ICLR

TAVAE: A VAE with adaptable priors explains contextual modulation in the visual cortex

Balázs Meszéna[1], Keith T. Murray[1,2], Julien Corbo[3], O. Batuhan Erkat[3], Márton A. Hajnal[1], Pierre-Olivier Polack[3], Gergő Orbán[1]

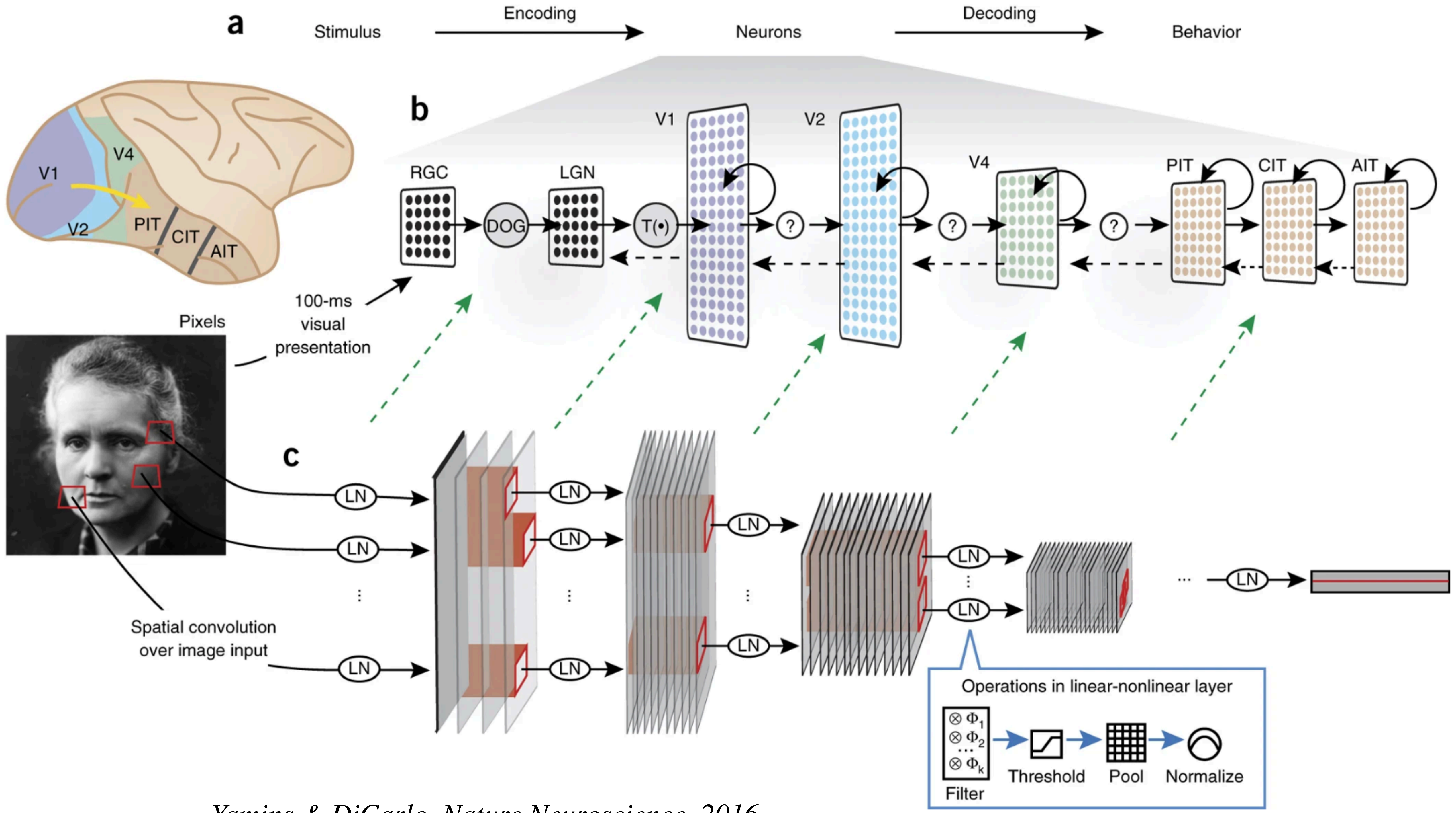
1) Computational Systems Neuroscience Lab, Department of Computational Sciences, HUN-REN Wigner Research Centre for Physics, Budapest, Hungary
2) Princeton Neuroscience Institute, New Jersey, United States
3) Center for Molecular and Behavioral Neuroscience, Rutgers University, Newark, United States

csnl HUN REN wigner

RUTGERS UNIVERSITY | NEWARK

http://golab.wigner.mta.hu/

Predicting neuronal responses in the visual cortex using deep neural networks

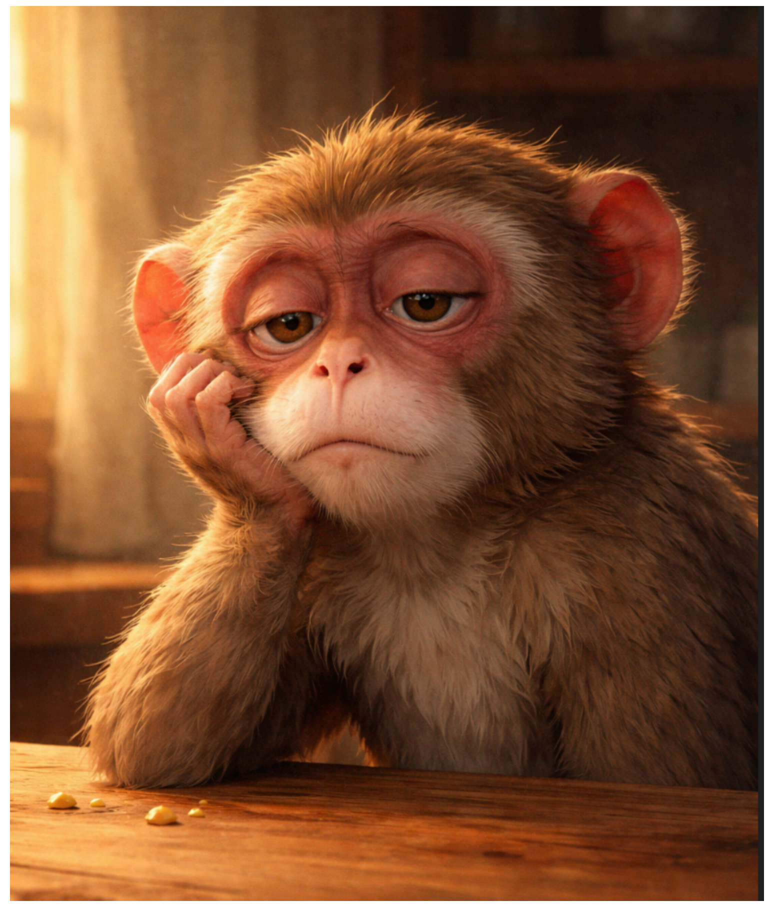


Yamins & DiCarlo, Nature Neuroscience, 2016

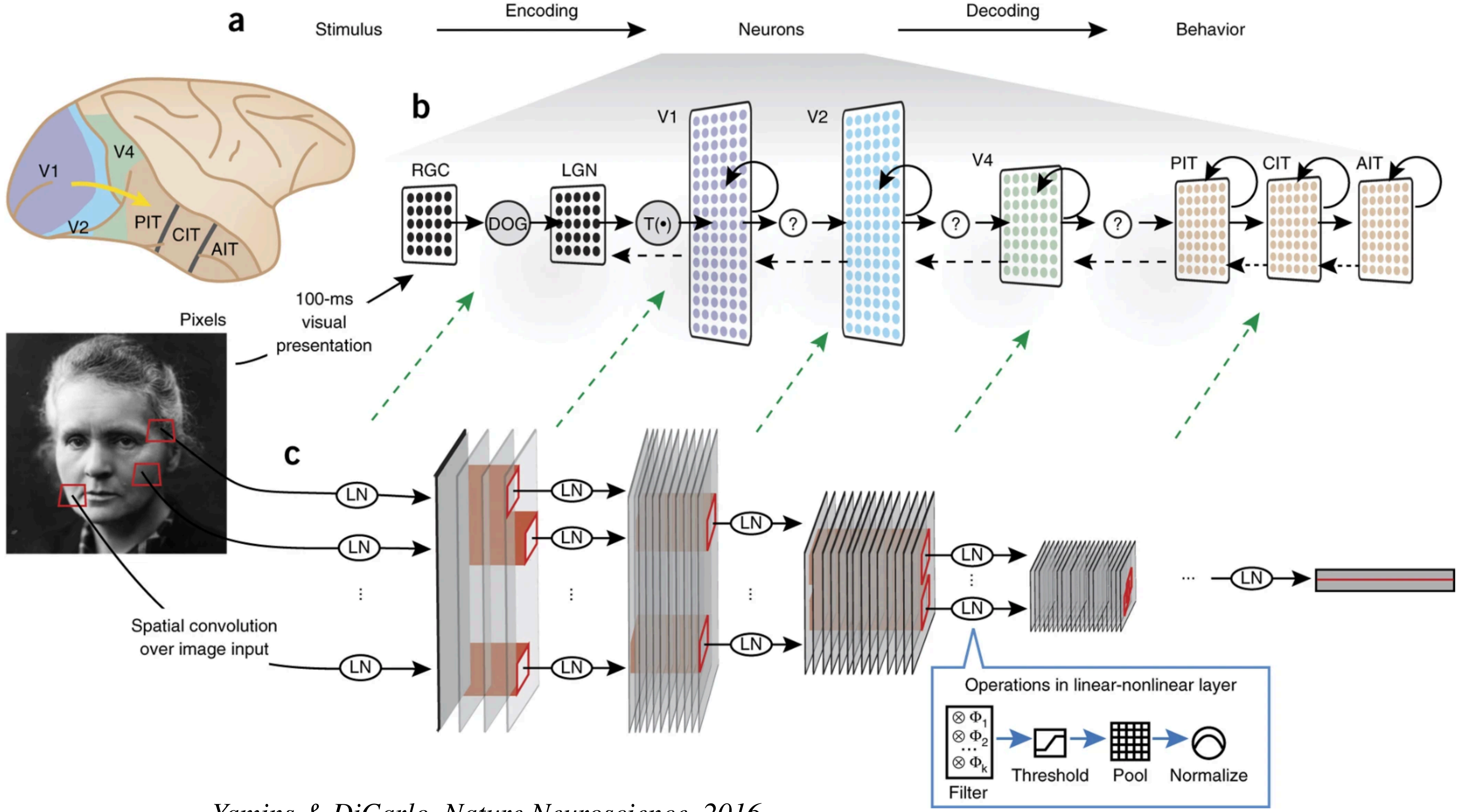
Biological neurons

Model CNN layers

Predicting neuronal responses in the visual cortex using deep neural networks



Passive viewing

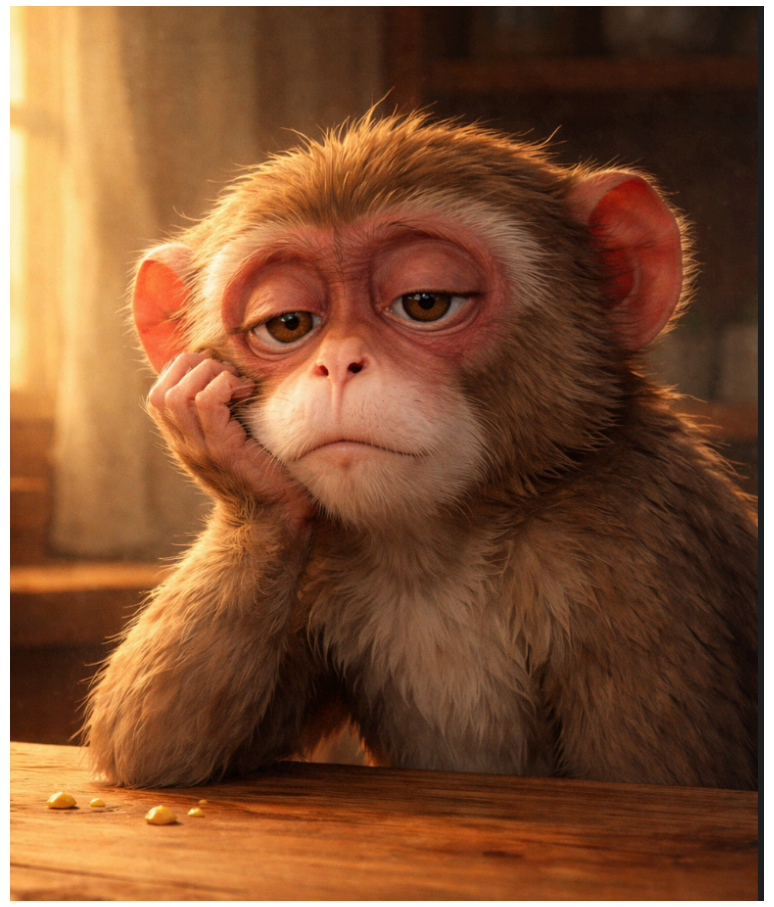


Yamins & DiCarlo, Nature Neuroscience, 2016

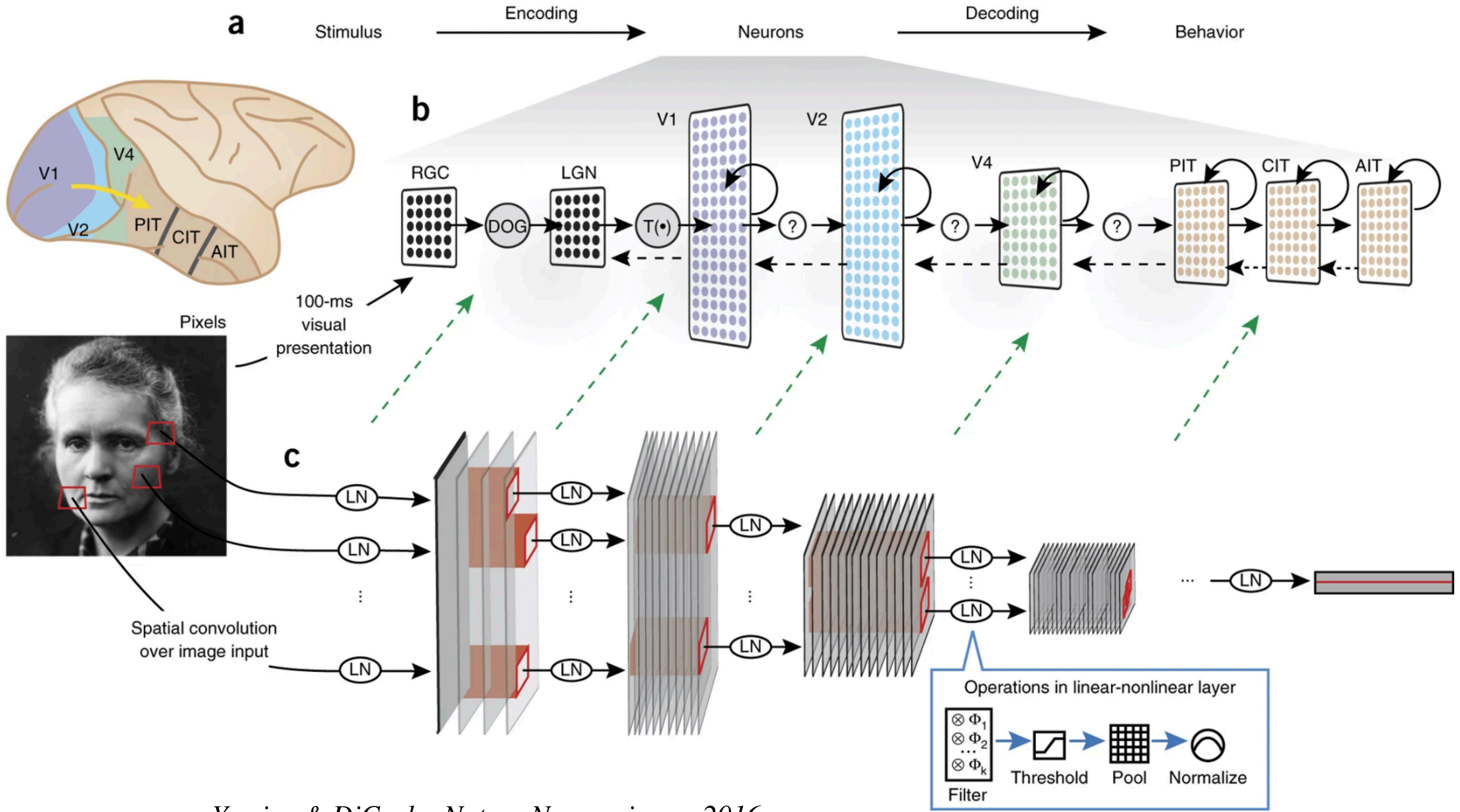
Biological neurons

Model CNN layers

Predicting neuronal responses in the visual cortex using deep neural networks



Passive viewing



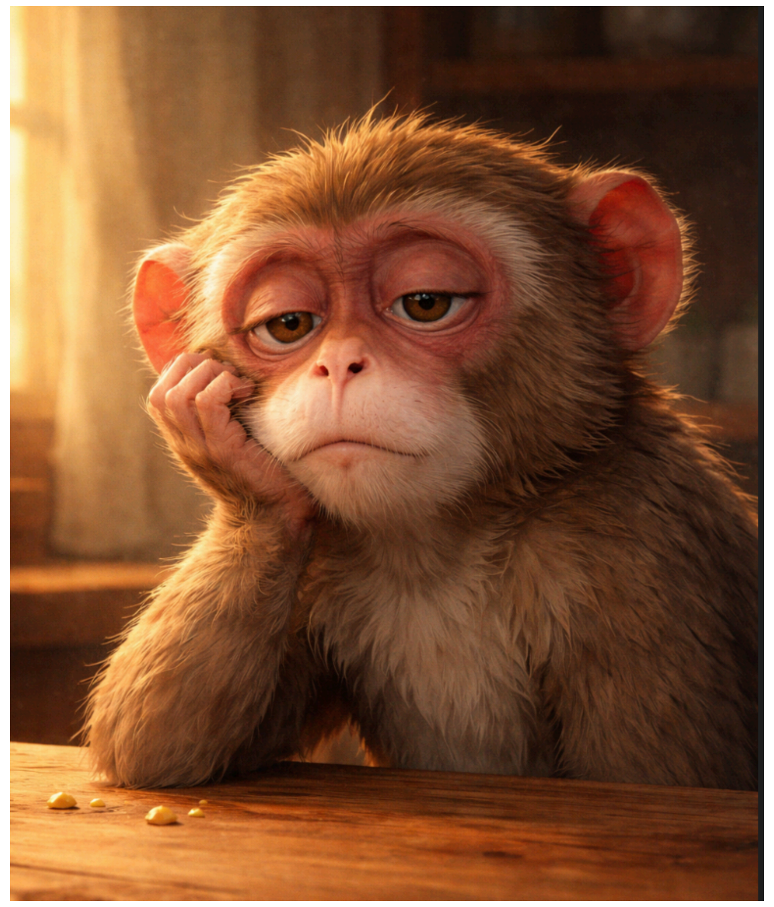
Yamins & DiCarlo, Nature Neuroscience, 2016

Biological neurons

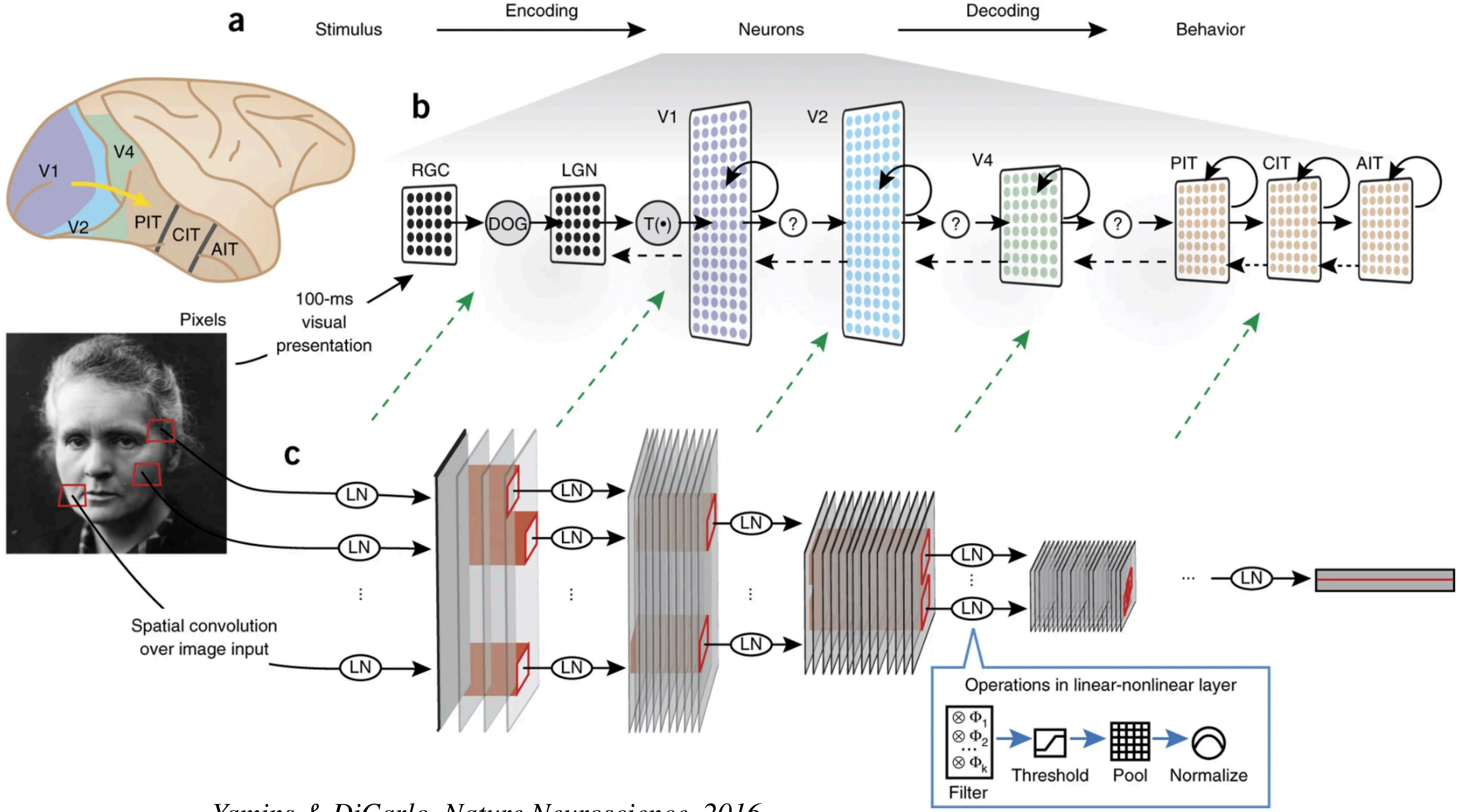
Model CNN layers

Neuronal response is context dependent

Predicting neuronal responses in the visual cortex using deep neural networks



Passive viewing



Yamins & DiCarlo, Nature Neuroscience, 2016

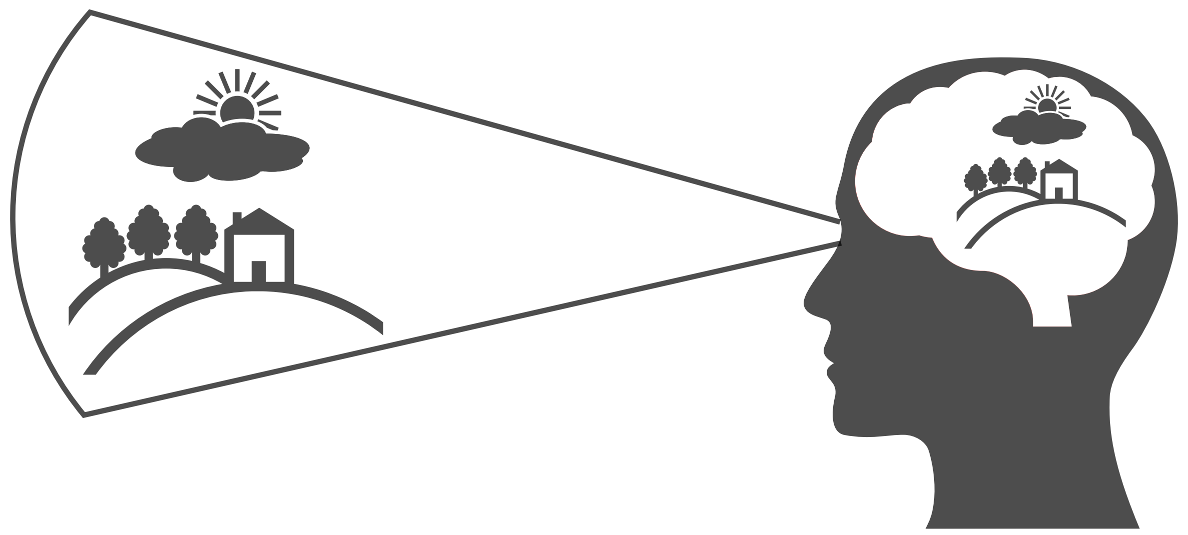
Biological neurons

Model CNN layers

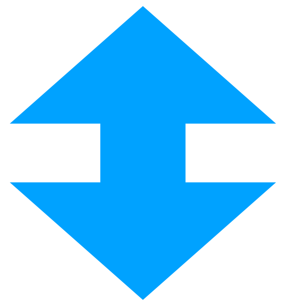
Neuronal response is context dependent

During task?

Modeling vision with latent generative model

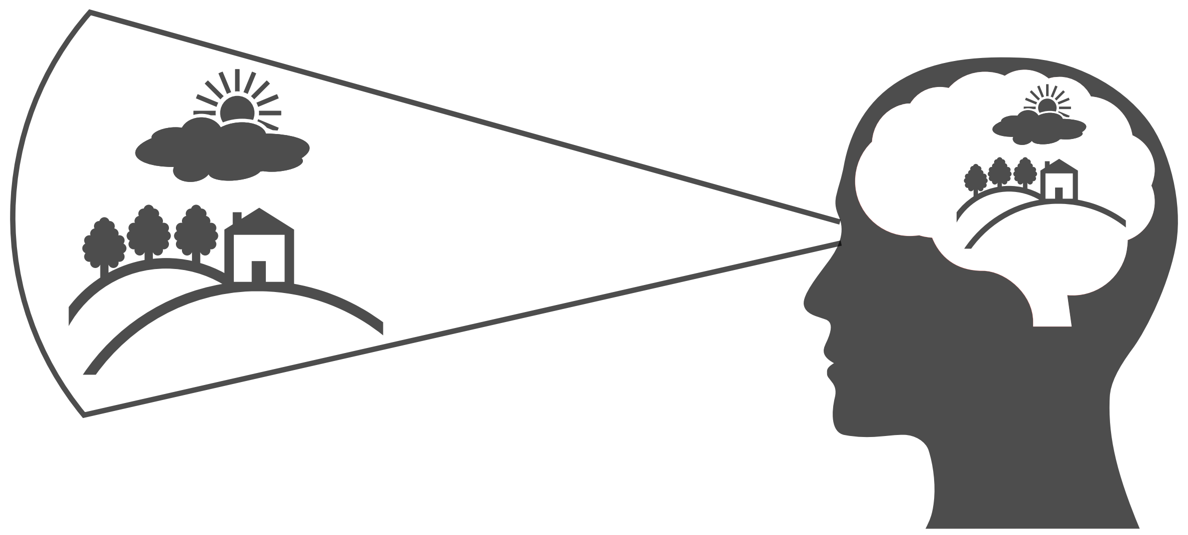


Visual cortex



Generative model

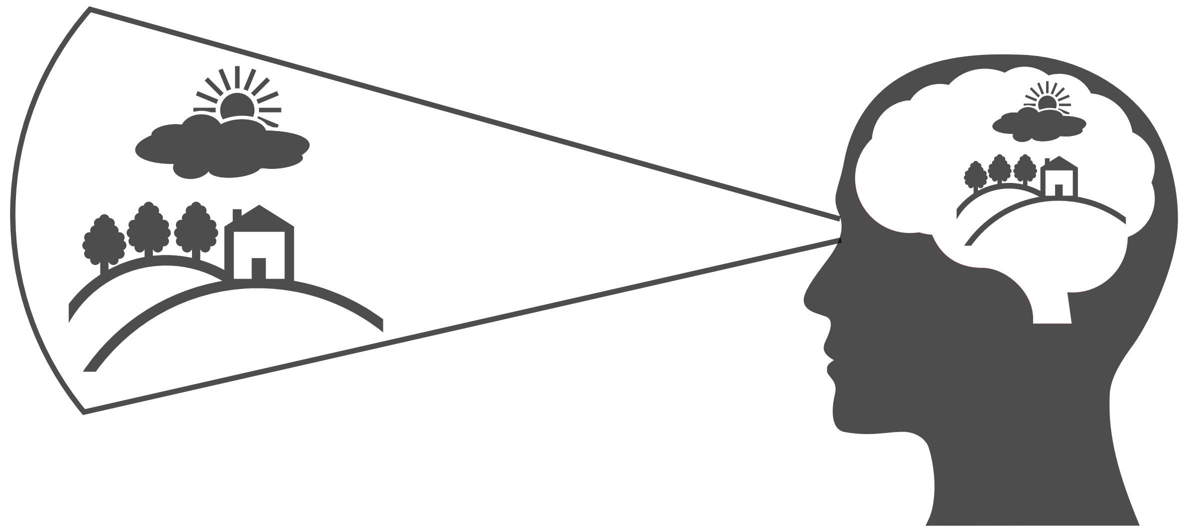
Modeling vision with latent generative model



Visual cortex ⇕ **Generative model**

Neuronal response ⇕ **Latent inference**

Modeling vision with latent generative model



Learning objective:
Learning regularities of natural images

$$p_0(\boldsymbol{x}) = \int p(\boldsymbol{x} \mid \boldsymbol{z}) p_0(\boldsymbol{z})\, d\boldsymbol{z},$$

Visual cortex

⇕

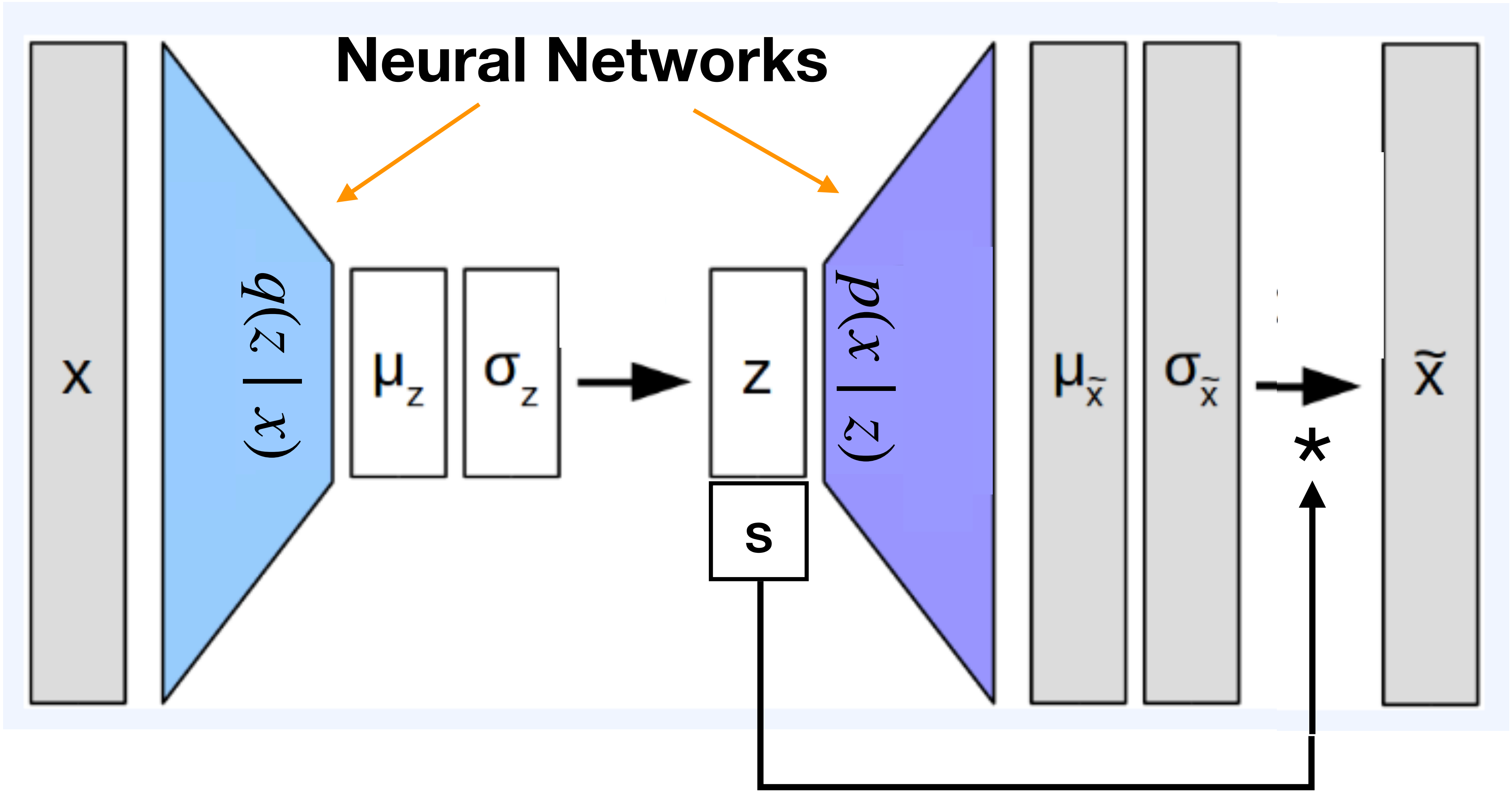
Generative model

Neuronal response

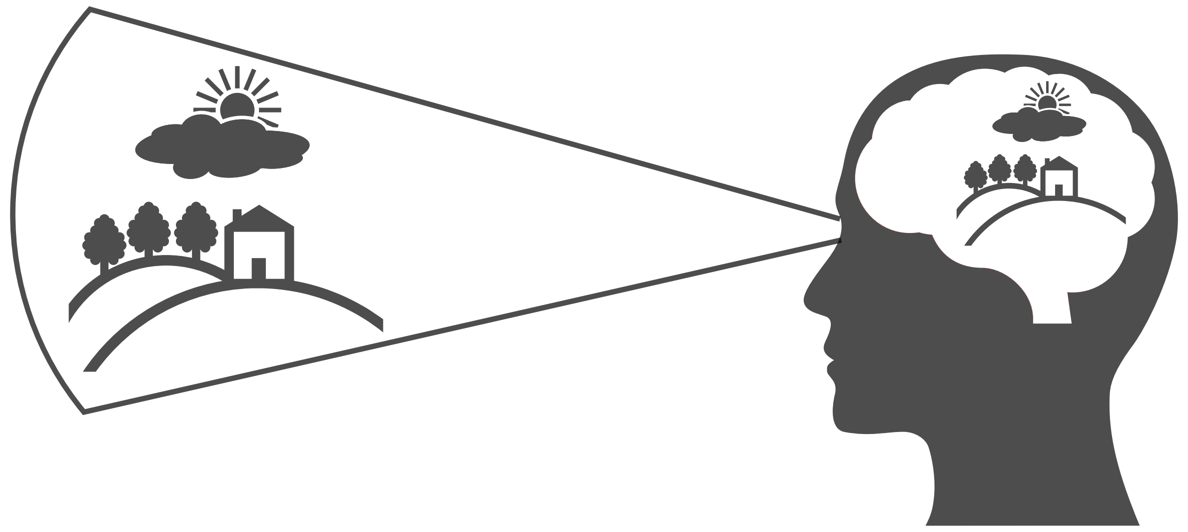
⇕

Latent inference

V1 VAE



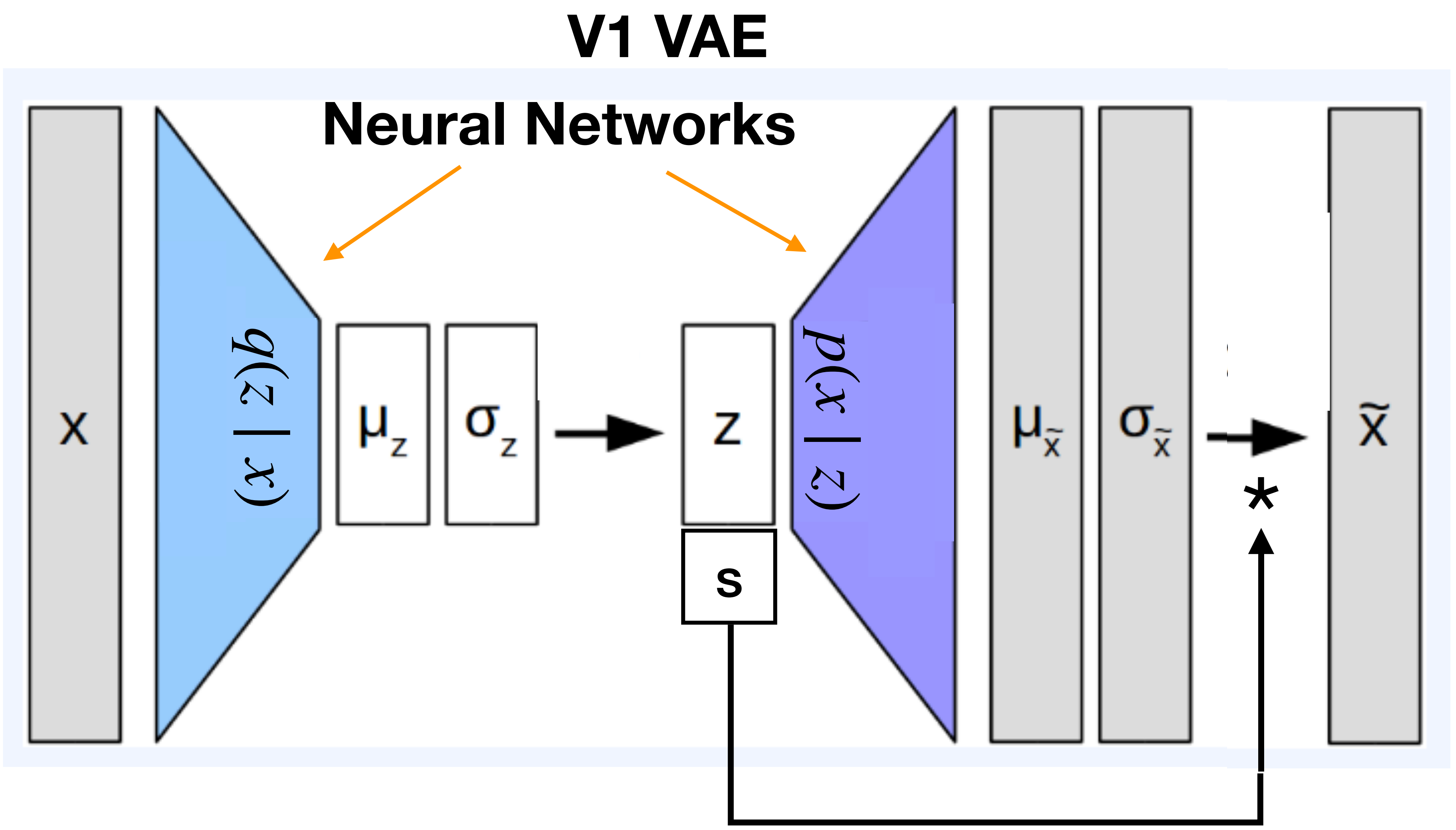
Modeling vision with latent generative model



Learning objective:
Learning regularities of natural images

$$p_0(\boldsymbol{x}) = \int p(\boldsymbol{x} \mid \boldsymbol{z}) p_0(\boldsymbol{z})\, d\boldsymbol{z},$$

Prior



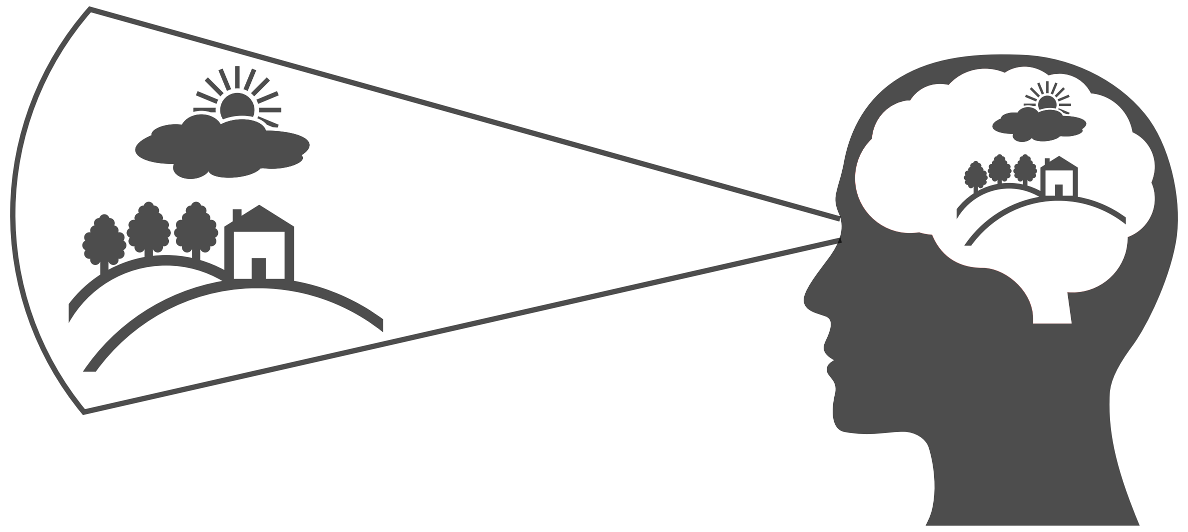
Visual cortex

Generative model

Neuronal response

Latent inference

Modeling vision with latent generative model



Learning objective:
Learning regularities of natural images

$$p_0(\boldsymbol{x}) = \int p(\boldsymbol{x} \mid \boldsymbol{z}) p_0(\boldsymbol{z}) \, d\boldsymbol{z},$$

Likelihood **Prior**

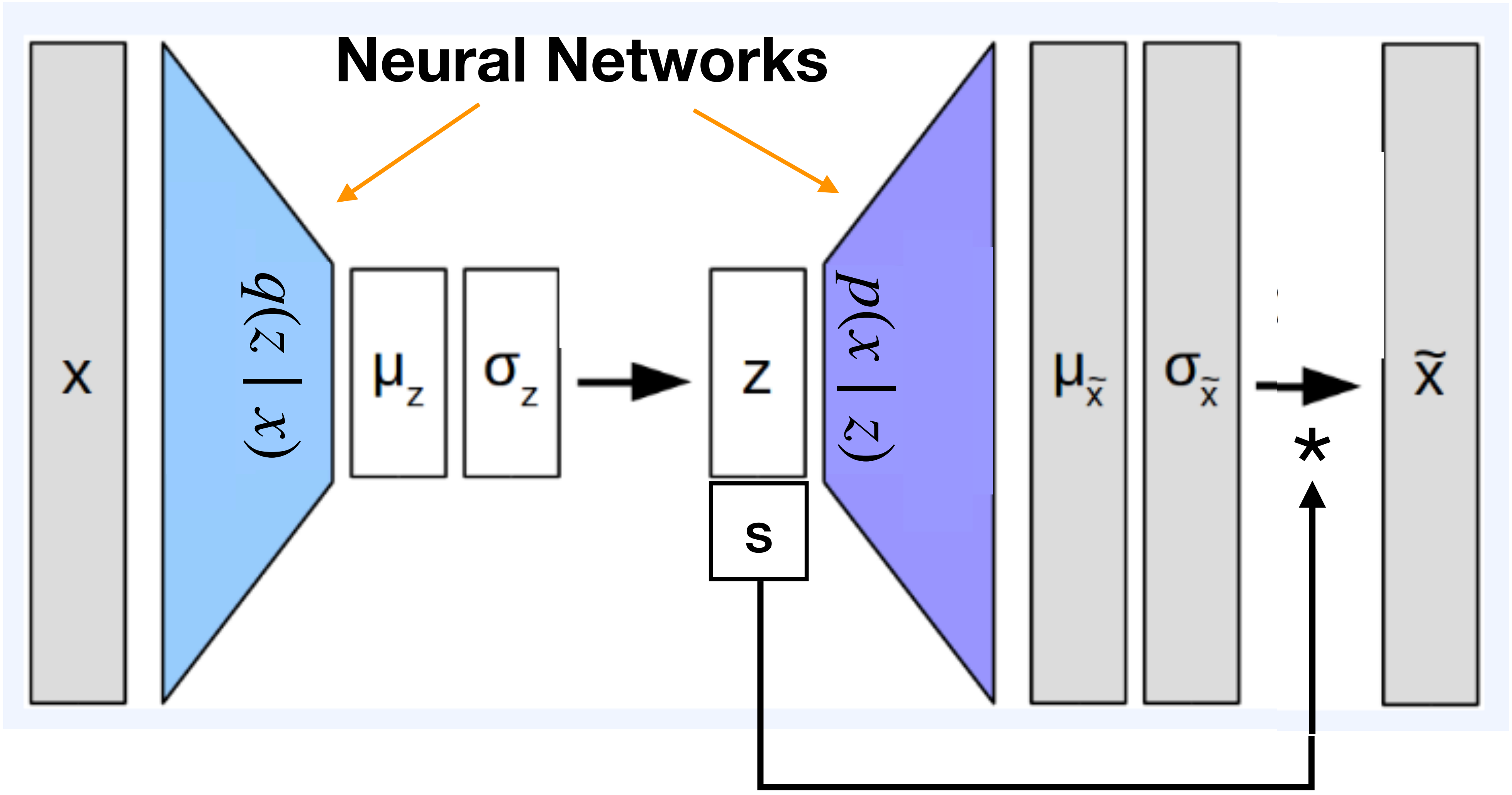
Visual cortex

Generative model

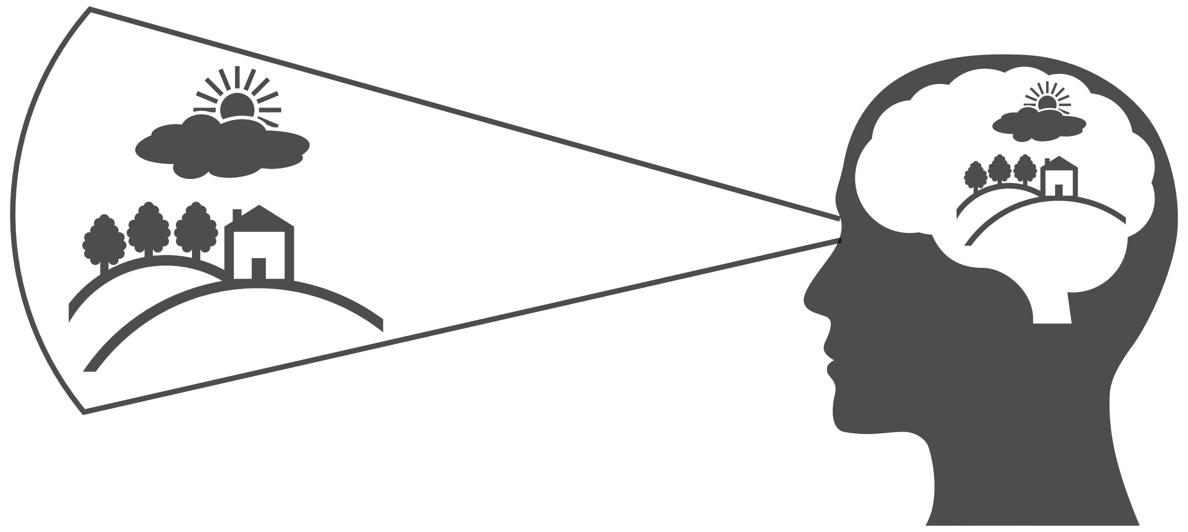
Neuronal response

Latent inference

V1 VAE



Modeling vision with latent generative model



Learning objective:
Learning regularities of natural images

$$p_0(\boldsymbol{x}) = \int p(\boldsymbol{x} \mid \boldsymbol{z}) p_0(\boldsymbol{z})\, d\boldsymbol{z}$$

Likelihood **Prior**

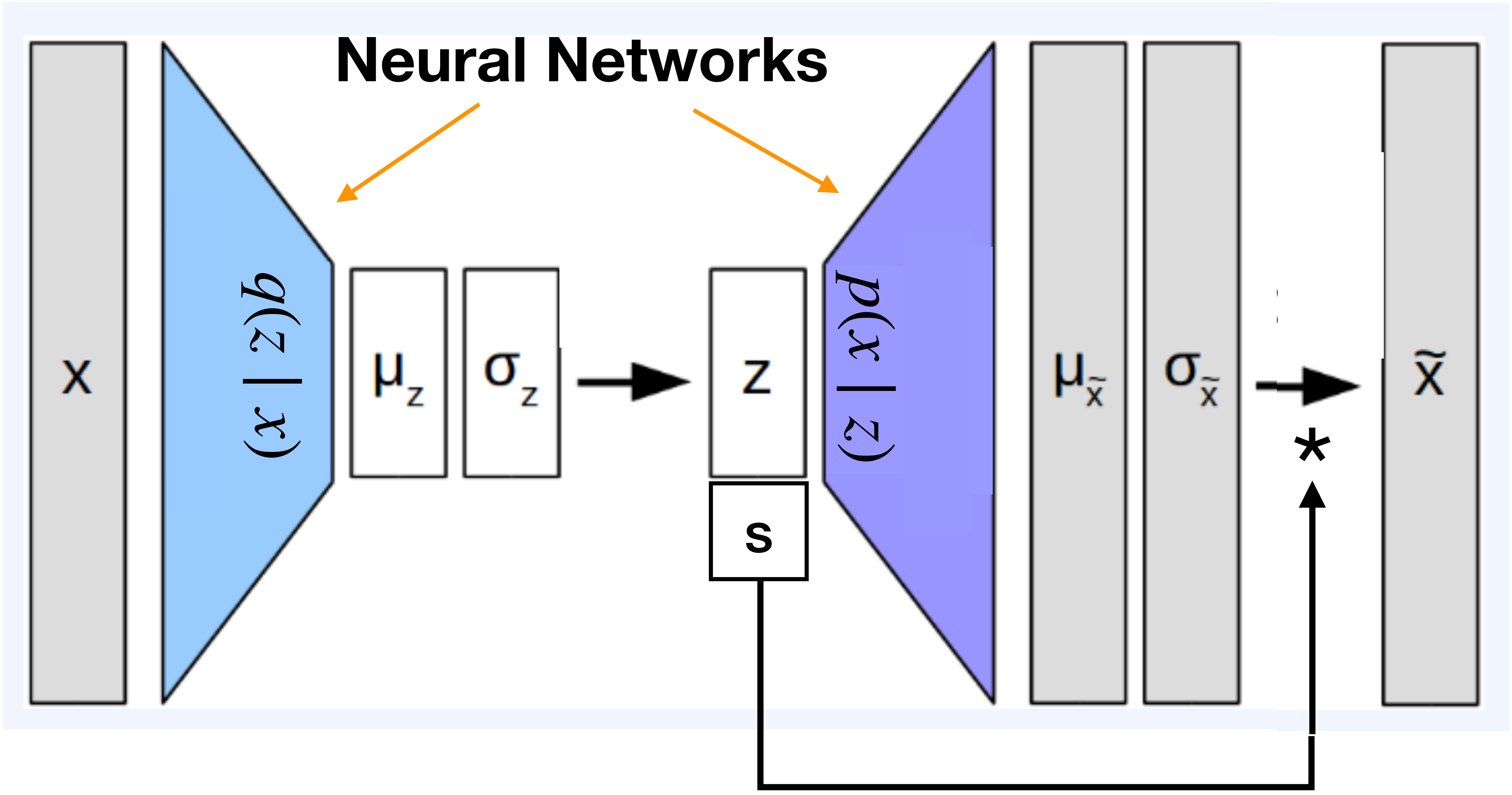
Visual cortex

Generative model

Neuronal response

Latent inference

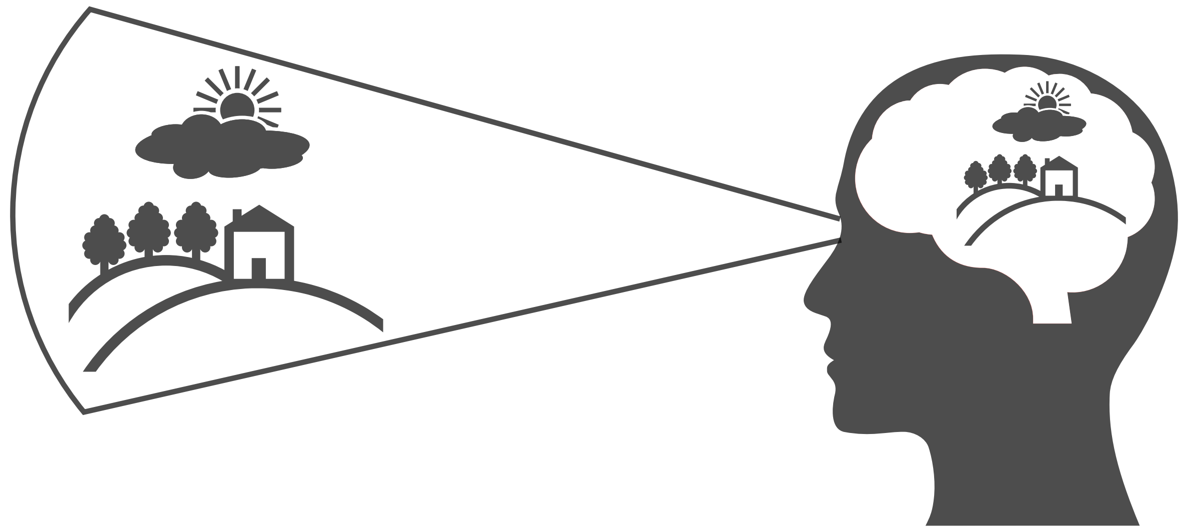
V1 VAE



natural images

Training

Modeling vision with latent generative model

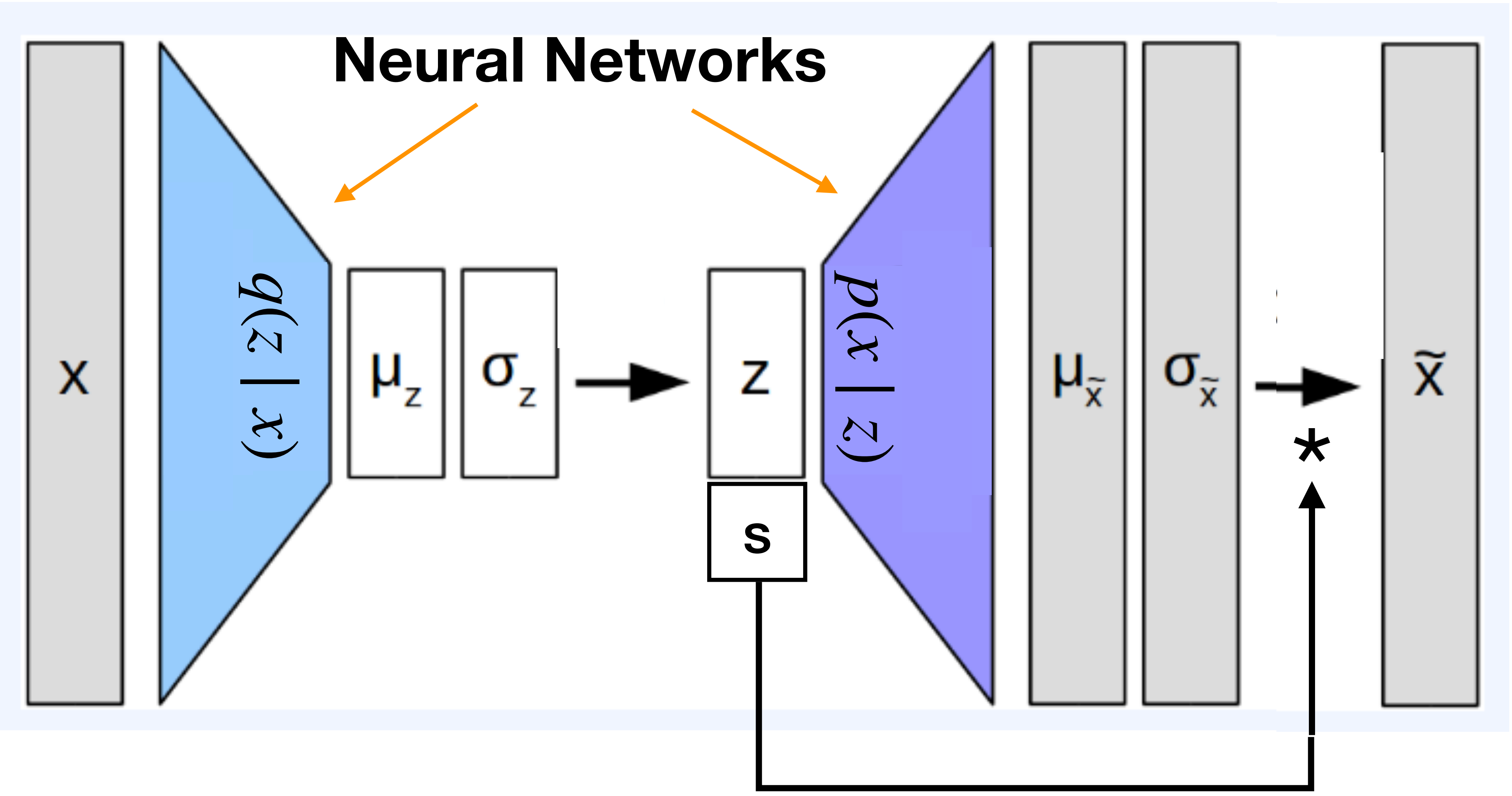


Learning objective:
Learning regularities of natural images

$$p_0(\boldsymbol{x}) = \int p(\boldsymbol{x} \mid \boldsymbol{z})p_0(\boldsymbol{z})\,d\boldsymbol{z},$$

Likelihood **Prior**

V1 VAE



Visual cortex

Generative model

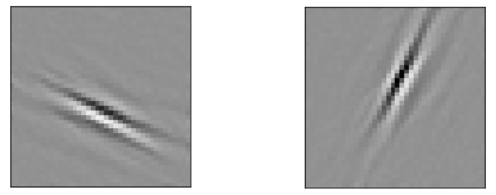
Neuronal response

Latent inference

natural images

Training

$q(z \mid x)$



Variational posterior

Neuronal response

Task adaptation in generative models

$p_0(\boldsymbol{x})$

Task adaptation in generative models

$p_0(\boldsymbol{x})$

Task adaptation in generative models

$p_0(\boldsymbol{x})$

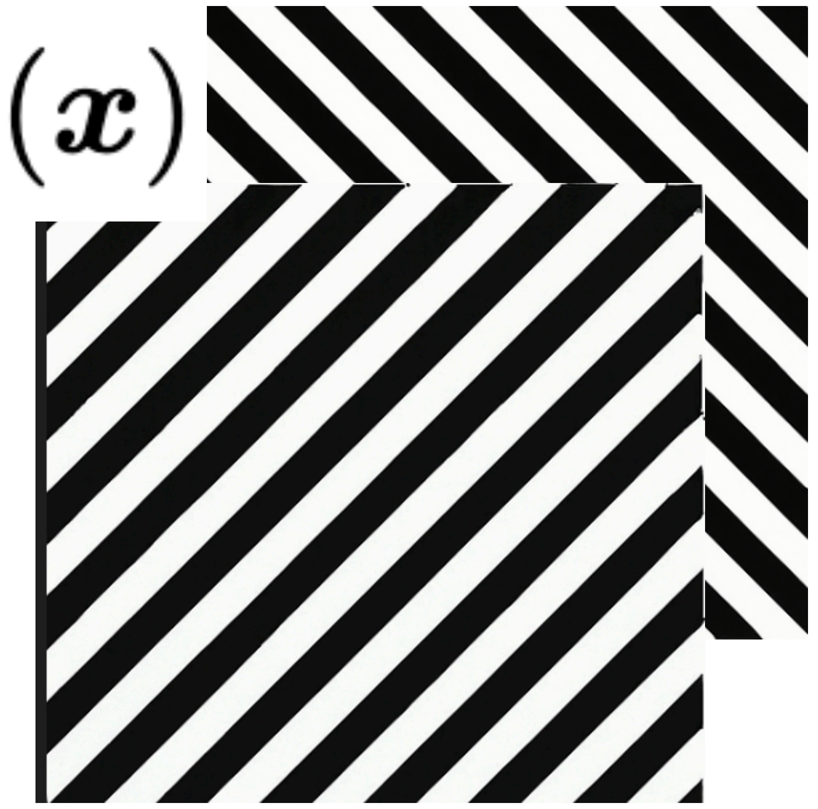
How do we incorporate task adaptation?

Task adaptation in generative models

$p_0(\boldsymbol{x})$

Change in stimulus statistics

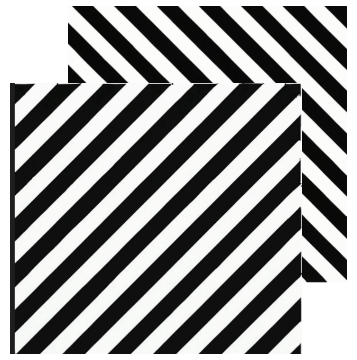
$p_T(\boldsymbol{x})$



How do we incorporate task adaptation?

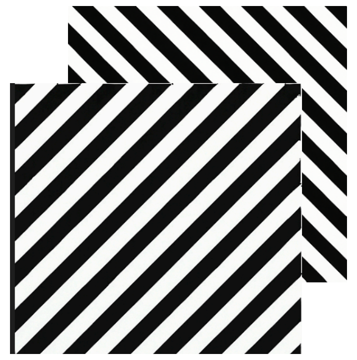
Task adaptation in generative models

Brute force: train new VAE for the new image set



Task adaptation in generative models

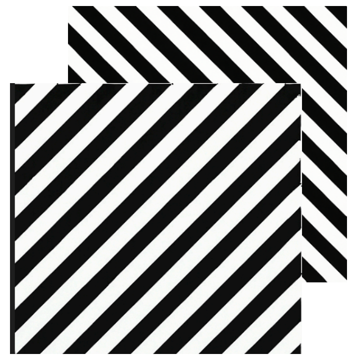
Brute force: train new VAE for the new image set



Not enough data

Task adaptation in generative models

Brute force: train new VAE for the new image set

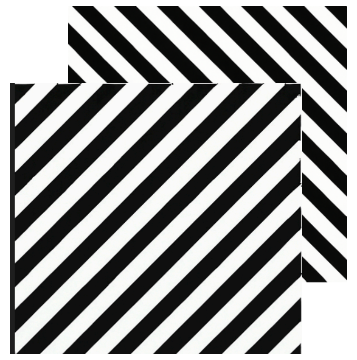


Not enough data ❌

Tune away important representation ❌

Task adaptation in generative models

Brute force: train new VAE for the new image set



Not enough data

Tune away important representation

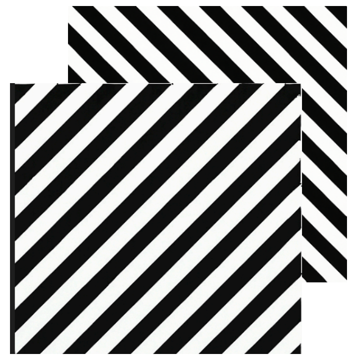
Instead: computational efficient, biologically plausible solution

Retain likelihood

$$p_T(\boldsymbol{x}) = \int p(\boldsymbol{x} \mid \boldsymbol{z}) p_T(\boldsymbol{z})\, d\boldsymbol{z}$$

Task adaptation in generative models

Brute force: train new VAE for the new image set



Not enough data

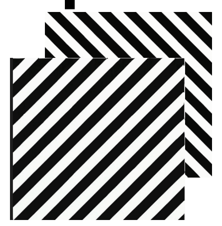
Tune away important representation

Instead: computational efficient, biologically plausible solution

Retain likelihood

$$p_T(\boldsymbol{x}) = \int p(\boldsymbol{x} \mid \boldsymbol{z}) \boxed{p_T(\boldsymbol{z})} d\boldsymbol{z}$$

Cheap optimization



Task adaptation in generative models

Brute force: train new VAE for the new image set

Not enough data

Tune away important representation

Instead: computational efficient, biologically plausible solution

Retain likelihood

$$p_T(\boldsymbol{x}) = \int p(\boldsymbol{x} \mid \boldsymbol{z}) \boxed{p_T(\boldsymbol{z})}\, d\boldsymbol{z}$$

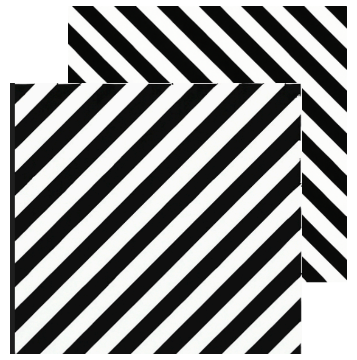
Cheap optimization

$$\boxed{q_{\mathrm{T}}(\boldsymbol{z}|\boldsymbol{x}) \propto \frac{q(\boldsymbol{z}|\boldsymbol{x})\, p_{\mathrm{T}}(\boldsymbol{z})}{p_0(\boldsymbol{z})}}$$

Inference in task

Task adaptation in generative models

Brute force: train new VAE for the new image set



Not enough data

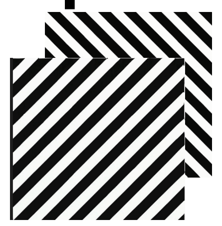
Tune away important representation

Instead: computational efficient, biologically plausible solution

Retain likelihood

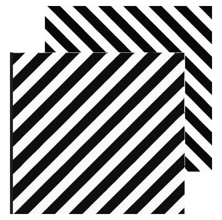
$$p_T(\boldsymbol{x}) = \int p(\boldsymbol{x} \mid \boldsymbol{z}) \boxed{p_T(\boldsymbol{z})} dz$$

Cheap optimization



$$\boxed{q_{\mathrm{T}}(\boldsymbol{z}|\boldsymbol{x}) \propto \frac{\boxed{q(\boldsymbol{z}|\boldsymbol{x})}\, p_{\mathrm{T}}(\boldsymbol{z})}{p_0(\boldsymbol{z})}}$$

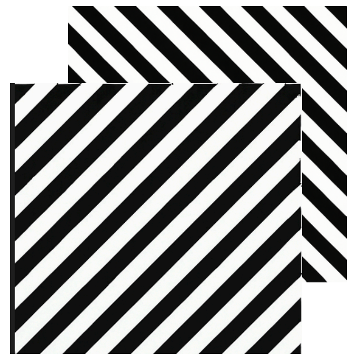
Inference in task



Context-free inference

Task adaptation in generative models

Brute force: train new VAE for the new image set



Not enough data

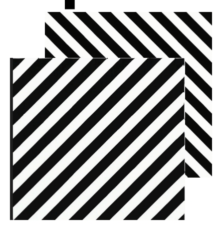
Tune away important representation

Instead: computational efficient, biologically plausible solution

Retain likelihood

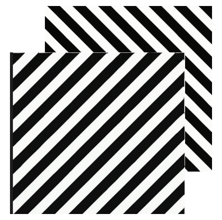
$$p_T(\boldsymbol{x}) = \int p(\boldsymbol{x} \mid \boldsymbol{z}) \boxed{p_T(\boldsymbol{z})} d\boldsymbol{z}$$

Cheap optimization



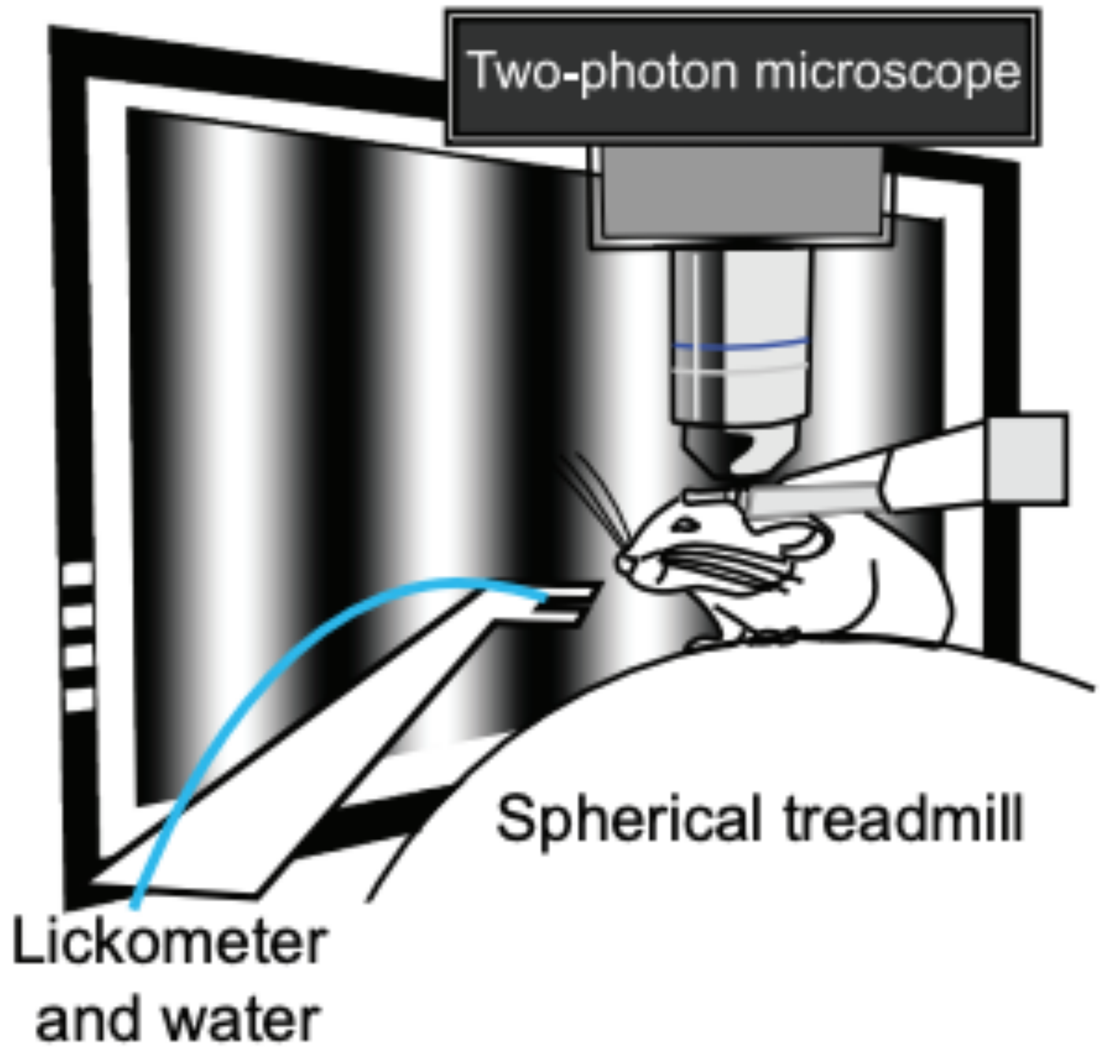
$$\boxed{q_{\mathrm{T}}(\boldsymbol{z}|\boldsymbol{x}) \propto \frac{\boxed{q(\boldsymbol{z}|\boldsymbol{x})}\boxed{p_{\mathrm{T}}(\boldsymbol{z})}}{p_0(\boldsymbol{z})}}$$

Inference in task

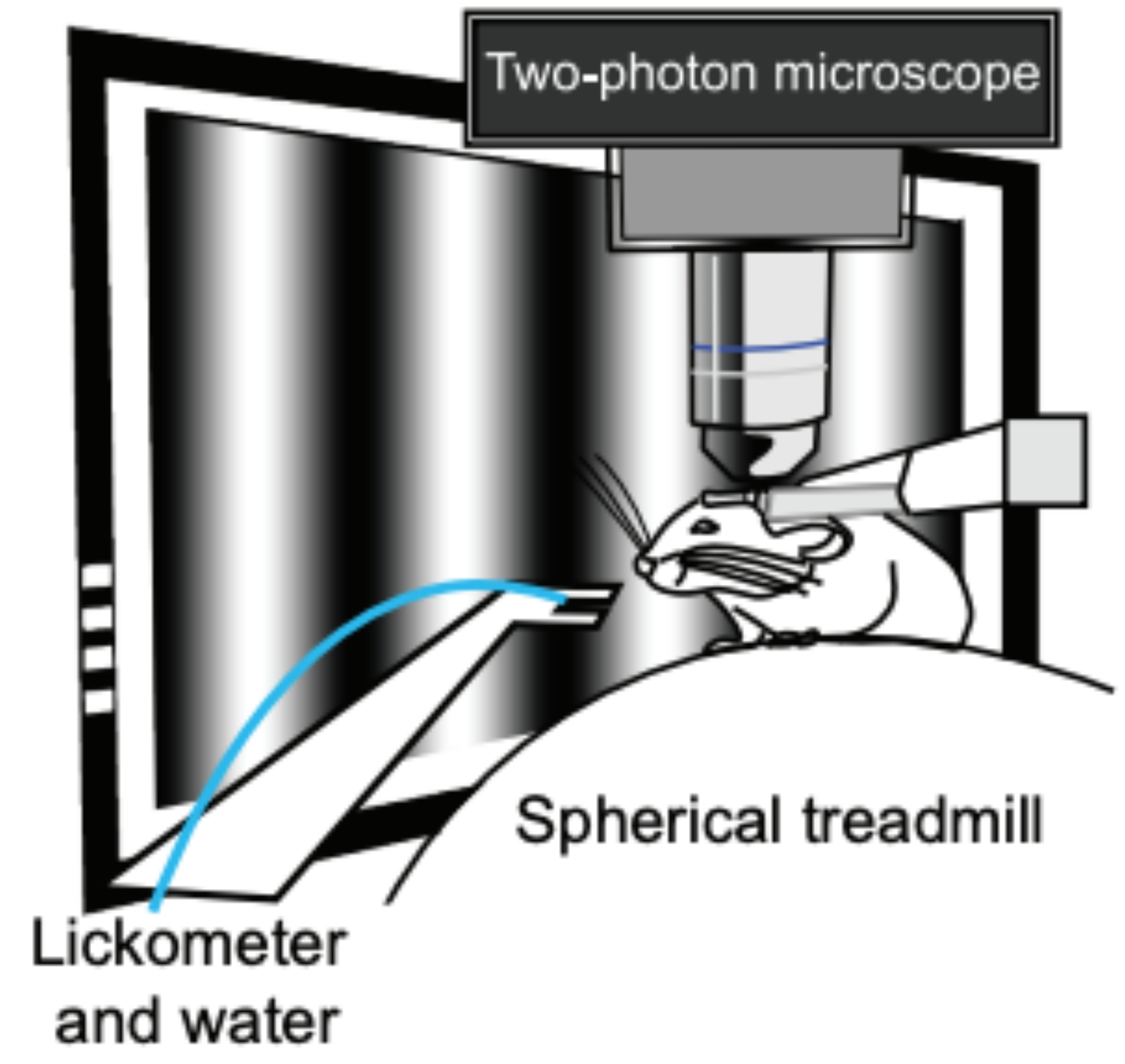


Context-free inference

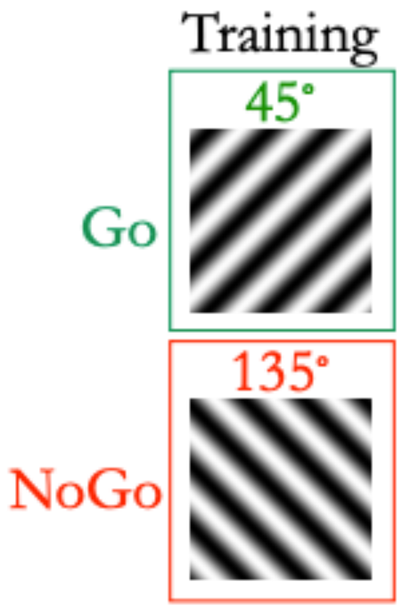
Experimental setup



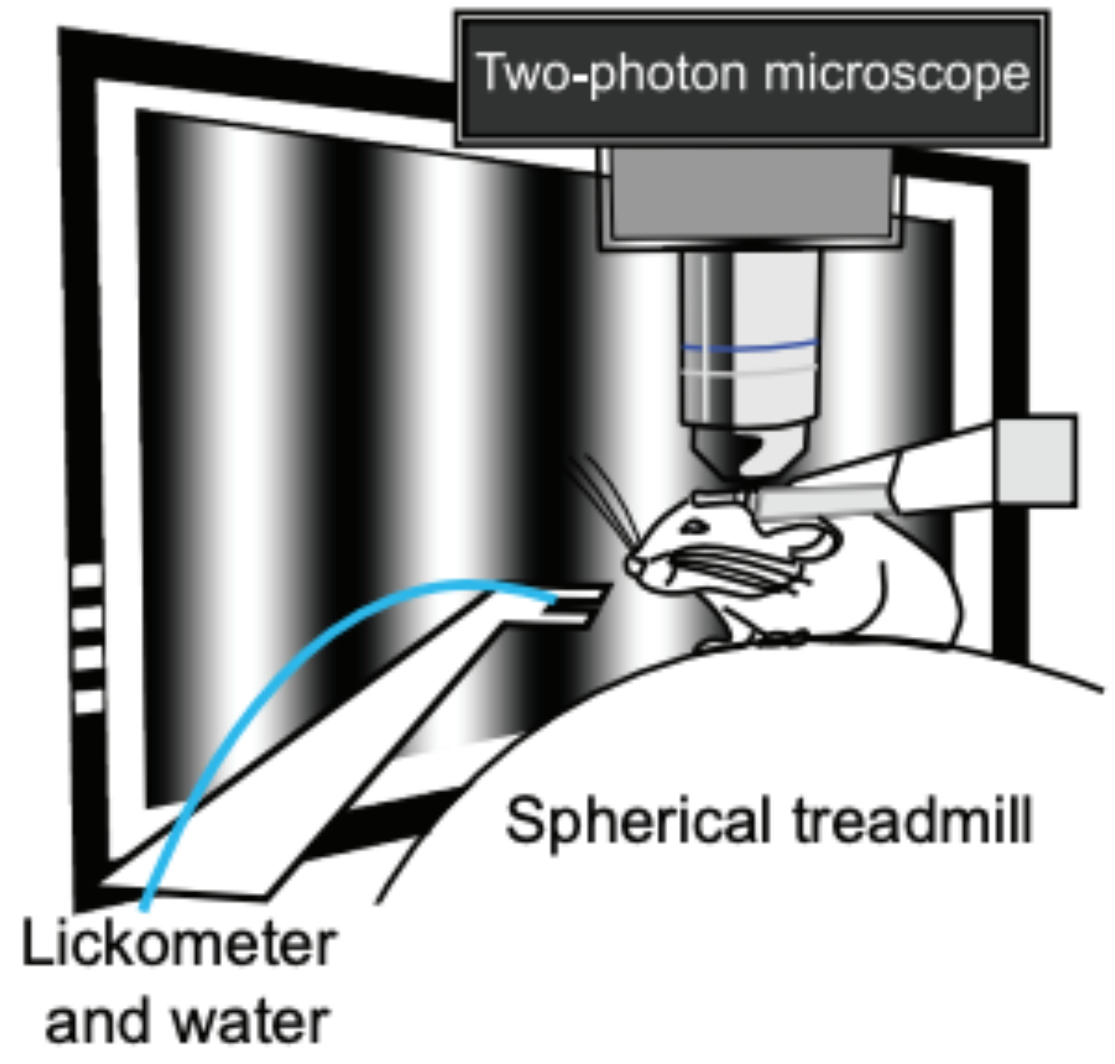
Experimental setup



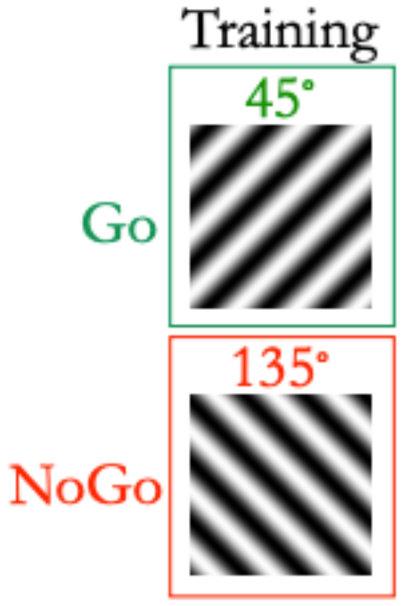
9-18 days



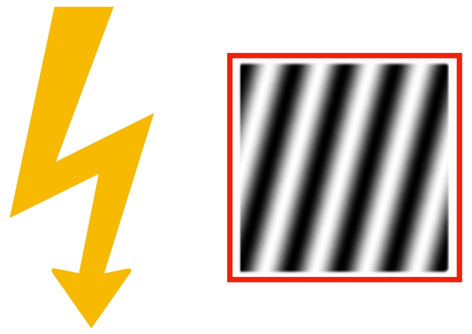
Experimental setup



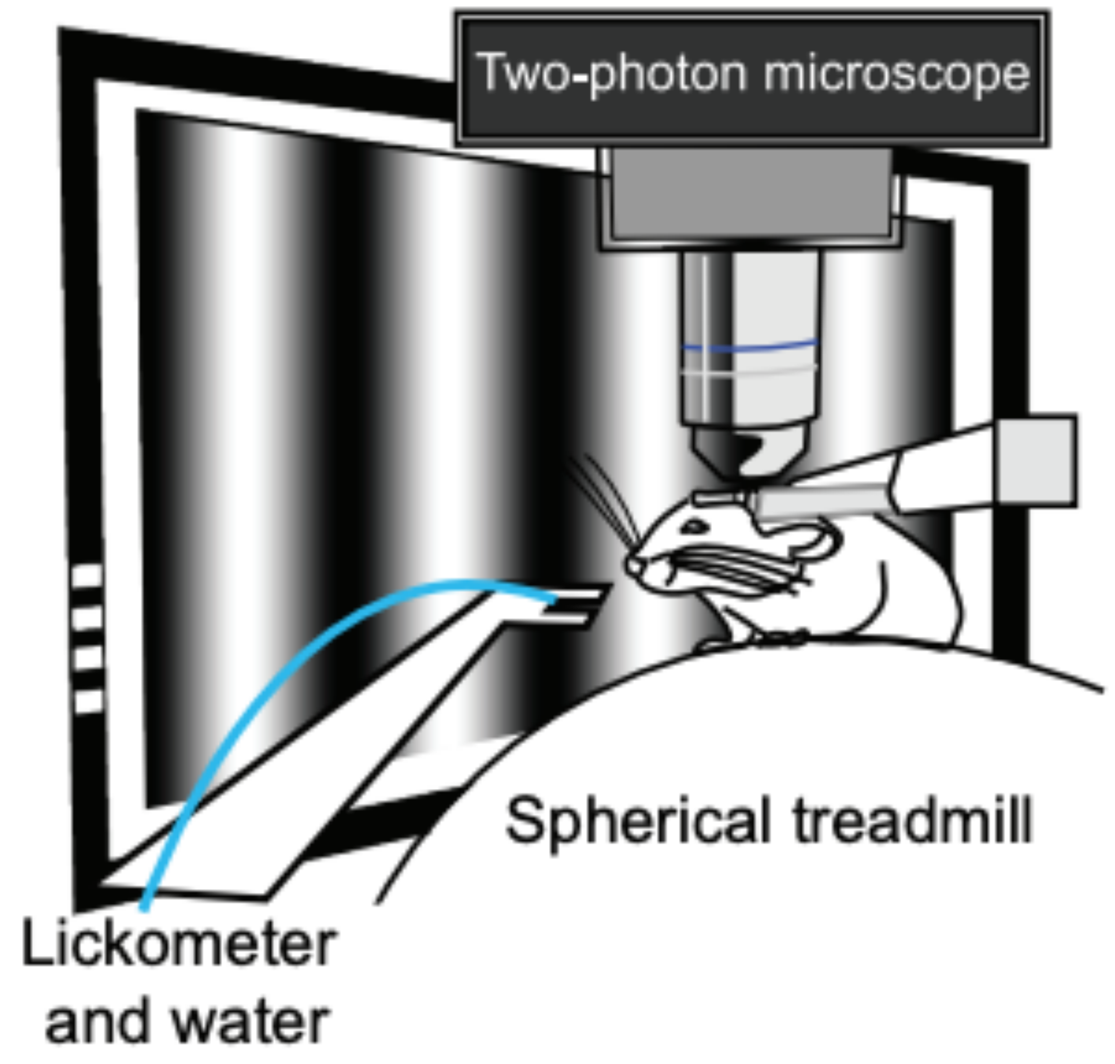
9-18 days



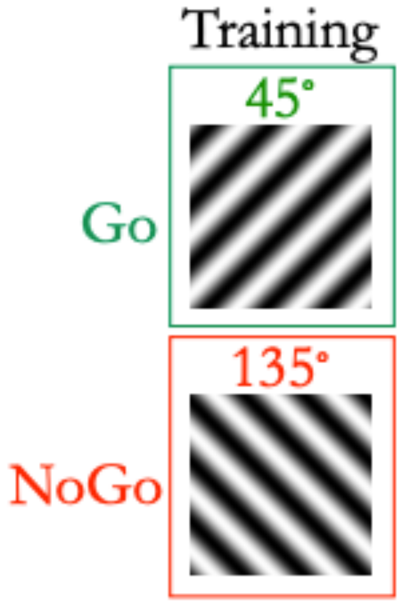
Test days



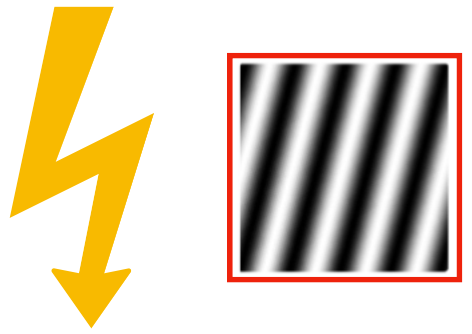
Experimental setup

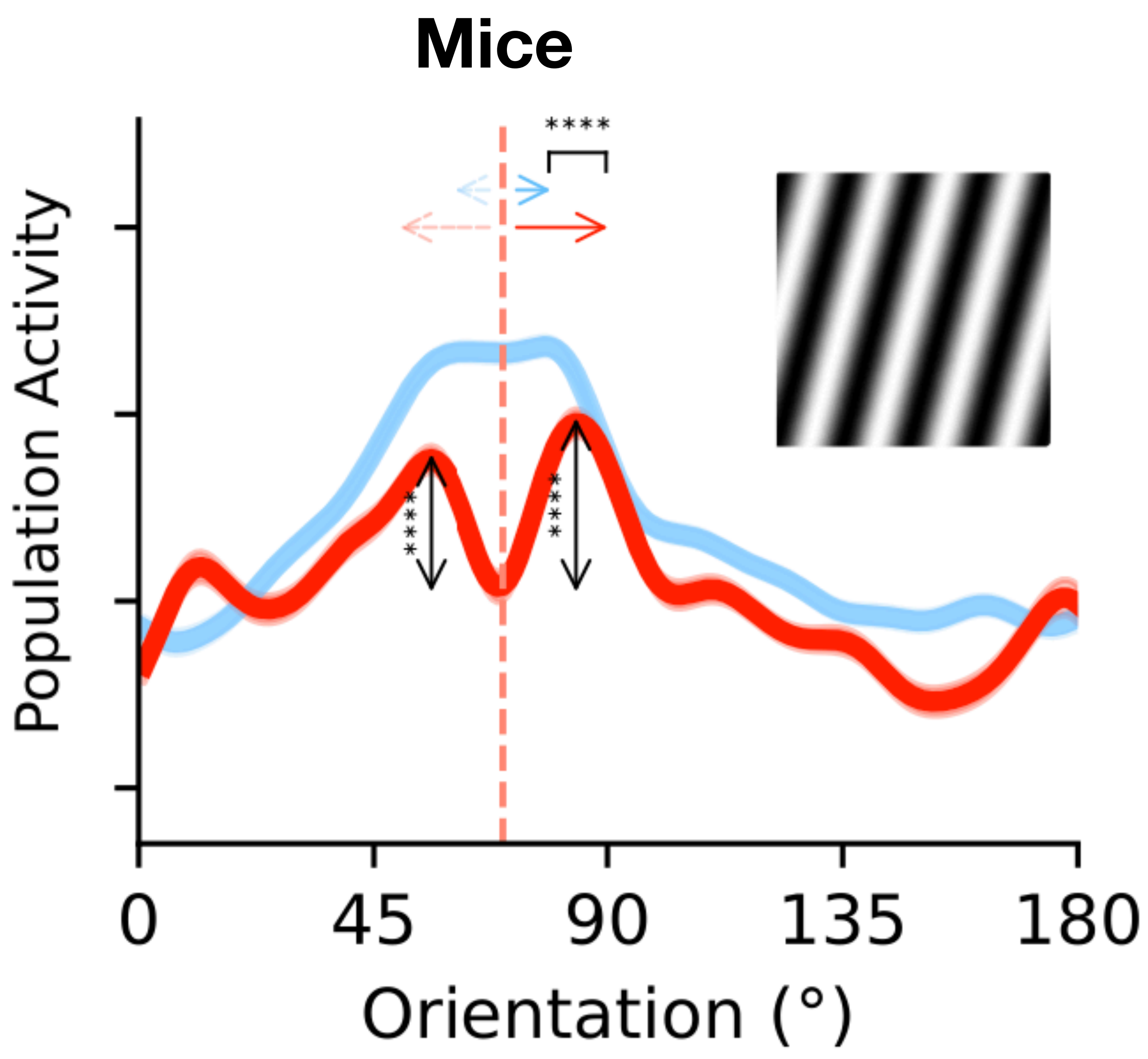


9-18 days

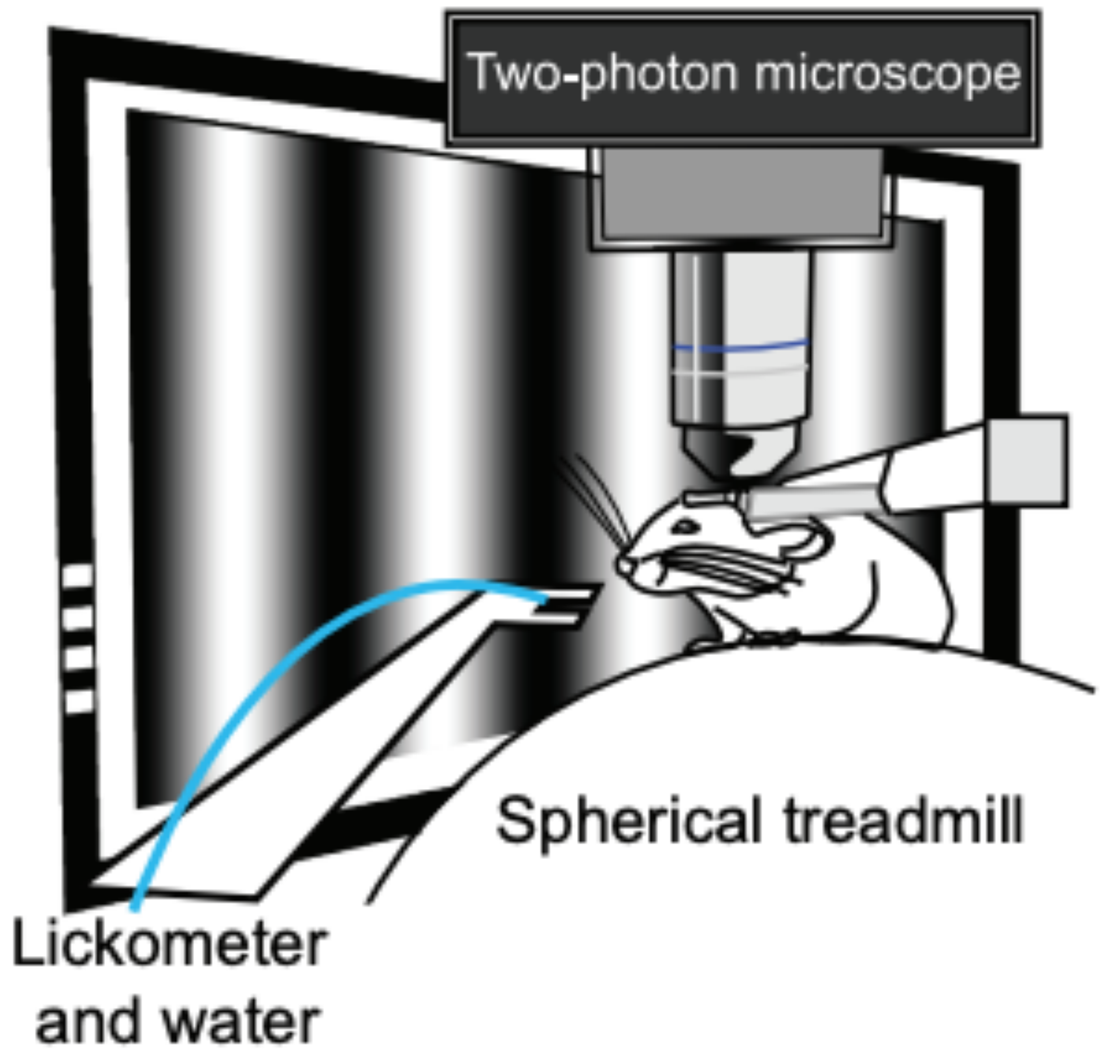


Test days

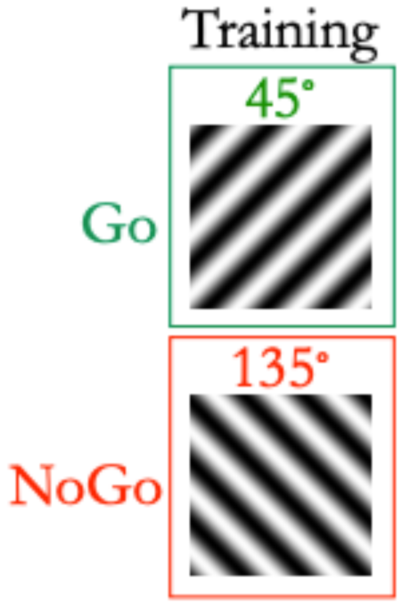




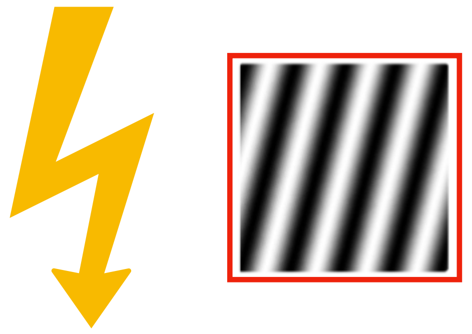
Experimental setup

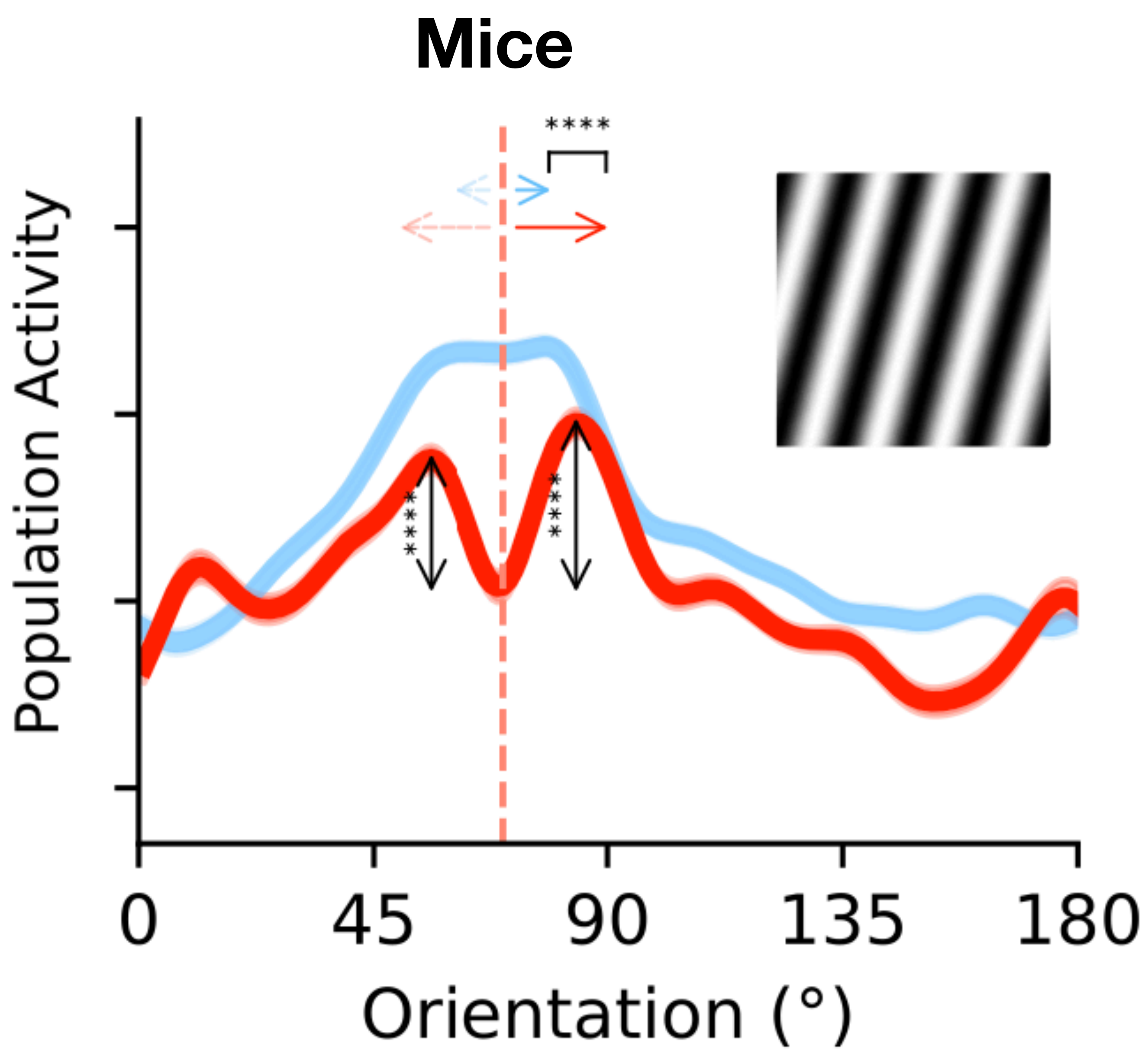


9-18 days

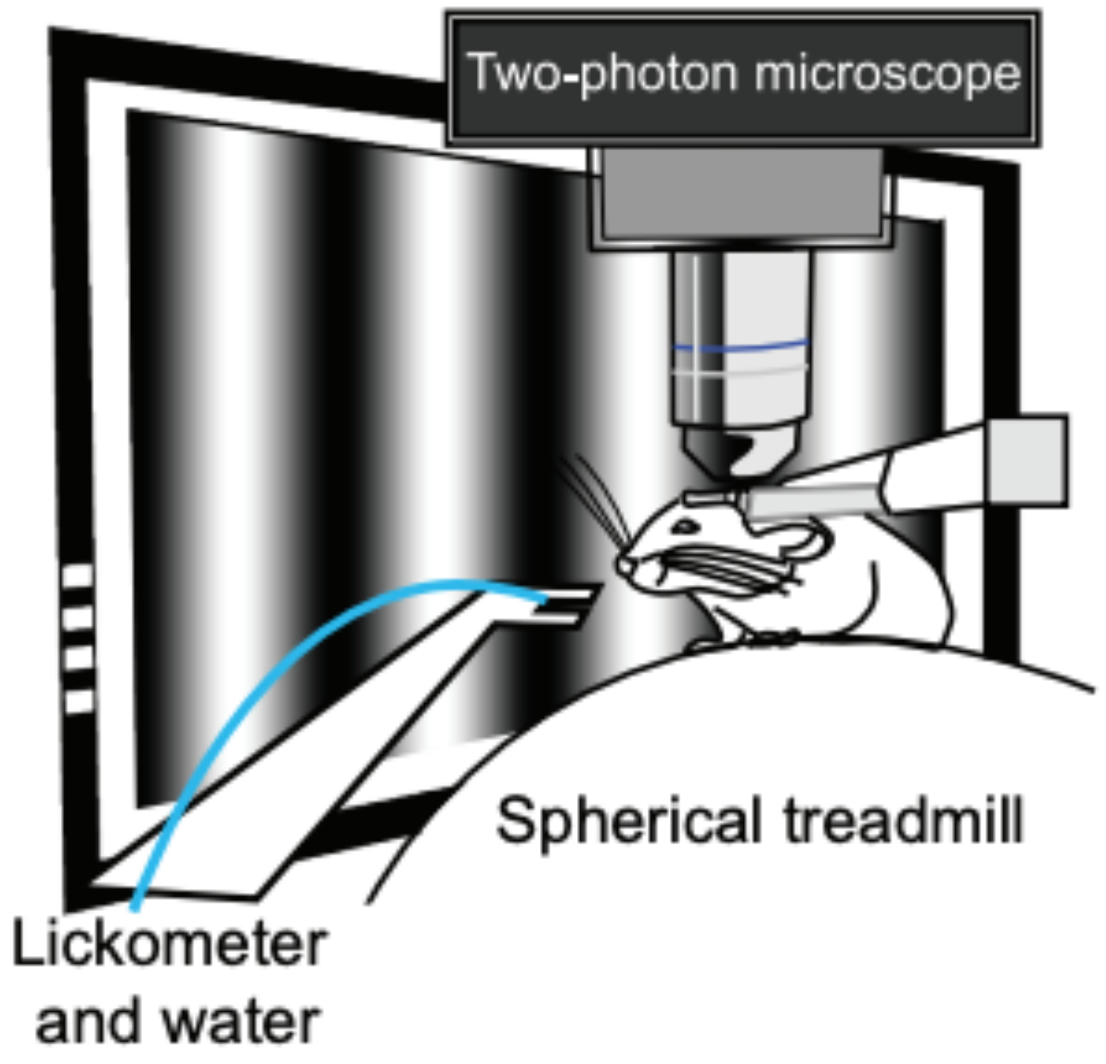


Test days

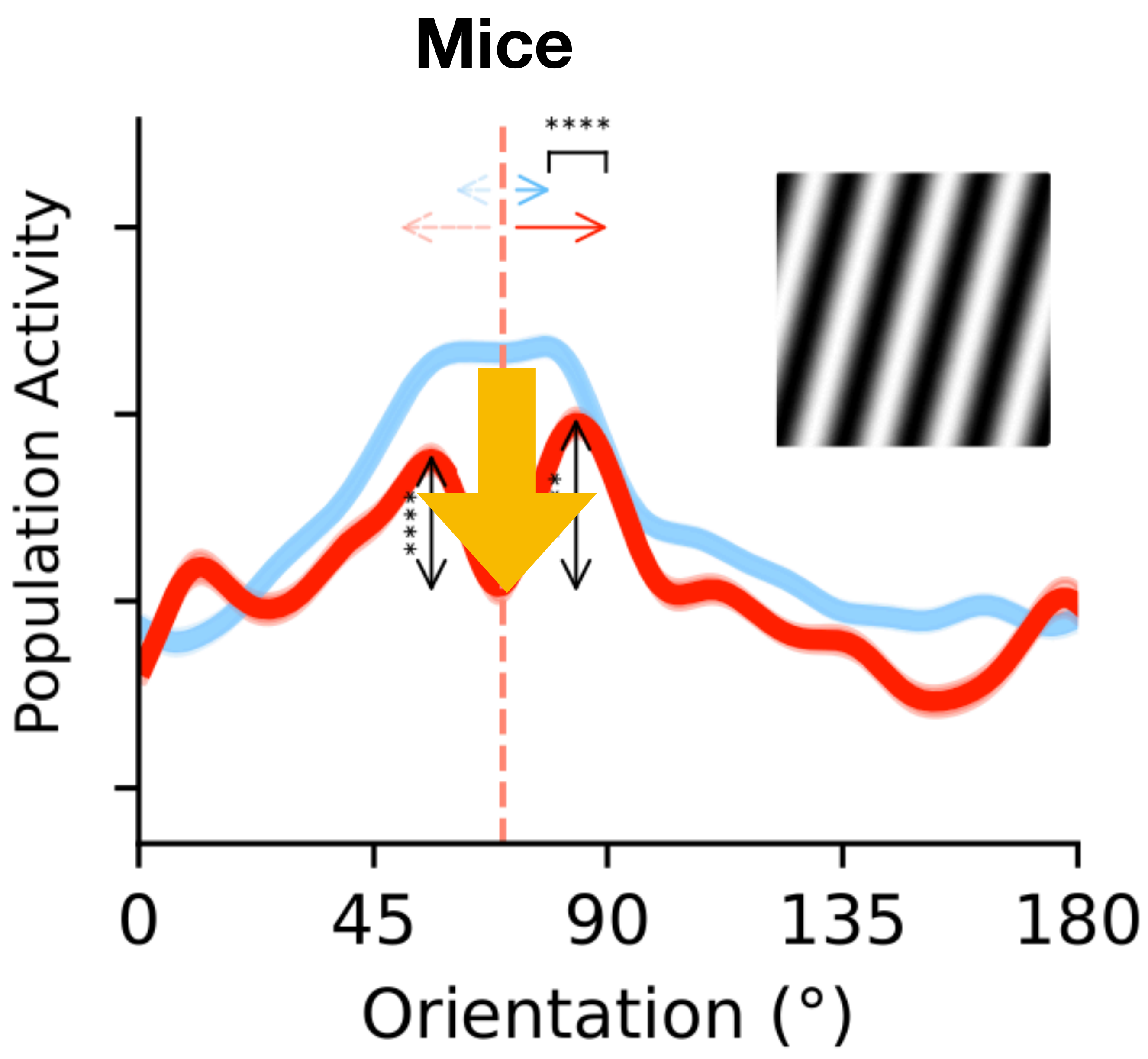




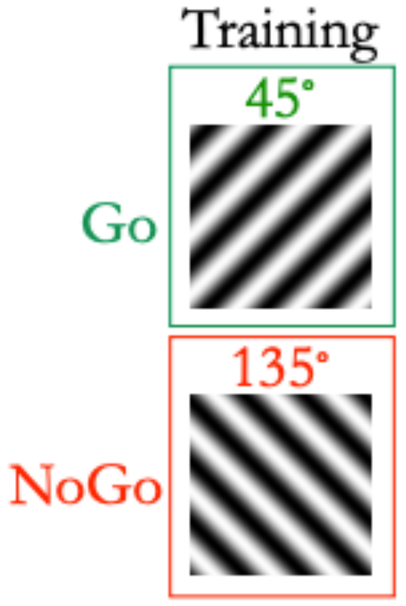
Experimental setup



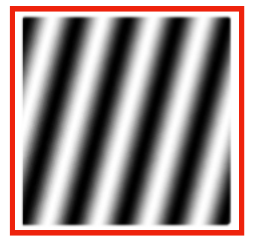
Bimodal response profile



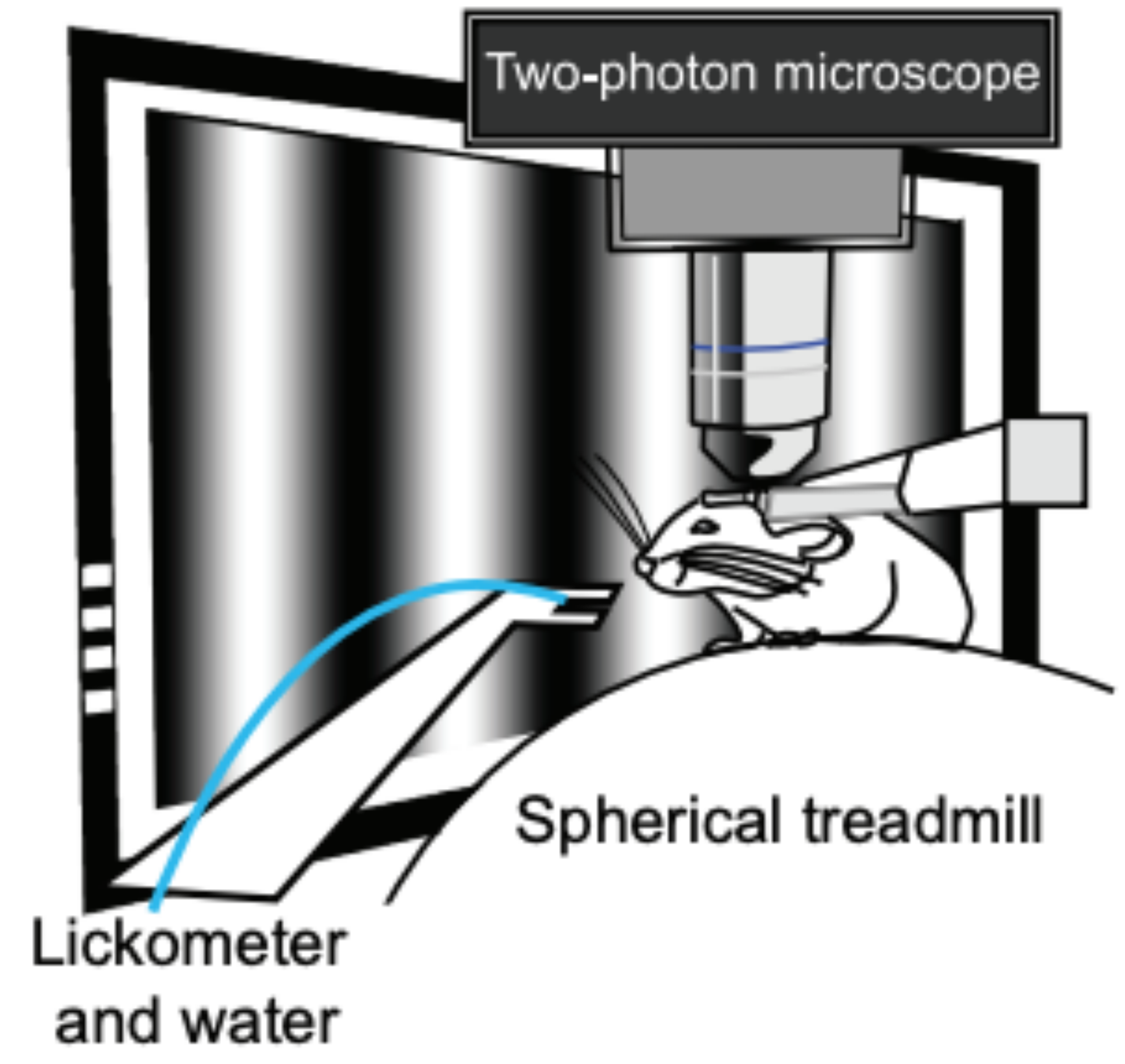
9-18 days



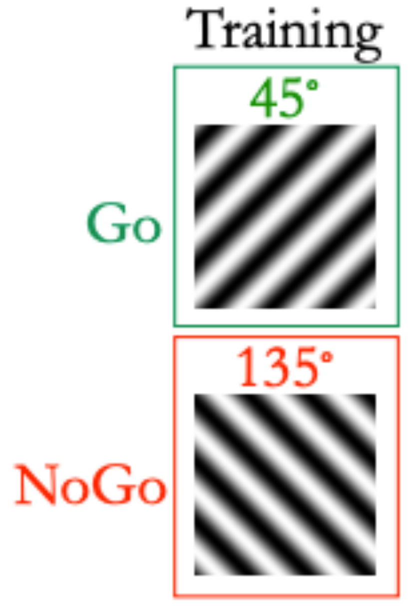
Test days



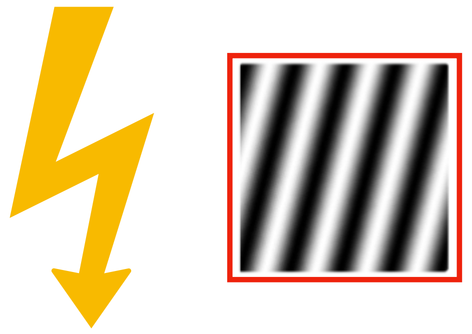
Experimental setup



9-18 days

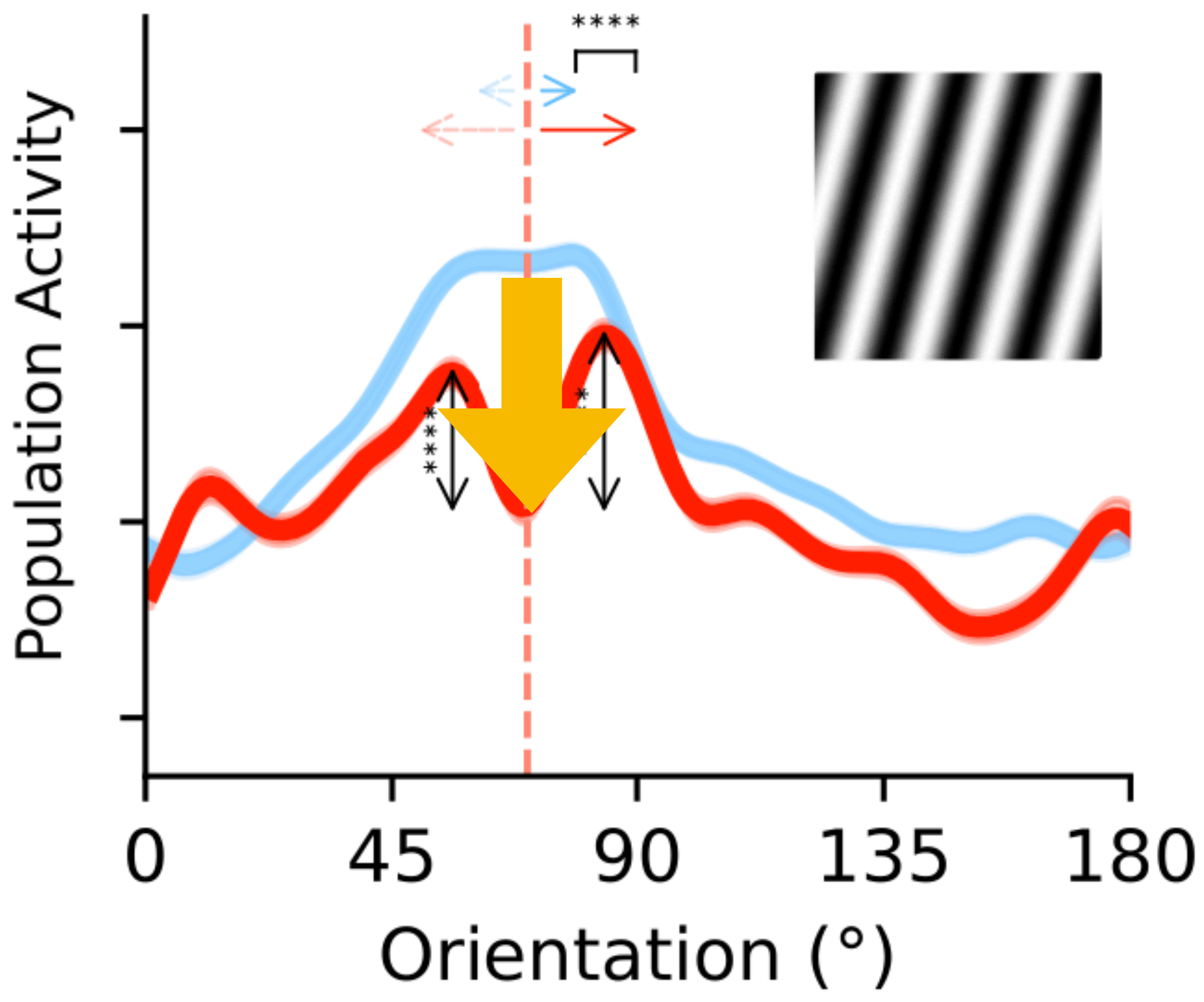


Test days



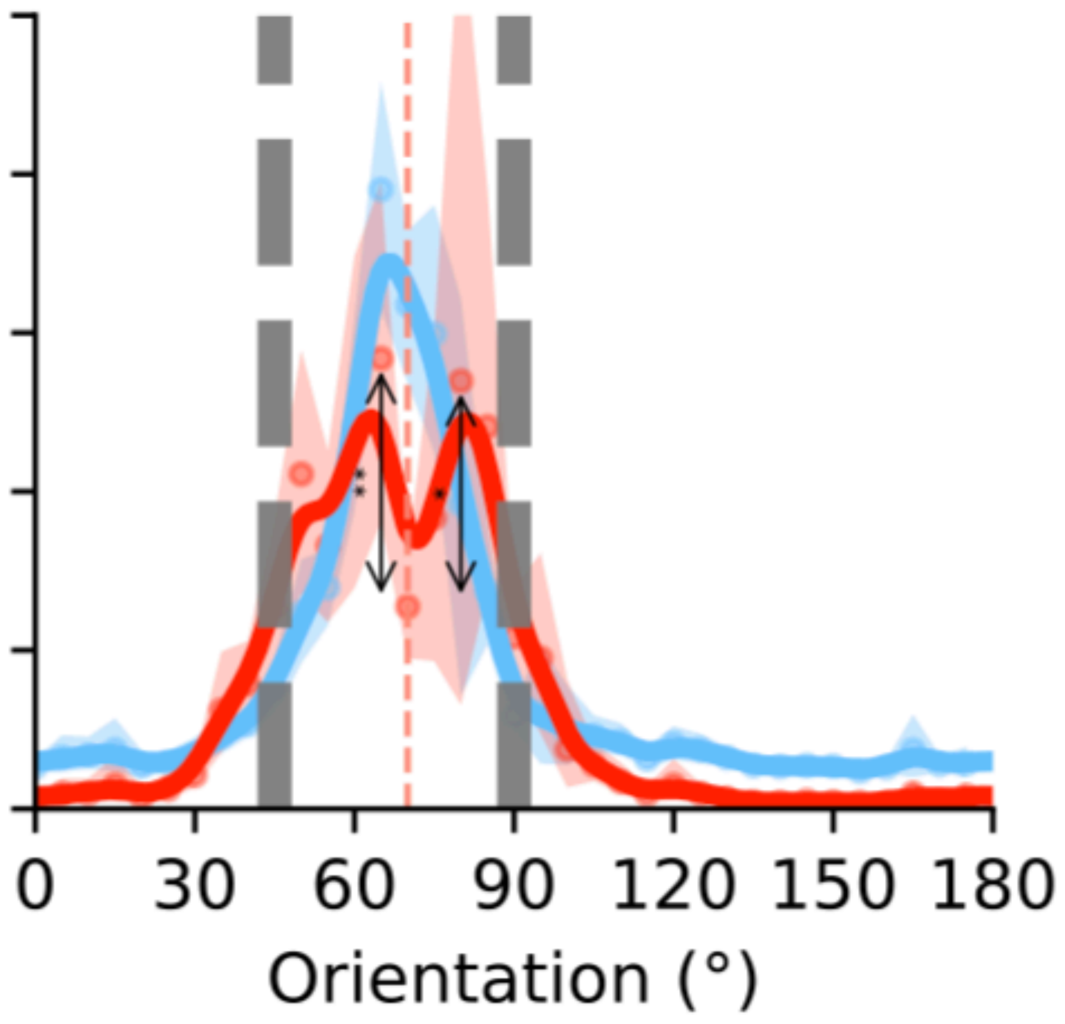
Bimodal response profile

Mice



Model explains bimodality

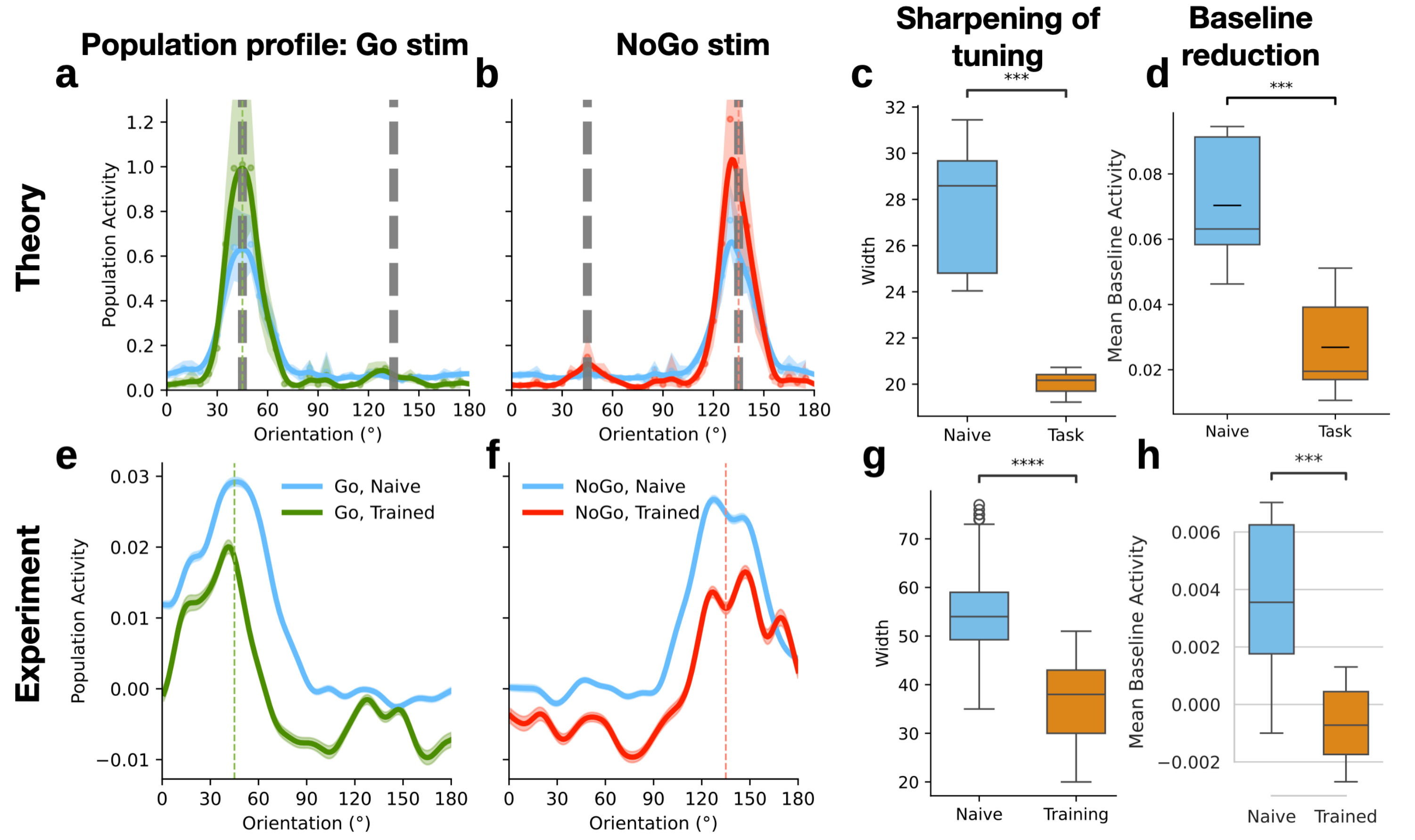
TAVAE



Detailed results in the paper!

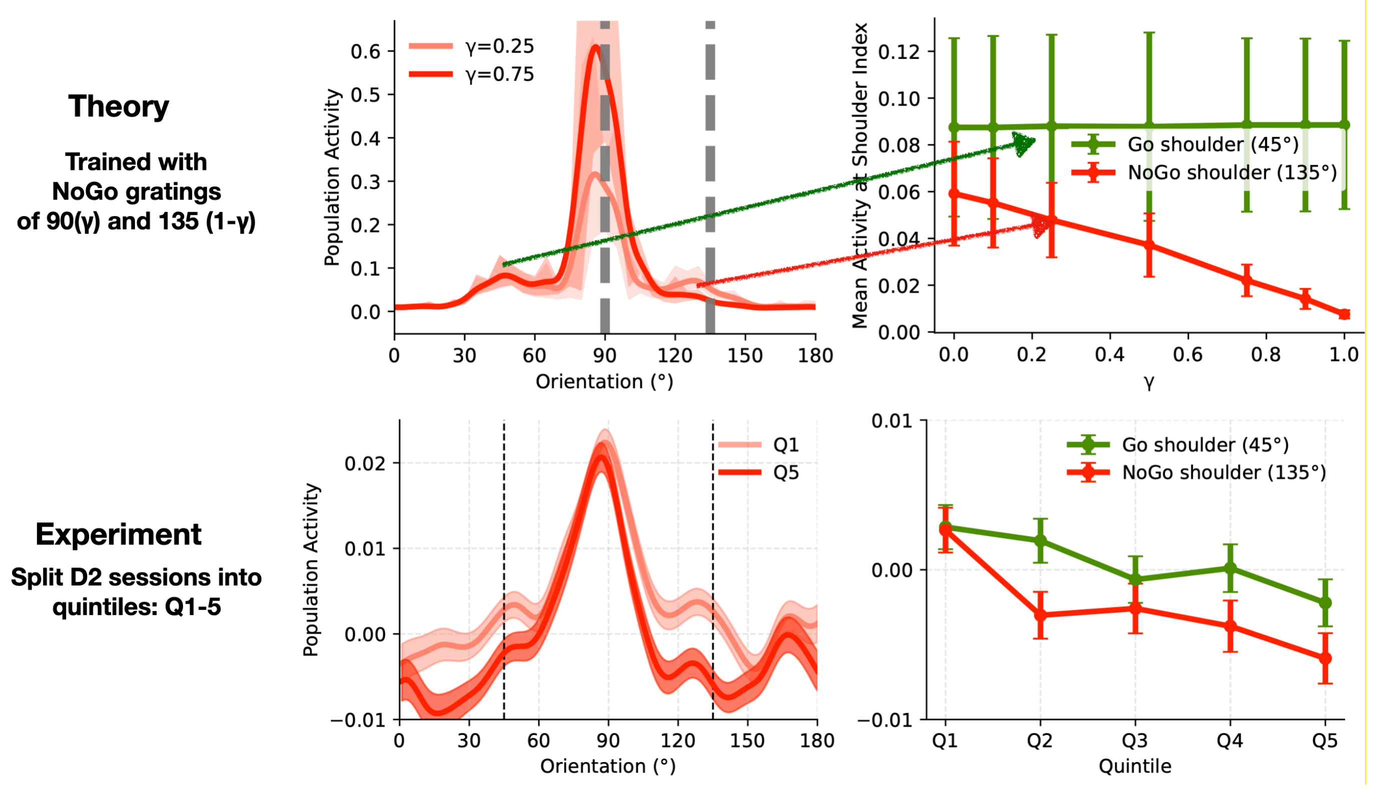
D1: task congruent stimuli

Adapting to the discrimination task: acquiring a contextual prior



Gradual update of the prior upon stimulus mismatch (D2)

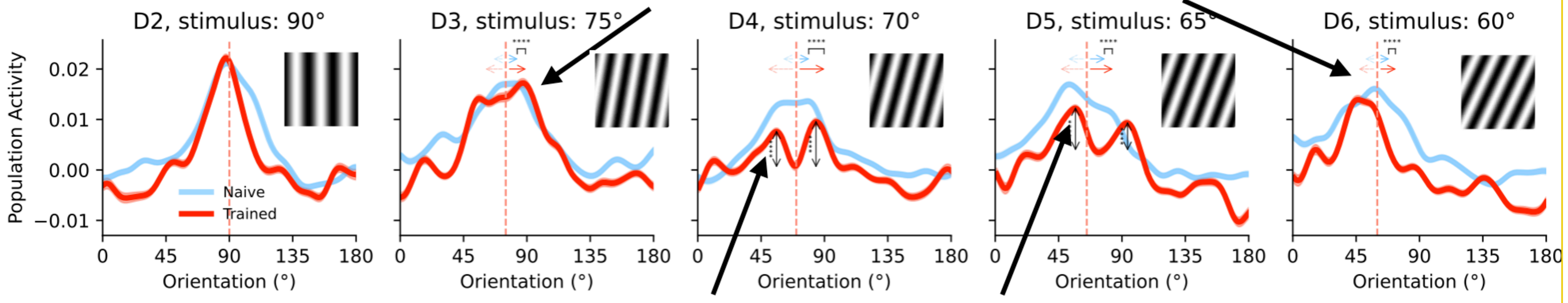
The contextual prior can adapt to clear violations of learned statistics: unfounded hypotheses early in a session are gradually discredited



D2-D6: task-stimulus mismatch

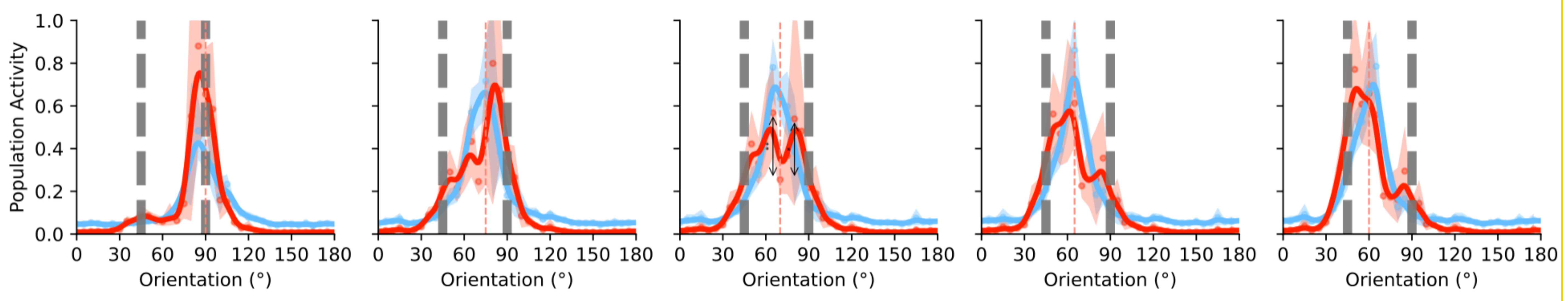
Violation of trained stimulus statistics: biased responses through contextual prior

Peak position shifts away from stimulus



Bimodal responses (minimum at stimulus angle)

- **Partial adapter model: Adaptation to D2 stimuli then retain D2 in later sessions**



Predictive power of alternative accounts

	TAVAE (45, 90)	VAE	TAVAE (45, 135)	TAVAE (EAGER)
$r(\text{TASK})$	$\mathbf{0.78 \pm 0.02}$	0.53 ± 0.12	0.54 ± 0.10	0.53 ± 0.11
$r(\text{TASK–NAIVE})$	$\mathbf{0.58 \pm 0.09}$	–	0.32 ± 0.17	-0.10 ± 0.23