Visual Dynamics in Human Brain and Artificial Neural Network

Cai Wingfield

Daniel MacSwayne

Friday 26th April 2024

SBA Workshop, University of Birmingham

Jianbo Jiao

Dan MacSwayne

Ole Jensen

Cai Wingfield

Alex Billingham

Syanah Wynn

Computer Science

Centre for Human Brain Health

Institute for Interdisciplinary Data Science and AI

Neuroimaging Center

University of Birmingham

Artist

Johannes Gutenberg University Medical Center Mainz

UNIVERSITY OF BIRMINGHAM

MIX

CHBH

THE INSTITUTE FOR INTERDISCIPLINARY DATA SCIENCE AND AI

SMQB Centre for Systems Modelling & Quantitative Biomedicine

Outline

1. The priming effect in visual processing.
2. Representational Similarity Analysis (RSA).
3. An Artificial Neural Network (ANN) model of object recognition.
4. Unpicking the dynamics of the priming effect.
5. Future directions: neuronal dynamics in ANNs.

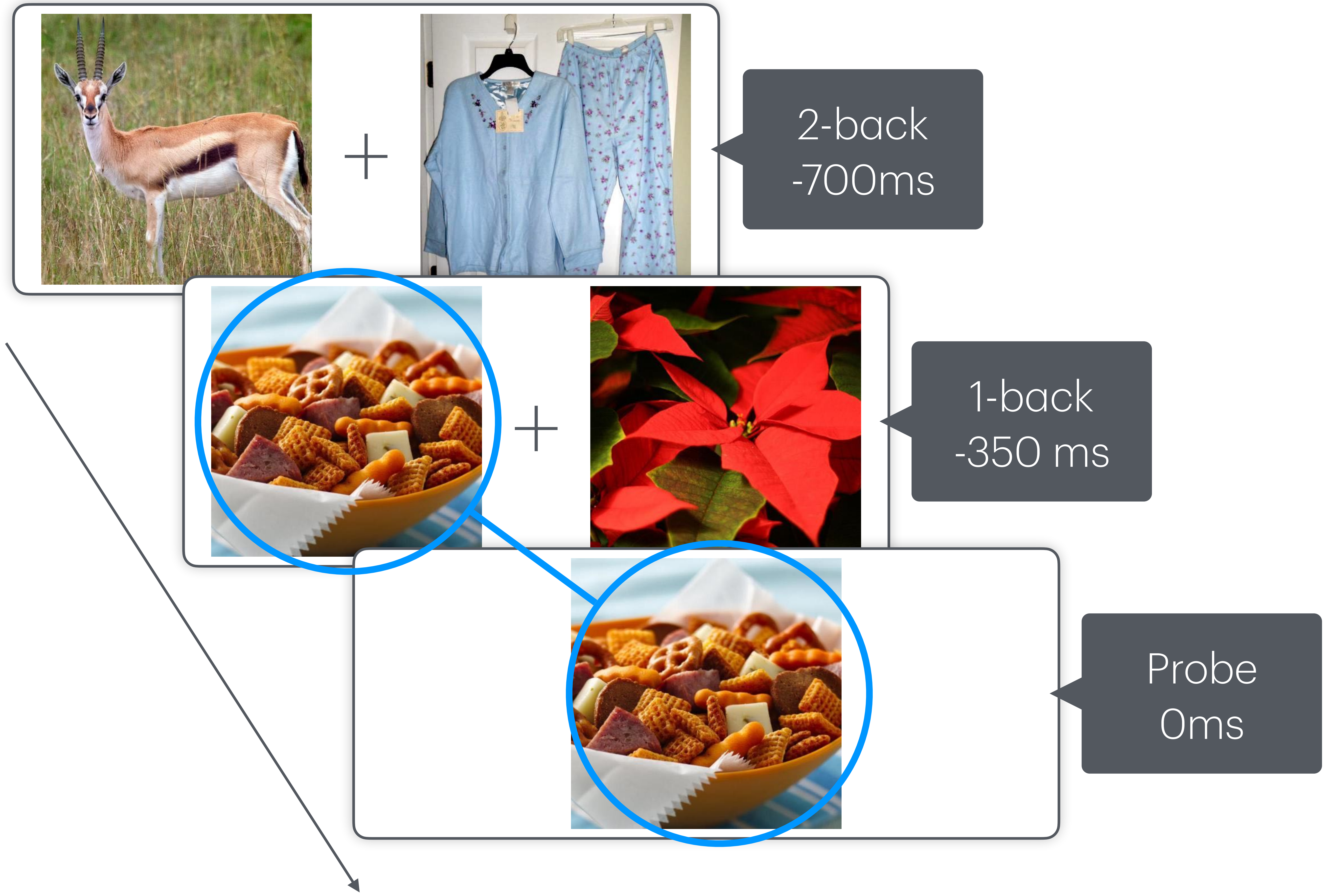
The priming effect in visual processing

Priming

Behaviour and brain response

- 28 participants included (7 more excluded due to ≥50% missing trials).
- Whole-scalp MEG data recorded.
- 7500 images = 5 categories × 100 objects × 15 examples.
- 1440 trials per participant: categories and priming conditions counterbalanced.

- In an experiment, participants viewed two pairs of rapidly presented parafoveal images, followed by a single foveal image.
- The foveal image was either **1-back primed** (same image seen in previous pair), **2-back primed** (seen in previous-but-one pair), or **unprimed** (previously unseen).
- The **priming effect** is any measured change in behaviour or brain signal when a stimulus is primed.
  - E.g. faster, more accurate behaviour.



Priming in the brain

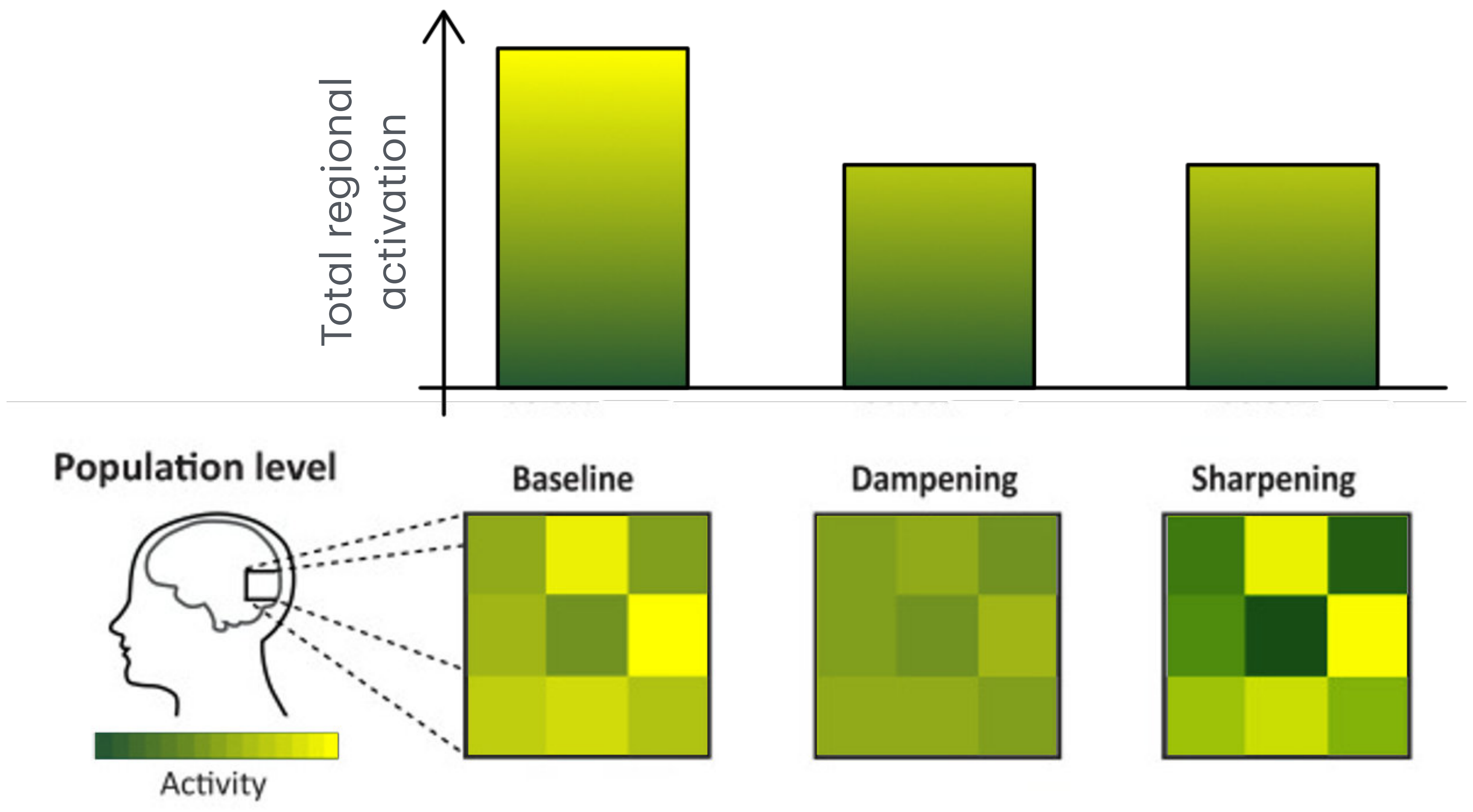
Repetition suppression

- One observable effect of priming is **repetition suppression**.
    - Can manifest as reduced activity when a stimulus is familiar, compared to novel.
    - Can manifest as a delay in activation level.
- We observed a 1-back priming "sweet spot".

Priming in the brain

Representational sharpening

- A **sharpening effect** is where activity is not just lower, but more "efficient".
- Previous activations leave some residual trace.
    - This can manifest as higher decoding accuracy despite lower activation.
    - This can manifest as representations resolving faster.
- We want to explore both of these effects using a deep model of image recognition.



Representational similarity analysis (RSA)

Apron

Blossom

Chips

Flower

Lilly

Recycling truck

Rooster

Salad

Snow mobile

Soufflé

Tie

Tiger

Tram

Turban

Whale

“Colour”

Animals

Clothing

Plants

Vehicles

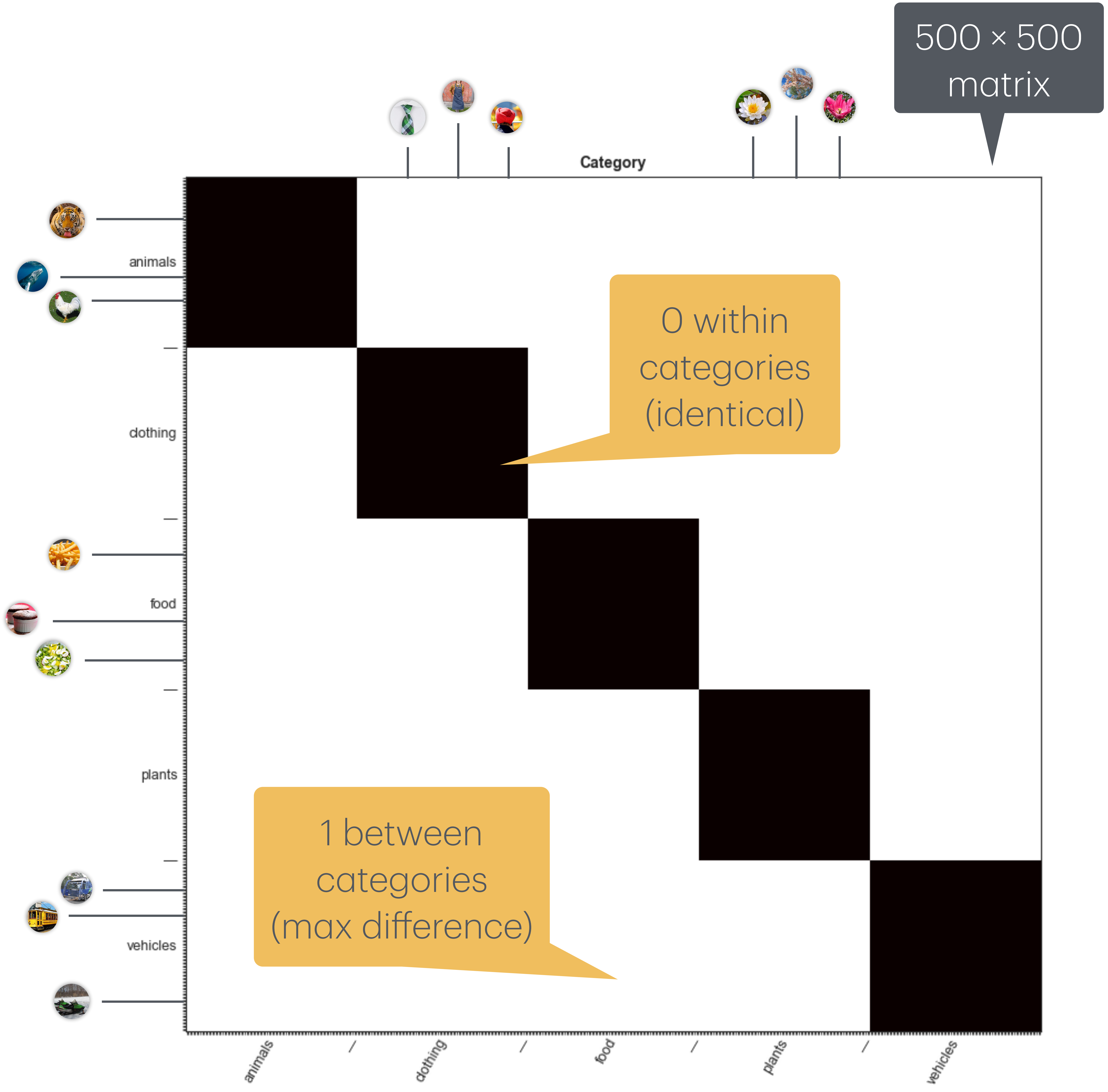
"Category"

Food

RSA

RDMs

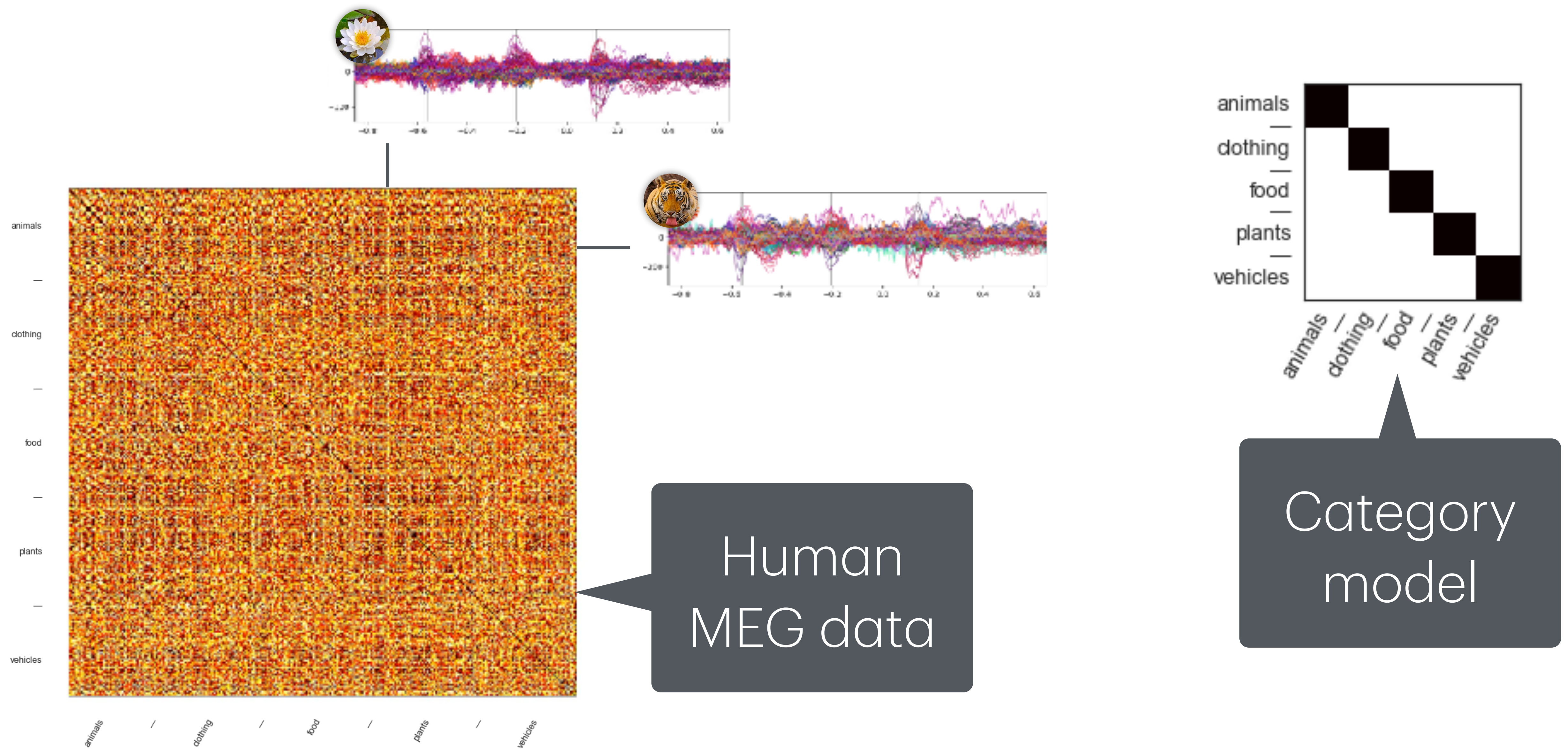
- These arrangements are captured in a **representational dissimilarity matrix** (RDM), which records how similar or different *each pair of stimuli is*.
- This category structure is very simple.
- The RDM captures how the model "sees" the stimulus space.
- RDMs can also come from brain data, or from more complex models.

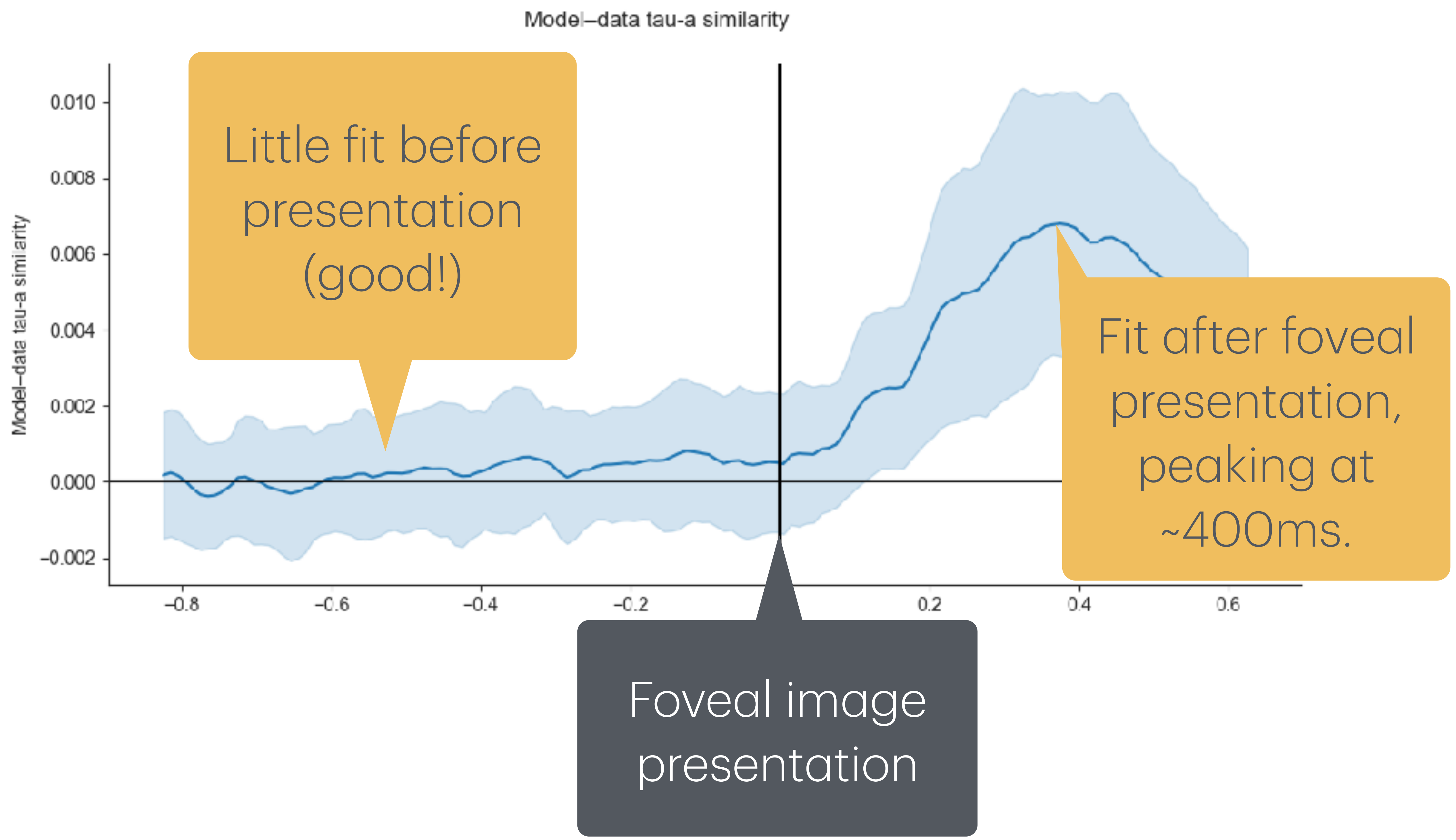


RSA

Category model results

- When does image category explain the MEG data?
- We similarly produce a (dynamic) RDM from the participants' MEG data.
- Correlate the category model RDM with the MEG-data RDM (Kendall's tau-a).



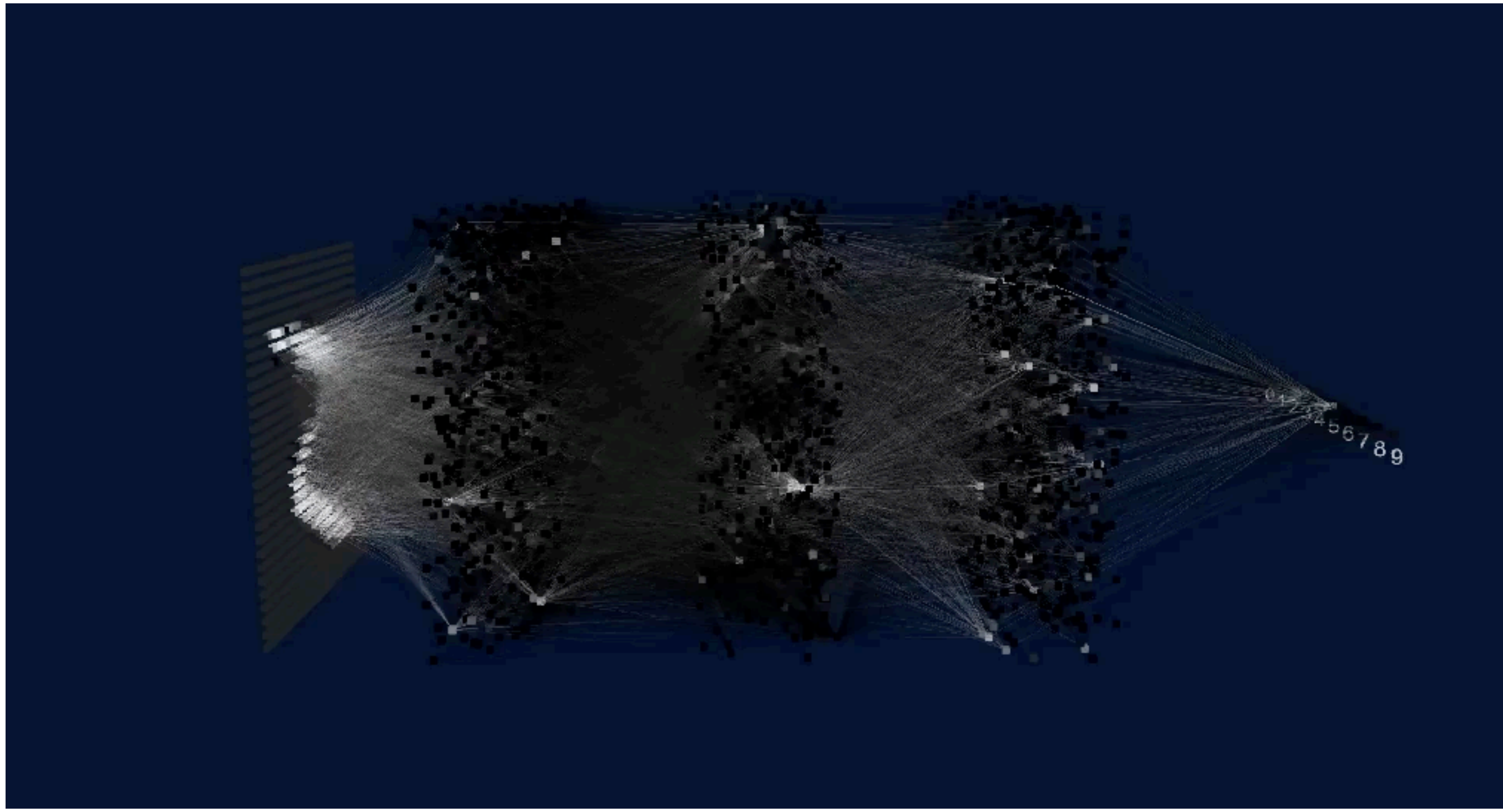


An Artificial Neural Network (ANN) model of object recognition

Artificial Neural Networks

The basics

- Artificial Neural Networks (ANNs) are a simple model of neurons in the brain.
- Virtual neurons are wired together in a hierarchically layered network.
- ANNs can be trained to classify the images they receive as input.
- Collections of neurons become activated when presented with specific signals.



https://www.youtube.com/watch?v=Tsvxx-GGlTg&t=1s&ab_channel=MarijnvanVliet

Artificial Neural Networks

Activation maps

- The network represents the image differently at each layer.
- The lower layers of the network tend to model granular details of the image such as edges
- The higher layers tend to model more abstract features such categories.
- We can build a dataset of neuron activation patterns at each layer of the network for each image.
- We want to understand how a network organises the representation space and whether it is similar to the brain.
- The animation shows the network activating according to the tiger as a stimulus

Supervised vs self-supervised learning

- **Supervised learning**
  - The model is trained to complete a task requiring a human-**labelled dataset**.
  - For example, learning to classify images from a set of labelled images.
- **Self-Supervised Learning**
  - The model is pre-trained on an auxiliary task using an **unlabelled dataset**. The model may learn representations which are useful for downstream tasks.

AlexNet

RotAlexNet

SimCLR

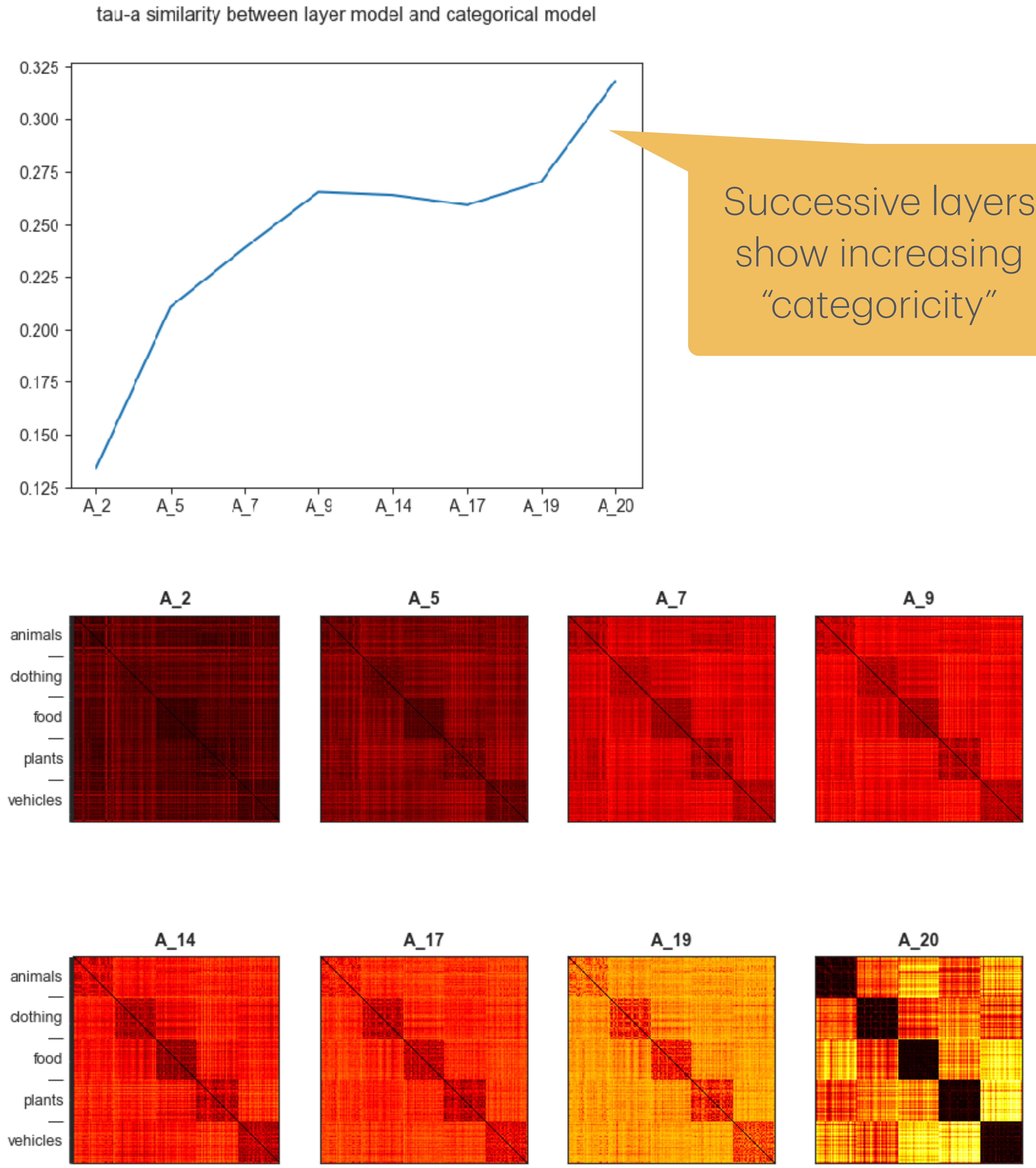
Two parallel lines of enquiry

- **Question 1**
  - How well do the ANN activations explain the brain data?
  - Does the evolving representations in the ANN model match those in the brain?
- **Question 2**
  - How can we implement the neuronal dynamics that we's observe in the brain into an ANN?

Unpicking the dynamics of the priming effect (Q1)

AlexNet models

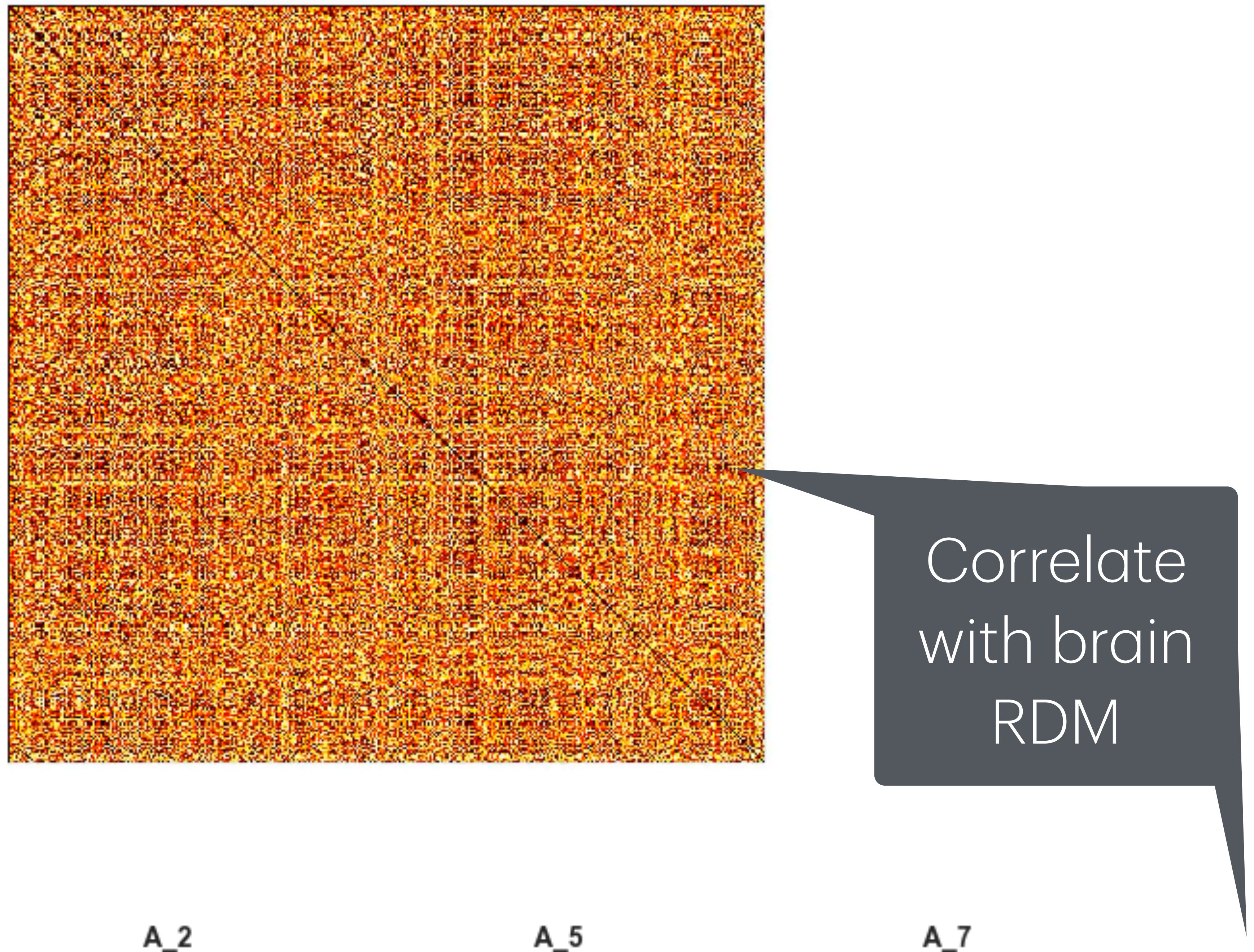
- Model RDMs computed using cosine distance between layer-activation vectors.

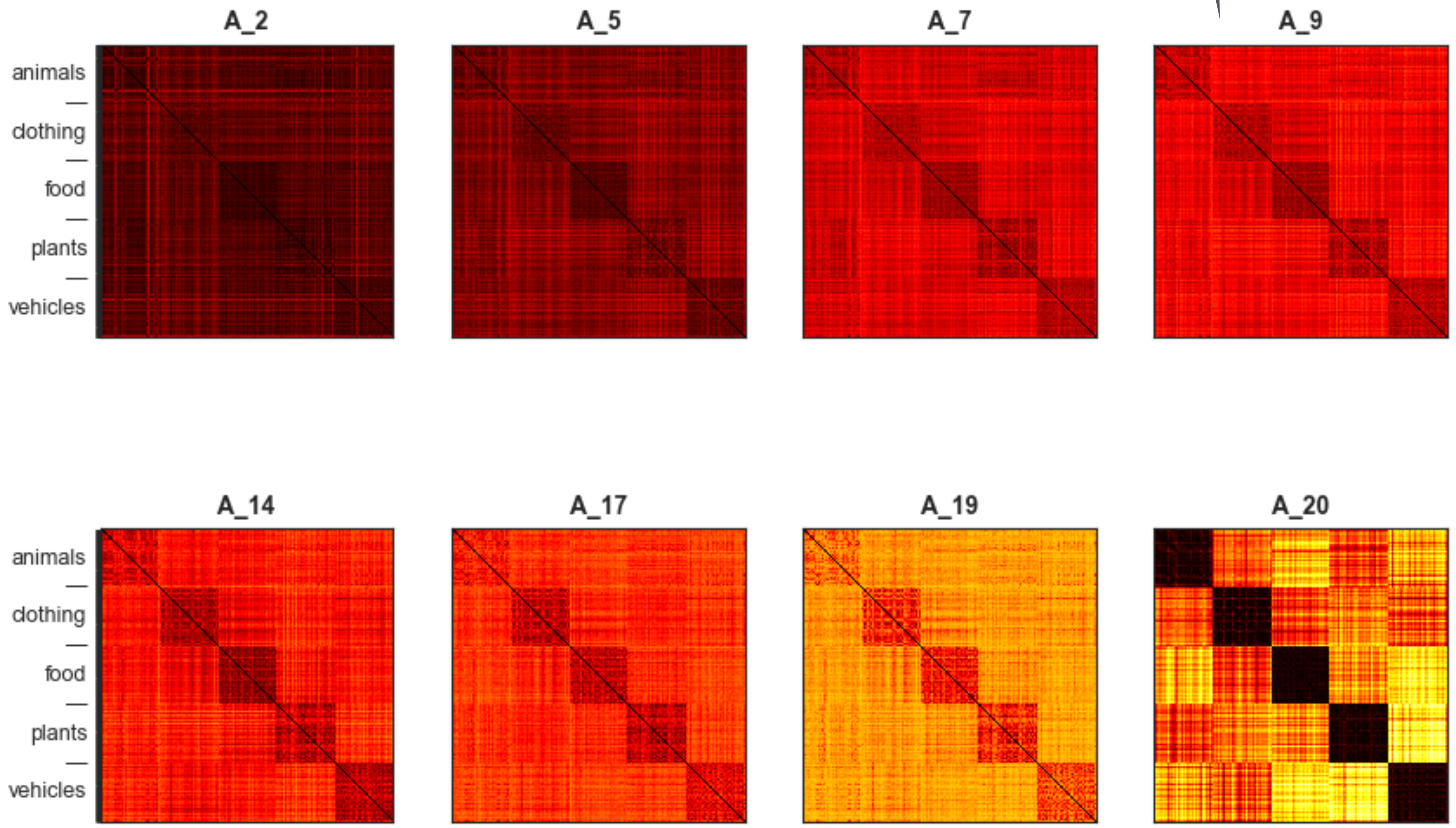


Successive layers show increasing "categoricity"

AlexNet models

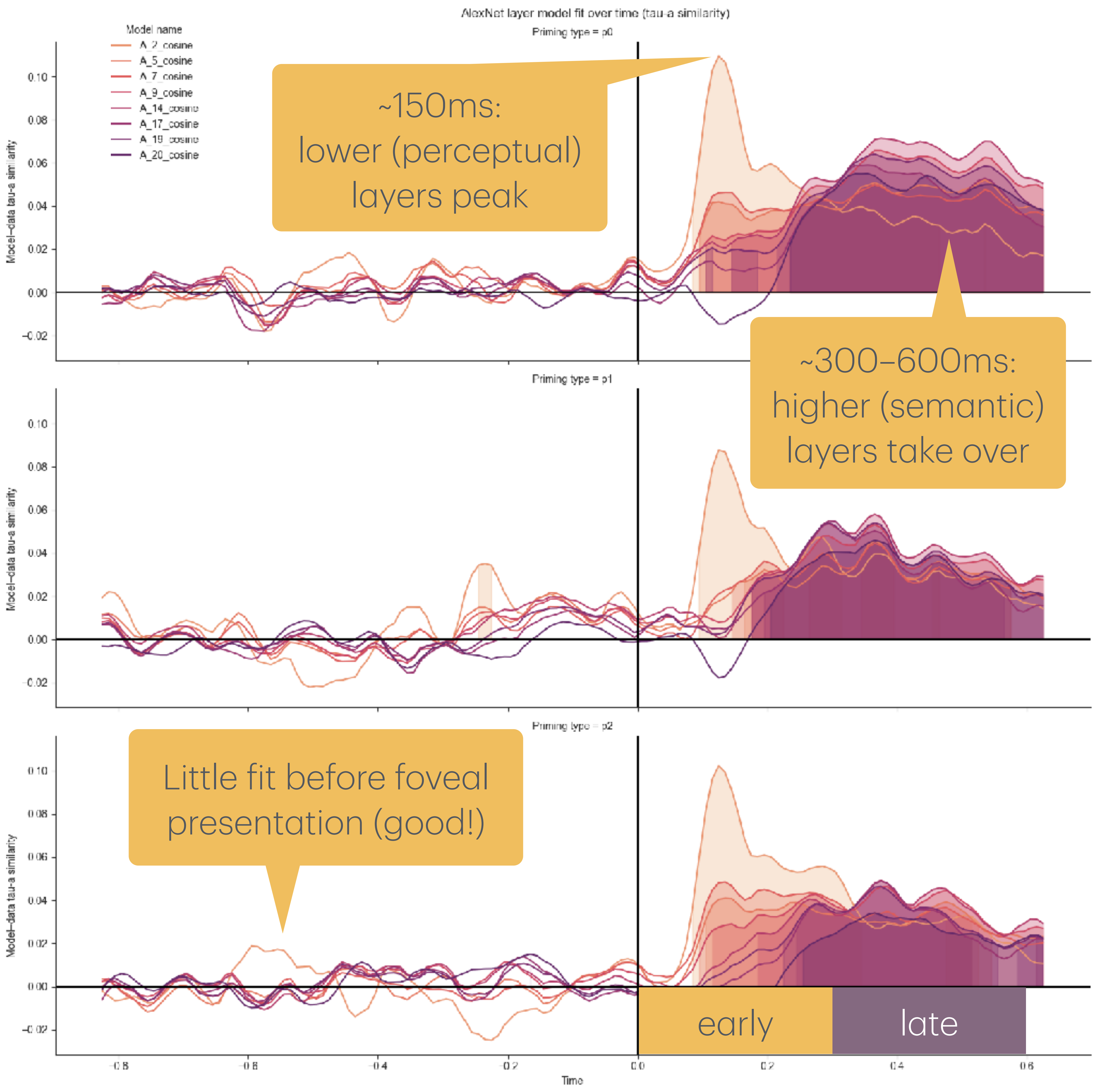
- Model RDMs computed using cosine distance between layer-activation vectors.
- We'll perform the same analysis as with the category model.





AlexNet results

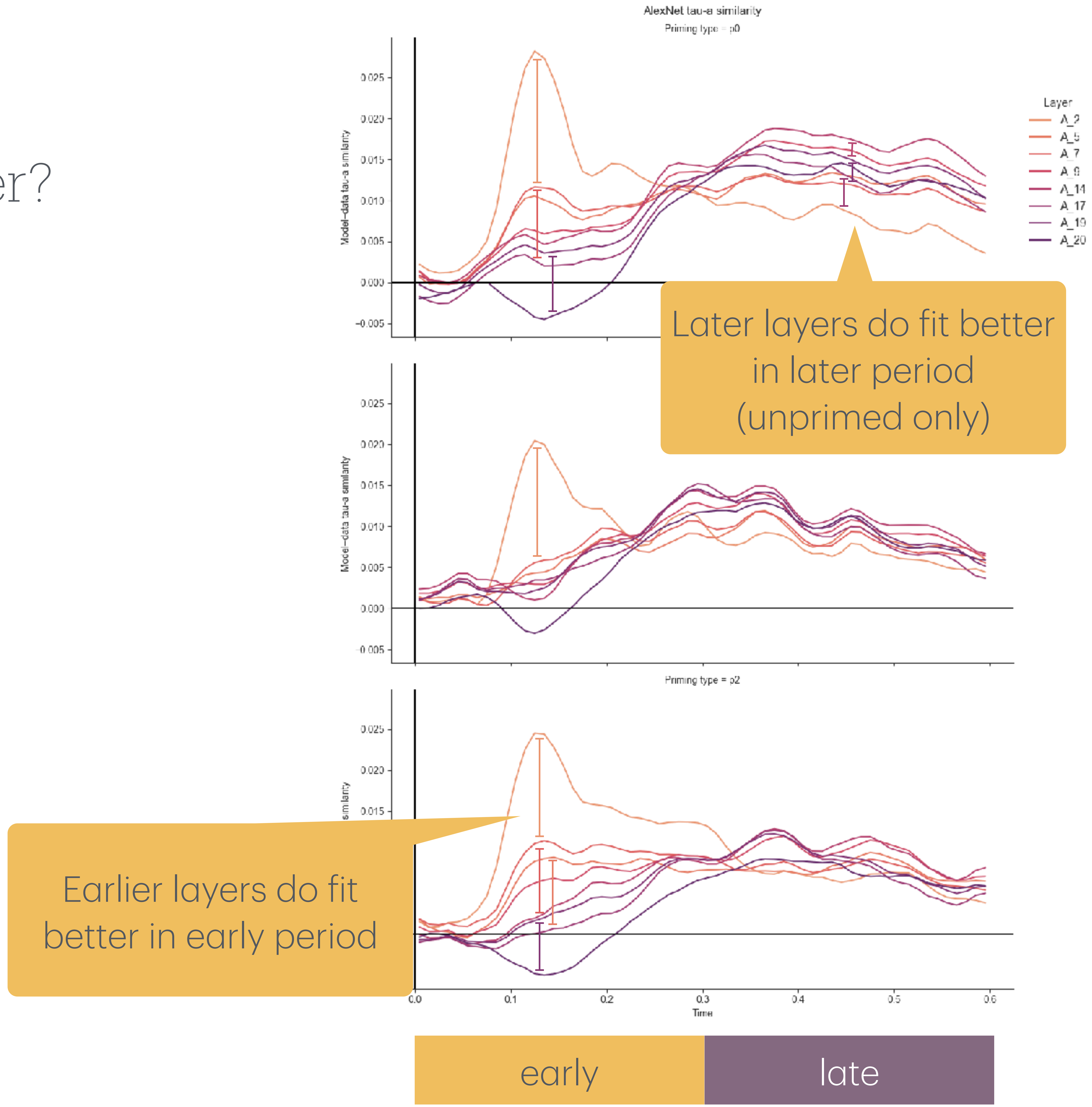
- All layers fit the model well, but not all at the same time.
    - Condition permutation test ($p < 0.05$; corrected for temporal multiple comparisons).
- Identify early (0–300ms) and late (300–600ms) periods for next analysis.



Between models

Are the higher peaks really better?

- We compared the individual models' fits in the early/late periods using an area-under-the-curve (AUC) analysis.
- Tests confirm observations ($p < 0.05$; select differences shown).

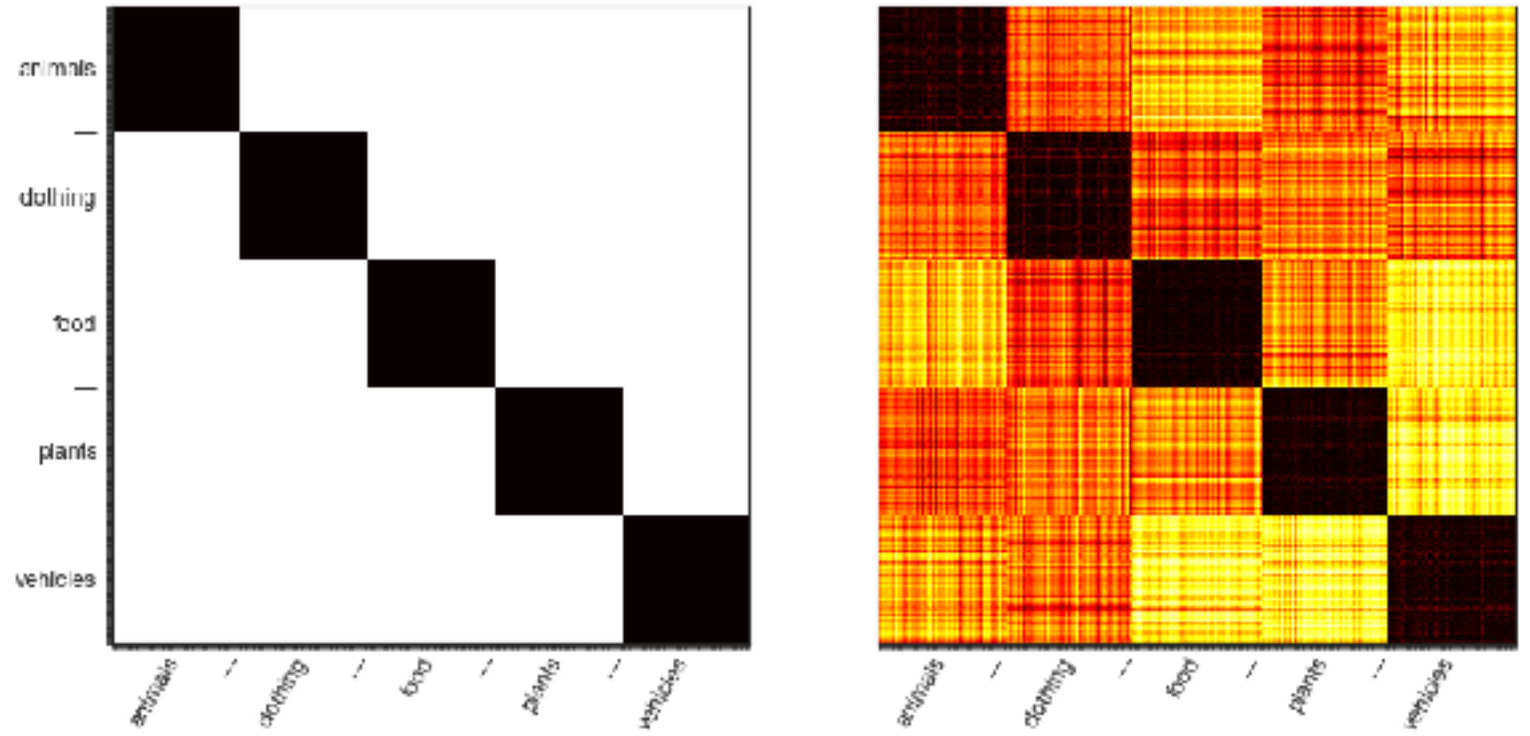


AlexNet A_{20} vs category

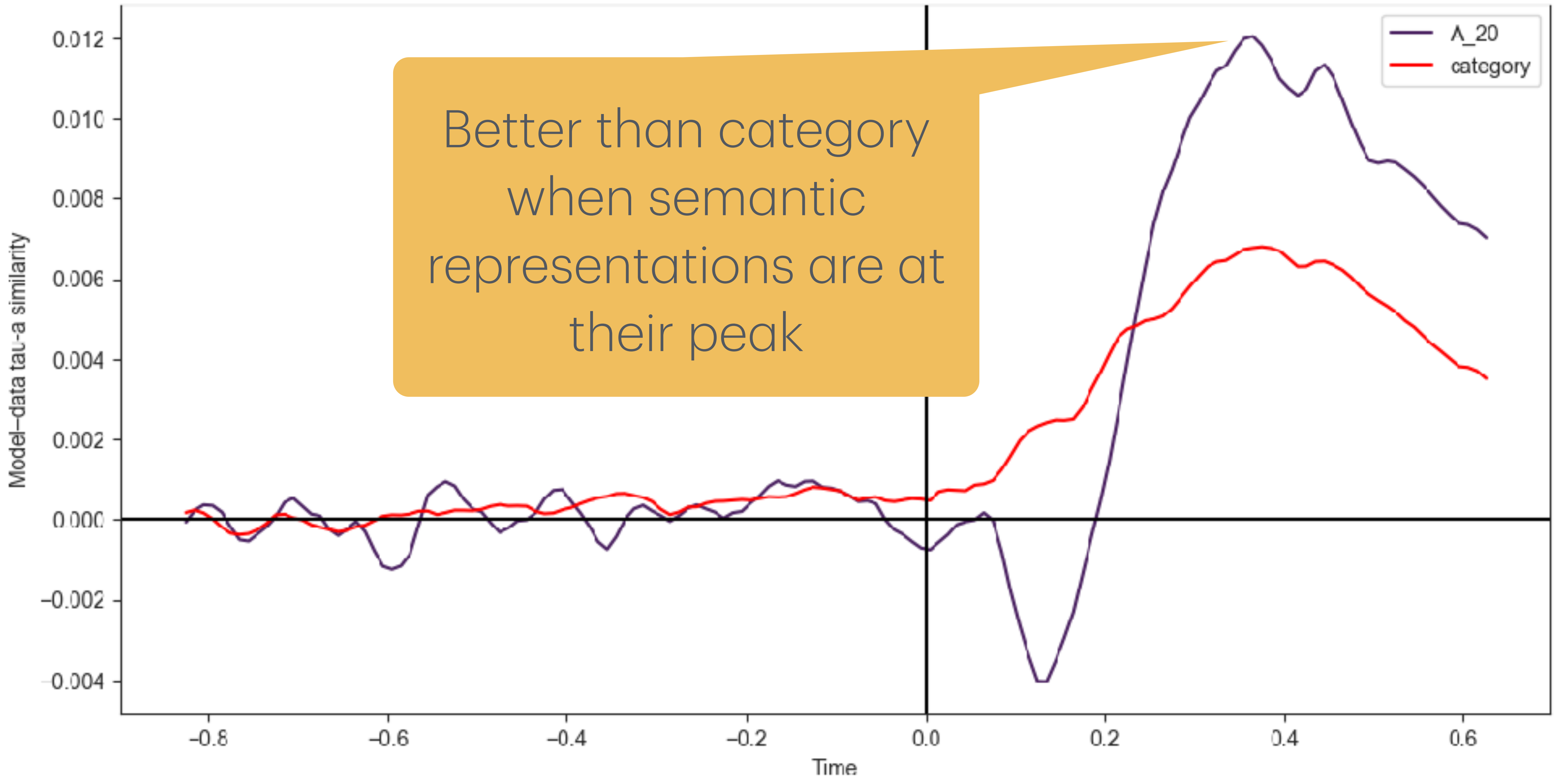
- A_{20} is the output activation prior to softmax.
- It captures pre-decision “confusion” when attempting to apply a category label.
- This is a better explanation for the MEG RDMs than the correct decision.

true category

pre-softmax activation

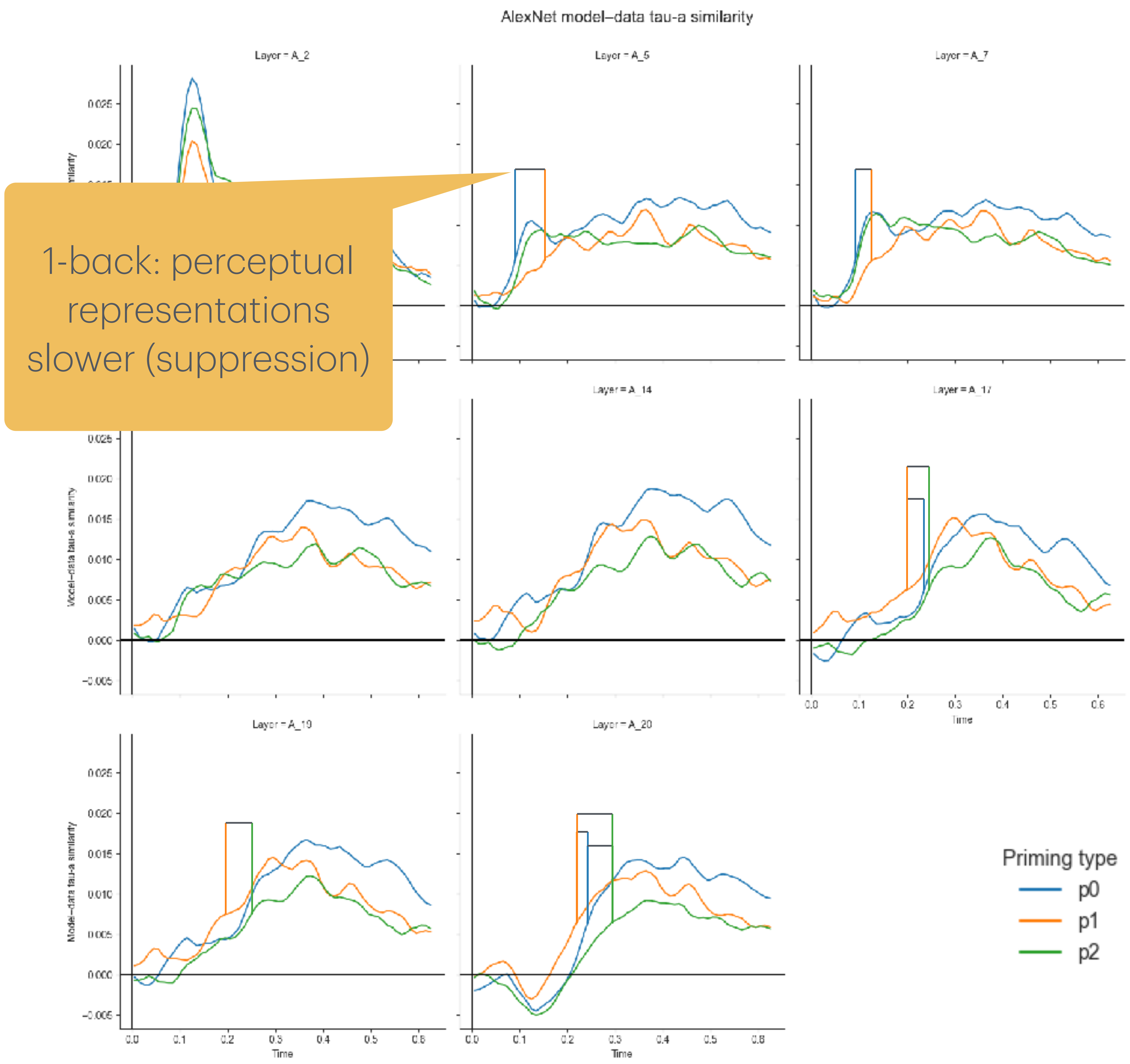


Better than category when semantic representations are at their peak



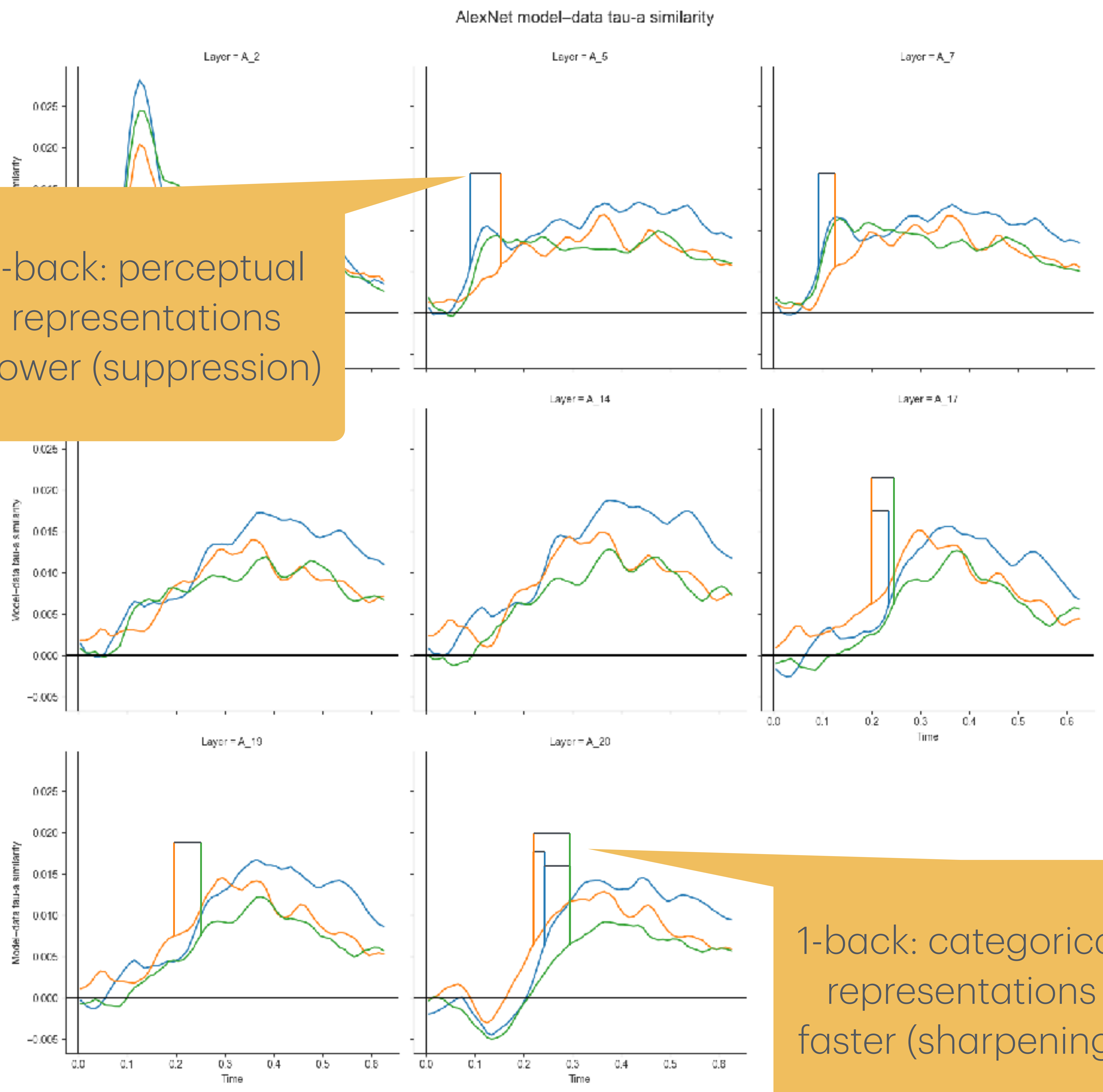
Evidence of sharpening

- Repetition suppression can manifest as lower/slower representations.
- Sharpening can manifest as representations arising faster when primed.
  - Operationalise this as the timing of the "leading edge" of the curve (time-to-half-max).
- Significant differences shown (subject-bootstrap 95% CIs).



Evidence of sharpening

- Repetition suppression can manifest as lower/slower representations.
- Sharpening can manifest as representations arising faster when primed.
  - Operationalise this as the timing of the "leading edge" of the curve (time-to-half-max).
- Significant differences shown (subject-bootstrap 95% CIs).



1-back: perceptual representations slower (suppression)

1-back: categorical representations faster (sharpening)

Summary

MEG results

- **We used a fine-tuned AlexNet** to model the timeline of categorical semantic processing in a priming experiment.
- The ANN was **a better model of categorical representation in the brain** than just the pure class label.
- Shortly after the foveal probe presentation (0–300ms), perceptual layers were a good match, better than the semantic layers, and were slowed by a recent (1-back) prime.
  - Evidence of **repetition suppression** for lower-level representations.
- In a later period (300–600ms), semantic layers were good, (sometimes) better than perceptual layers, and were accelerated by a recent (1-back) prime.
  - Evidence of **representational sharpening** for higher-level representations.
- **Future work**: whole-scalp sensor-space → spatiotemporal source-space searchlight.

Future directions: neuronal dynamics in ANNs (Q2)

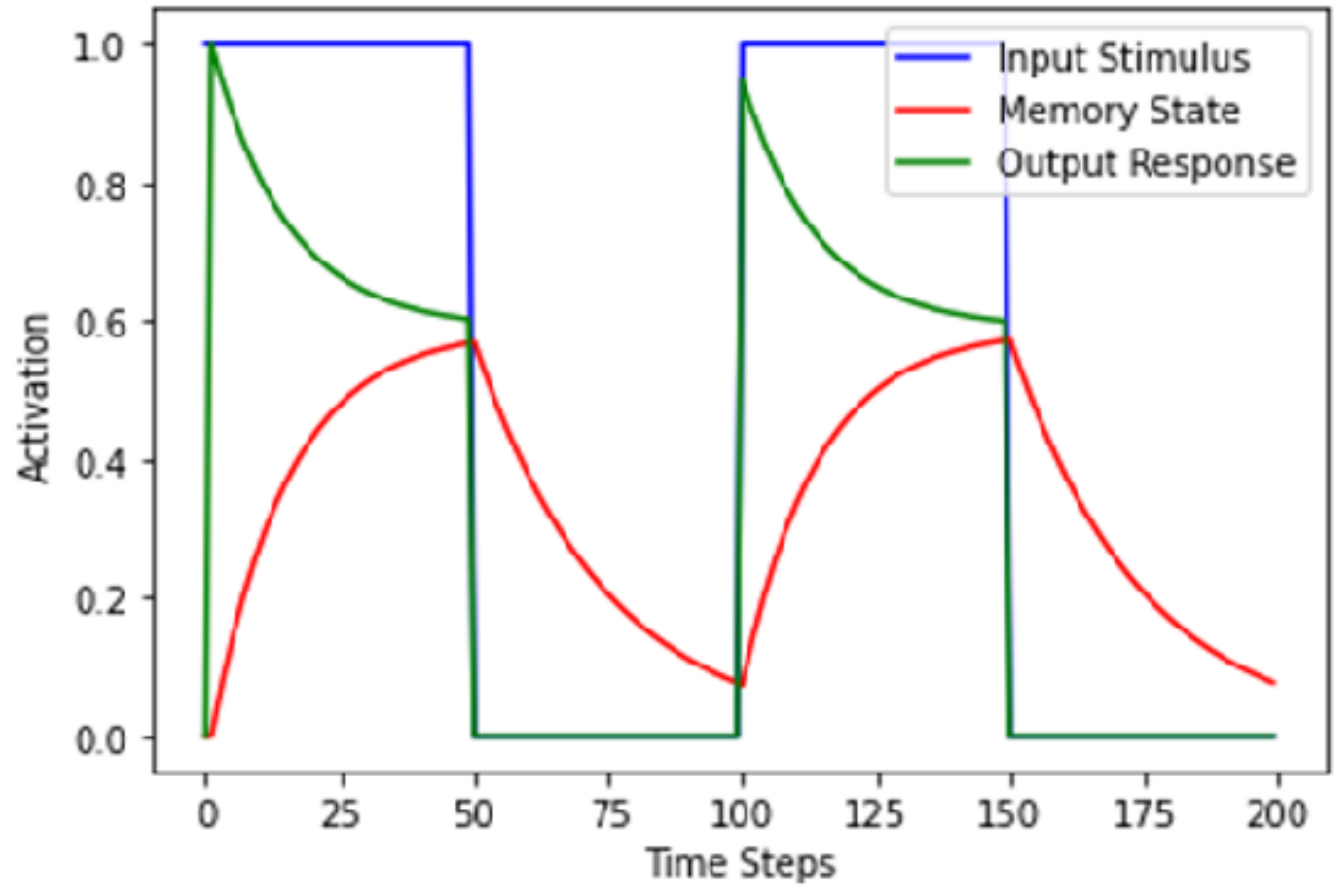
Priming effects in neuronal dynamics

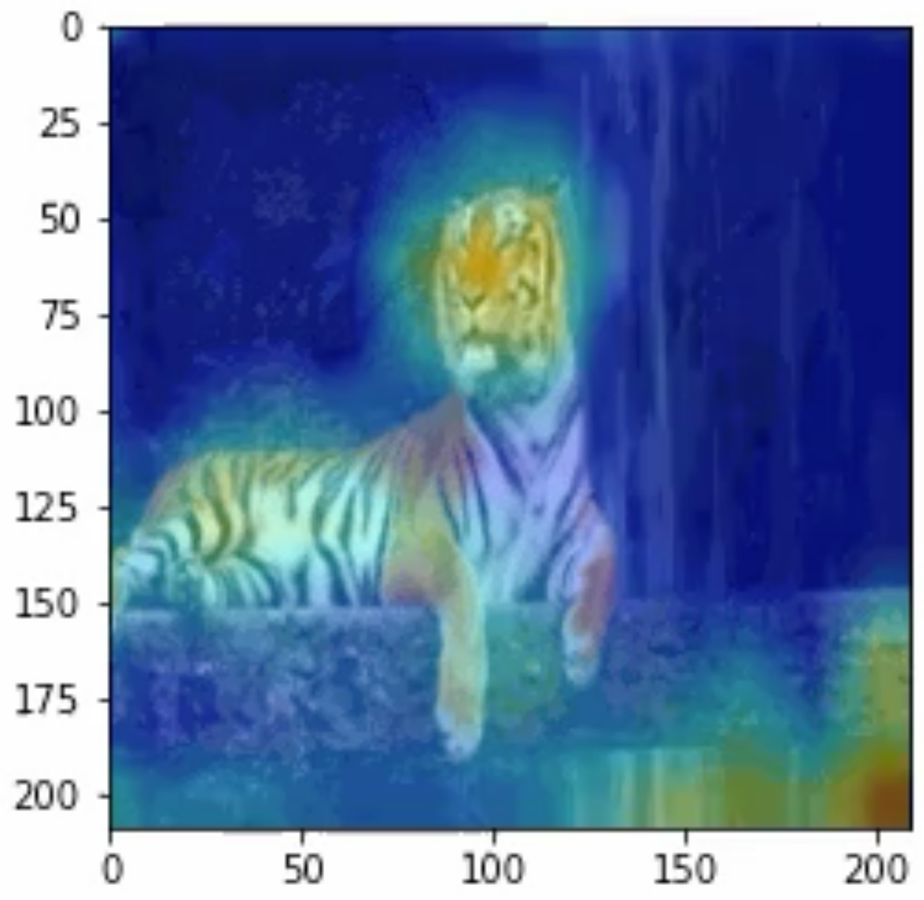
Suppression and sharpening

- Recall: we've talked about two forms of priming effect:
  - **Suppression**: Neurons become 'bored' of repeated similar stimuli and produce a weakened activation. They reactivate when they are 'surprised'.
  - **Sharpening**: If the stimulus has been processed previously, it is easier to process it upon re-exposure. The activation is weaker but the signal is more salient.

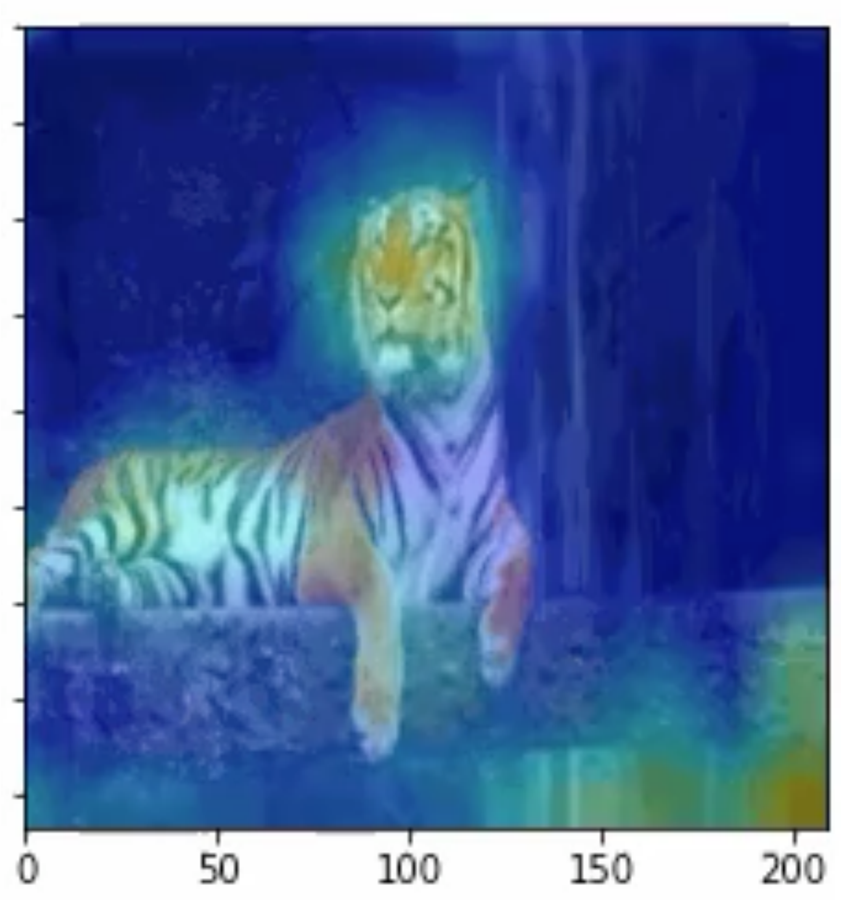
Modelling suppression

- A memory state builds up to match the input signal
- The memory state is subtracted from the activation
- Long exposure to a stimulus will cause the activation to decrease over time





No Suppression

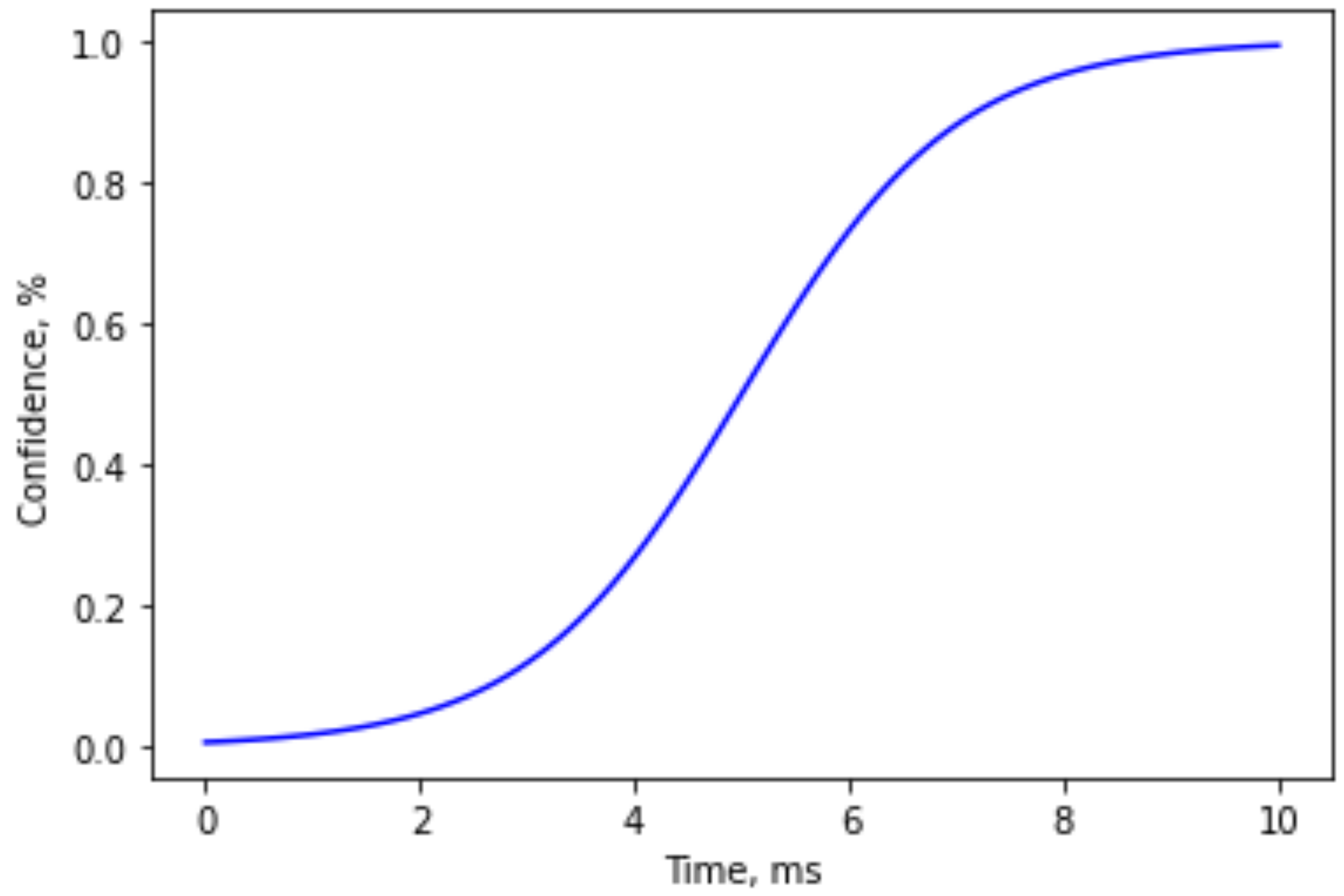


With Suppression

Vinken, K., Boix, X., & Kreiman, G. (2020). Science Advances, 6(42), eabd4205.

Modelling sharpening

- How can the neurons process the input more quickly?
- Currently, ANNs process images one at a time. They do not accumulate evidence over time and they have no memory state.
- The brain is building some kind of short term memory state. How can we incorporate a memory state into ANNs?



Thanks