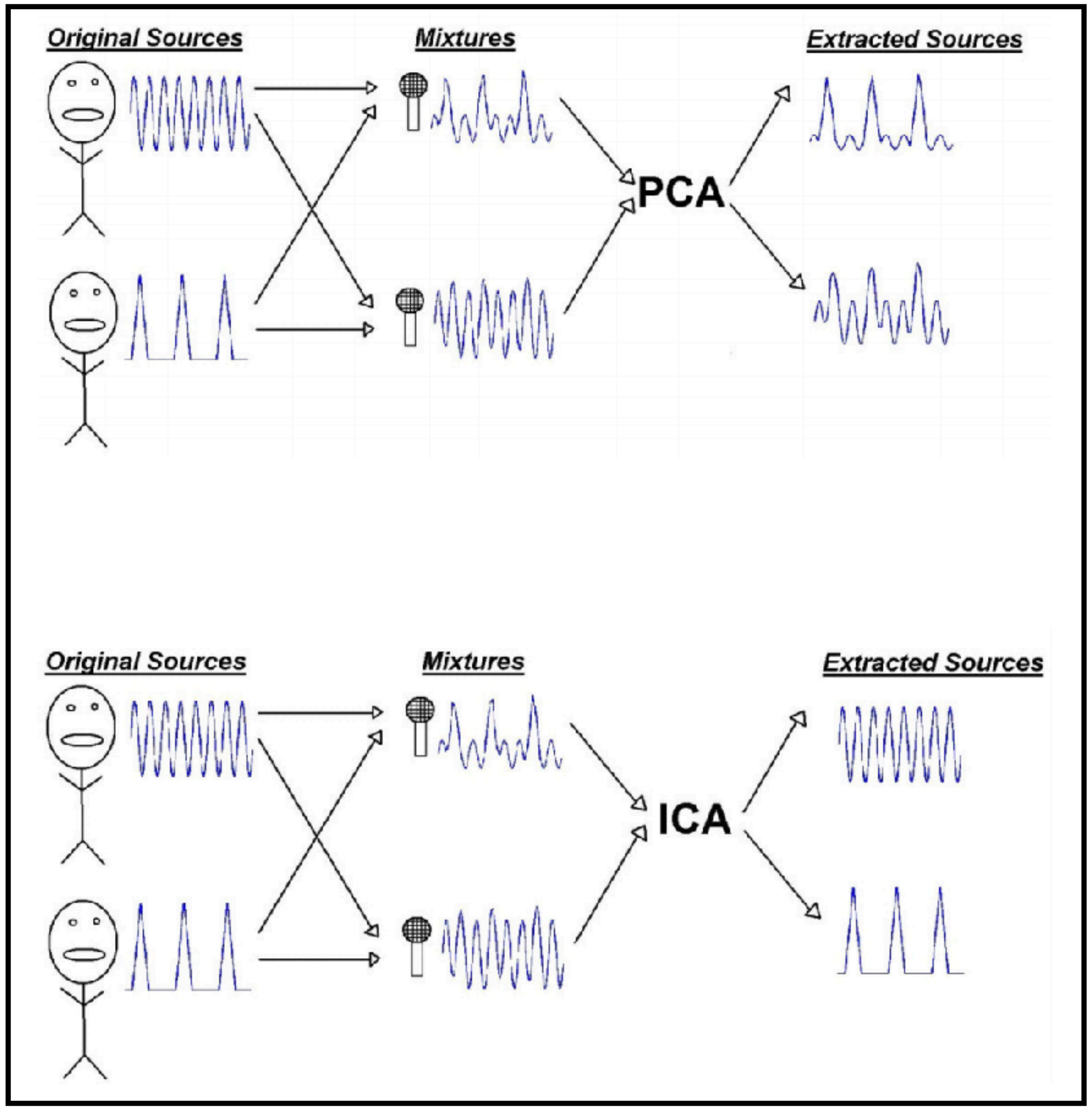
Dominic Langlois, Sylvain Chartier, and Dominique Gosselin
University of Ottawa

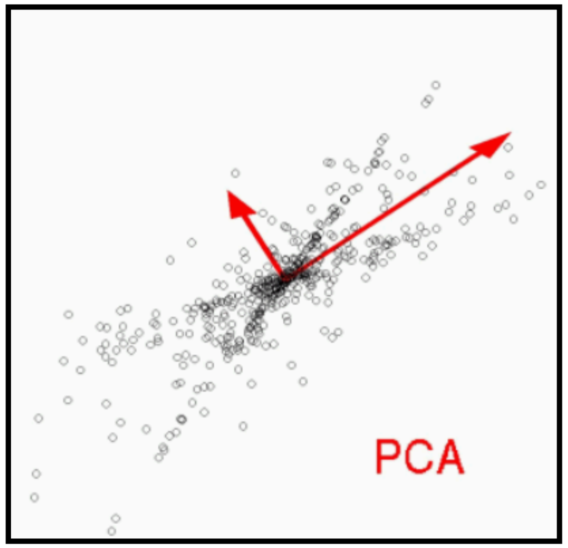
ICA is a family of algorithms, e.g.,:

InfoMax minimizes mutual information between latent components.

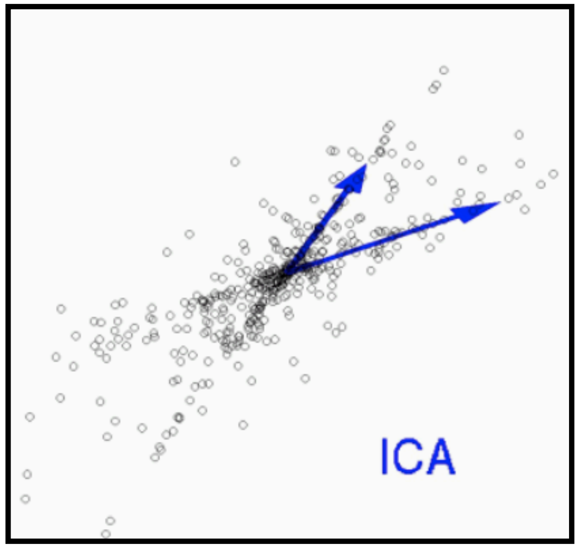
FastICA maximizes entropy of components (encourages non-Gaussianity)

Independent Component Analysis



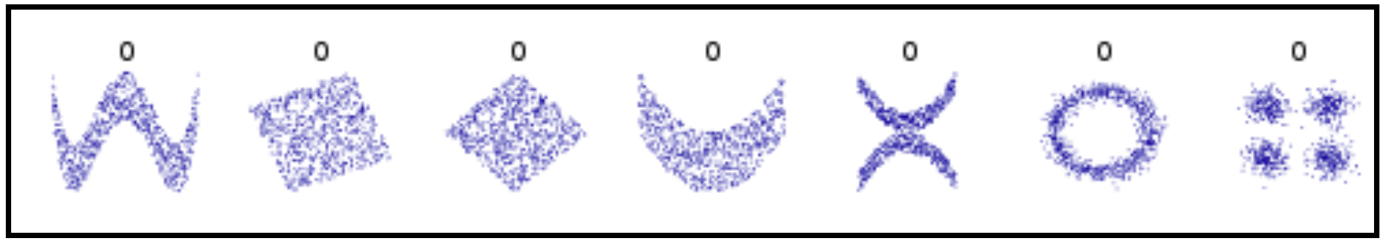


Orthogonal (Linear Independence)
Finds directions of maximal variance



These are all linearly independent

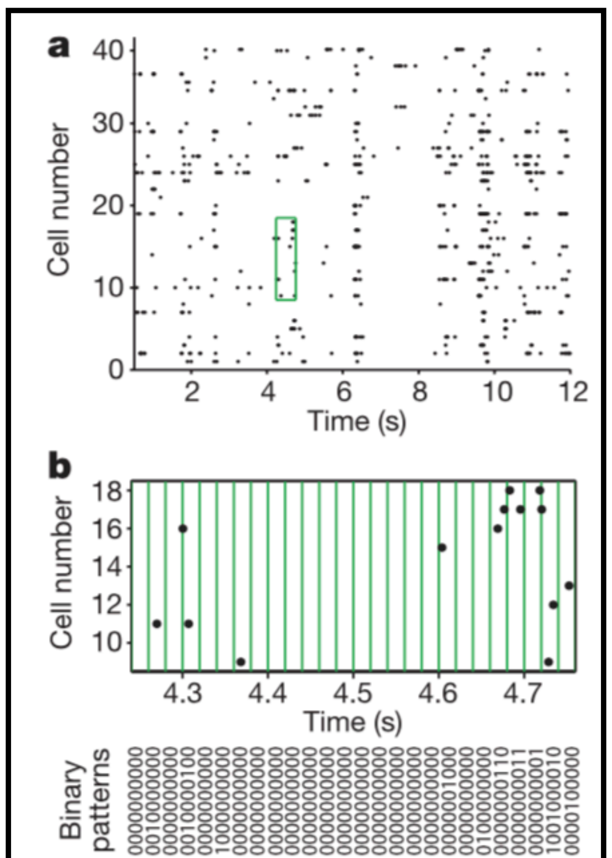
Statistical Independence

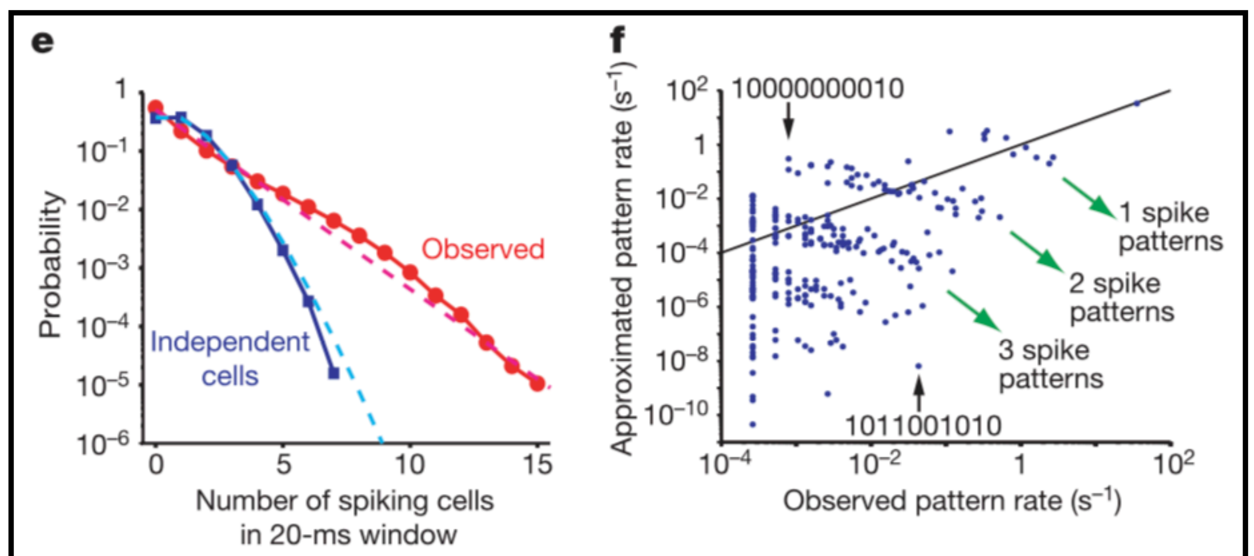


Information Theory of Neural Coding

ARTICLES

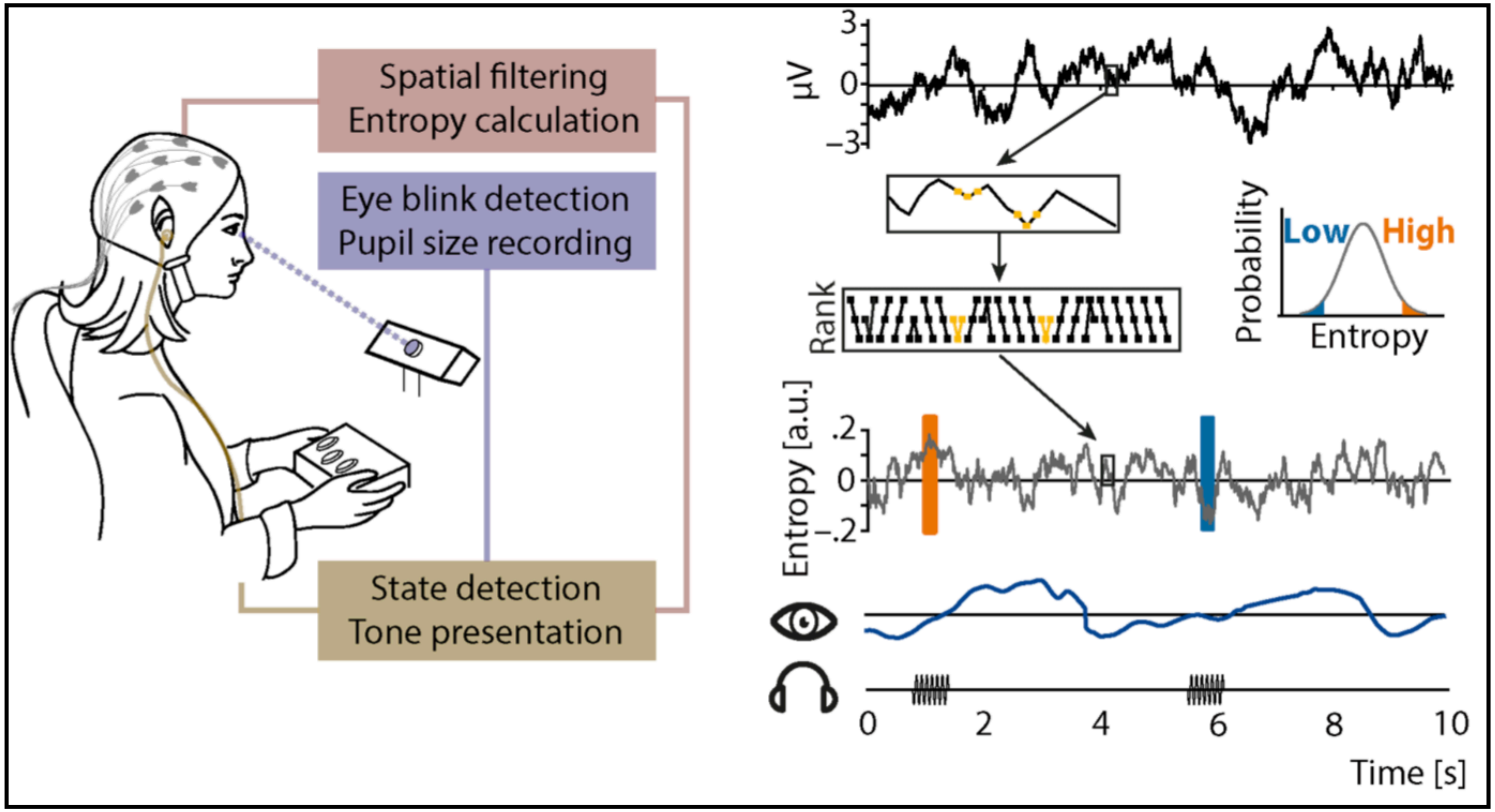
Weak pairwise correlations imply strongly correlated network states in a neural population





RESEARCH ARTICLE

Changes in EEG multiscale entropy and power-law frequency scaling during the human sleep cycle



Waschke et al., 2019

1. More PCA intuition & examples
2. Define information & common quantities
3. Examples in neuroscience

https://tinyurl.com/cogs118c-att