

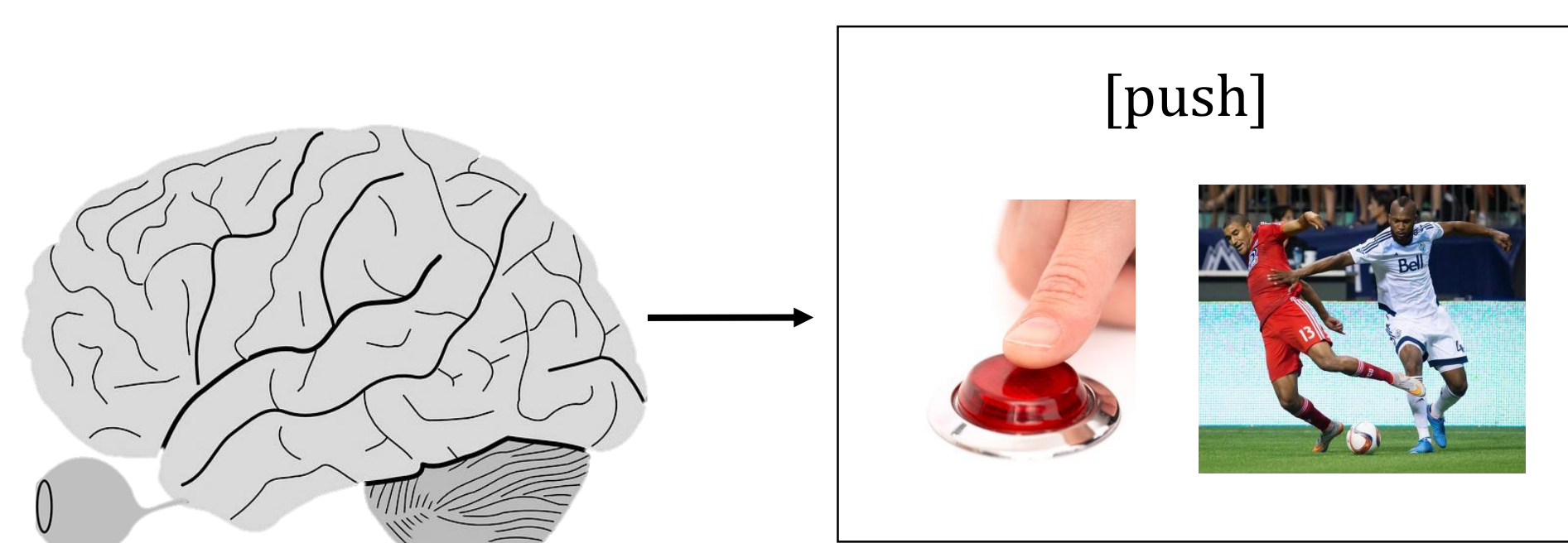
# Modeling the Neural Structure Underlying Human Action Perception

Leyla Tarhan & Talia Konkle

Department of Psychology, Harvard University

## Introduction

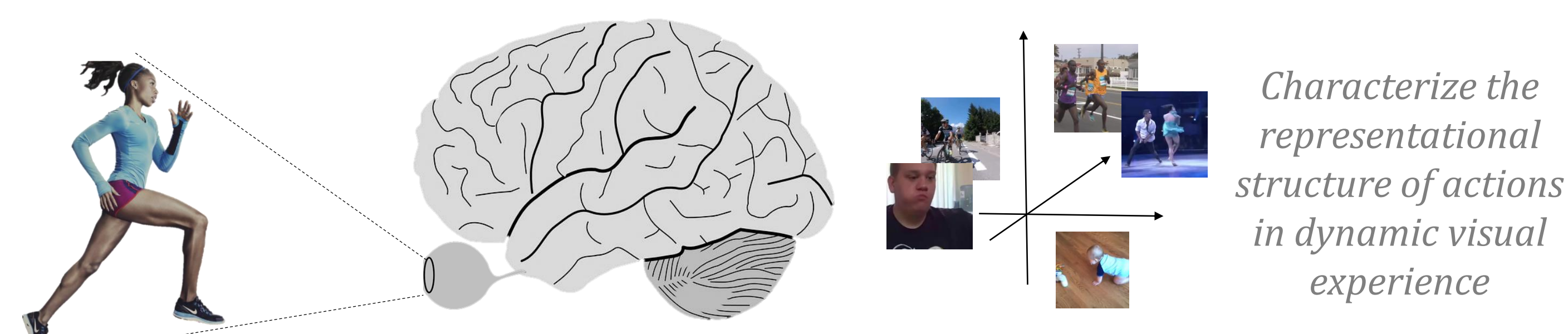
Prior work often approaches the concept of “actions” within a *linguistic framework*.



Search for high-level representations that link movements by verb.

Hafri et al. (2017); Wurm et al. (2017); Pulvermuller et al. (2005); Watson et al. (2011)

Instead, we investigate actions from a *perceptual stance* reflecting variation in what we see people do.

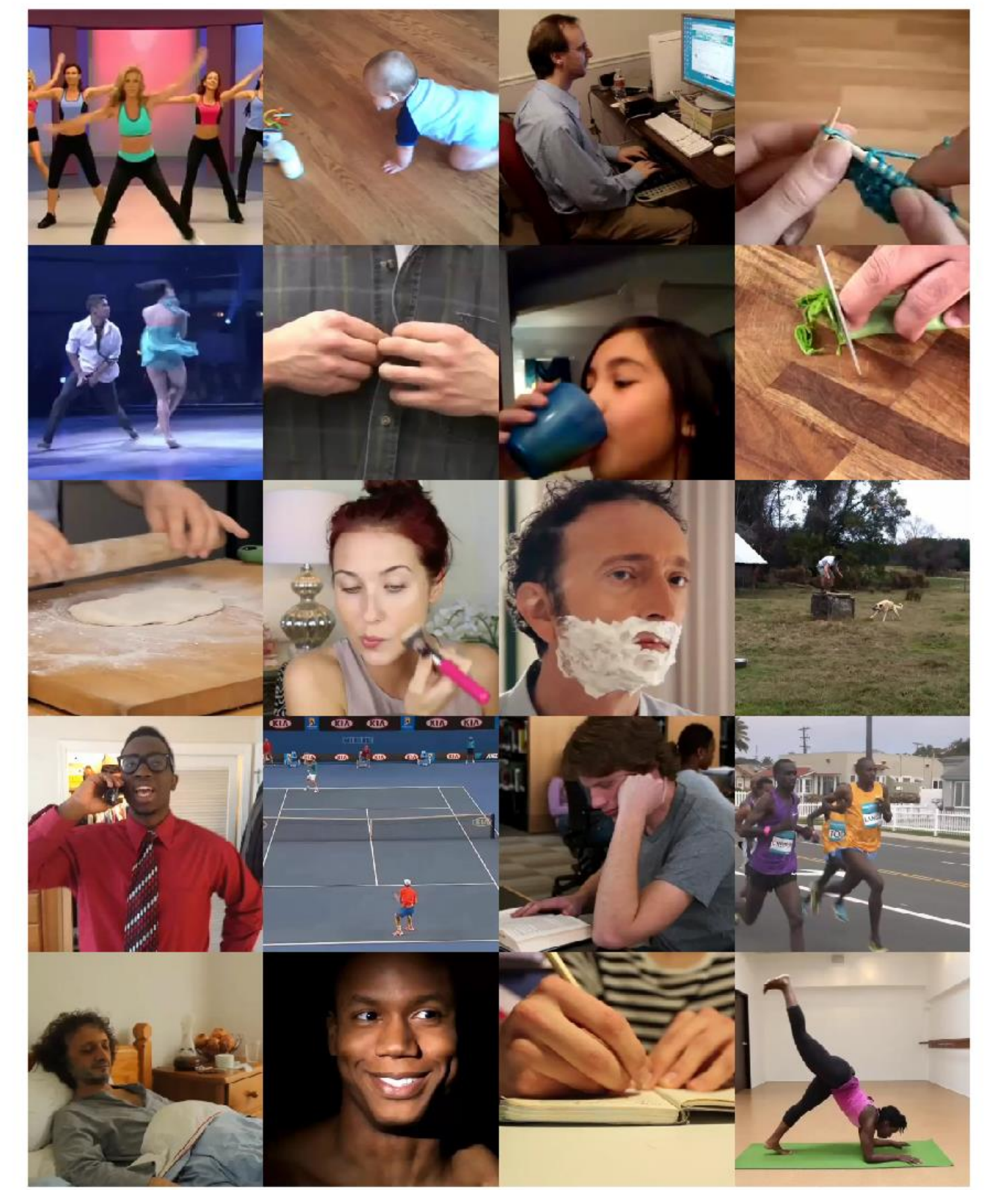


Characterize the representational structure of actions in dynamic visual experience

Lingnau & Downing (2015); Giese & Poggio (2003); Isik et al. (2016);

## Stimuli

A wide sample of everyday actions



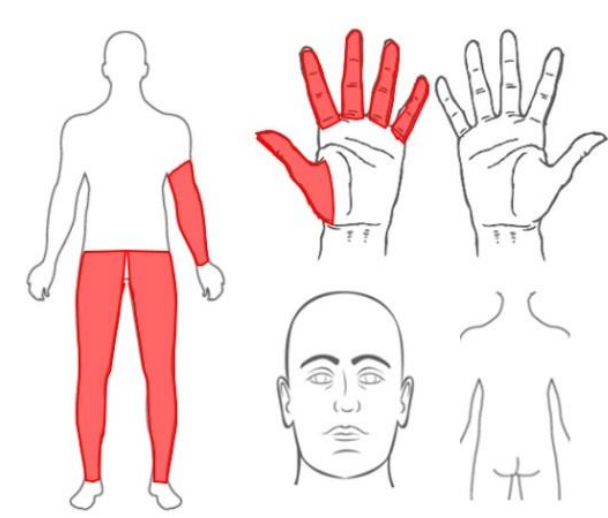
120 2.5s videos of 60 actions  
Sampled from American Time Use Survey

## Results 1: Encoding Models

What kinds of features predict action responses, and where?

### Body Parts

Select the body parts involved in the action



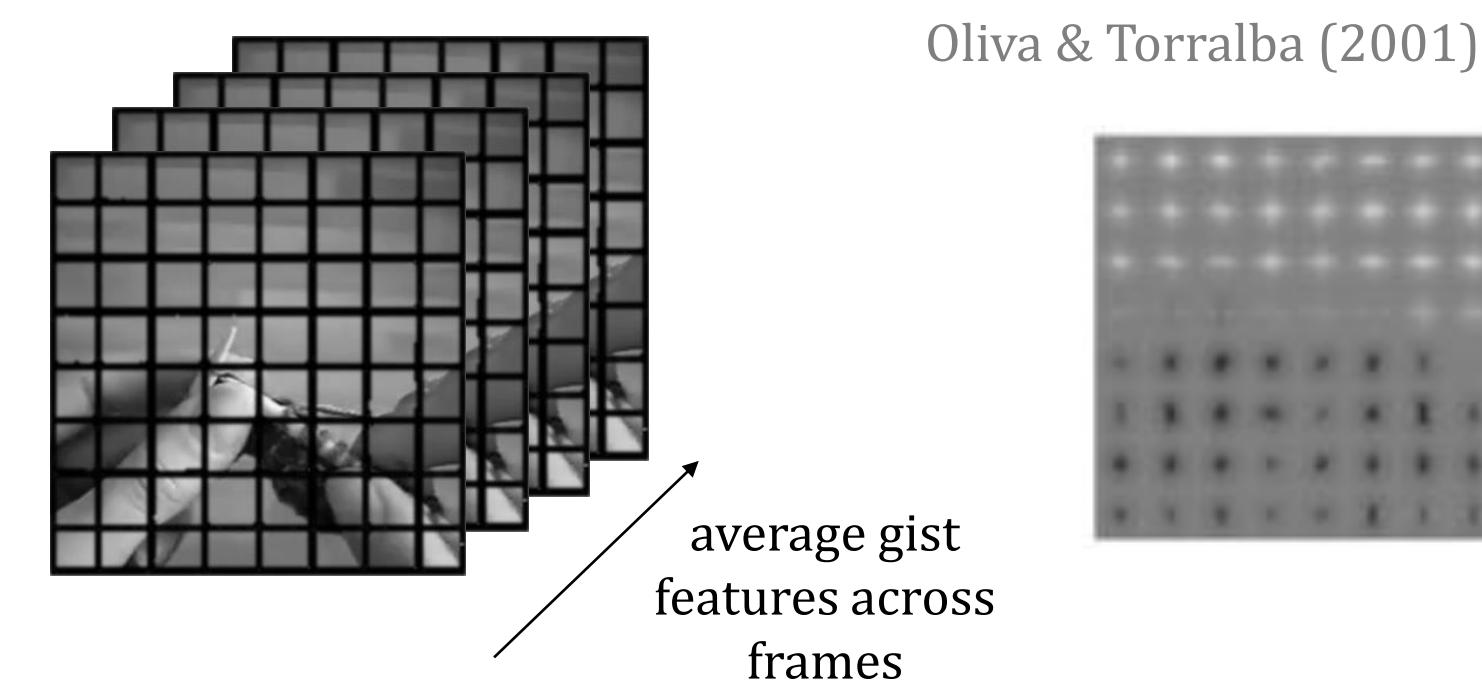
### Action Target

What is this action directed at?

- an object
- another person
- your own body
- space

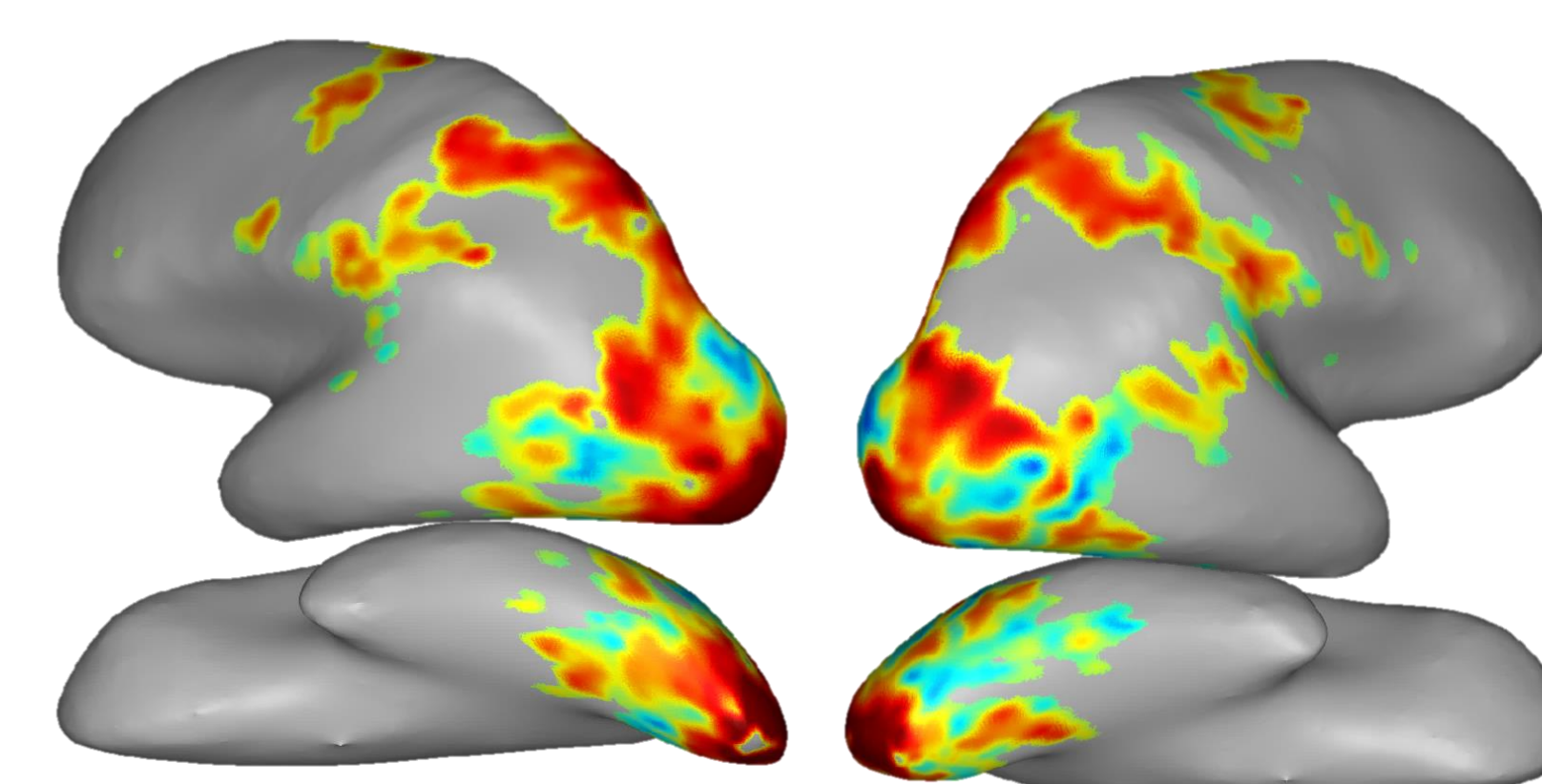
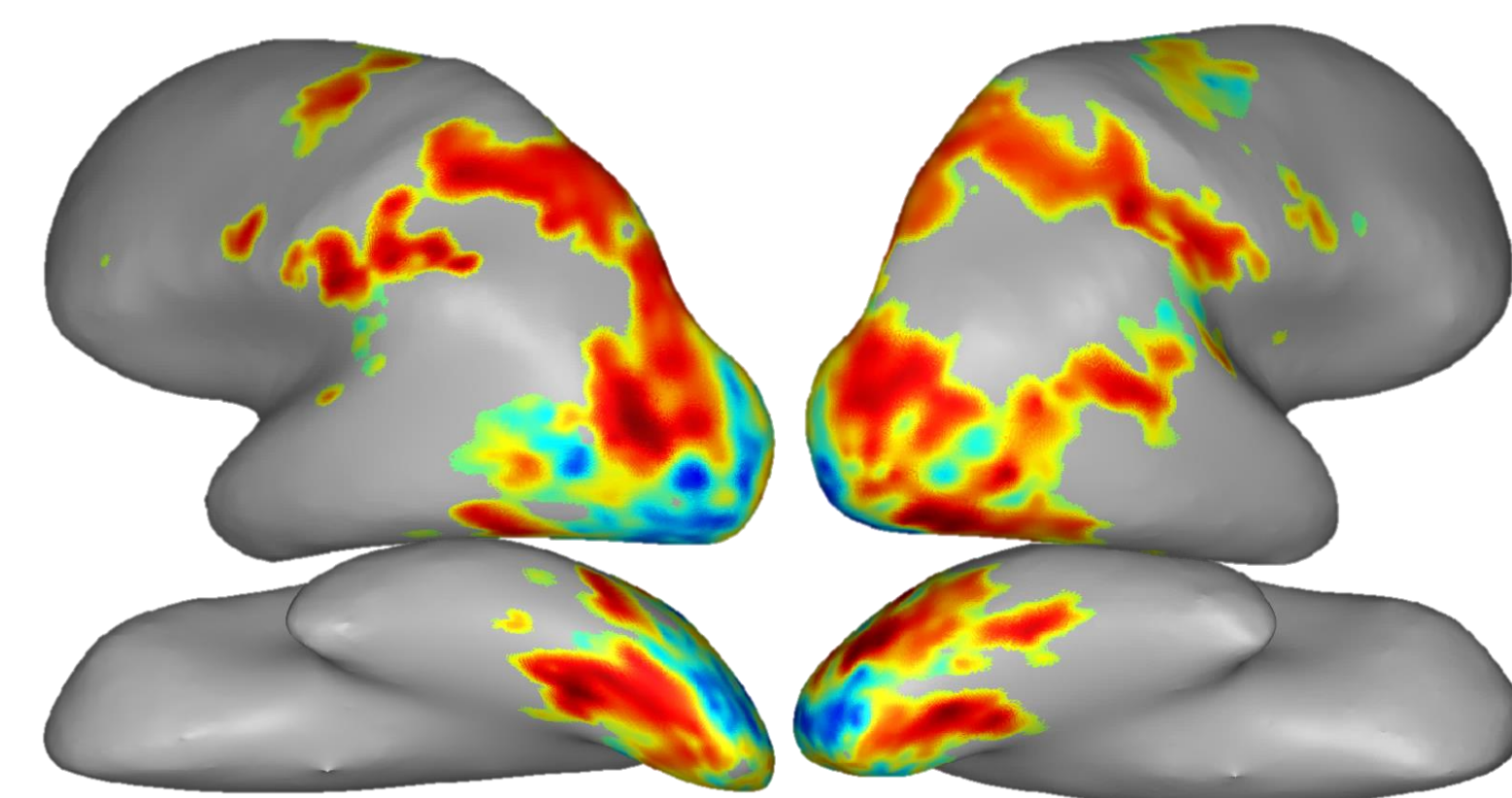
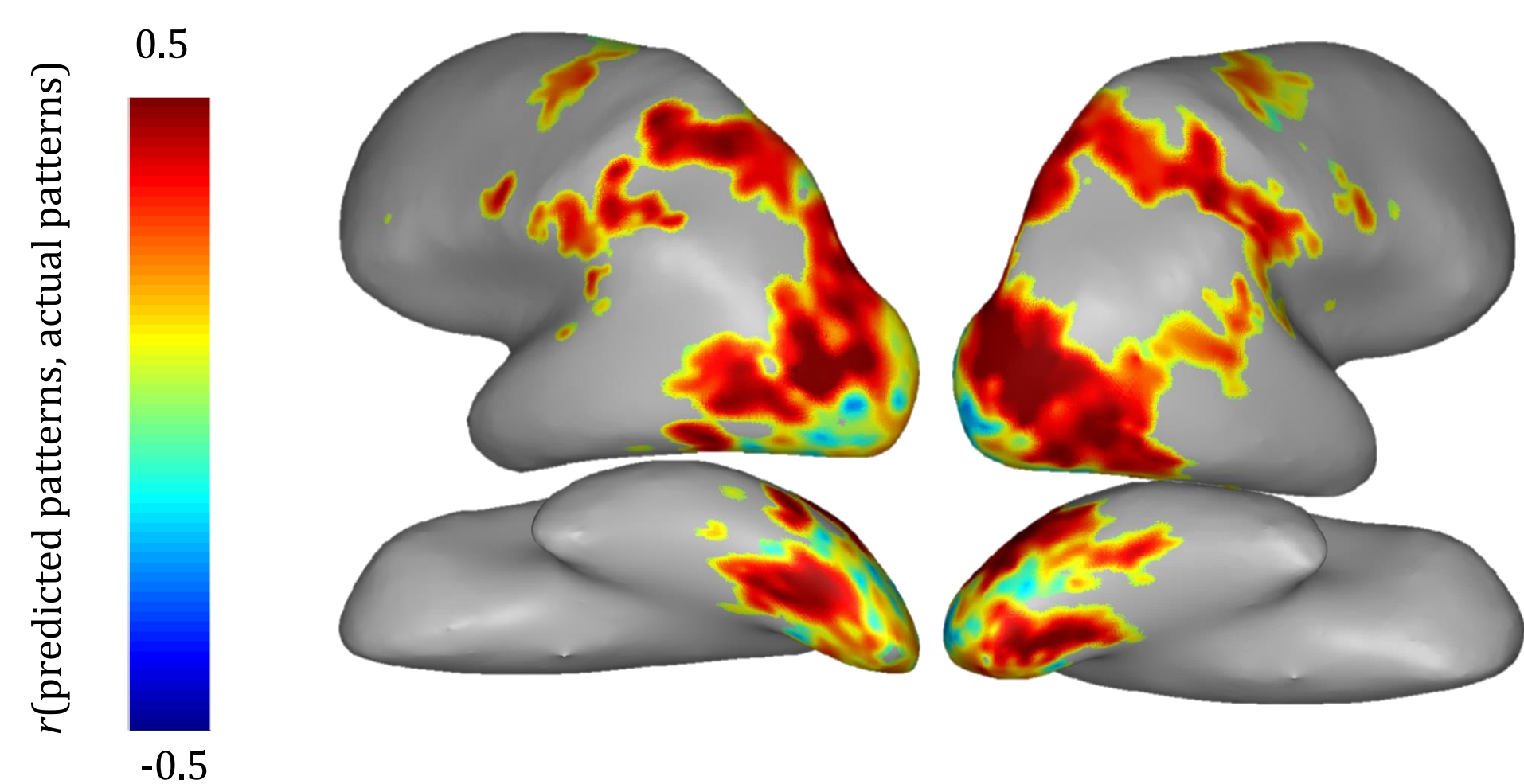
### Gist

Capture global form and low-level visual properties

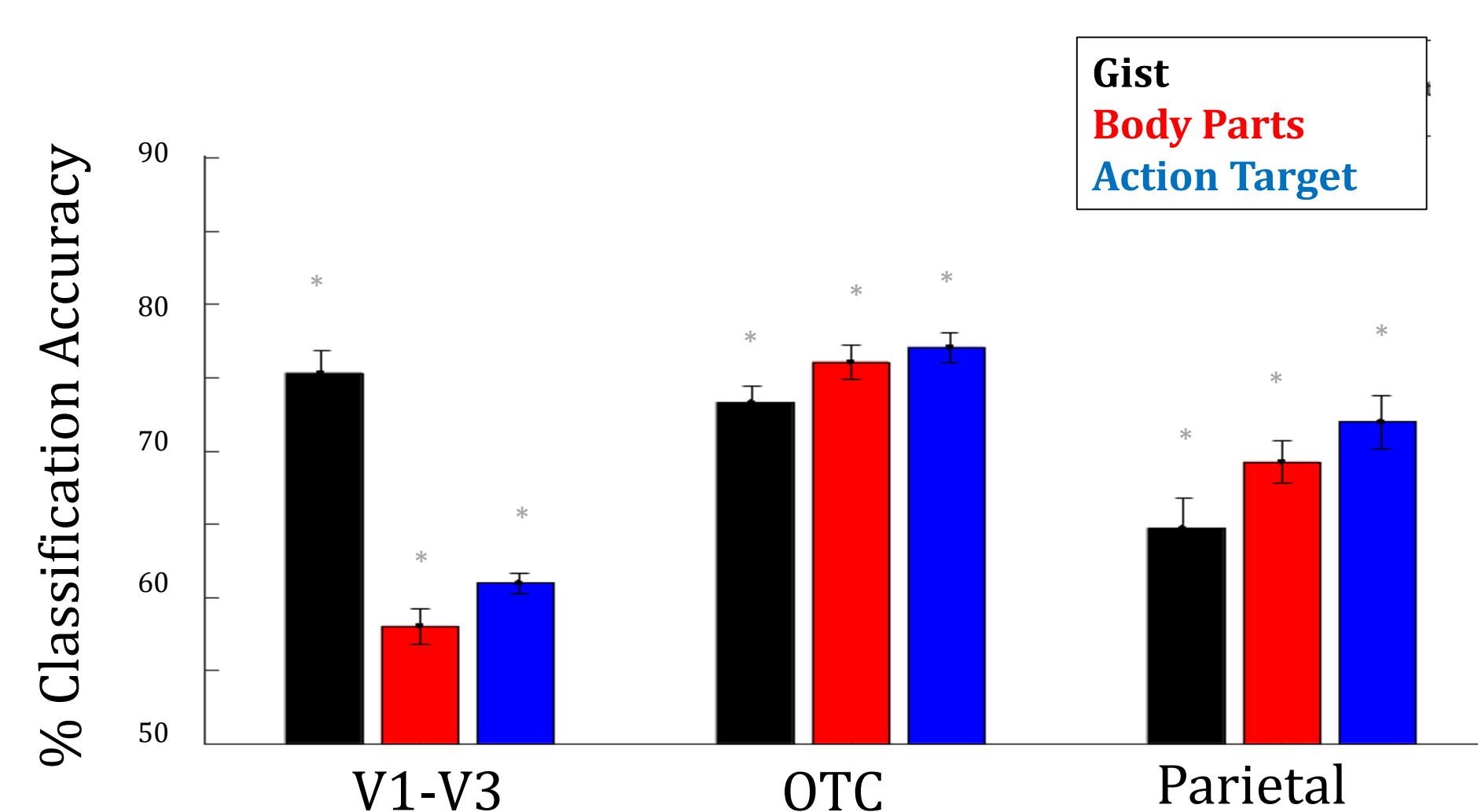


Oliva & Torralba (2001)

average gist features across frames



Prediction Accuracy across visual stream divisions:



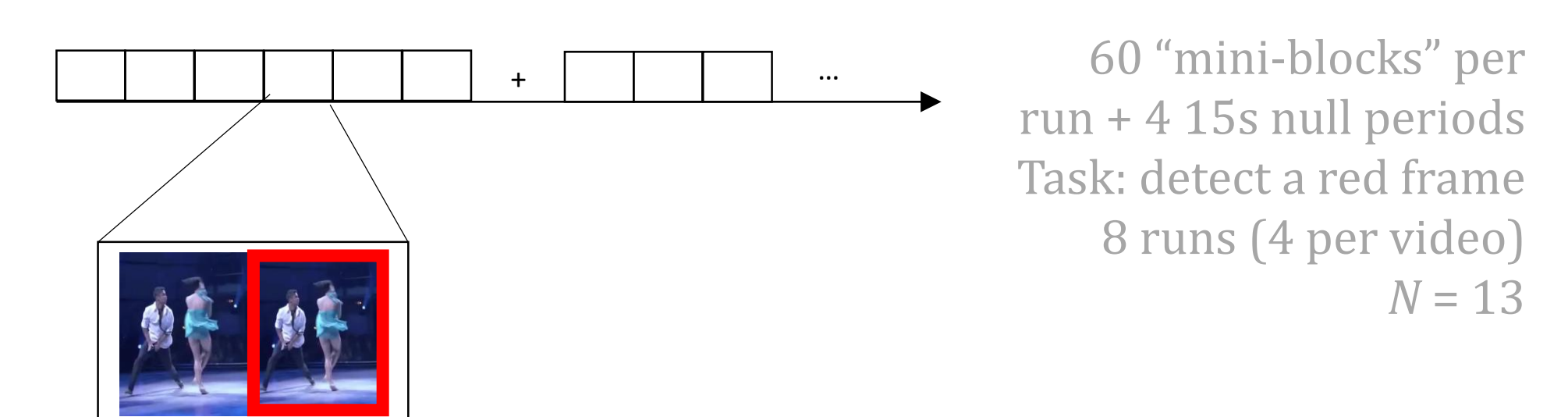
All models perform well across visual cortex – up to 49% of the variance is accounted for by these combined feature dimensions ( $\leq 33\%$  shared).

Gist features perform best in V1 – V3, and do surprisingly well along the ventral and parietal streams.

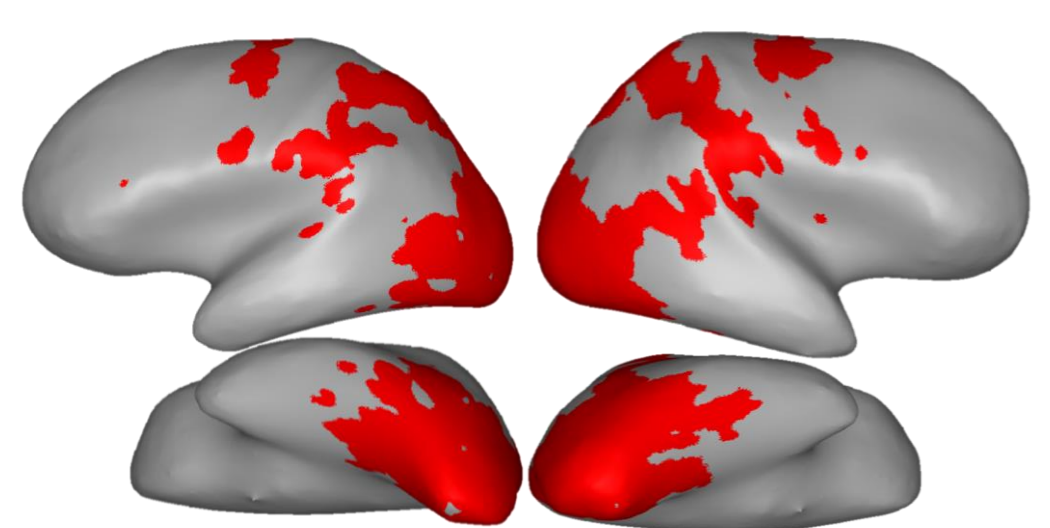
High-level models do best outside of V1 – V3.

## Neural Measures

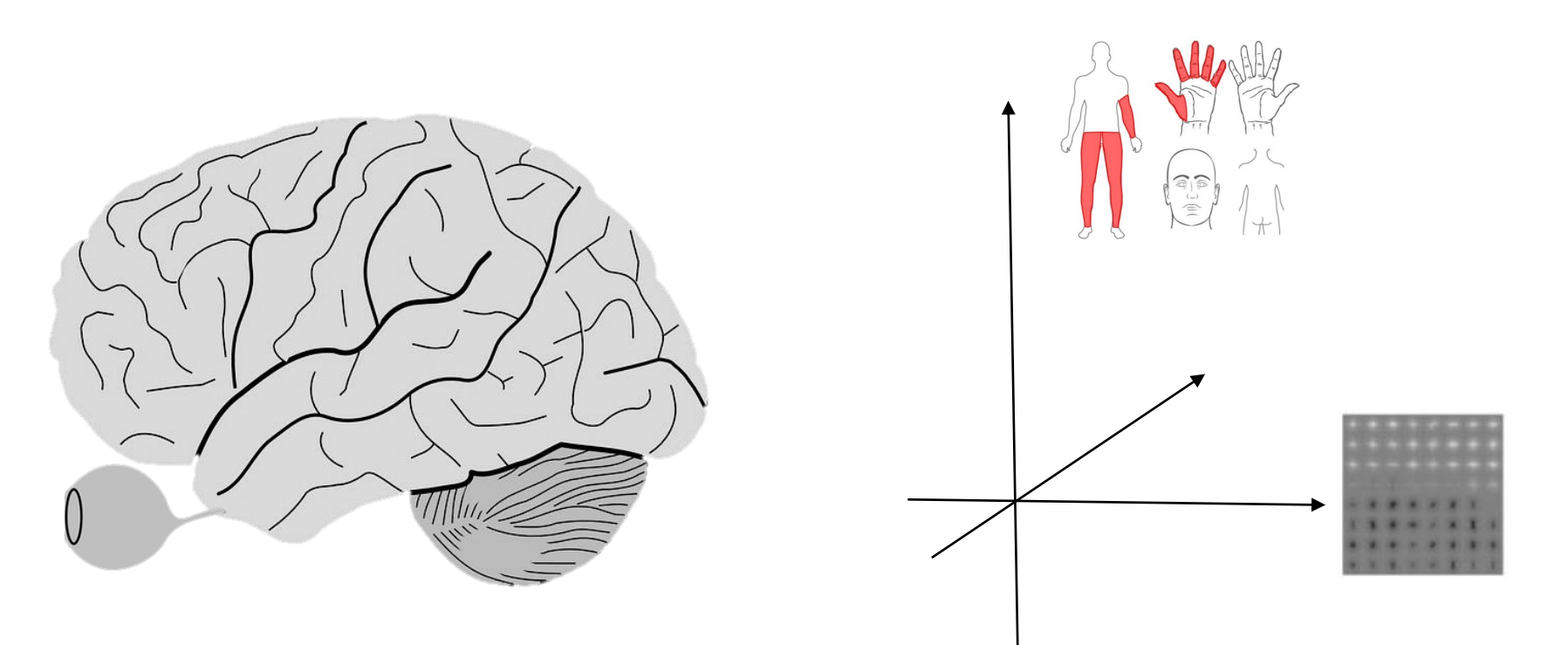
1. Collect neural responses to each video (fMRI)



2. Restrict analyses to reliable voxels across runs

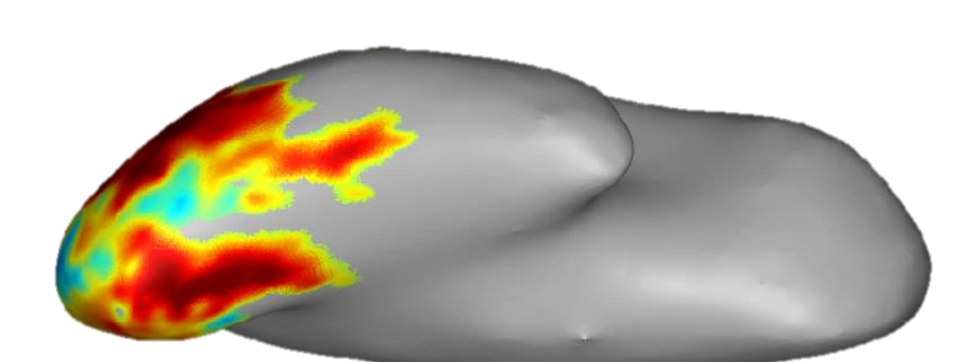


## Conclusions



Observing actions drives the full visual hierarchy and neural responses are well-fit by multiple kinds of feature models.

Are action and object representations separate domains ... with distinct neural substrates?



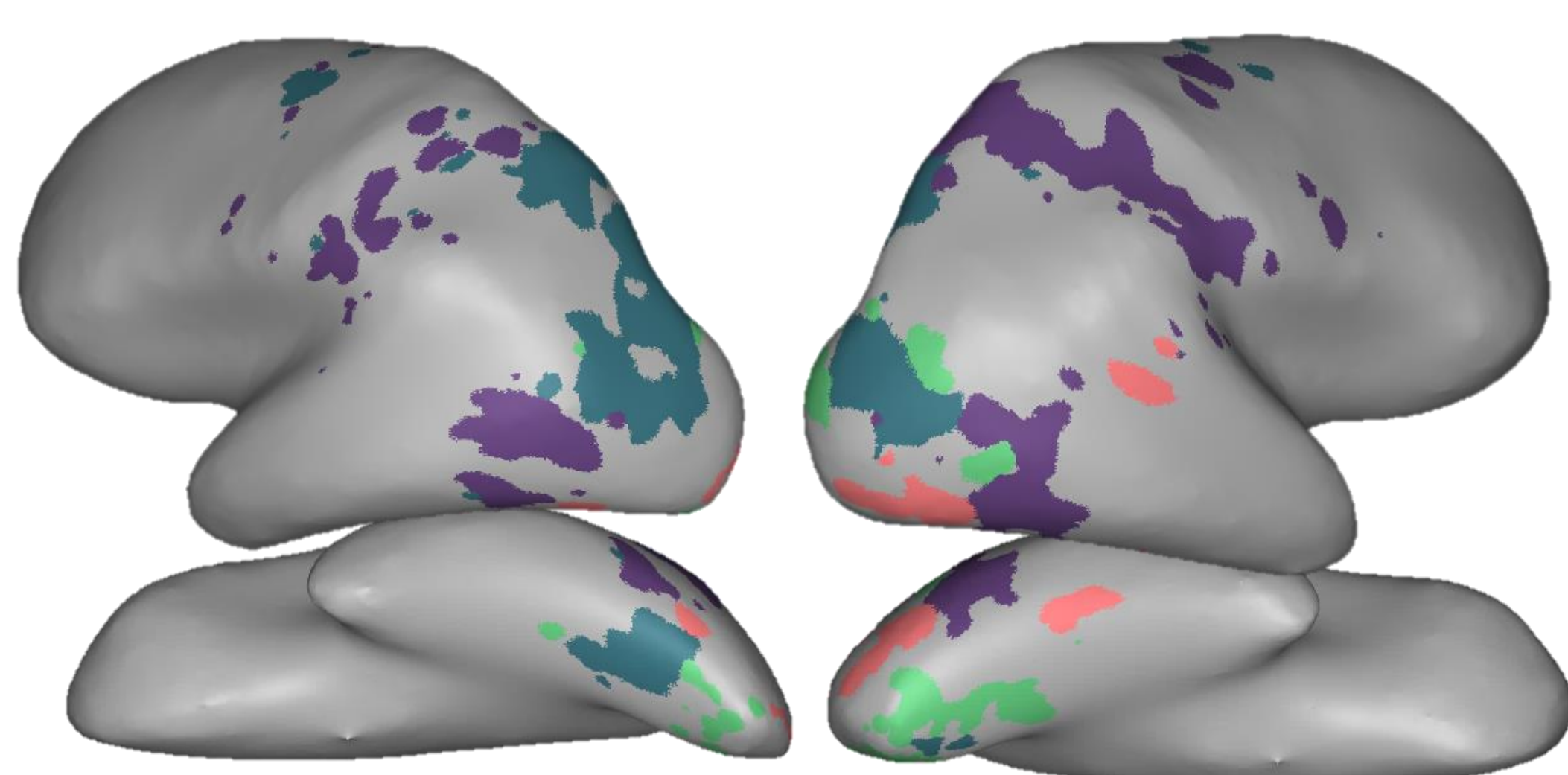
The ventral stream (“what” pathway) has a larger role that includes action and object representation

(Lingnau & Downing, 2015)

## Results 2: Voxel Tuning

How are the voxels tuned? Cluster voxels based on their model-specific weights.

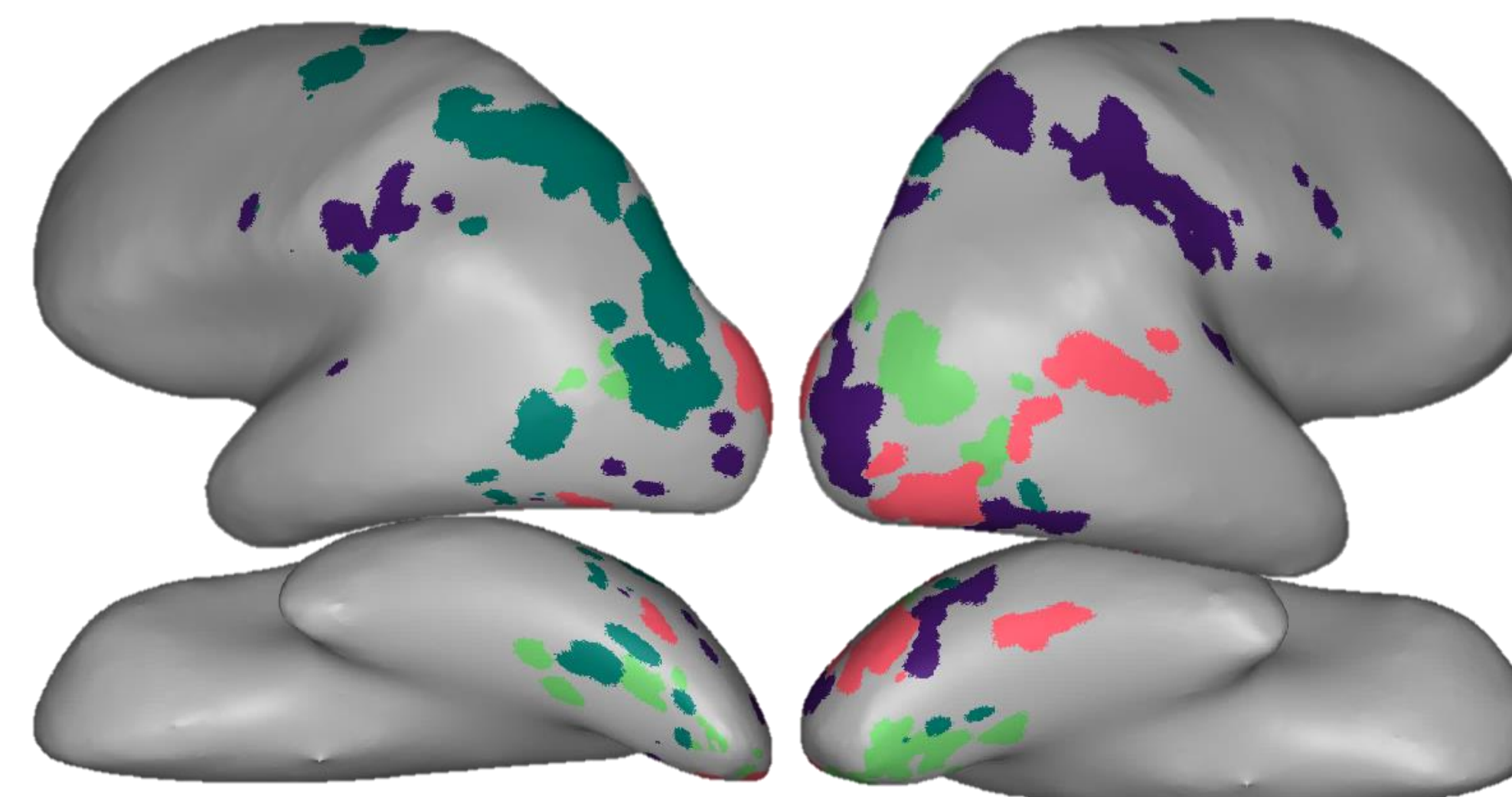
### Body Parts



Voxels cluster into 4 groups:

- 1) feet & legs
- 2) core & hands
- 3) head & rear
- 4) everything but face

### Action Target



Voxels cluster into 4 groups:

- 1) objects
- 2) objects & self
- 3) objects & near space
- 4) far space & people