Brain decoding with Machine Learning

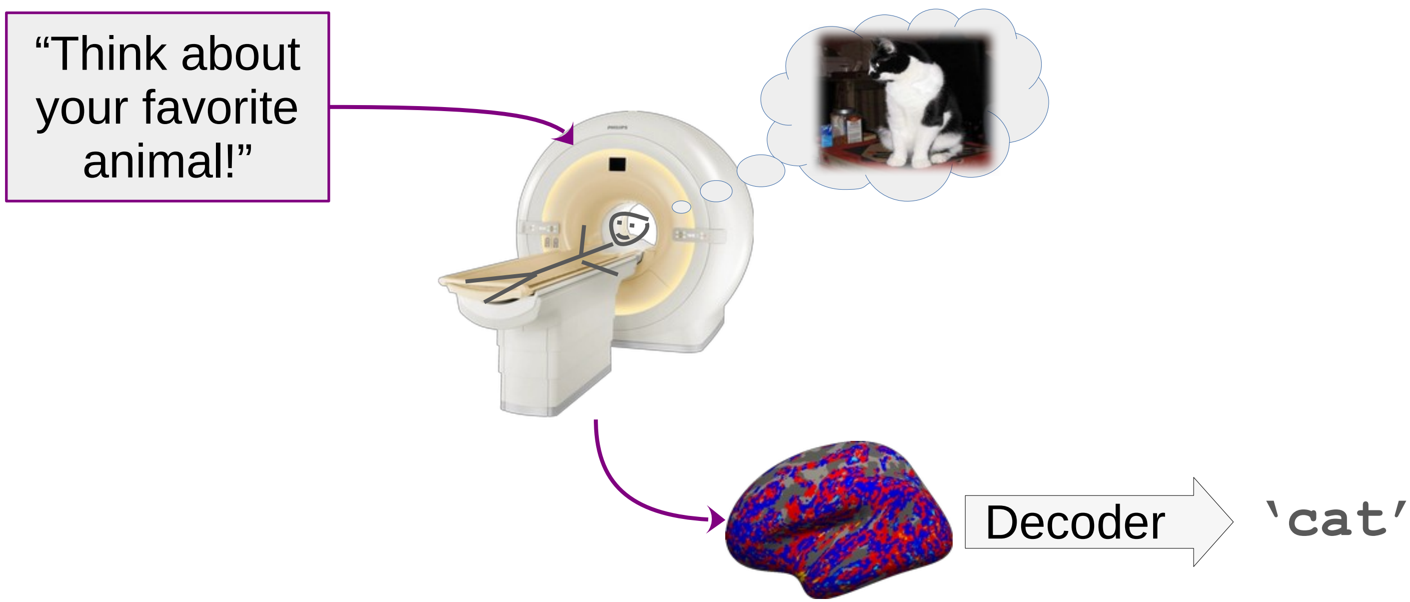
Mitja Nikolaus
Slides adapted from Leila Reddy
CerCo/CNRS

Brain Decoding

- Using brain imaging techniques to decode what people see, imagine, remember, etc.

Brain Decoding

- Using brain imaging techniques to decode what people see, imagine, remember, etc.



Outline

- Brief introduction to brain recording methods
- The “old” way: decoding stimulus info using classification methods (e.g. SVM)
- Decoding visual and language stimuli using Deep Neural Network (DNN) features
- Face and scene reconstruction from fMRI

Brain recording methods

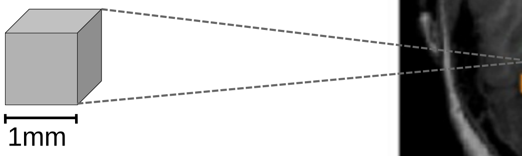
- **Single neuron electrophysiology:**
  - invasive, mostly in animals, rare in humans

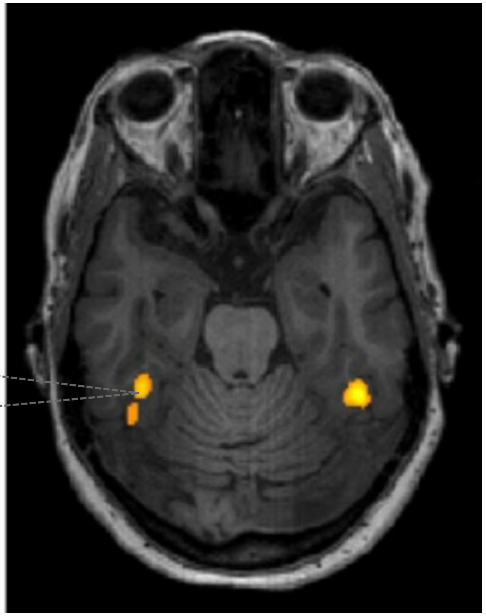
- **EEG (Electroencephalography)**
  - Measure electrical activity from the scalp
  - Good temporal resolution, poor spatial resolution

- **MEG (Magnetoencephalography)**
  - Similar to EEG but records magnetic fields
- **fMRI (functional Magnetic Resonance Imaging)**
  - measures brain activity by detecting changes in blood flow
  - Good spatial resolution, poor temporal resolution

Brain recording methods

- **functional Magnetic Resonance Imaging (fMRI):**
  - powerful tool to study the human brain non-invasively
  - spatial resolution:
    - voxels (“3D pixels”)
    - voxel size: ~$1mm^3$
    - ~10^5 neurons in one $1mm^3$ voxel



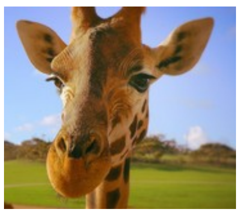


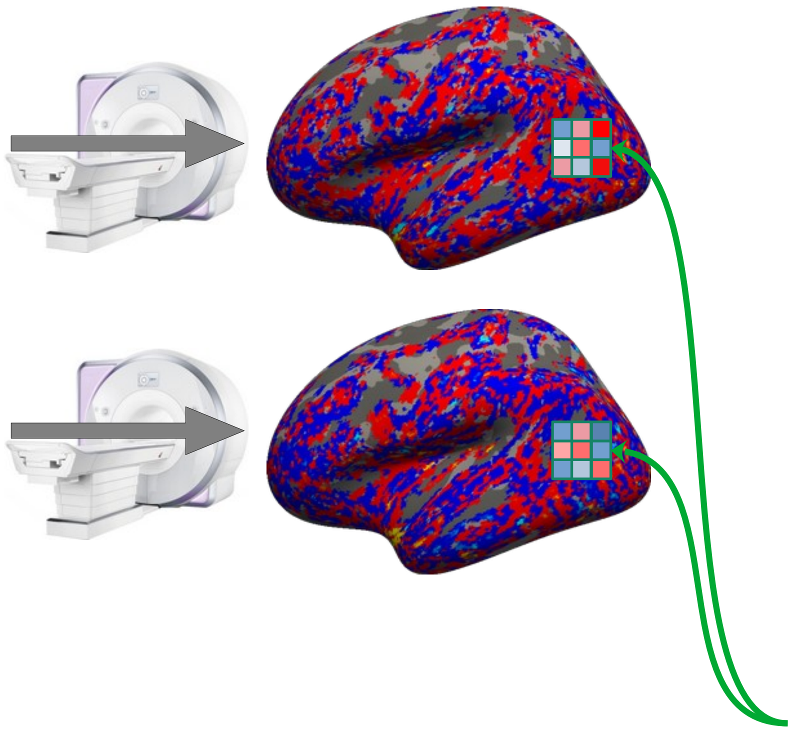
Brain Decoding: fMRI patterns

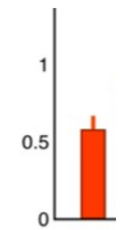
Image seen by participant in the scanner

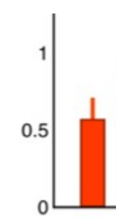
Measured fMRI brain response

fMRI response averaged across voxels







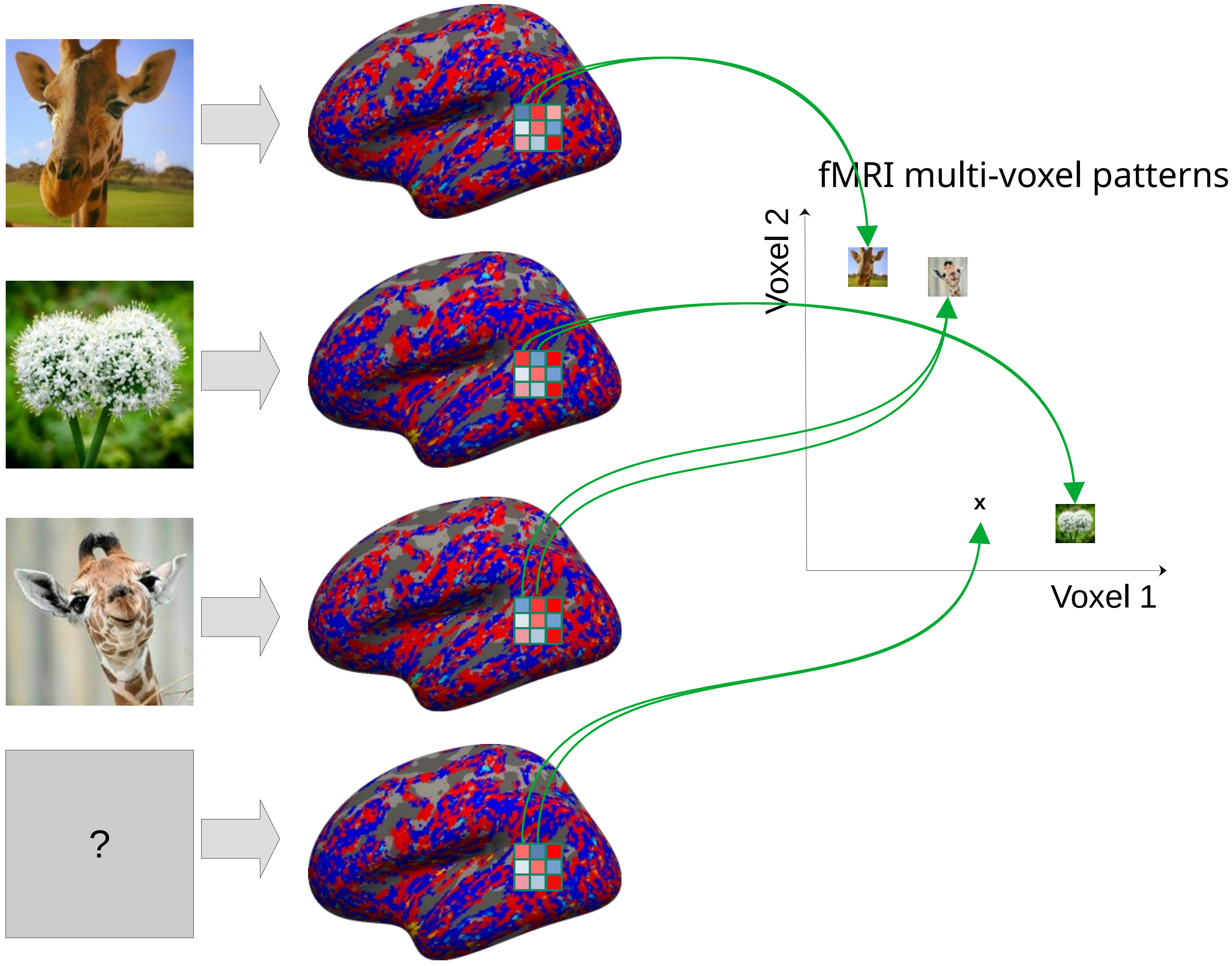


The averaged response may be identical in different conditions..

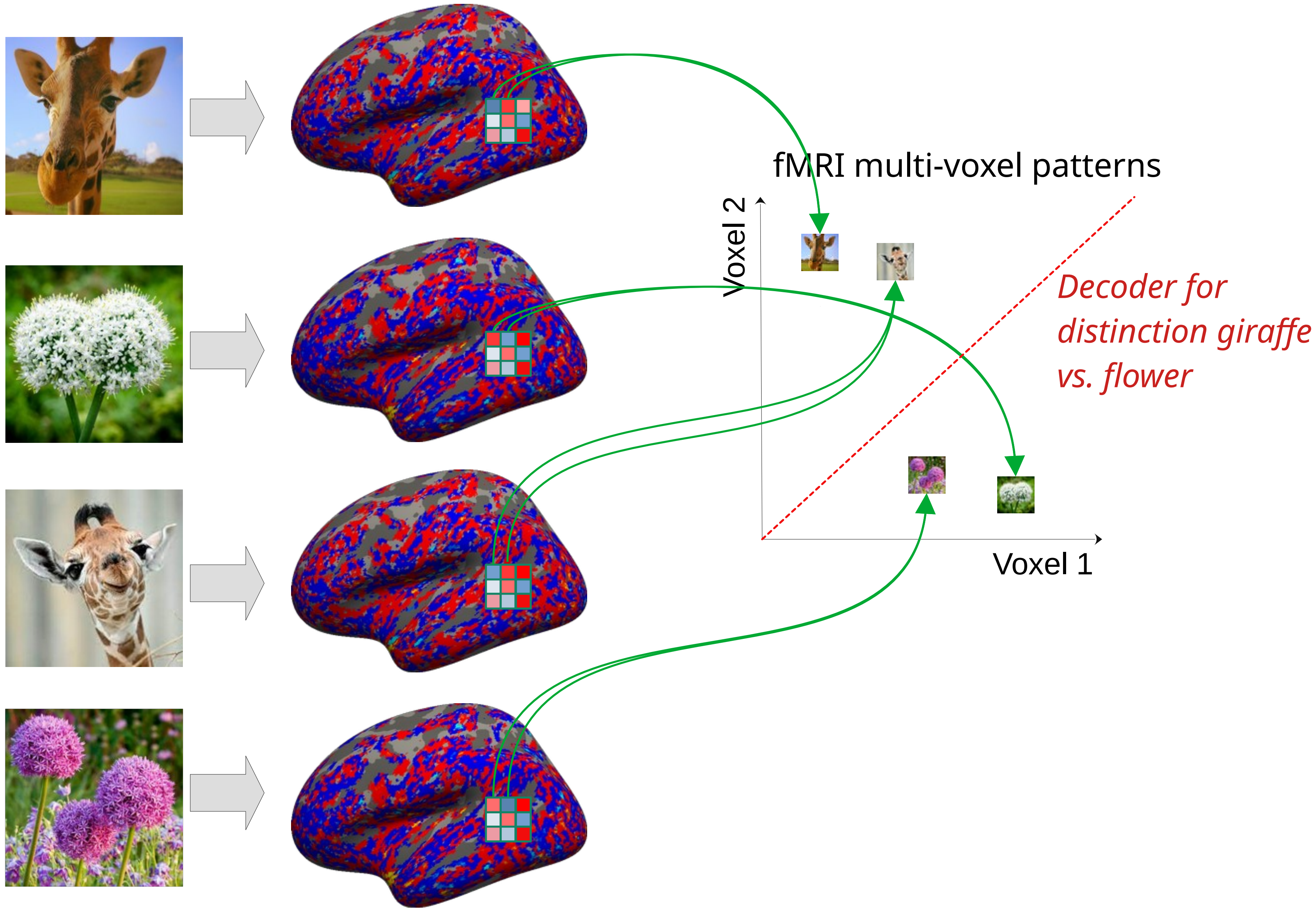
but the fMRI multi-voxel patterns may still carry information!

→ Learn a relationship between the fMRI patterns and the stimuli

Brain Decoding: Training



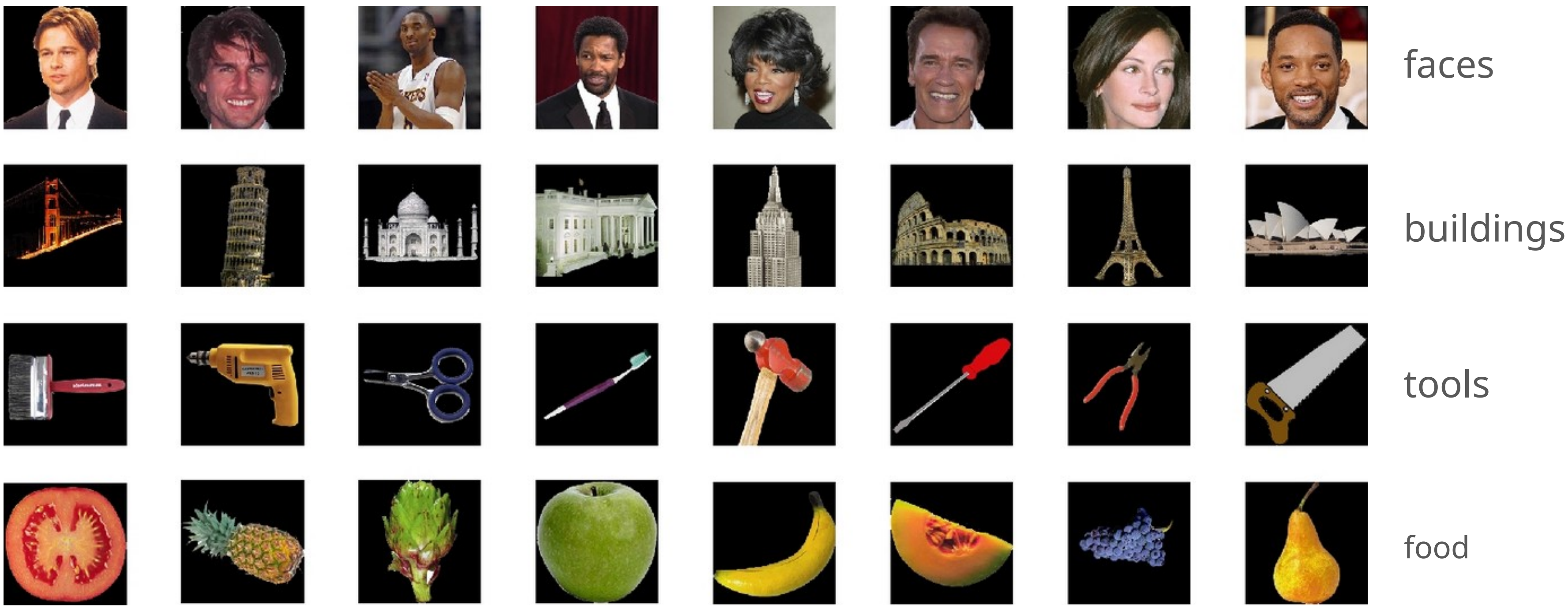
Brain Decoding: Training



Decoding perception and imagery

Example study: Reddy et al., 2010

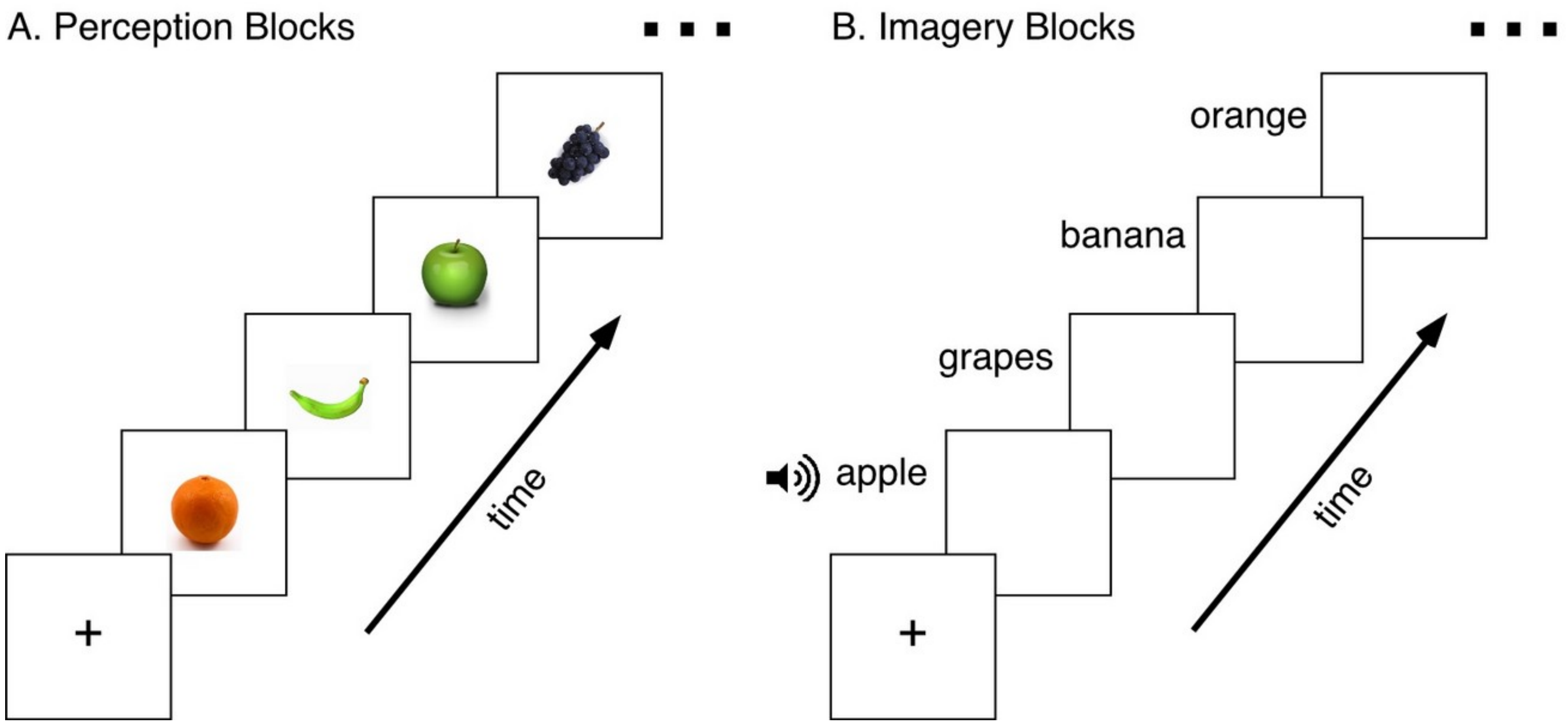
Decoding object categories of perception and imagery



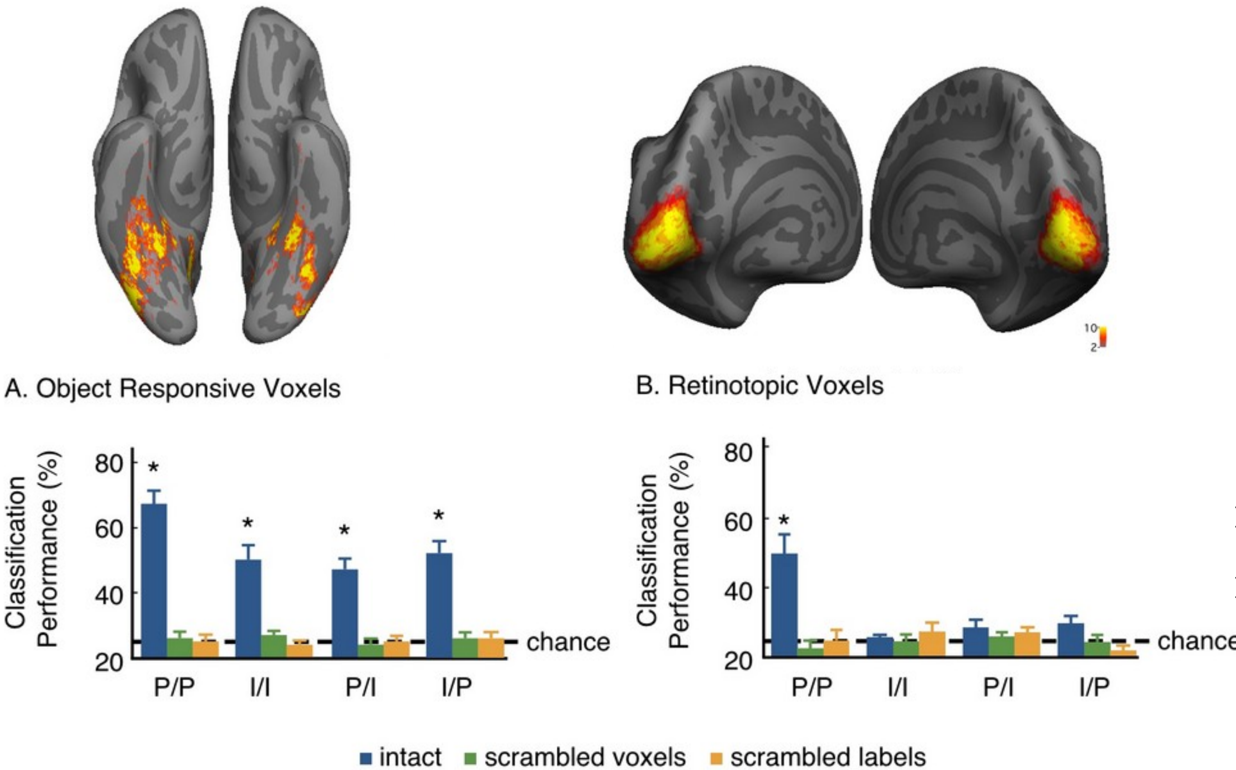
Decoding perception and imagery

Example study: Reddy et al., 2010

Decoding object categories of perception and imagery



Decoding perception and imagery



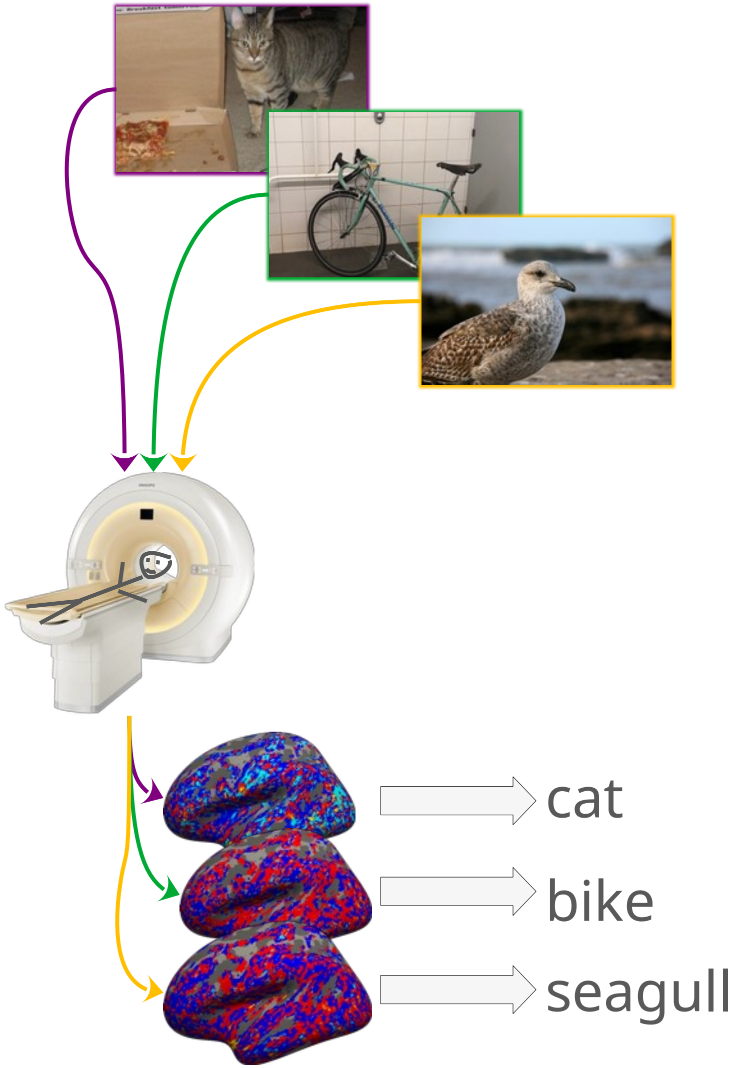
Decoder model: SVM

P/P: train and test on perception
I/I: train and test on imagery
P/I: train on perception, test on imagery
I/P: train on imagery, test on perception

- Perception decoding from low- and high-level areas
- Imagery decoding only in high-level regions
- Perception and imagery share representations in high-level regions

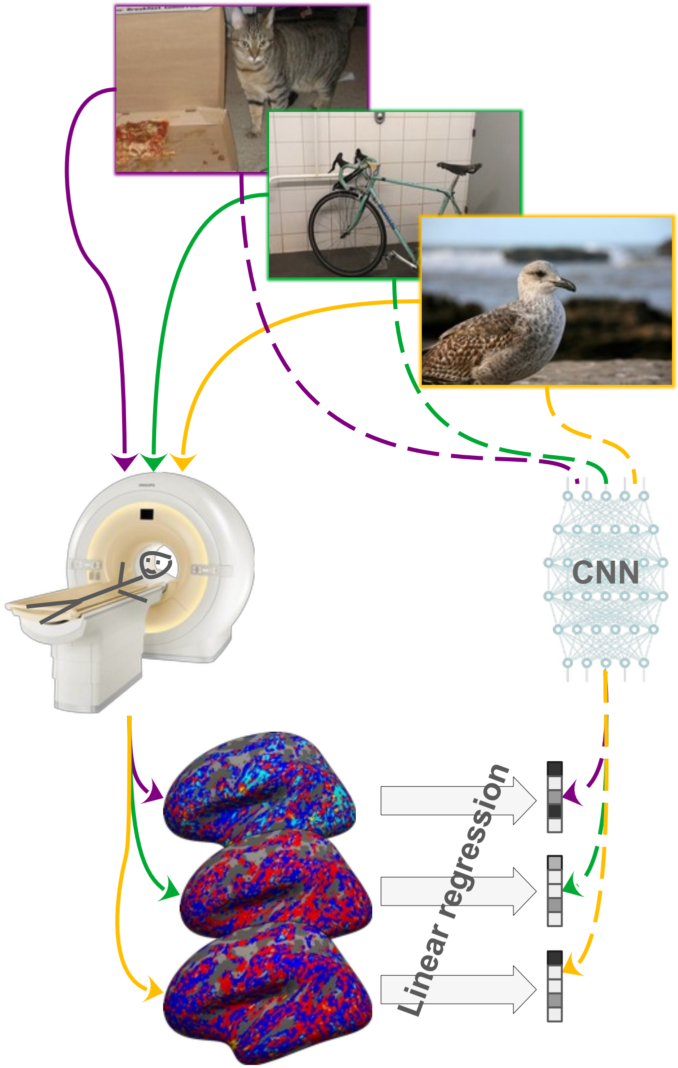
Reddy et al., 2010

Brain decoding methods



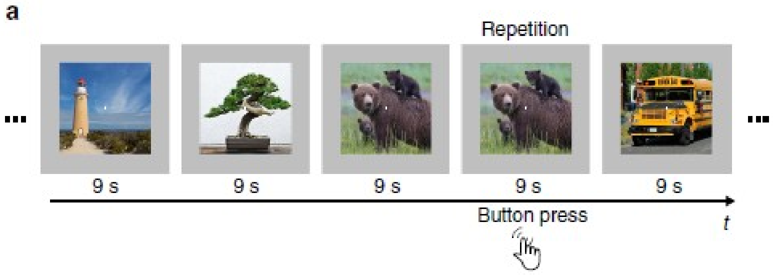
- Classification-based approach (as in Reddy et al., 2010):
  - Learn mapping between brain patterns and object classes
  - Limited to decoding the classes used for training
  - Cannot generalize to novel classes

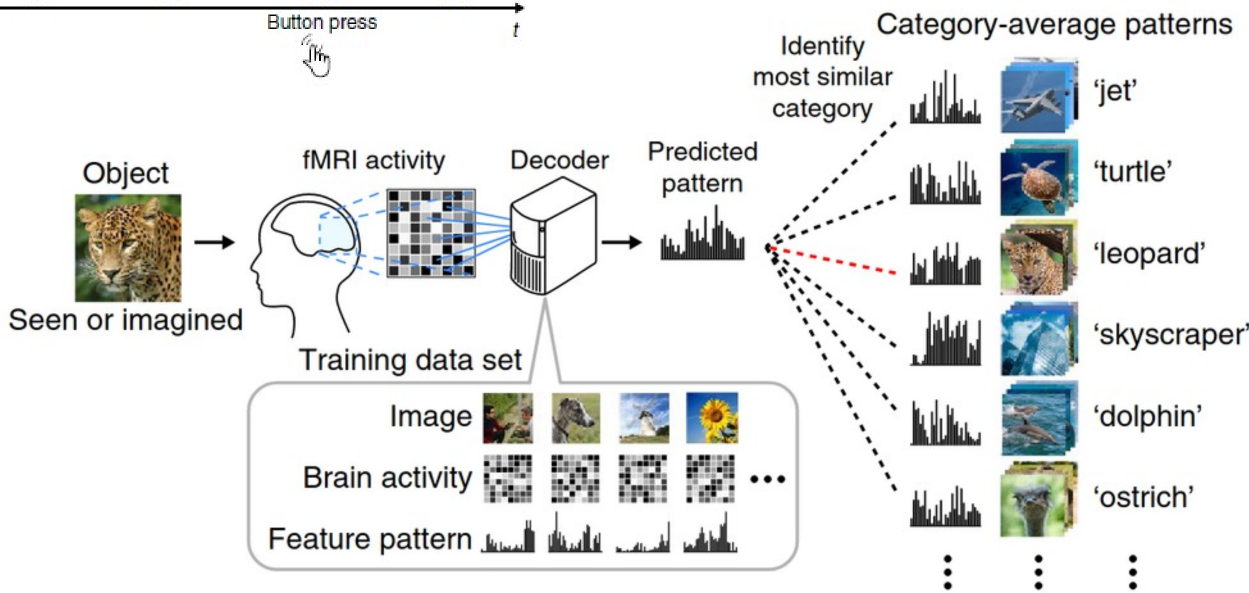
Brain decoding methods

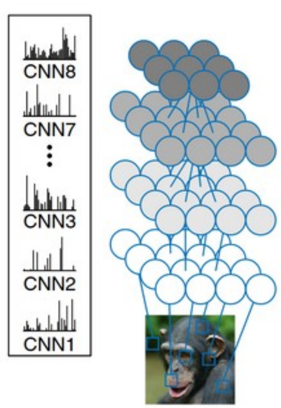


- Regression-based approach with Deep Neural Network (DNN) features
  - Use DNNs (e.g. CNN) to extract features from image
  - Learn mapping (e.g. Linear regression) between brain patterns and representations in the DNN feature space
  - Allows generalization to new classes

Decoding with DNN features







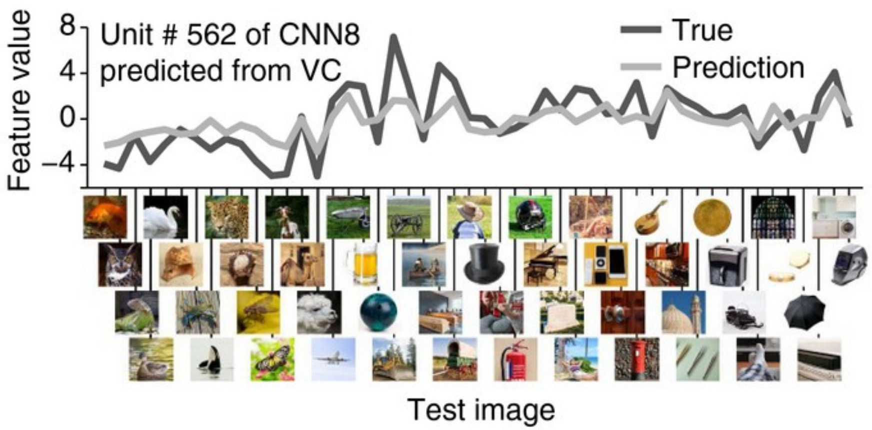
Decoding accuracy = correlation between true and predicted feature vectors

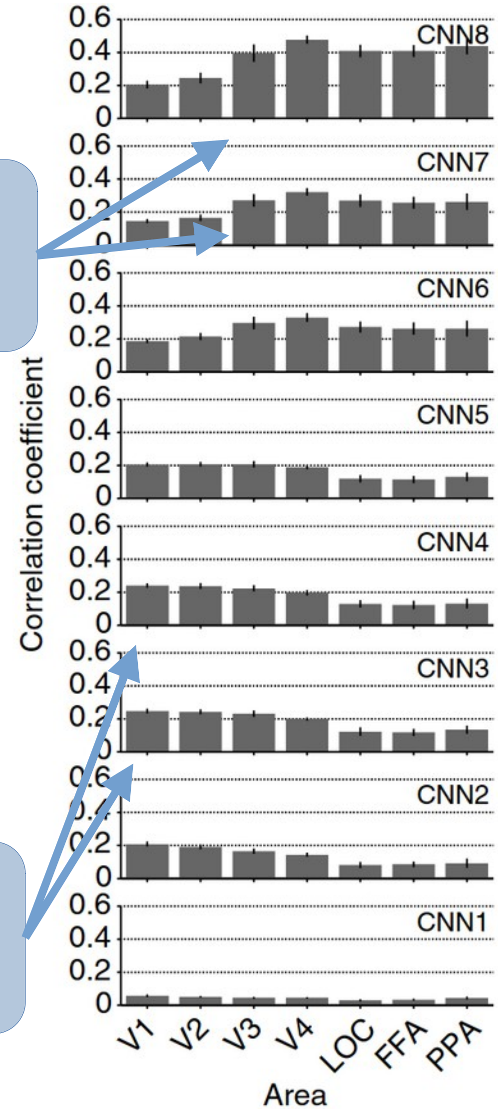
Horikawa & Kamitani, 2017

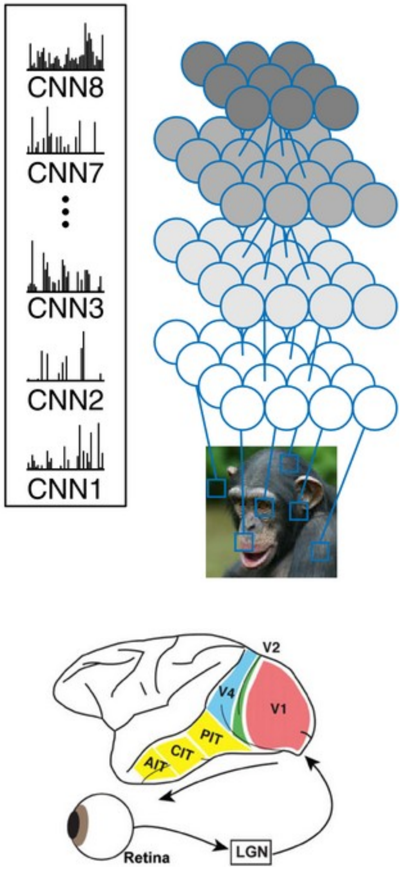
Decoding with DNN features

Decoding of seen objects

higher-order features are better predicted from fMRI signals in higher ROIs





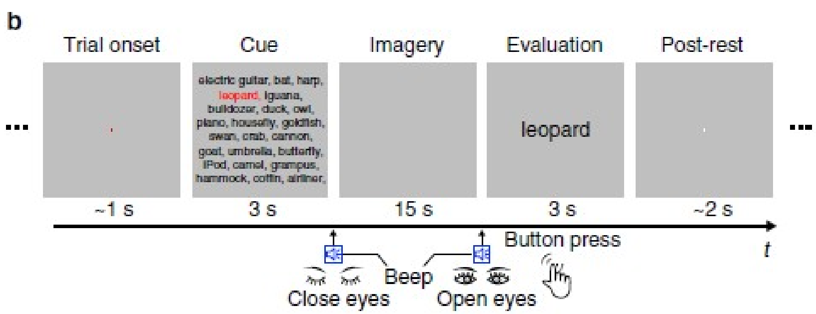


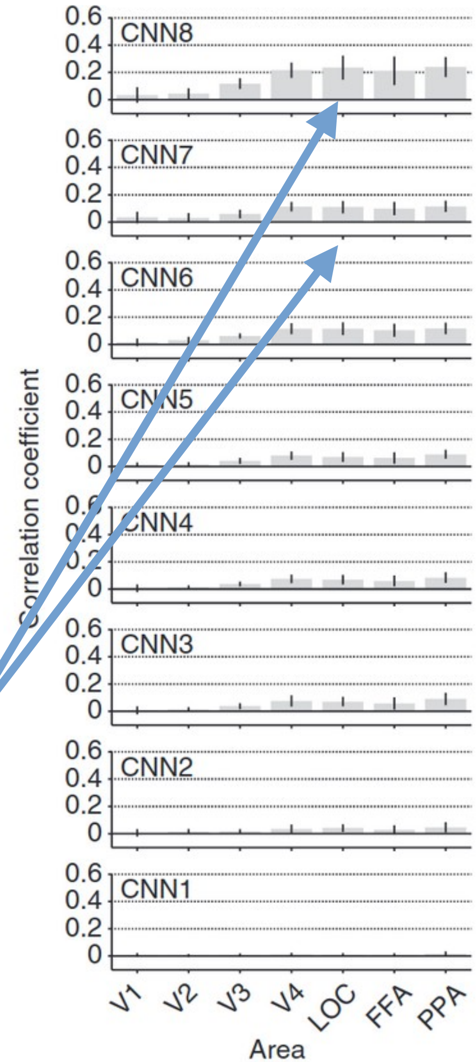
lower-order features are better predicted from fMRI signals in lower ROIs

Horikawa & Kamitani, 2017

Decoding with DNN features

Decoding of imagined objects





Imagery decoding is only possible from fMRI signals in higher ROIs

Horikawa & Kamitani, 2017

DNNs and brain decoding

- Horikawa & Kamitani (2017):
  - Mapping between brain and CNN features, rather than brain and image pixel space
  - Allows for generalizing to new images/categories not seen during training
- DNNs can be used to decode mental content of images
- What about the mental contents of language?
  - Can current Natural Language Processing (NLP) models help?

Decoding linguistic information

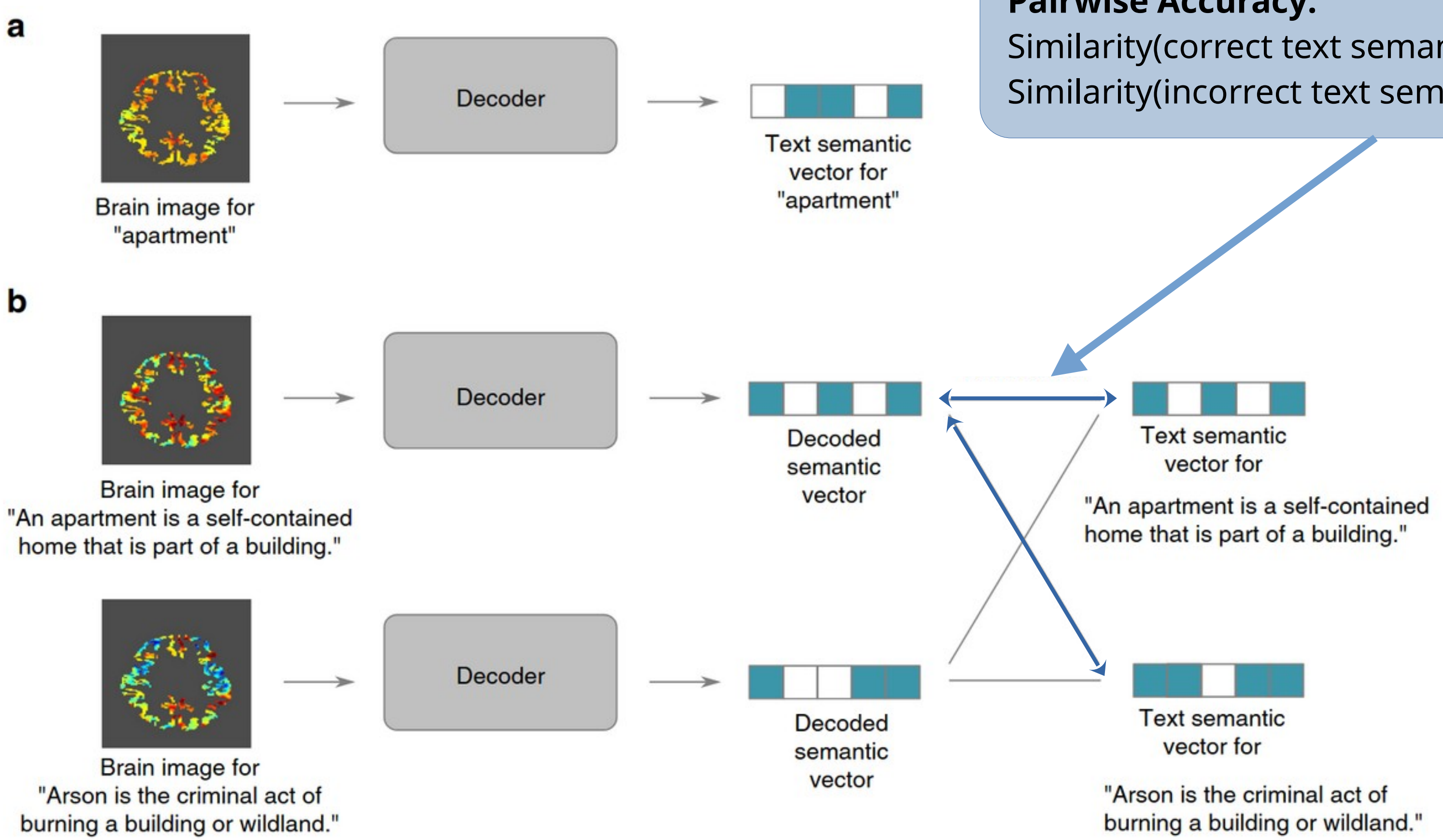
- Word embeddings (a linguistic “latent space”)
  - Fairly low dimensional (e.g., 300 or 500 dimensions)
  - A word/sentence is represented by a vector in this space
  - Vector operations:

vector $\vec{x}$ defined as:	Example 1	Example 2
$\vec{x}$ = Paris – France		

Mikolov *et al.* (2013)

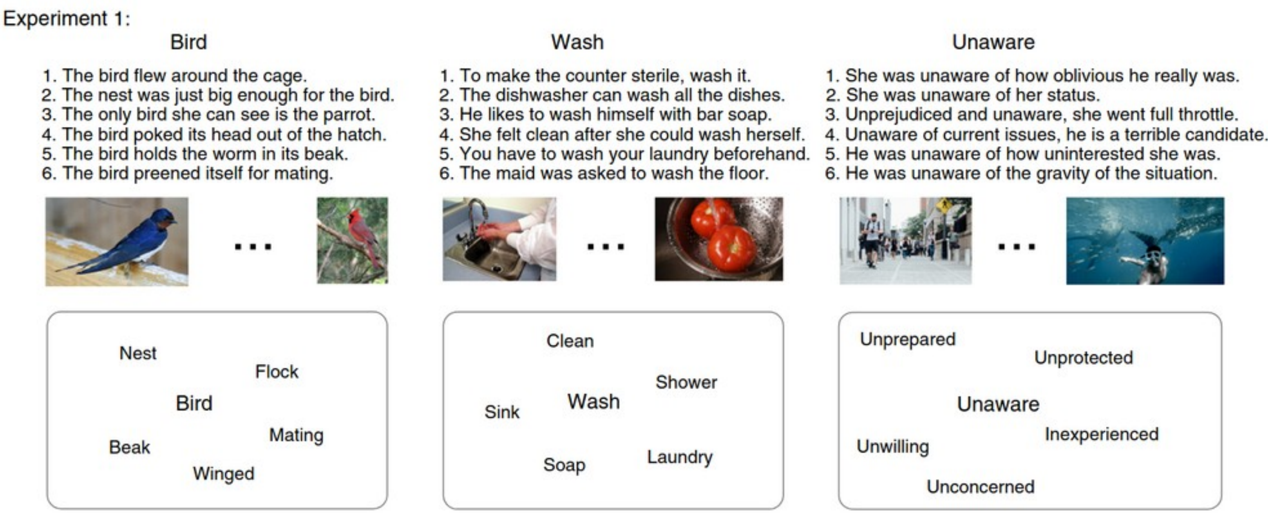
Decoding linguistic information

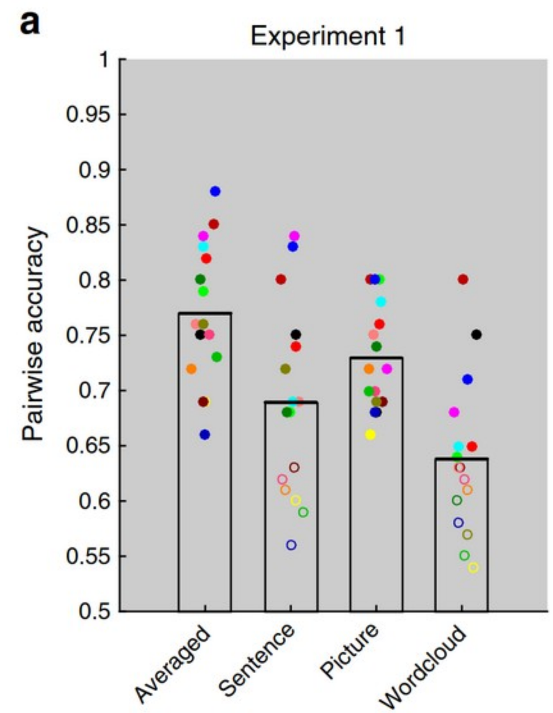
Pairwise Accuracy:
Similarity(correct text semantic vector) > Similarity(incorrect text semantic vector)



Pereira et al. (2018)

Decoding linguistic information





Pereira et al. (2018)

Decoding linguistic information

Experiment 2:

Musical instruments (clarinet)

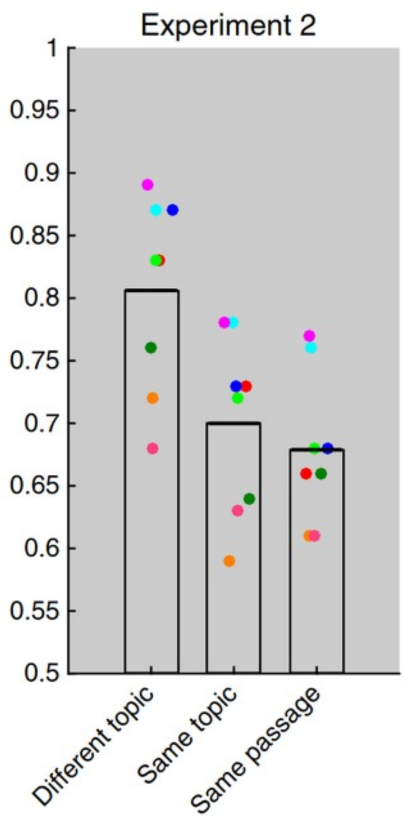
A clarinet is a woodwind musical instrument.
It is a long black tube with a flare at the bottom.
The player chooses notes by pressing keys and holes.
The clarinet is used both in jazz and classical music.

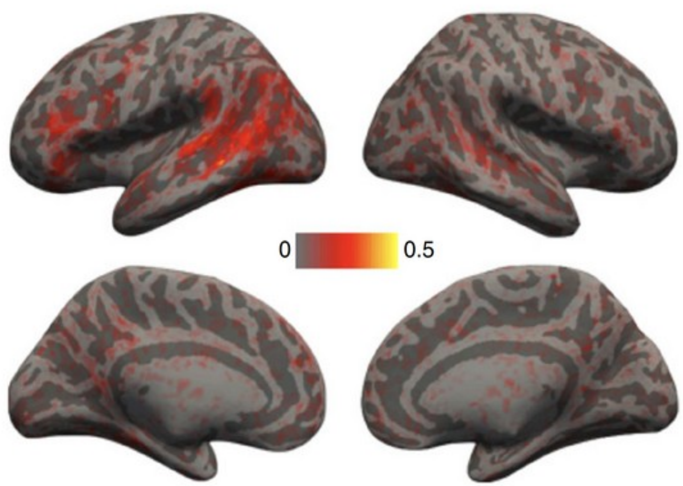
Musical instruments (accordion)

An accordion is a portable musical instrument with two keyboards. One keyboard is used for individual notes, the other for chords. Accordions produce sound with bellow that blow air through reeds. An accordionist plays both keyboards while opening and closing the bellows.

Musical instruments (piano)

The piano is a popular musical instrument played by means of a keyboard. Pressing a piano key causes a felt-tipped hammer to hit a vibrating steel string. The piano has an enormous note range, and pedals to change the sound quality. The piano repertoire is large, and famous pianists can give solo concerts.





Fraction of subjects for which a voxel was among the 5000 most informative voxels

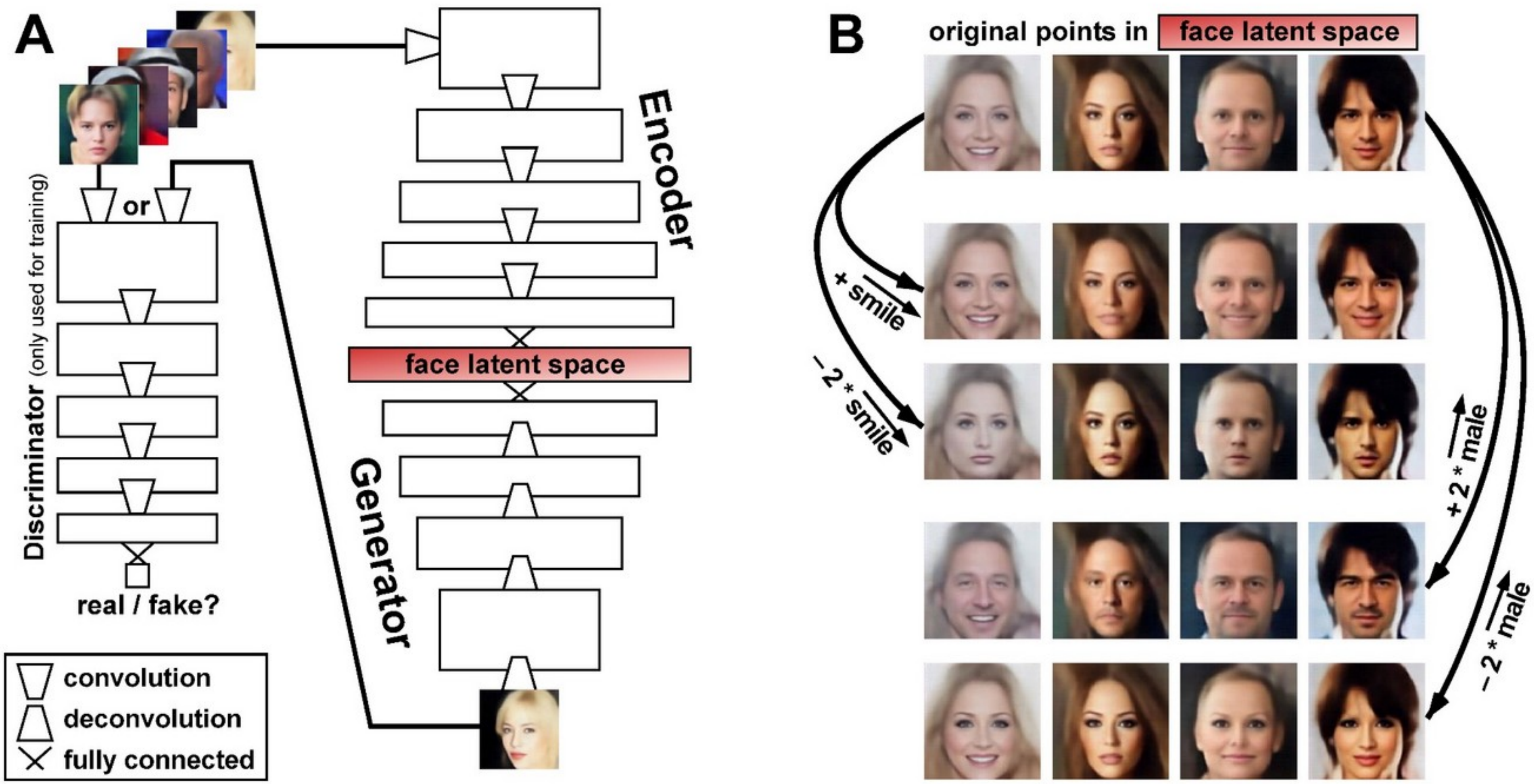
Pereira et al. (2018)

Stimulus reconstruction

- So far:
  - Decoding of stimulus category
  - Decoding of stimulus features (using DNNs)
    - Discriminate true stimulus from set of possible stimuli
- Next:
  - Stimulus Reconstruction
    - Leverage generative models to *reconstruct* the stimulus a subject was looking at

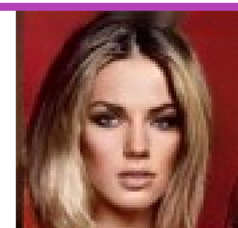
VAE-GAN model

- GAN: Generative model that generates images from vectors
  - Training on a celebrity dataset (200,000 images)
  - Encoder defines a "**face latent space**" of 1024 dimensions
  - A point/vector in this space corresponds to a face



Larsen, Sønderby, Larochelle, & Winther (2016) *ICML*

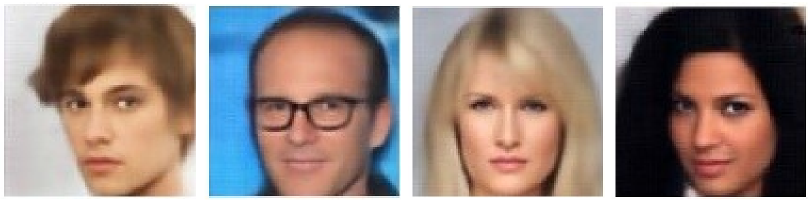
Face latent space

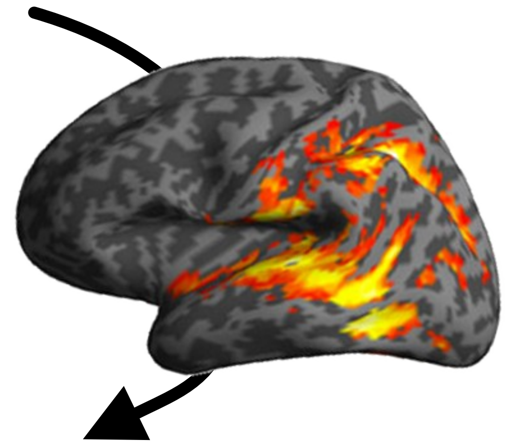


White, T. (2016) *arXiv:1609.04468*

Decoding faces

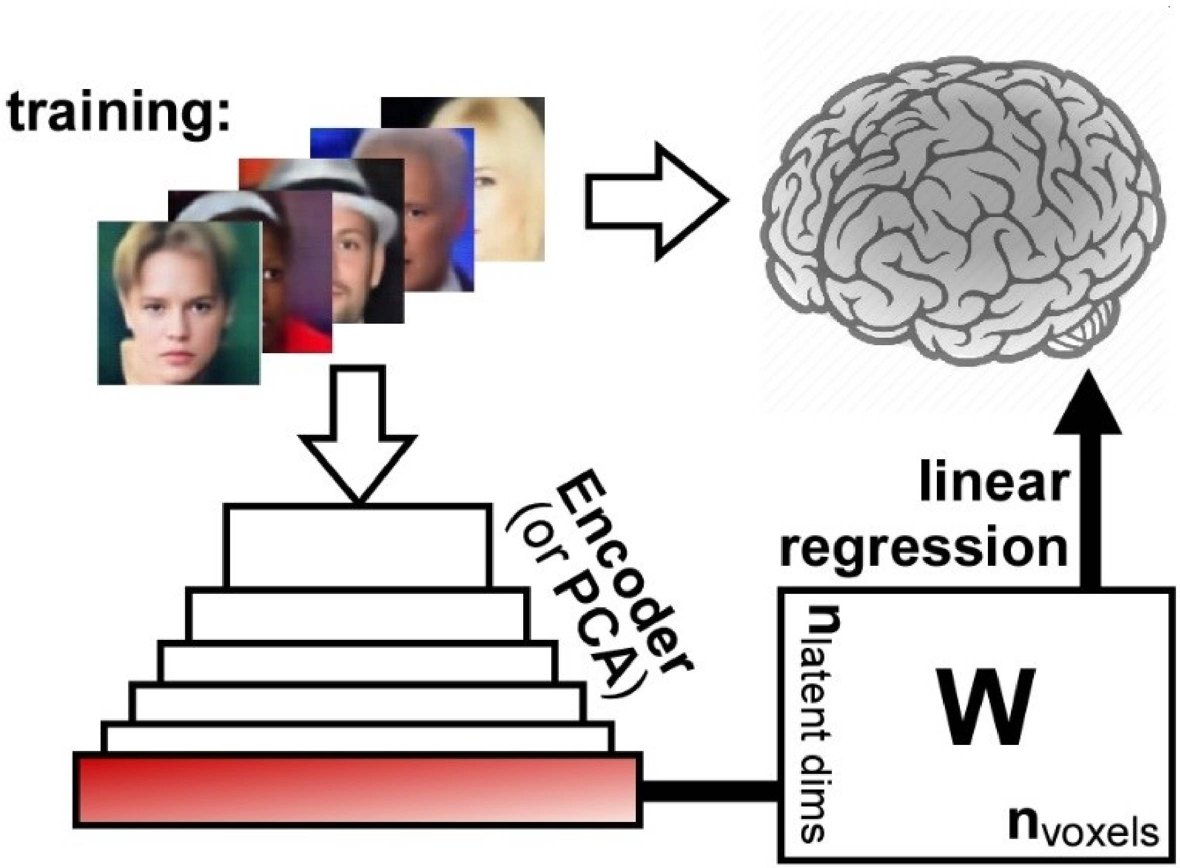
images seen by subject in MRI scanner





VanRullen & Reddy (2019) *Communications Biology*

Training the decoder



- Data from 4 subjects
- ~8000 faces per subject
- 16 hours of scanning/subject

➔ Freely available on OpenNeuro

$$Y = XW \Leftrightarrow X^TY = X^TXW$$

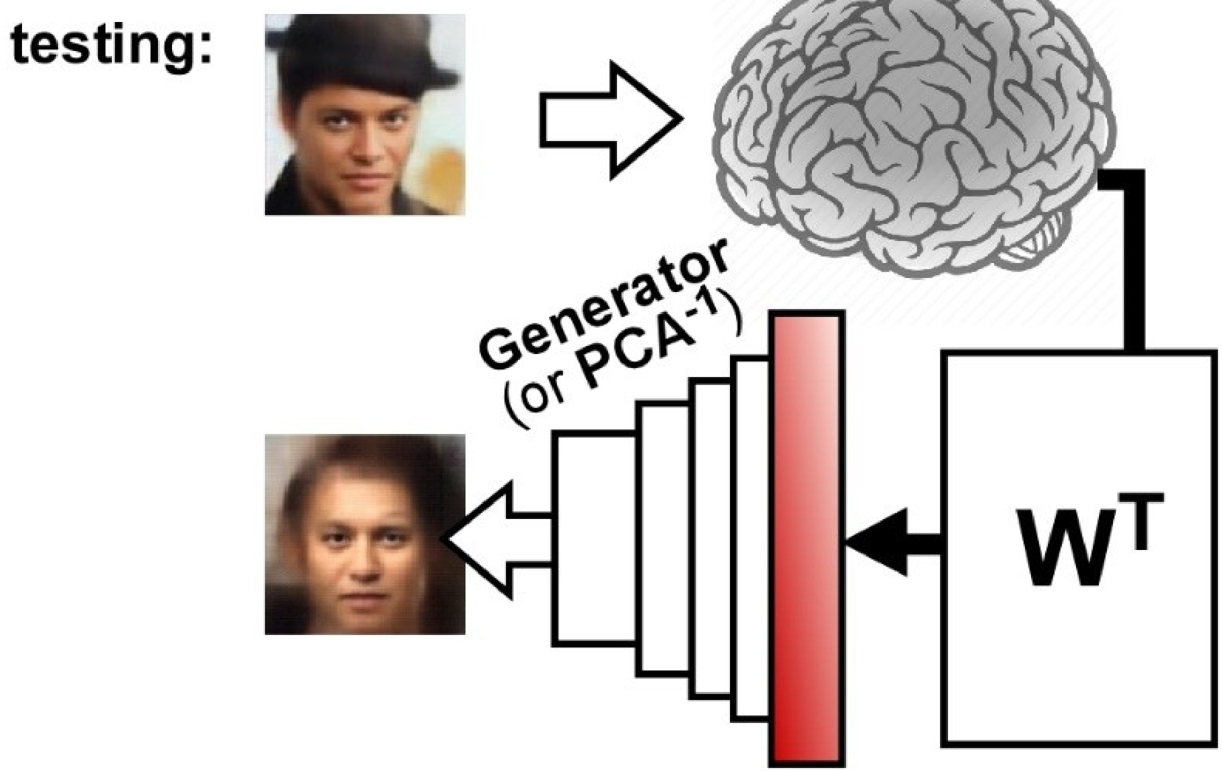
Y: $(8000, n_{voxels})$ X: $(8000, 1024)$ W: $(1024, n_{voxels})$

$$\boxed{W = (X^TX)^{-1}X^TY}$$

VanRullen & Reddy, 2019. *Communications Biology*

Testing the decoder

20 test images (x 45 repeats)



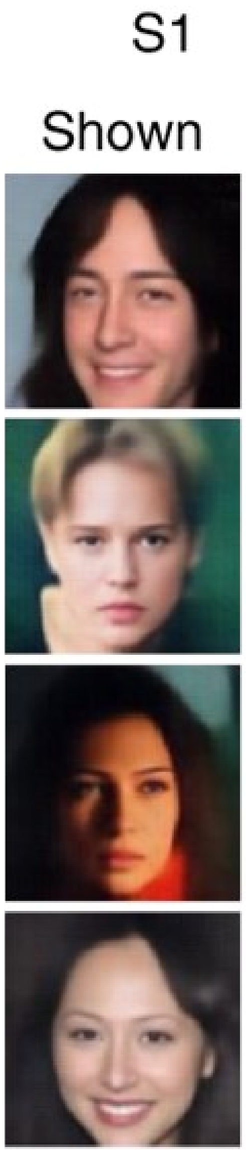
$$Y = XW \quad \Leftrightarrow \quad YW^T = XWW^T$$

$(20, n_{voxels})$ $(20,1024)$ $(1024, n_{voxels})$

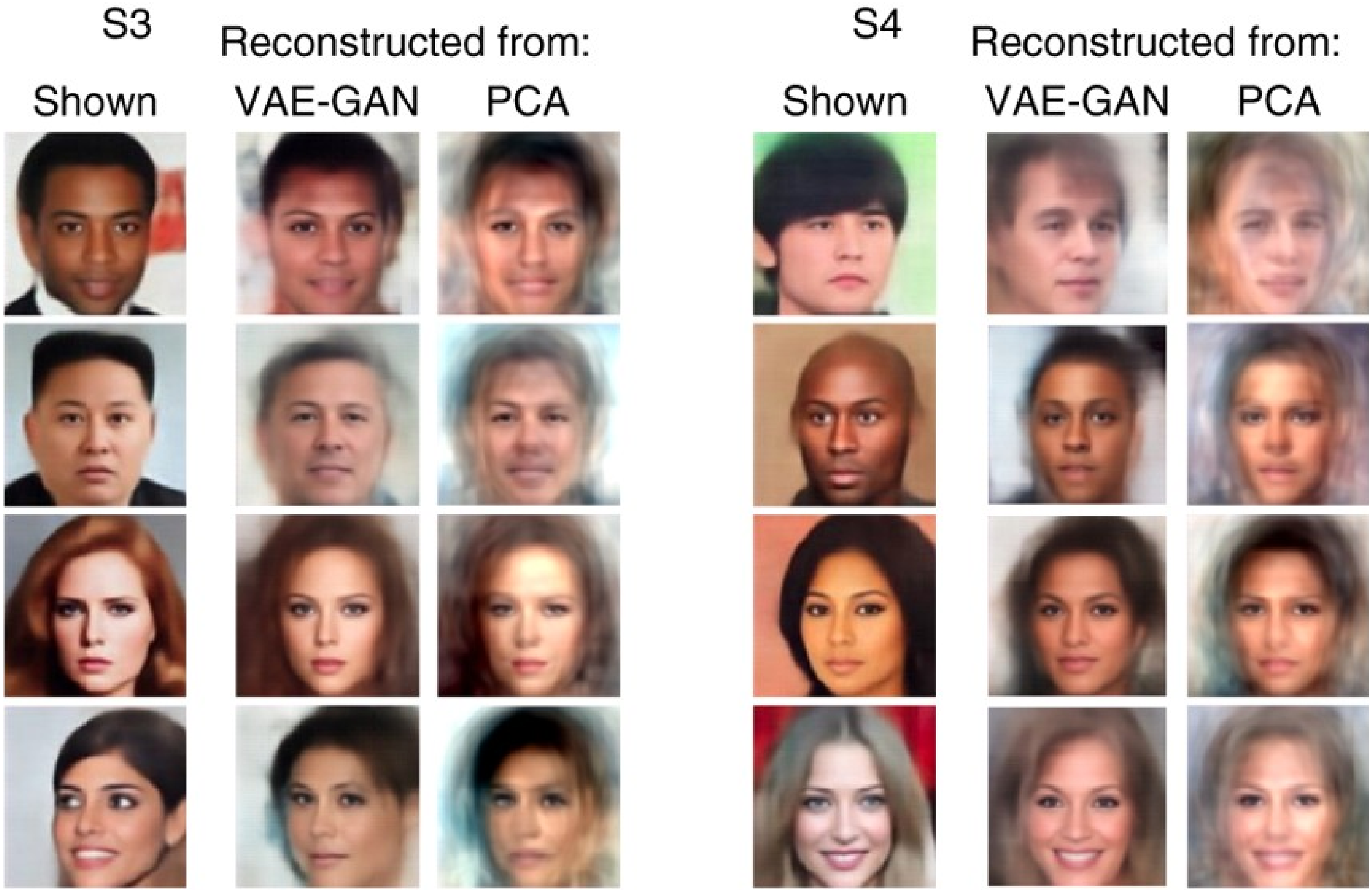
$$\boxed{X = YW^T(WW^T)^{-1}}$$

VanRullen & Reddy, 2019. *Communications Biology*

Face Reconstructions

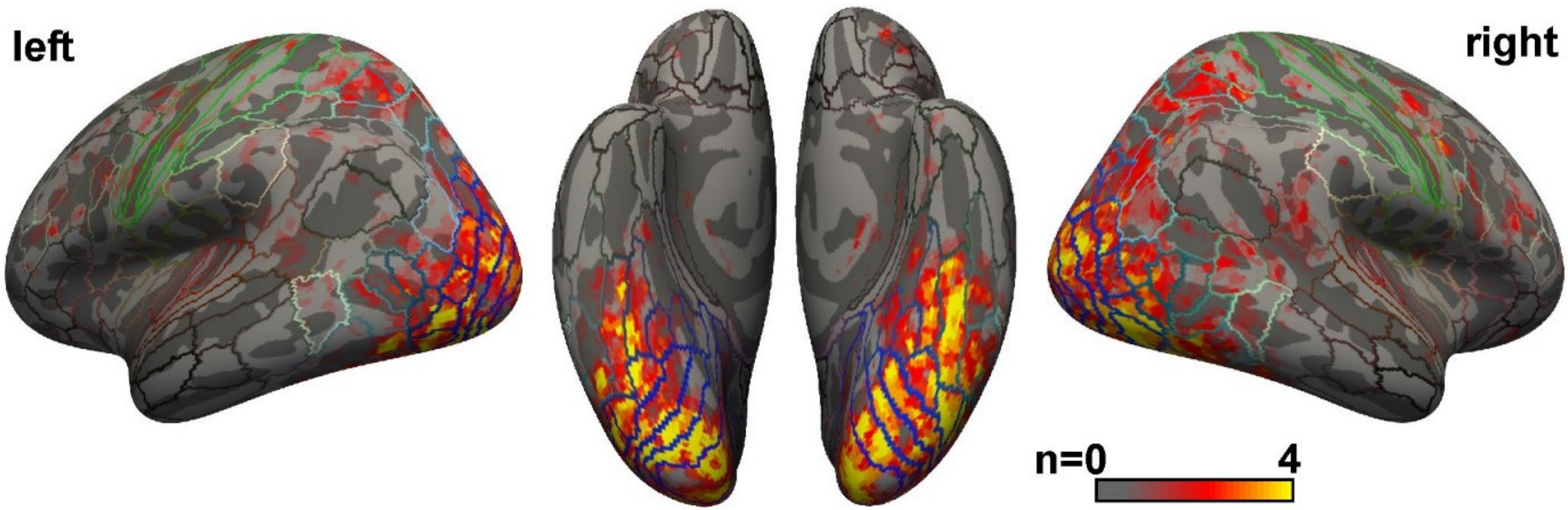


Face Reconstructions



VanRullen & Reddy, 2019. *Communications Biology*

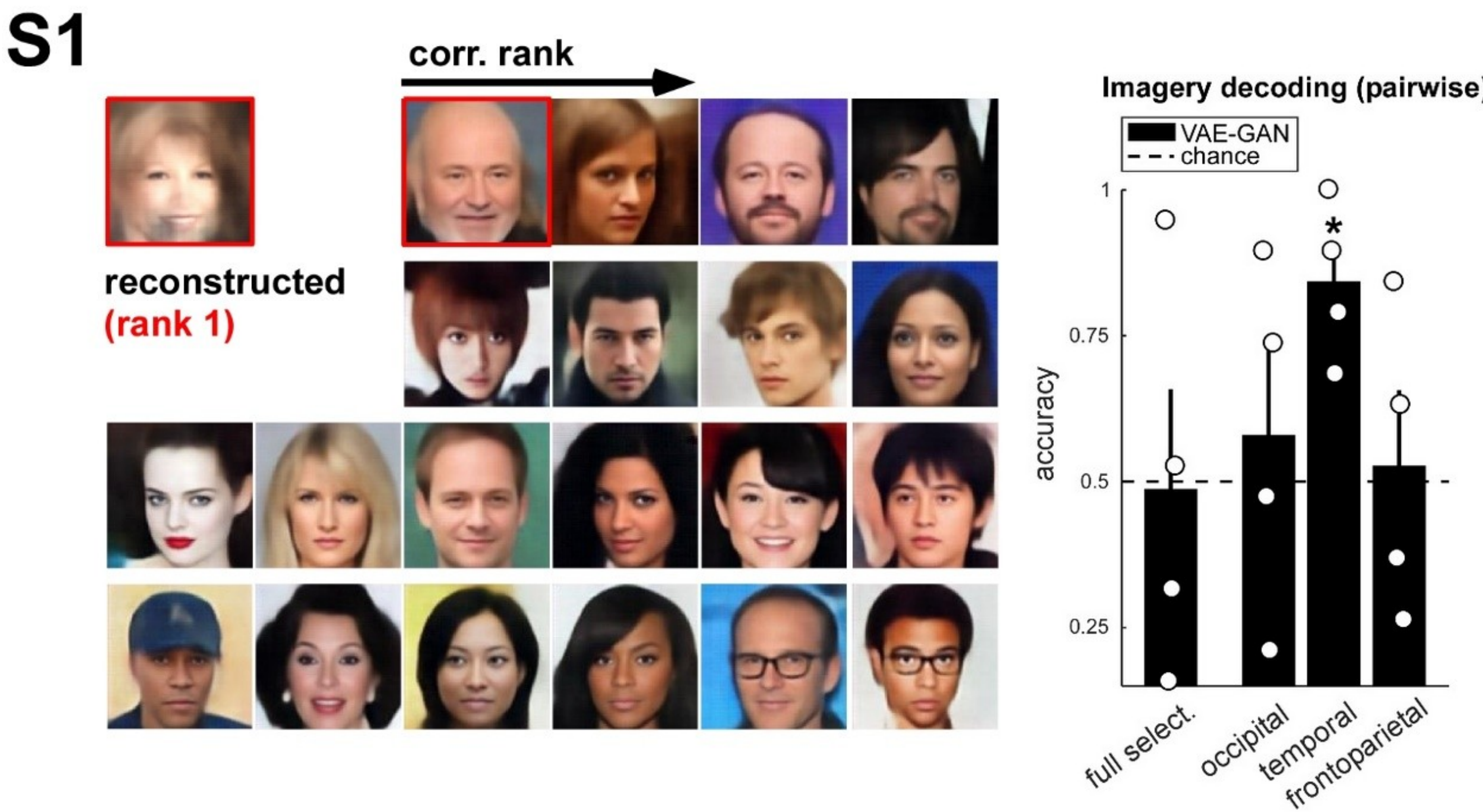
Which brain regions are involved?



Number of subjects for whom a particular voxel was informative (based on a combination of their visual responsiveness and their GLM goodness-of-fit during brain decoder training)

VanRullen & Reddy, 2019. *Communications Biology*

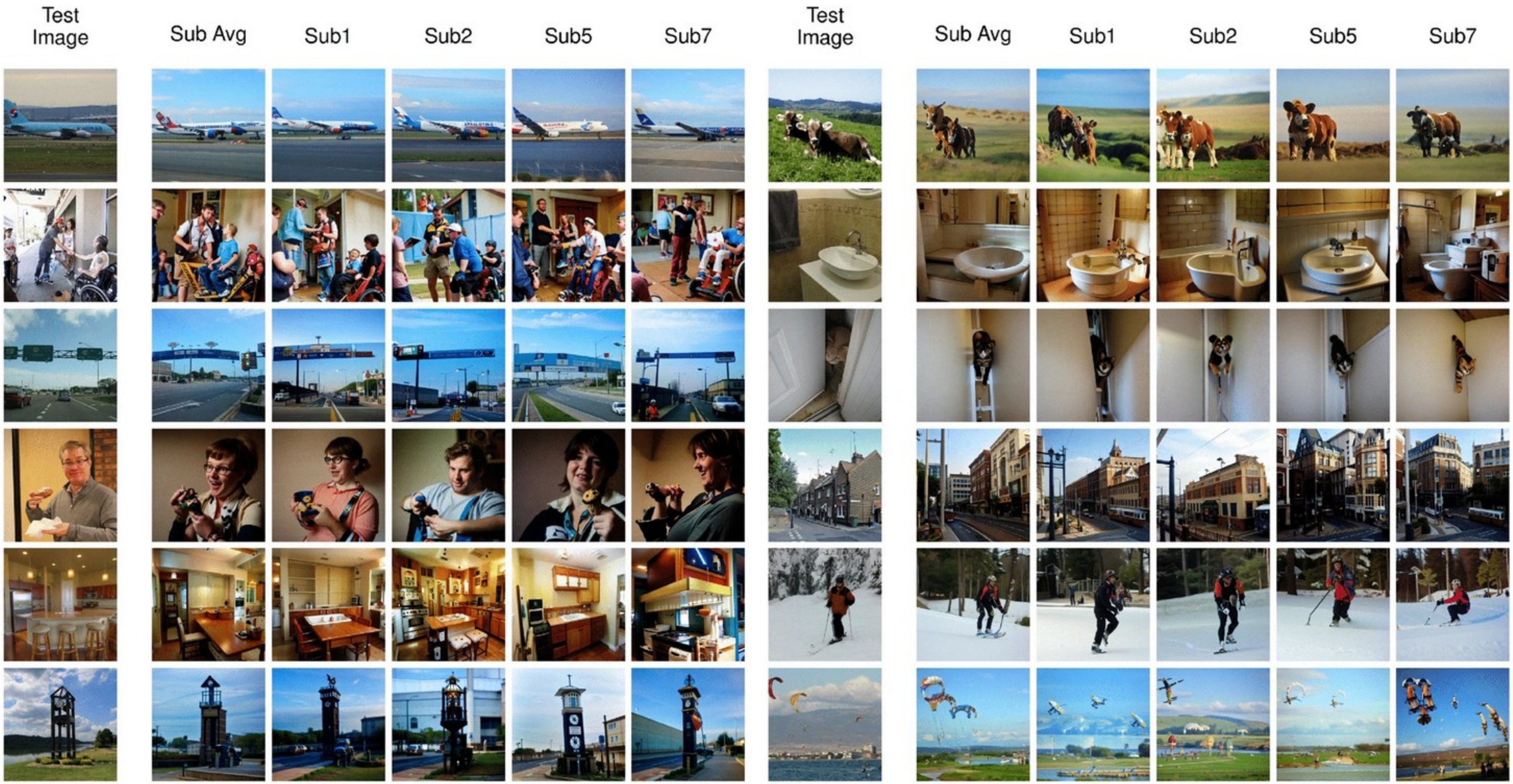
Decoding mental imagery



VanRullen & Reddy (2019) *Communications Biology*

Natural scene reconstruction

Leverage more recent generative models (diffusion models) to reconstruct natural scenes



Ozcelik & VanRullen (2023)

Summary

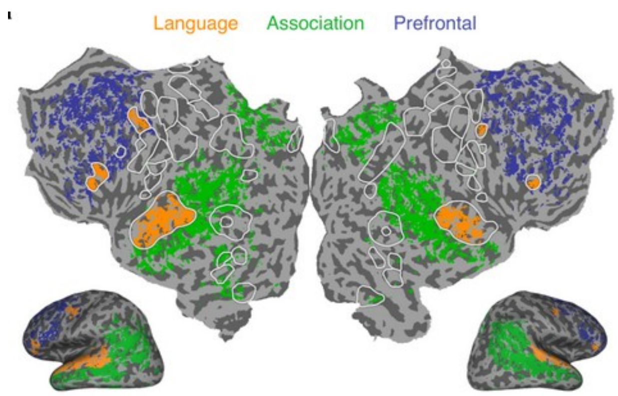
- Exploit patterns of brain activation to decode stimulus information
- Classification methods (e.g., SVM) that allow us to guess the stimulus (restricted to training stimuli)
- Using DNN feature space for decoding allows us to generalize to new images/categories/sentences
- Generative models open up a new range of possibilities: reconstruction of what the subject was looking at

Mental Privacy

Tang et al. 2023: Decoding of continuous language (subjects listening to audio books)

Actual stimulus	Left lang	Left assoc	Left PFC	Right lang	Right assoc	Right PFC
i was like no i'm out of here this is great and i went and hid behind a cabana and he left	they drove off they didn't even look back as i sat there thinking what the hell i should do	tell me to leave i said ok and ran out to the parking lot i was like wait is that a cop car	i told them to leave but they insisted and kept saying i can't stay so i got up to go	ran away and didn't look back at me and said you can go on without me i'm leaving now	in the driveway i told him to leave me alone and went inside i ran out into the cold	let me through i don't know where he is right now but i will get there soon enough

- Can we decode information against the subjects will?
- Subject cooperation is required (both to train and to apply the decoder):
  - Applying decoders trained on data from other subjects → performance barely above chance
  - Subject is performing an additional task while listening (e.g. calculations): → decoding performance drops substantially



- Further Reading: Rainey et al. (2020) *Science and Engineering Ethics*

Speech Decoding of paralyzed person

 https://www.youtube.com/watch?v=_GMcf1fXdW8

Moses et al. (2021) *The New England Journal of Medicine*

Questions?