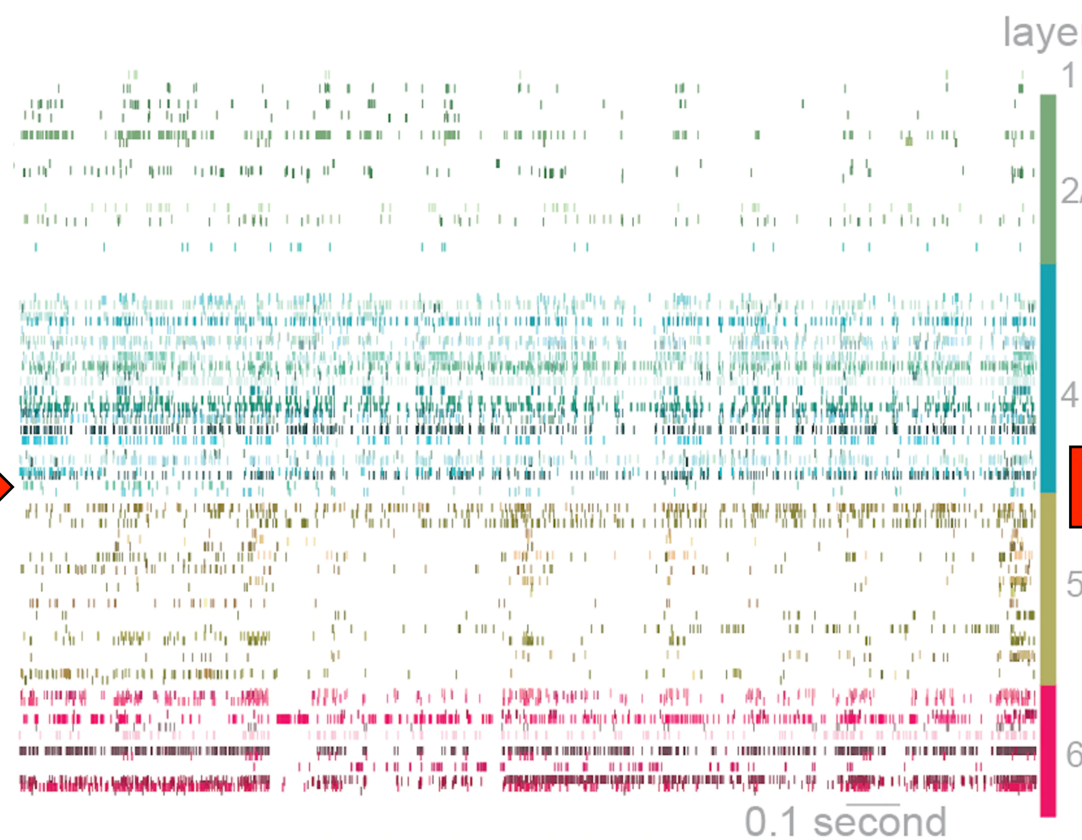
NeuroAI:
How to harness Artificial Intelligence Research for understanding the brain

GERGŐ ORBÁN

computational systems neuroscience lab
Dept Computational Sciences

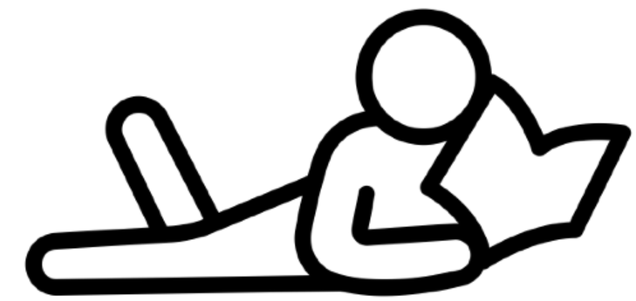
Wigner Research Center for Physics

Reading the neural code

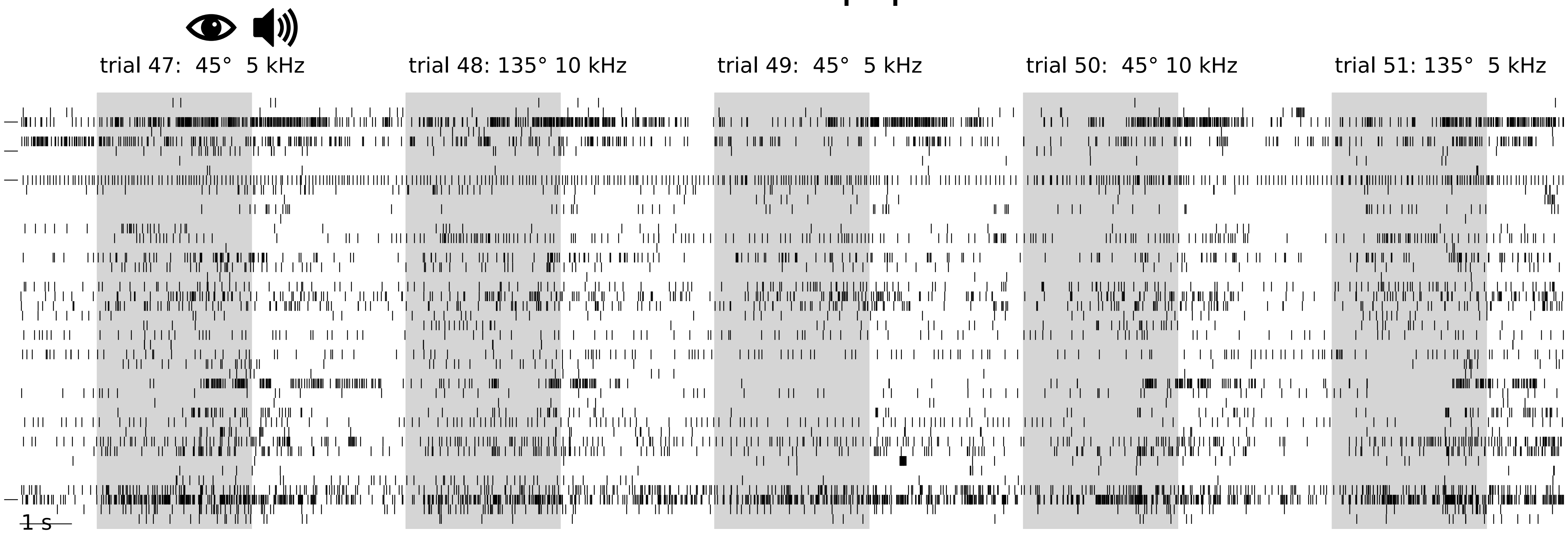


source: Allen Brain Observatory

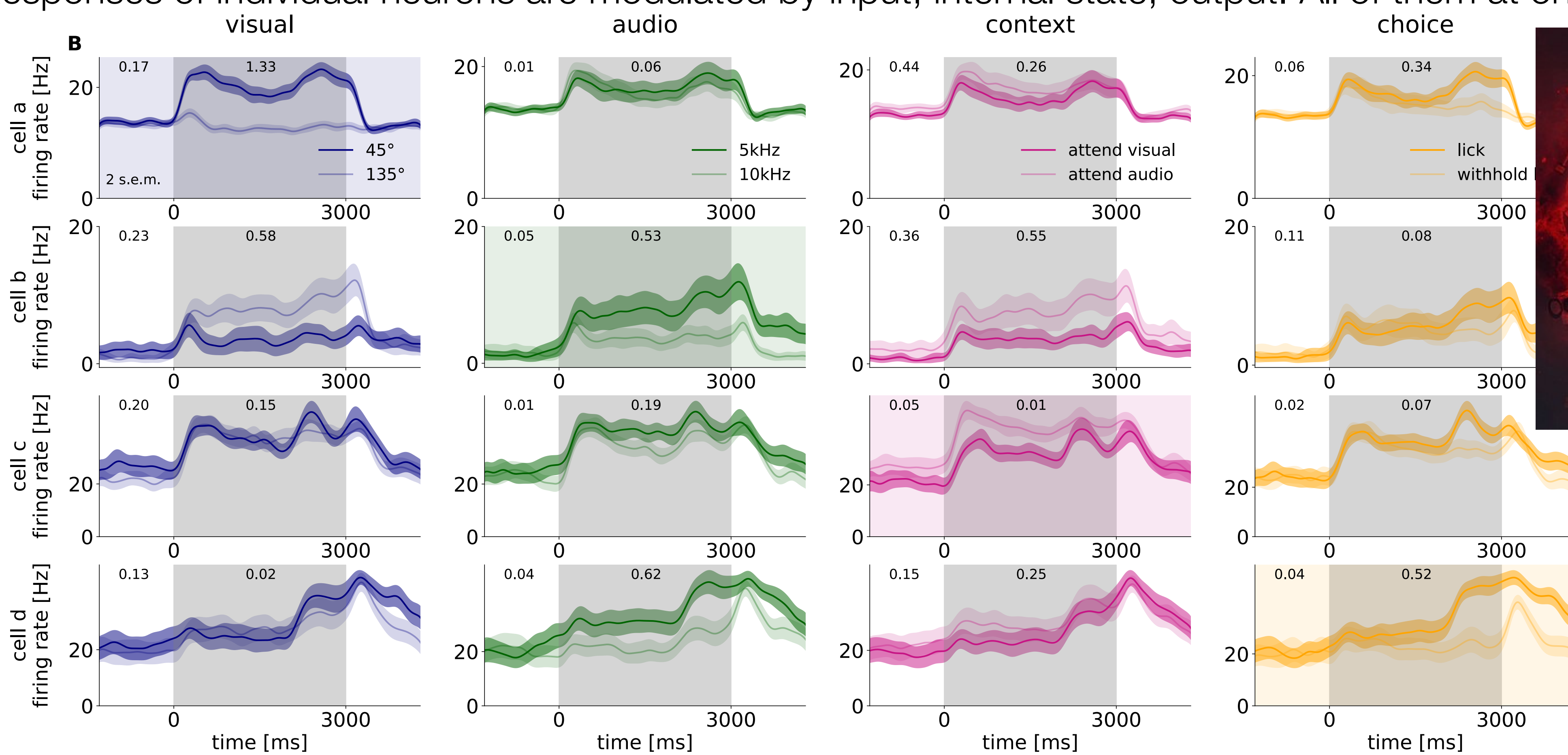
?



Data-driven approaches

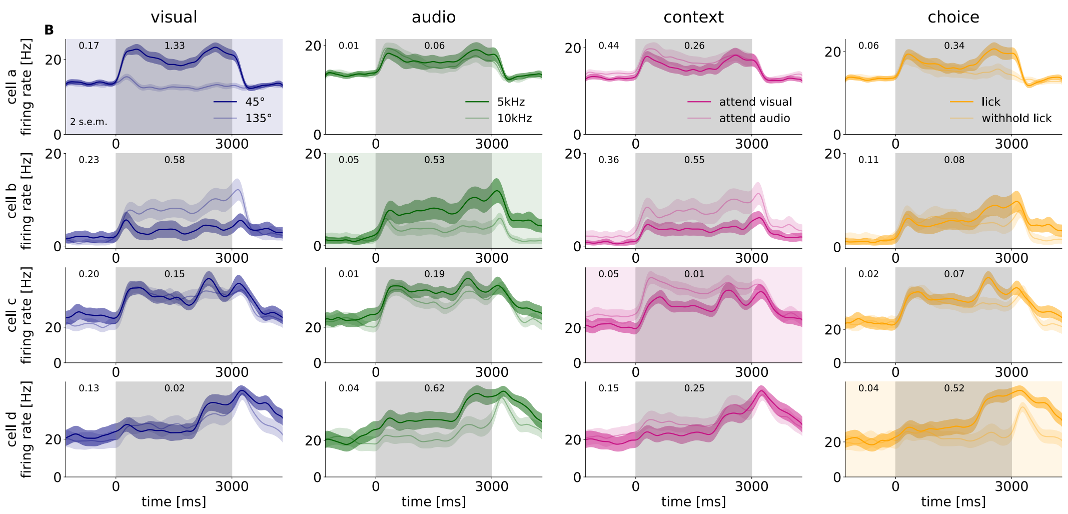


Responses of individual neurons are modulated by input, internal state, output. All of them at once.



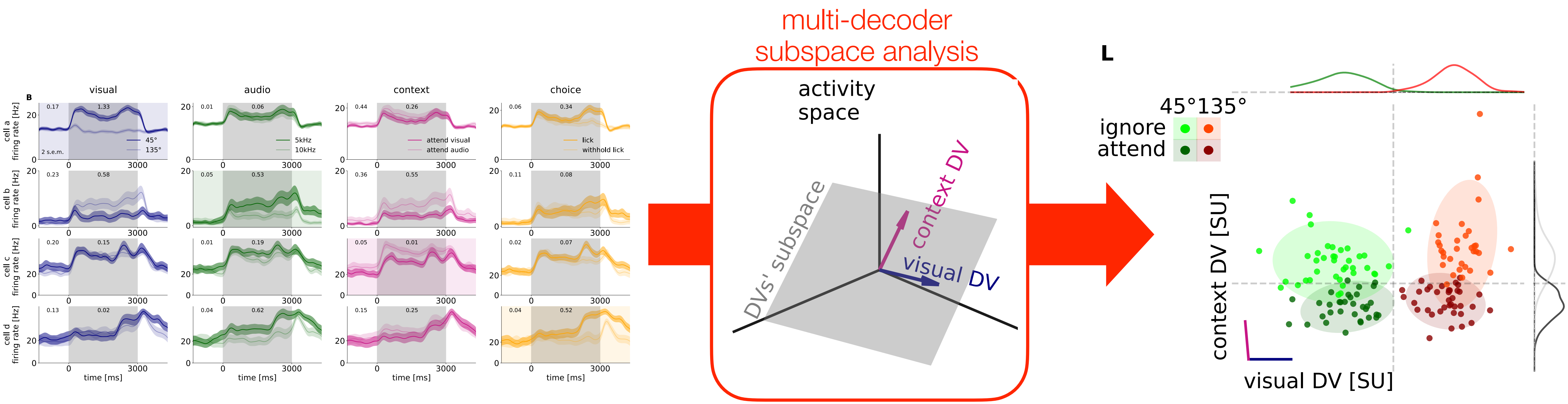
Hajnal et al (2023) Nature Communications

Data-driven approaches



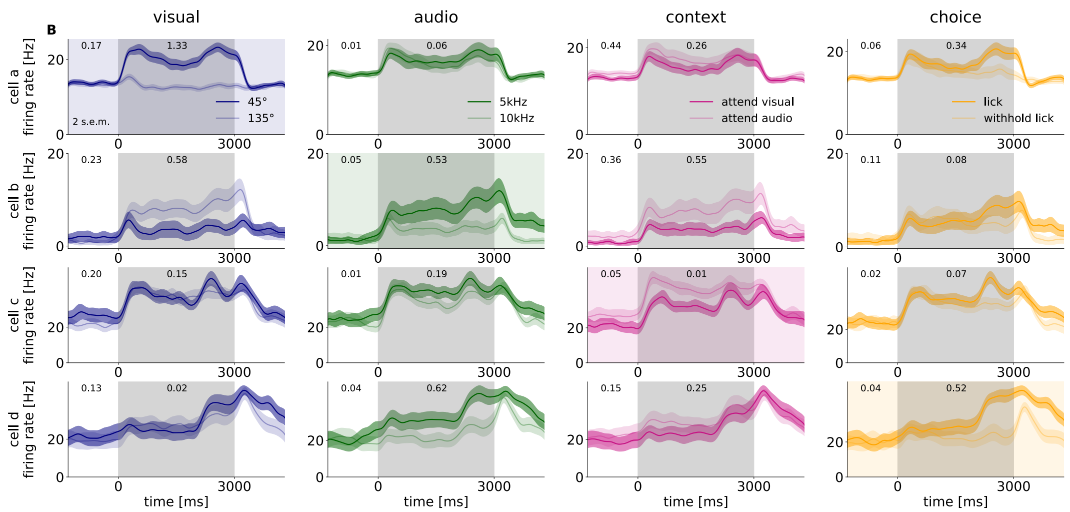
Hajnal et al (2023) Nature Communications

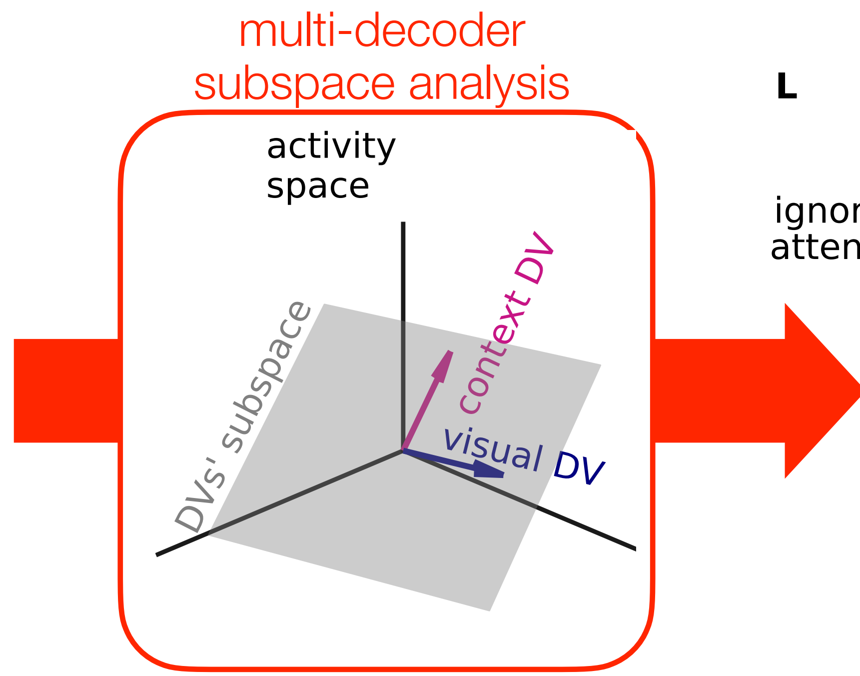
Data-driven approaches

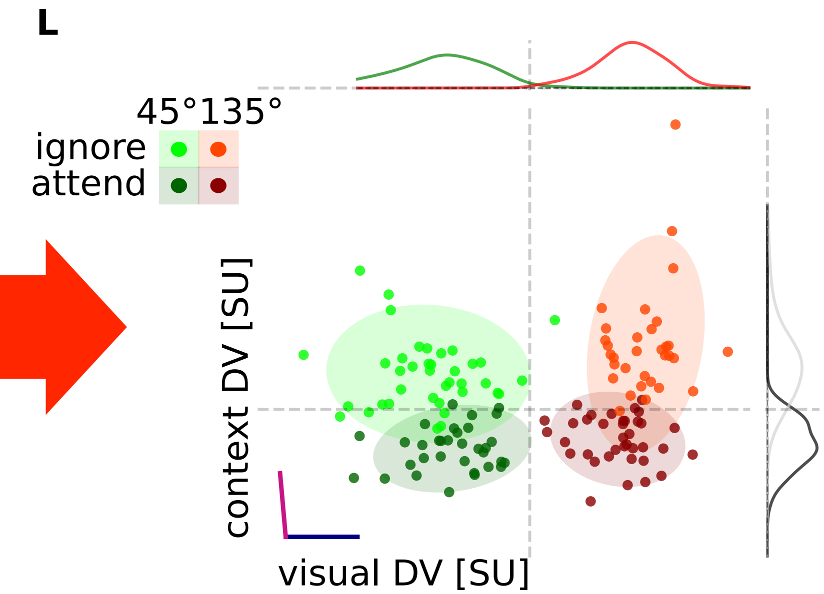


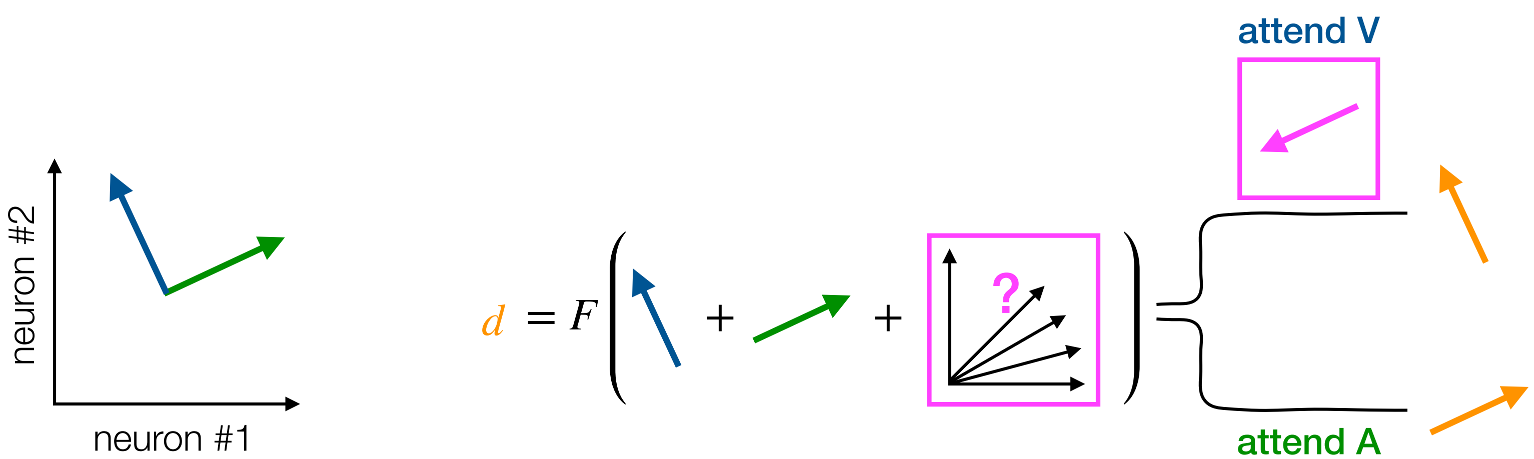
Hajnal et al (2023) Nature Communications

Data-driven approaches

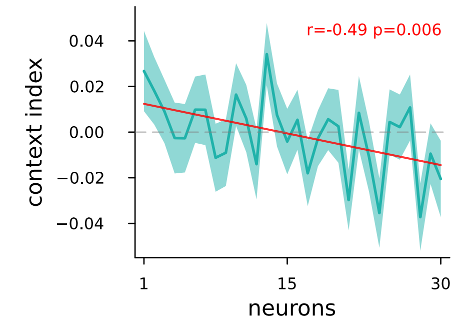




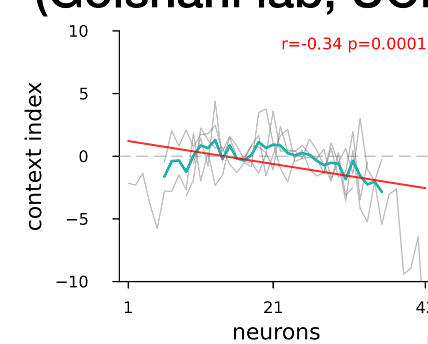




PREDICTION



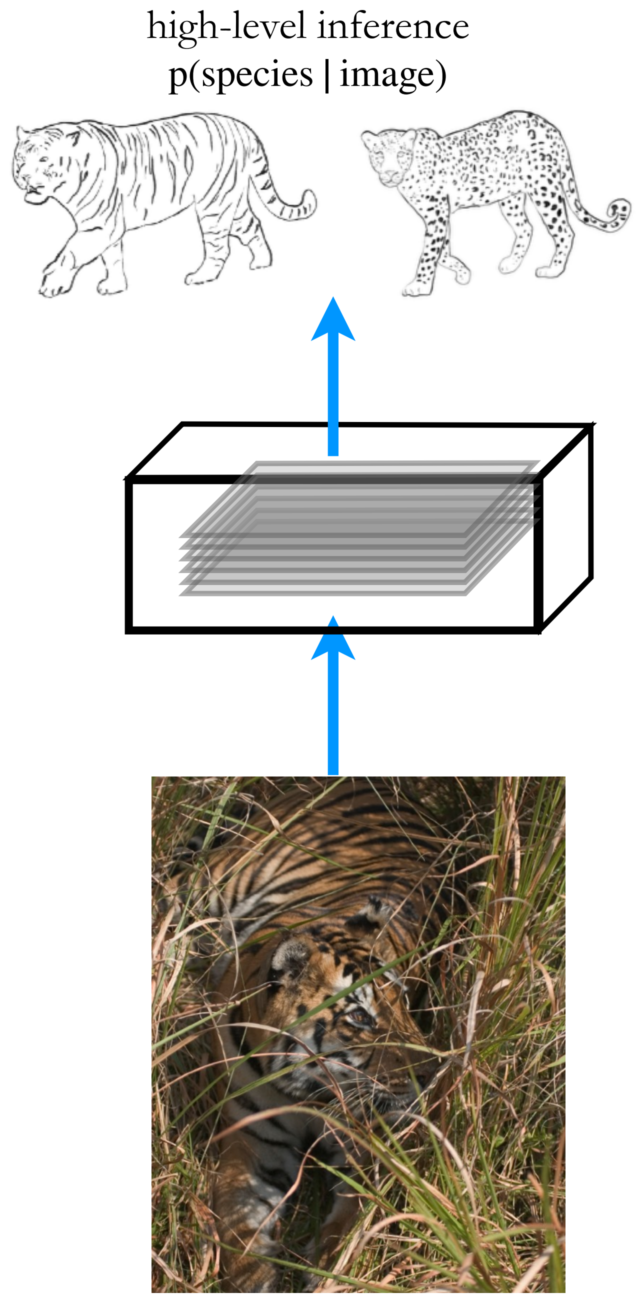
MOUSE RECORDINGS (Golshani lab, UCLA)

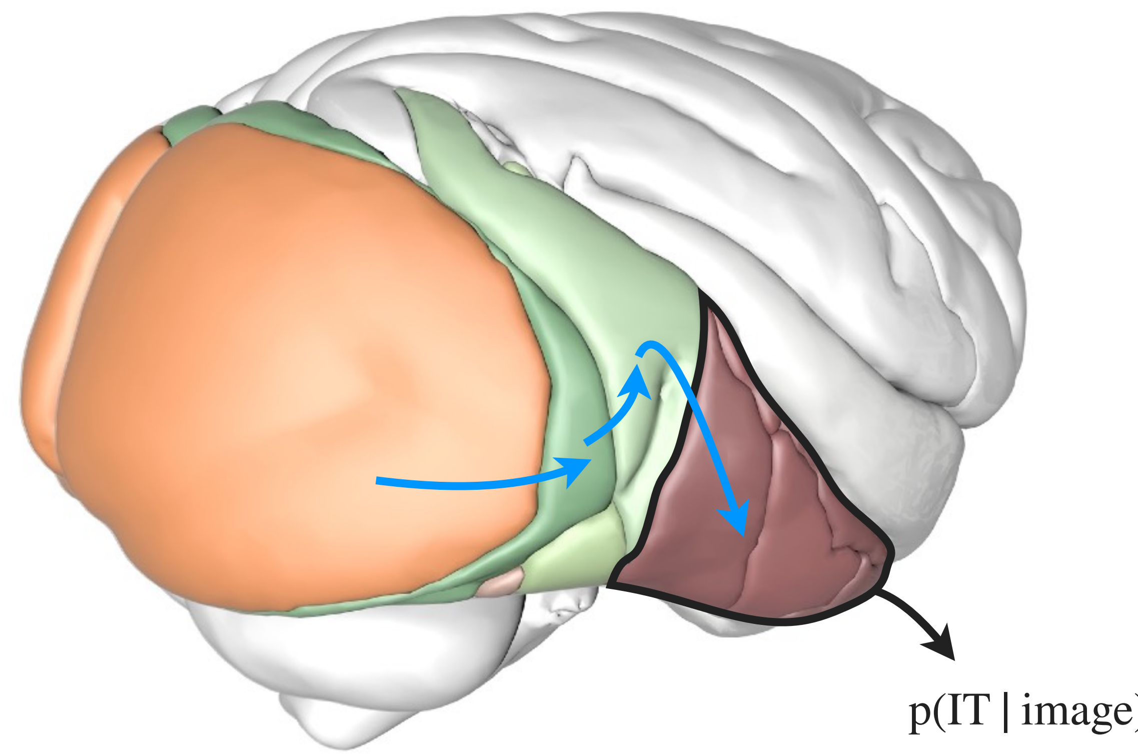


Hajnal et al, in prep

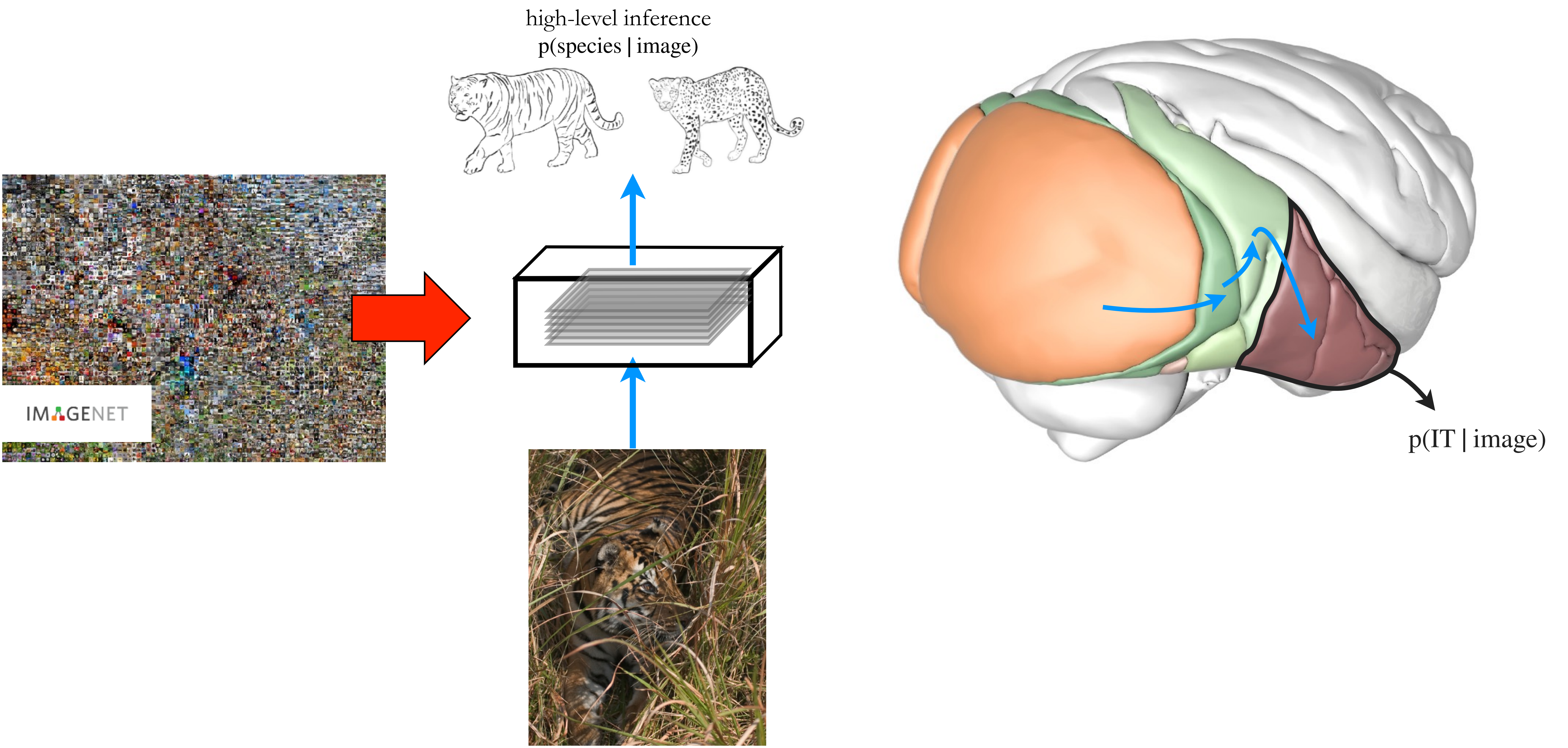
Hajnal et al (2023) Nature Communications

Adaptation-based approaches
deep **discriminative** models



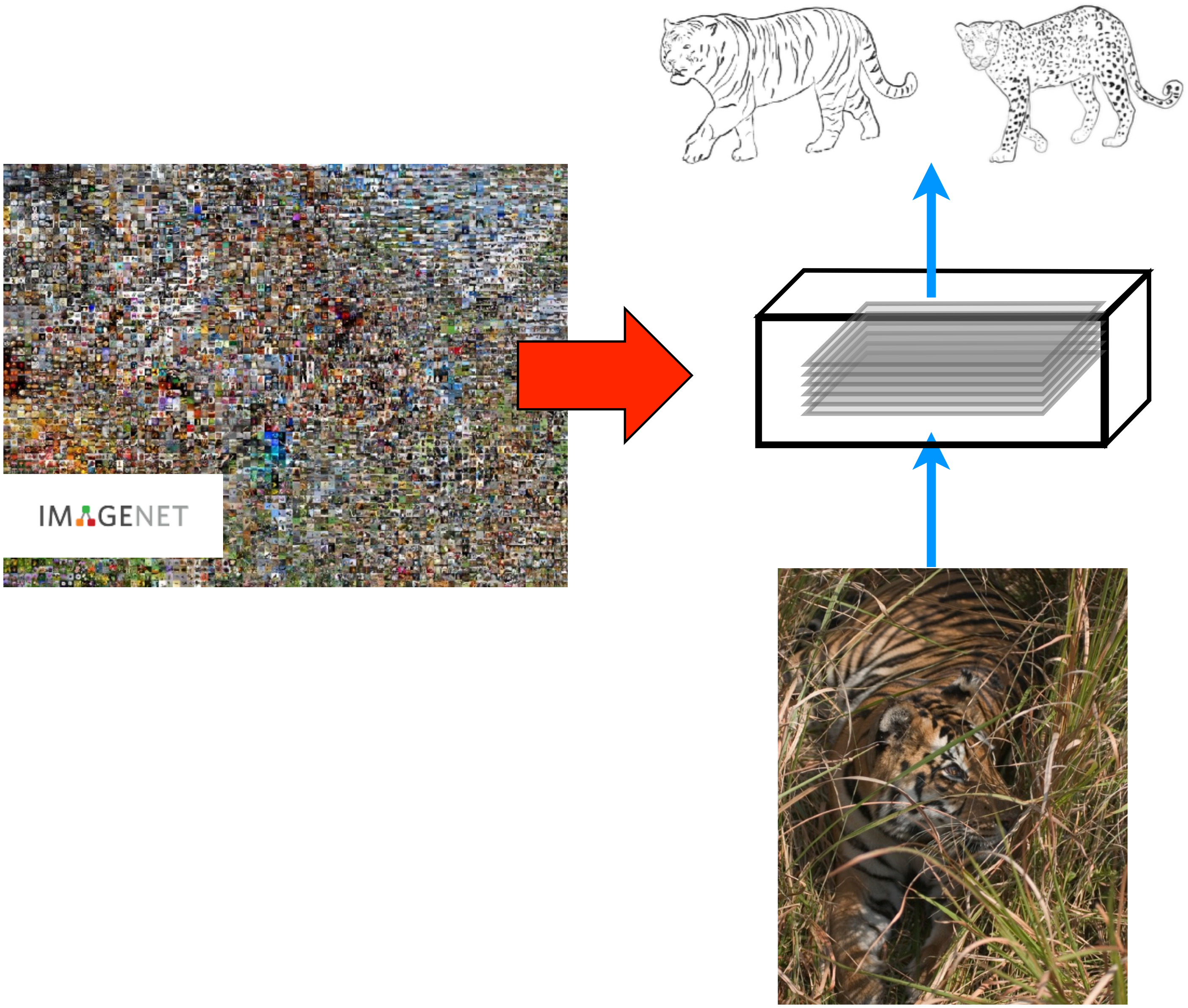


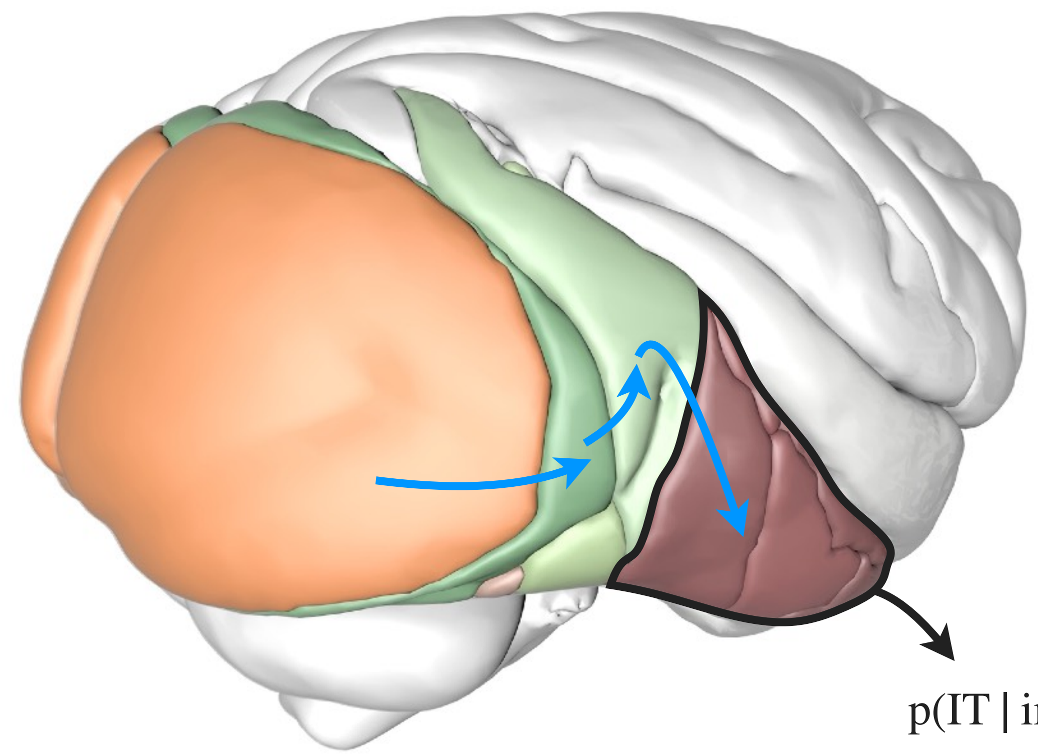
Adaptation-based approaches
deep **discriminative** models



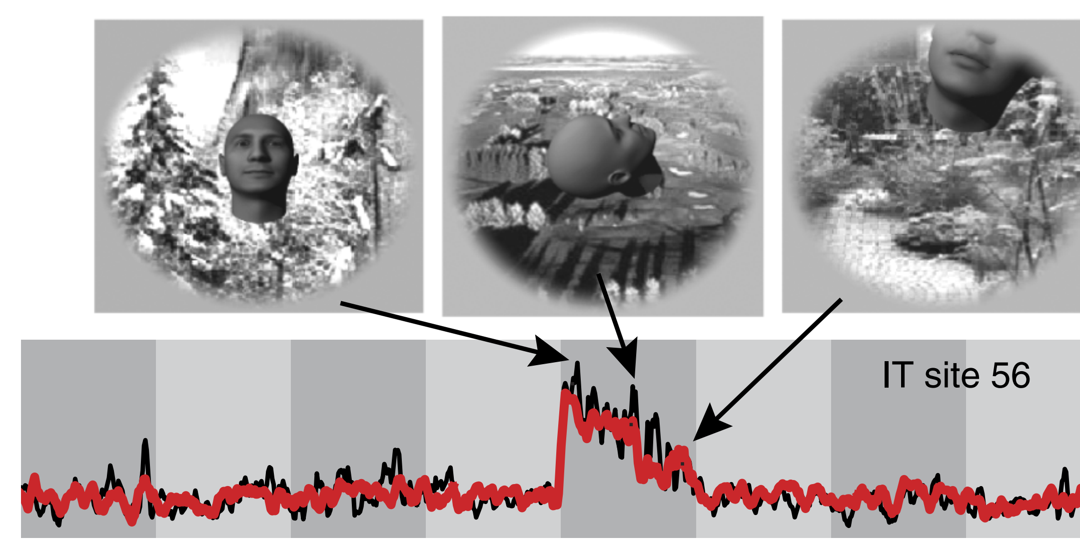
Adaptation-based approaches
deep **discriminative** models

high-level inference
p(species | image)





p(IT | image)

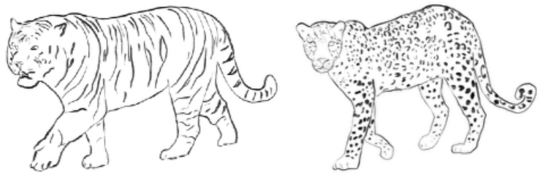


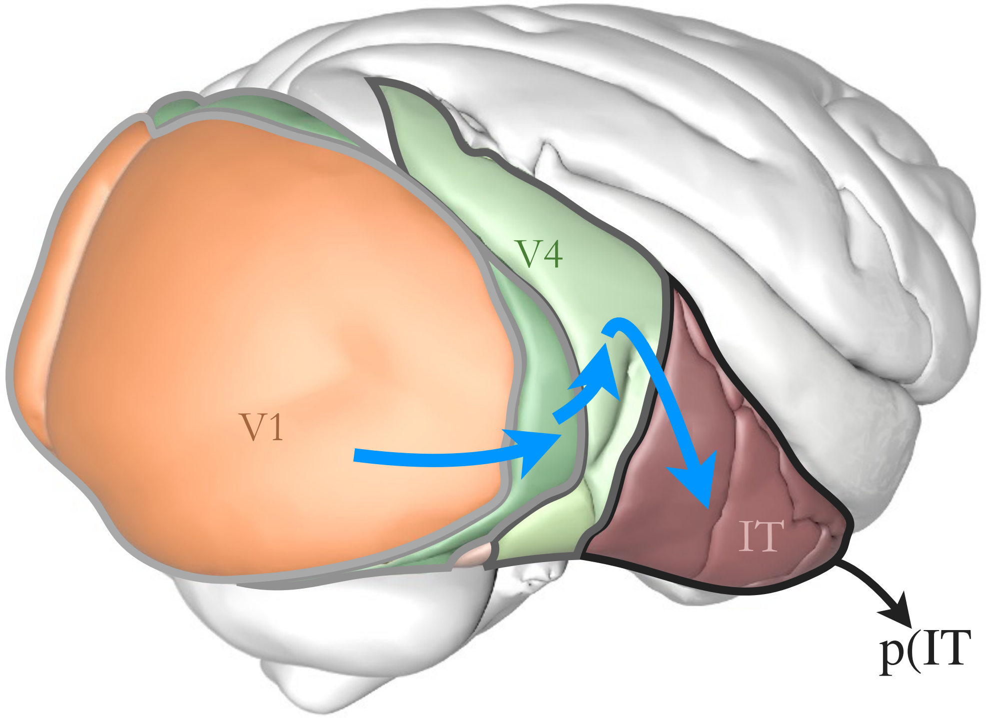
1600 unseen test image (sorted by category)

Yamins et al (2016) Nat Neurosci

Adaptation-based approaches
deep **generative** models

high-level inference
p(species | image)

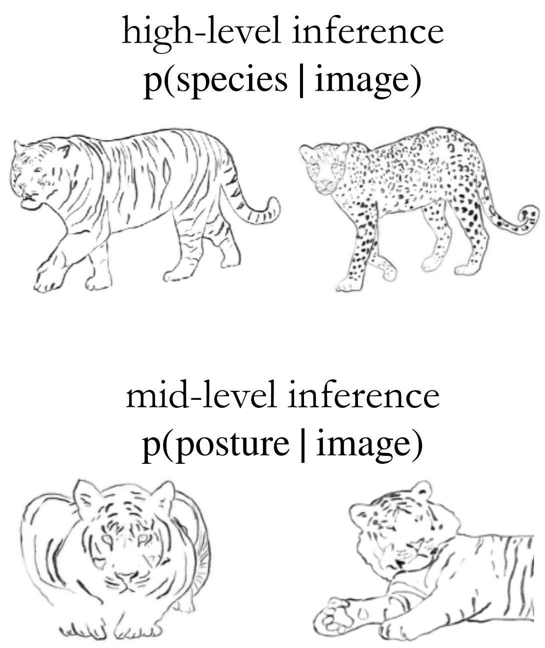


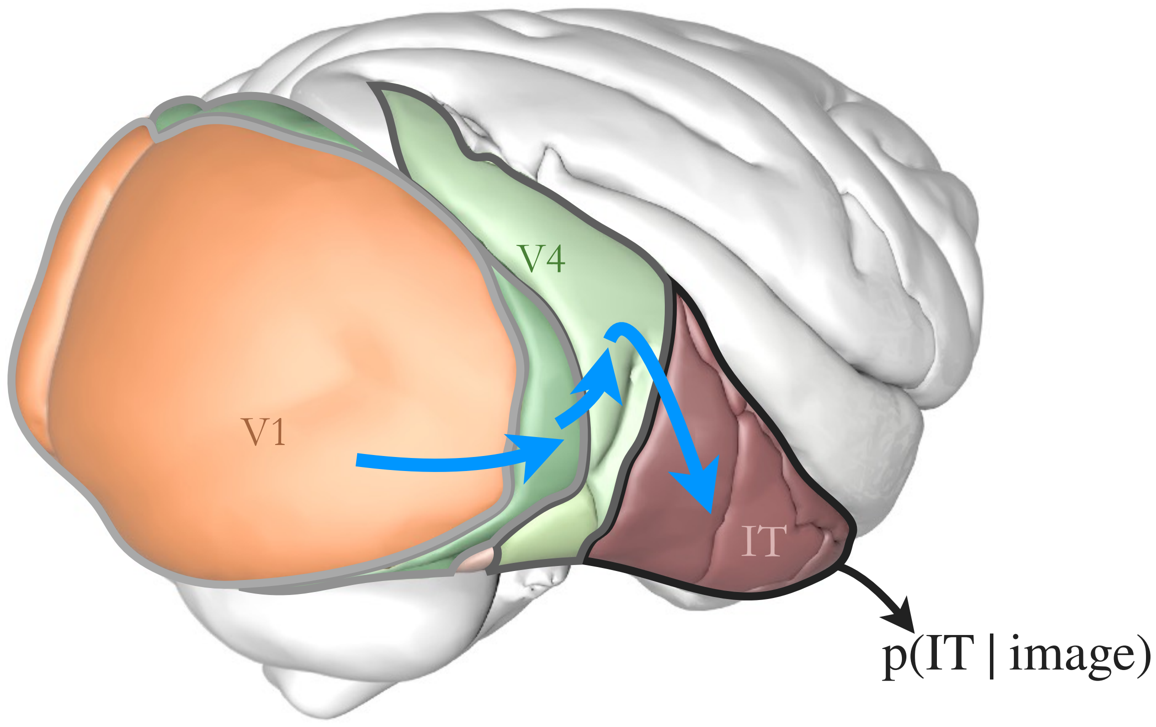


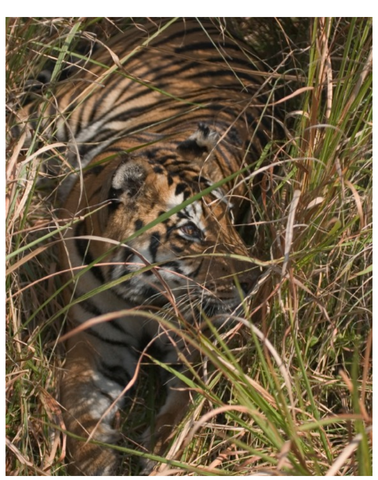
Meszéna et al (2022) NeurIPS
Csikor et al, in prep

Adaptation-based approaches

deep **generative** models



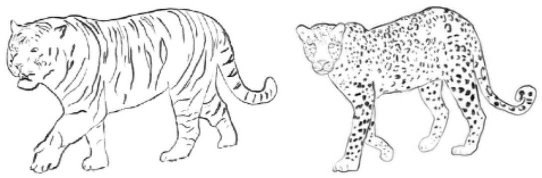




Meszéna et al (2022) NeurIPS
Csikor et al, in prep

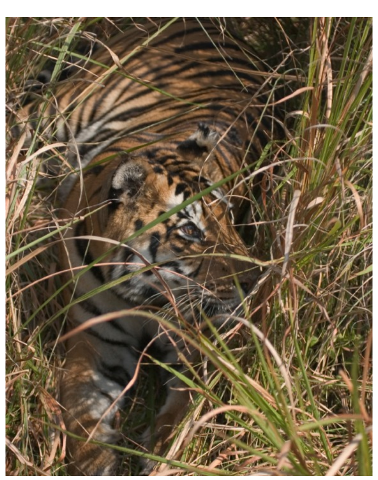
Adaptation-based approaches
deep **generative** models

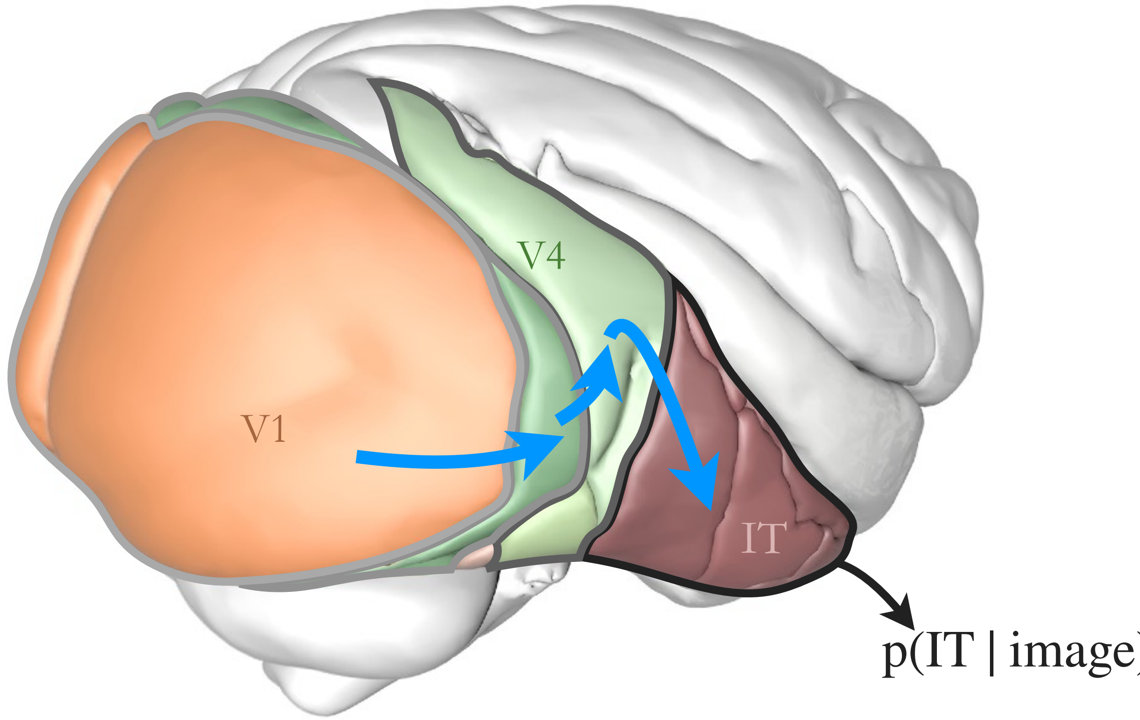
high-level inference
p(species | image)



mid-level inference
p(posture | image)

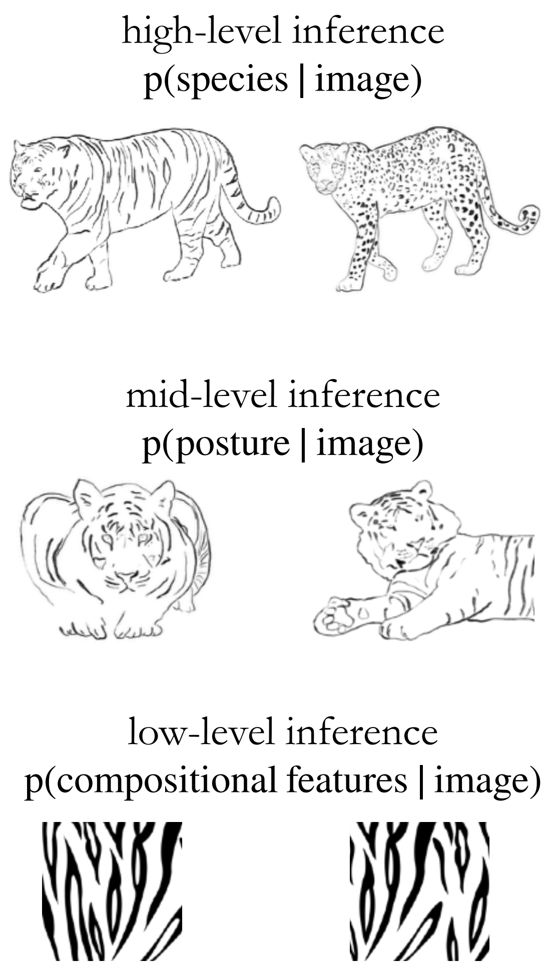
low-level inference
p(compositional features | image)

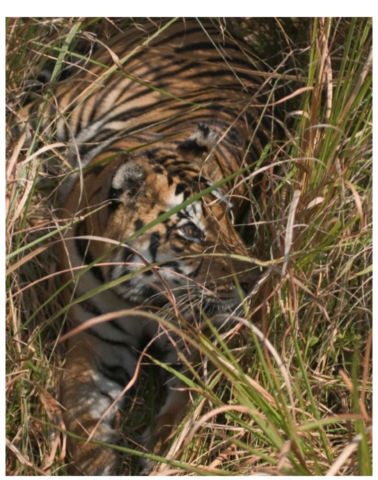


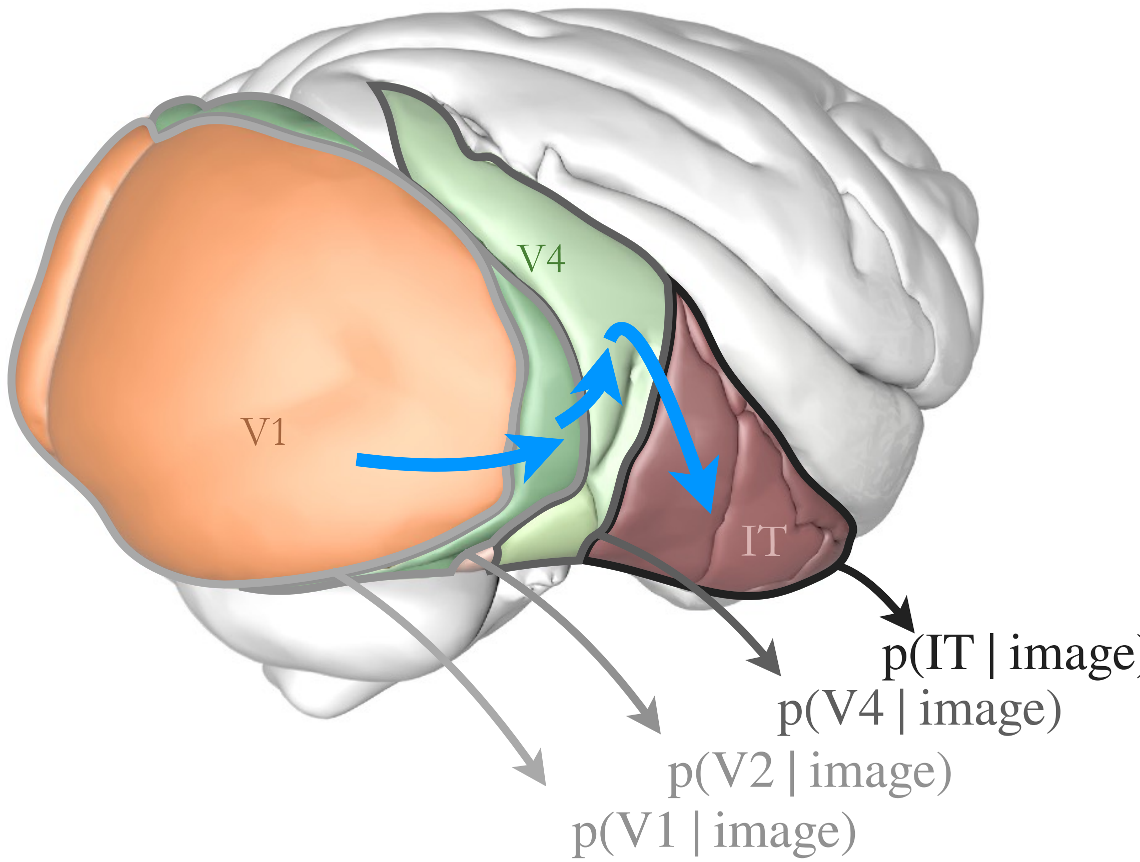


Meszéna et al (2022) NeurIPS
Csikor et al, in prep

Adaptation-based approaches
deep **generative** models

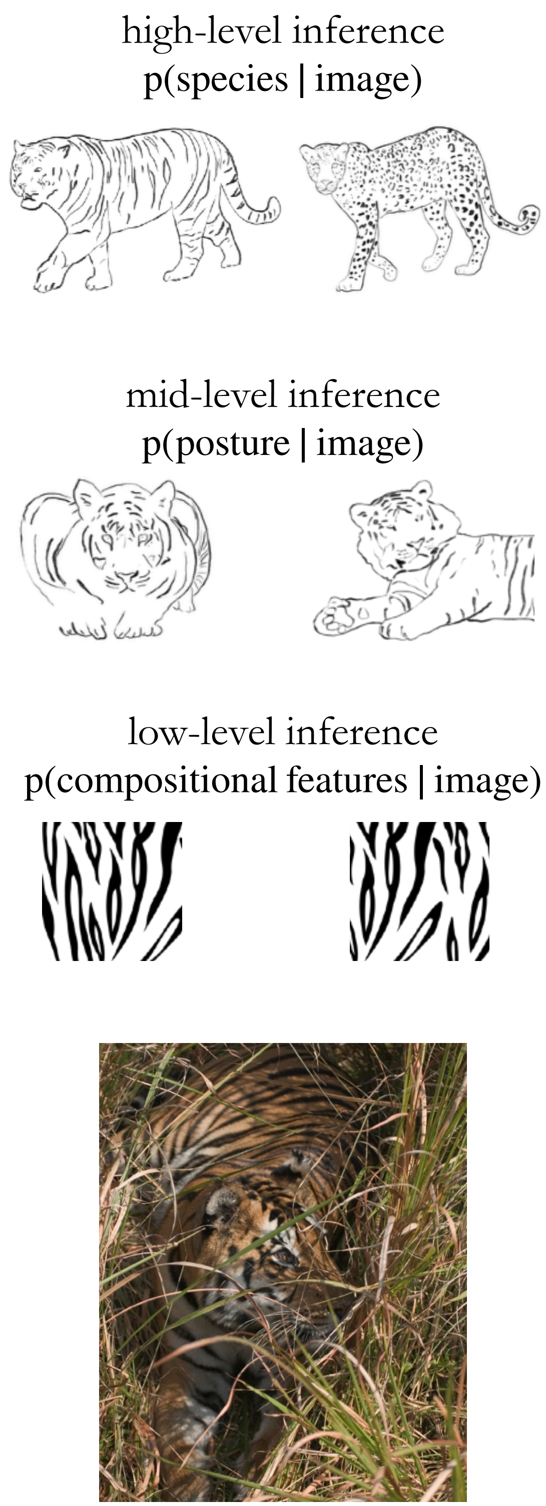


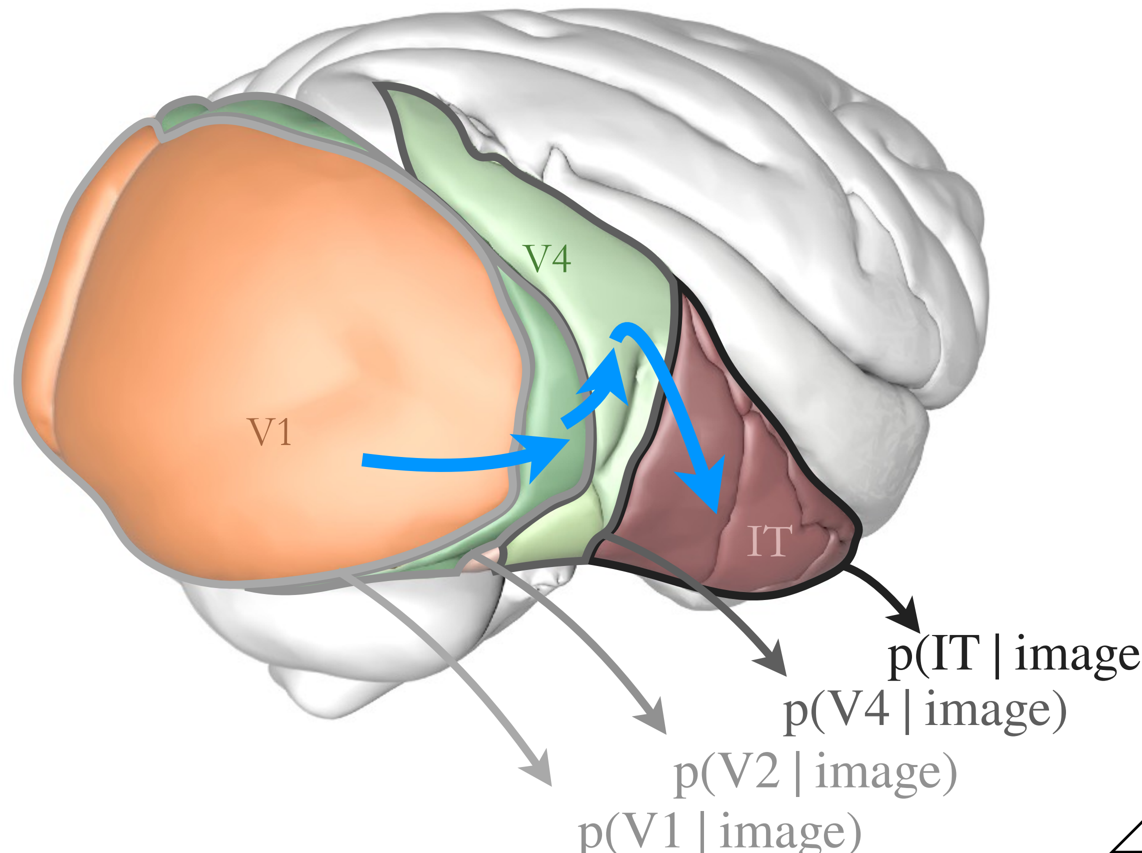


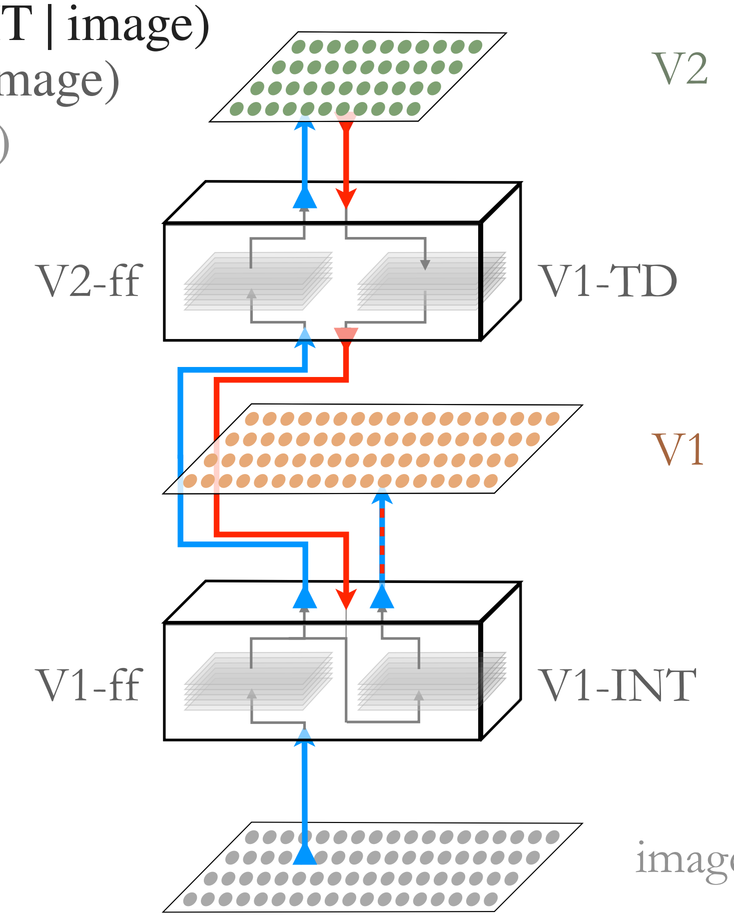


Meszéna et al (2022) NeurIPS
Csikor et al, in prep

Adaptation-based approaches
deep **generative** models



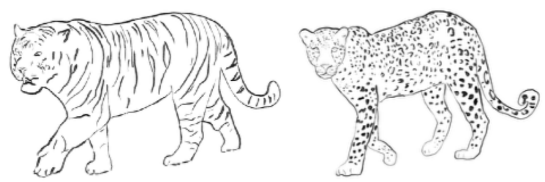




Meszéna et al (2022) NeurIPS
Csikor et al, in prep

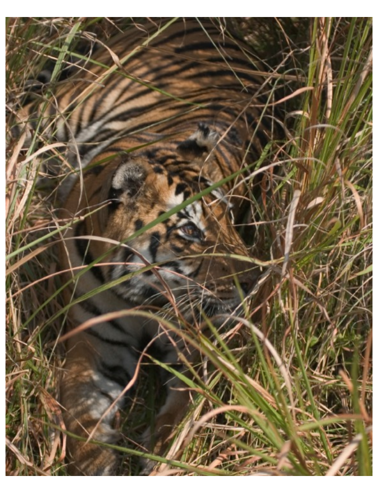
Adaptation-based approaches
deep **generative** models

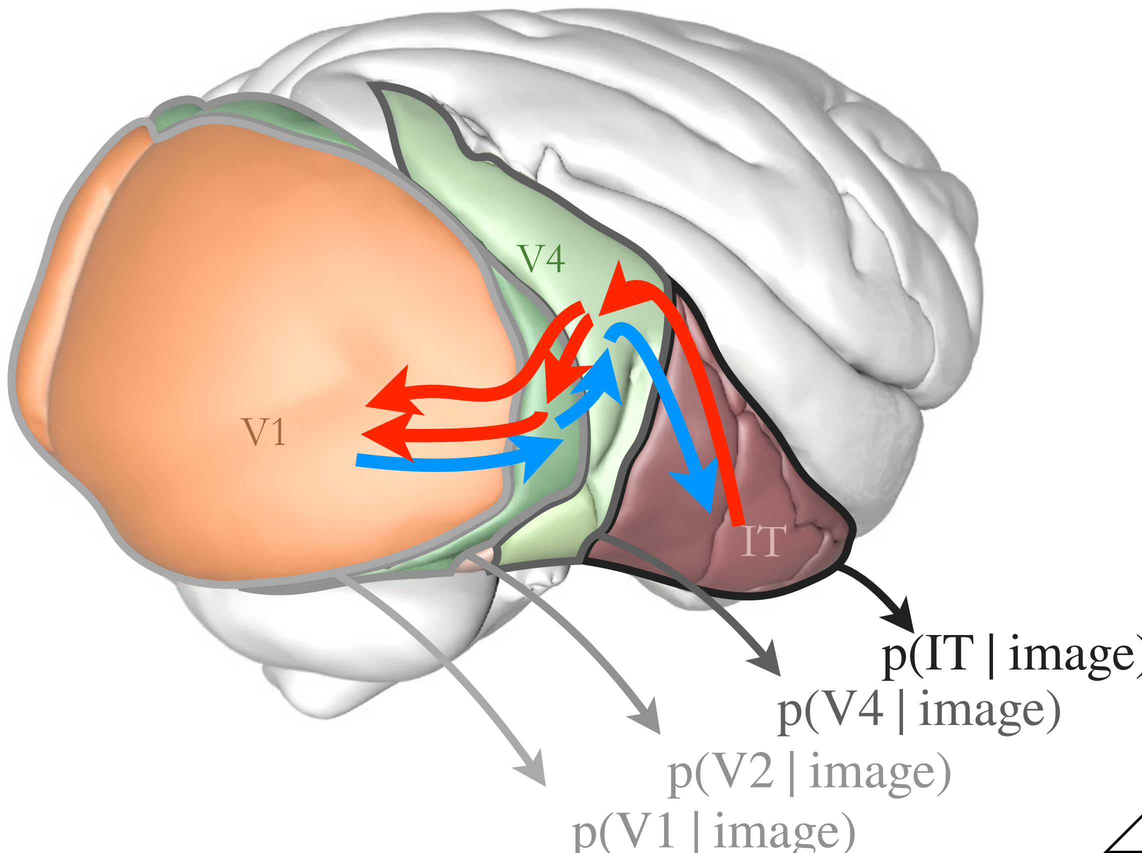
high-level inference
p(species | image)

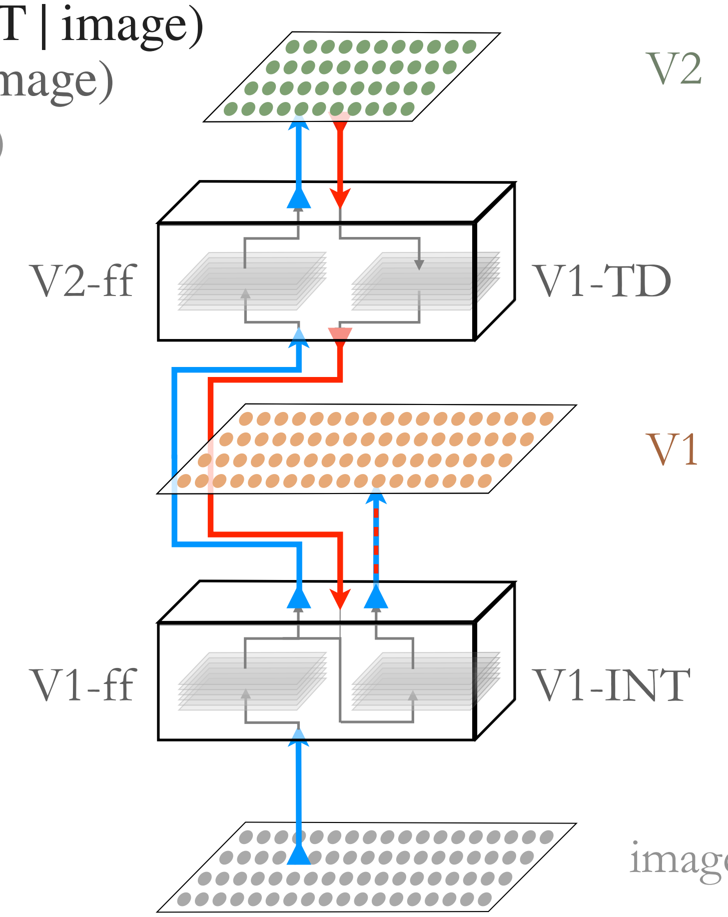


mid-level inference
p(posture | image)

low-level inference
p(compositional features | image)







Meszéna et al (2022) NeurIPS
Csikor et al, in prep

Interpreting the neural code: ML insights

Approximate inference

probability distributions need to be represented

Orbán et al (2016) Neuron

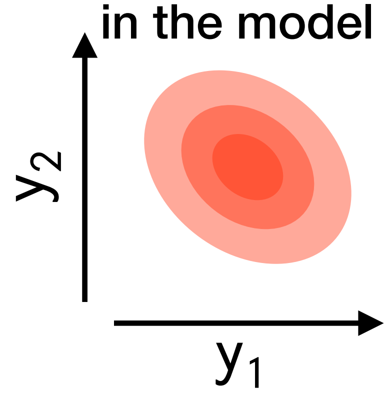
Bányai et al (2019) PNAS

Interpreting the neural code: ML insights

Approximate inference

probability distributions need to be represented

inference in the model



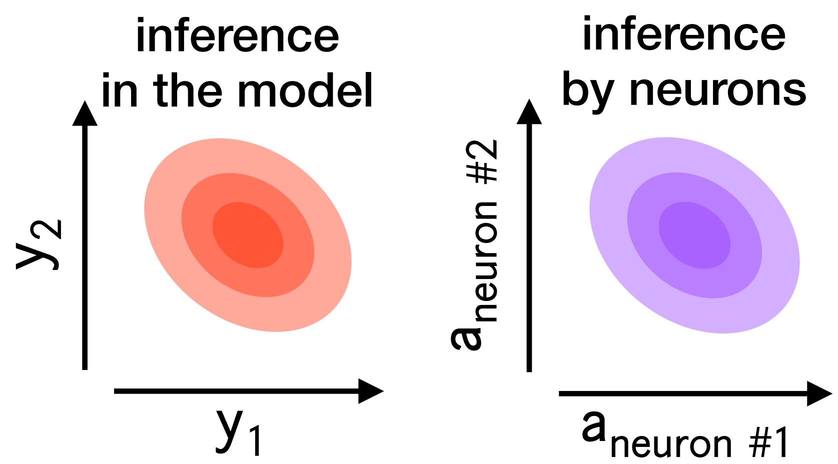
Orbán et al (2016) Neuron

Bányai et al (2019) PNAS

Interpreting the neural code: ML insights

Approximate inference

probability distributions need to be represented



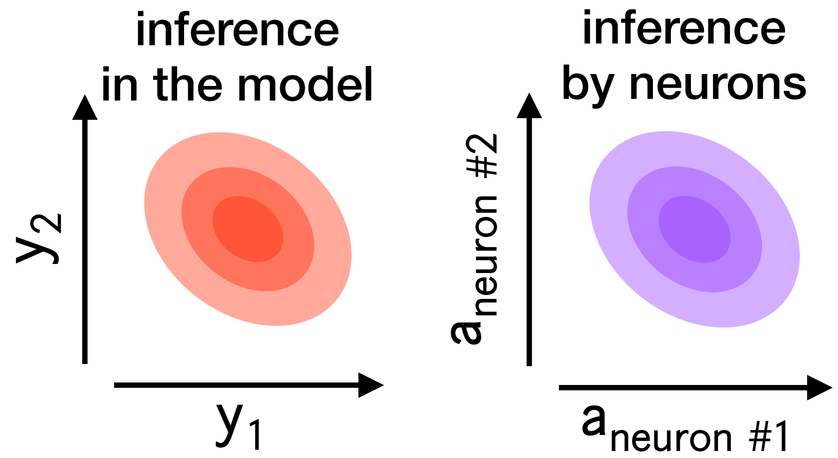
Orbán et al (2016) Neuron

Bányai et al (2019) PNAS

Interpreting the neural code: ML insights

Approximate inference

probability distributions need to be represented



stochastic sampling ⤳ variable membrane potential

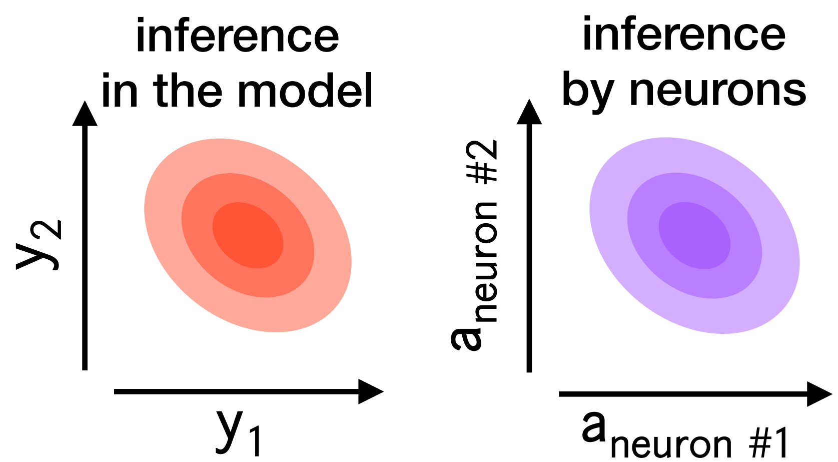
Orbán et al (2016) Neuron

Bányai et al (2019) PNAS

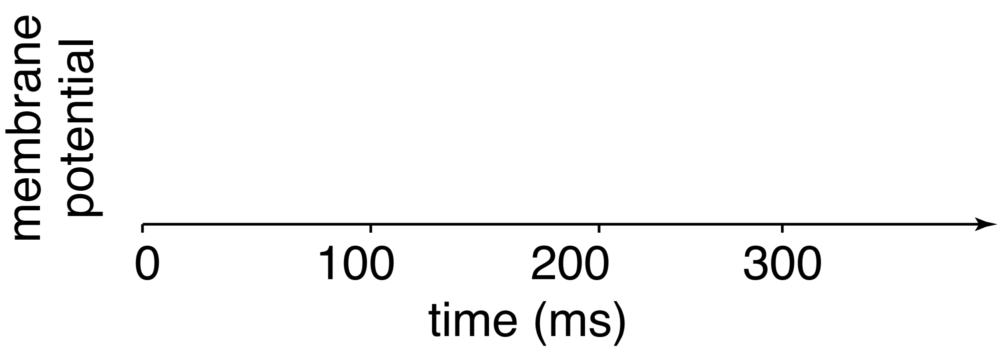
Interpreting the neural code: ML insights

Approximate inference

probability distributions need to be represented



stochastic sampling ⤳ variable membrane potential



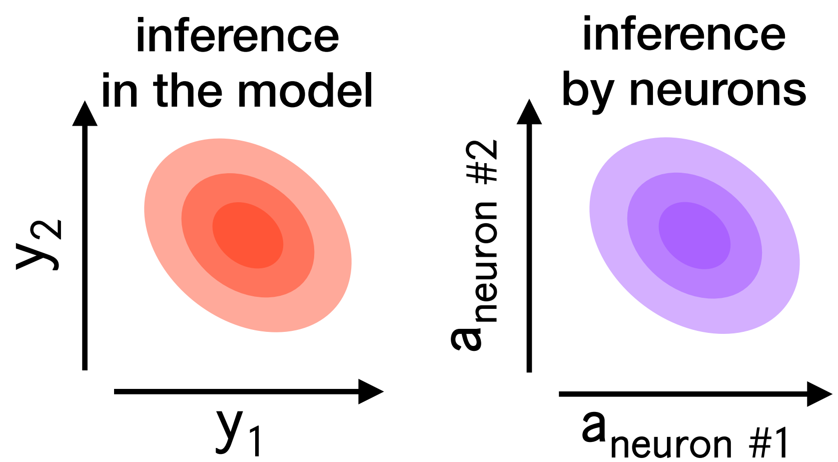
Orbán et al (2016) Neuron

Bányai et al (2019) PNAS

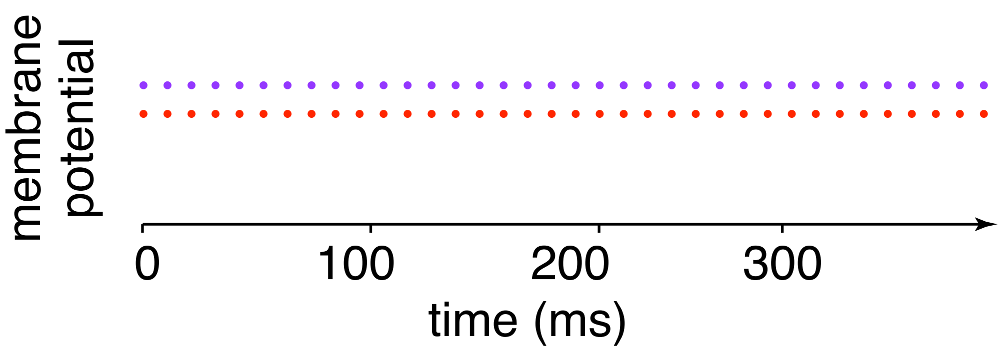
Interpreting the neural code: ML insights

Approximate inference

probability distributions need to be represented



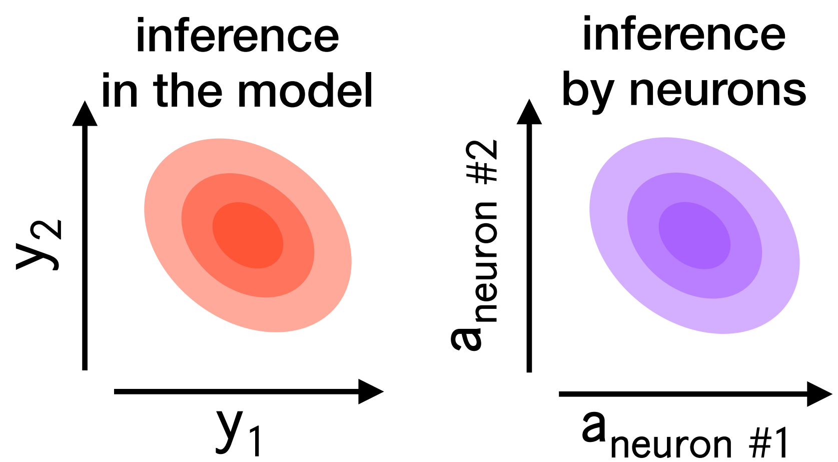
stochastic sampling ⇝ variable membrane potential



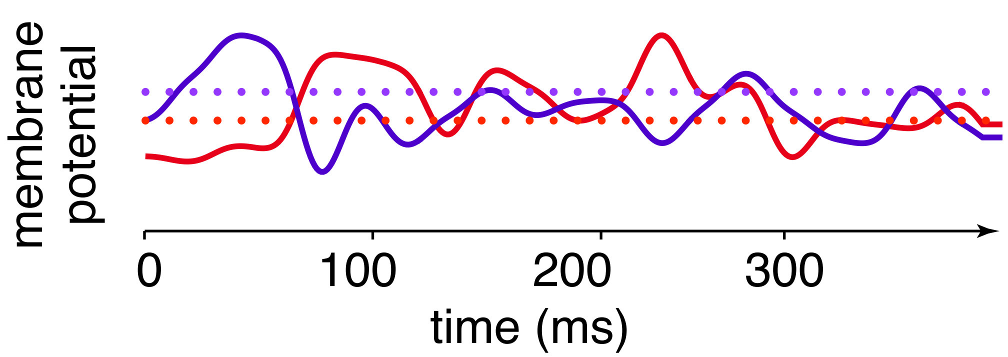
Interpreting the neural code: ML insights

Approximate inference

probability distributions need to be represented



stochastic sampling ⇝ variable membrane potential



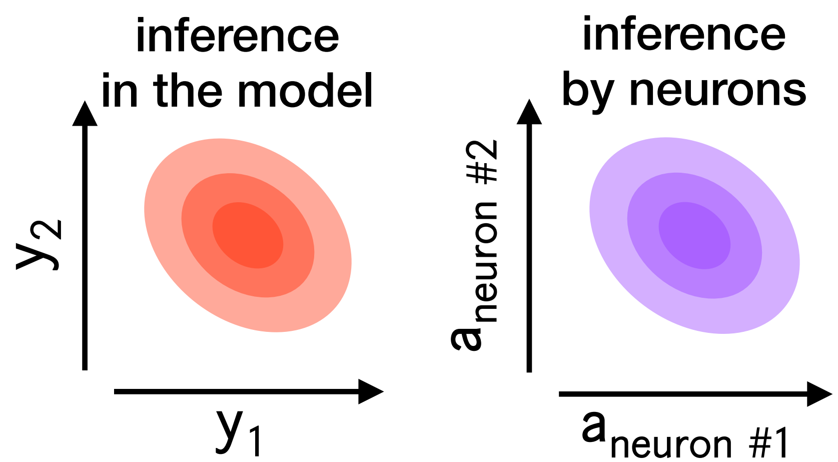
Orbán et al (2016) Neuron

Bányai et al (2019) PNAS

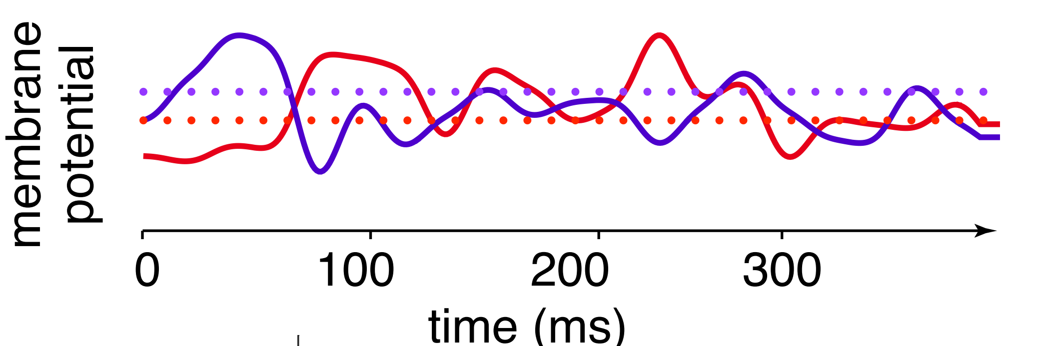
Interpreting the neural code: ML insights

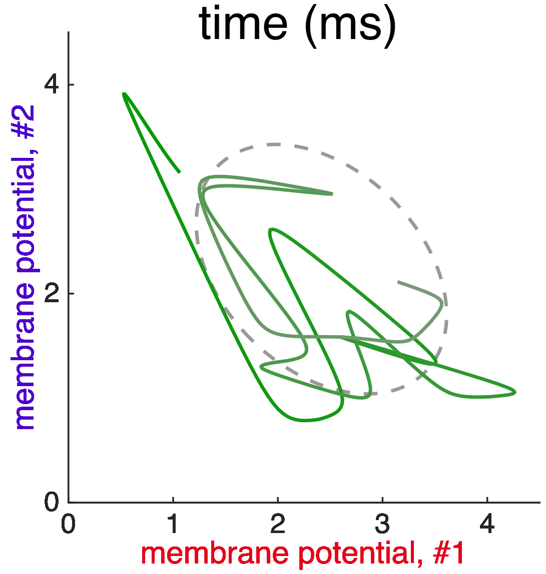
Approximate inference

probability distributions need to be represented



stochastic sampling ⤳ variable membrane potential





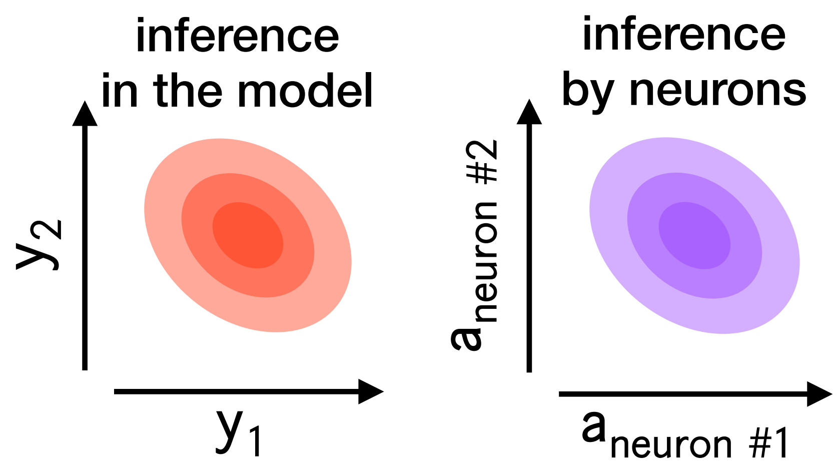
Orbán et al (2016) Neuron

Bányai et al (2019) PNAS

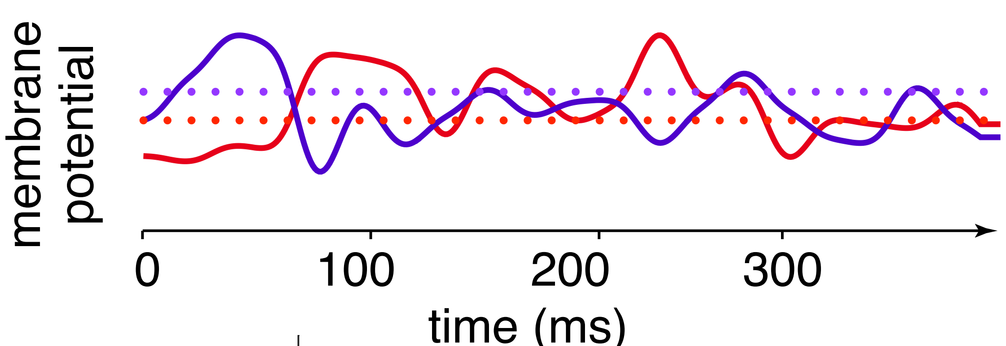
Interpreting the neural code: ML insights

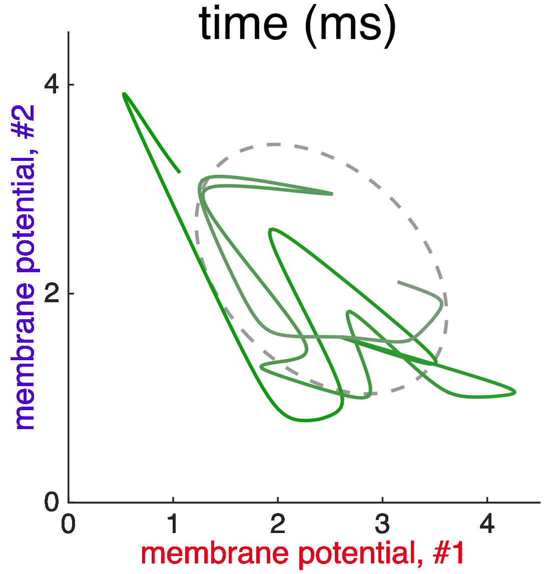
Approximate inference

probability distributions need to be represented



stochastic sampling ⤳ variable membrane potential





response variability ⇔ subjective uncertainty

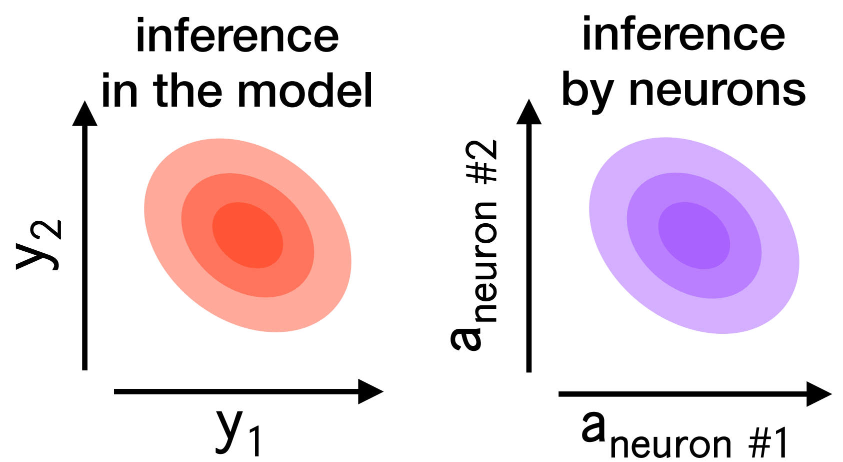
Orbán et al (2016) Neuron

Bányai et al (2019) PNAS

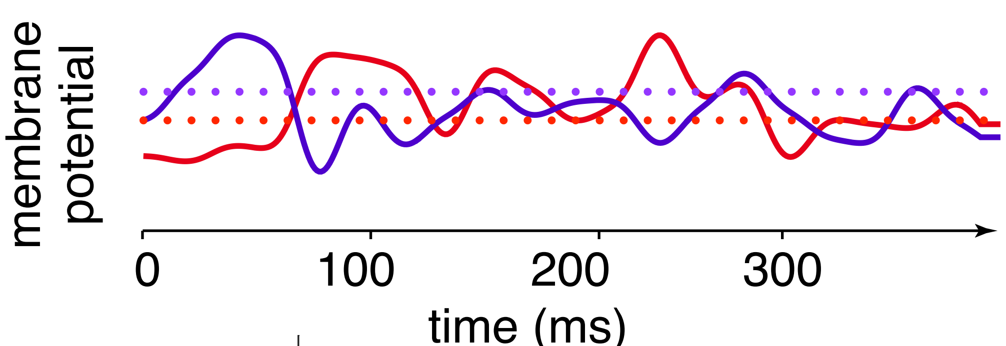
Interpreting the neural code: ML insights

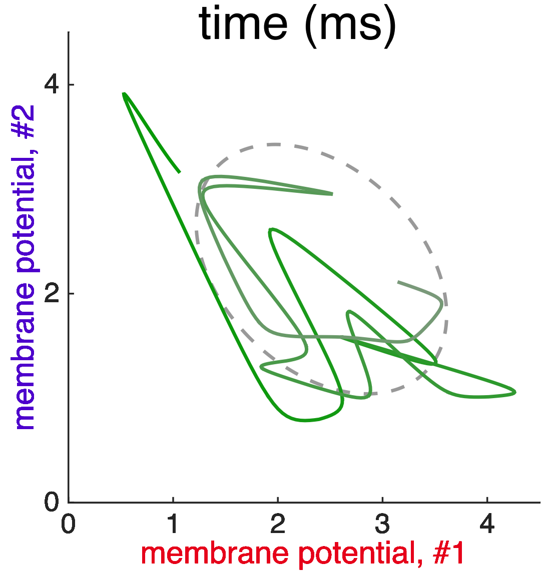
Approximate inference

probability distributions need to be represented



stochastic sampling ⤳ variable membrane potential





response variability ⇔ subjective uncertainty

Orbán et al (2016) Neuron

Contextual prior

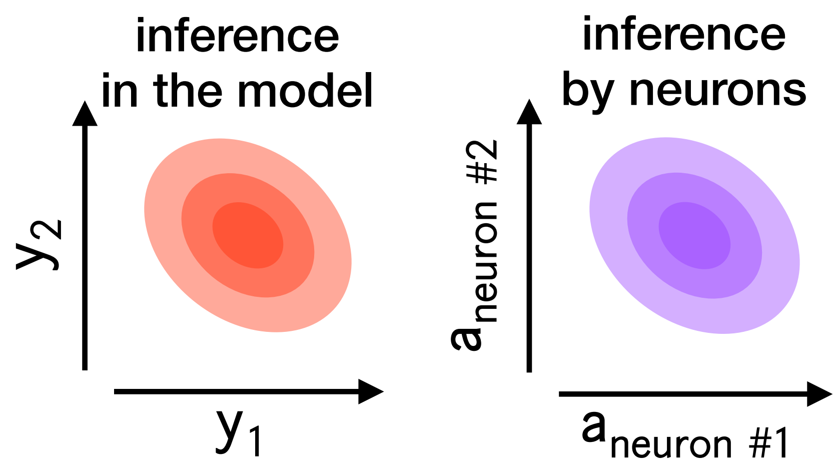
learning exploits regularities, regularities change from context-to context

Bányai et al (2019) PNAS

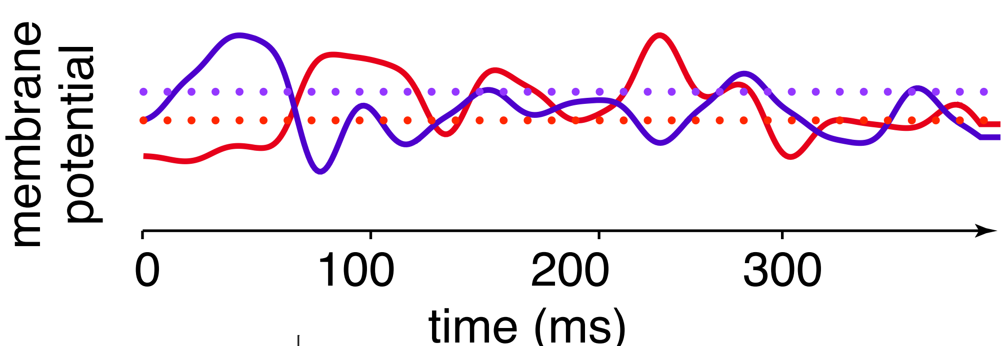
Interpreting the neural code: ML insights

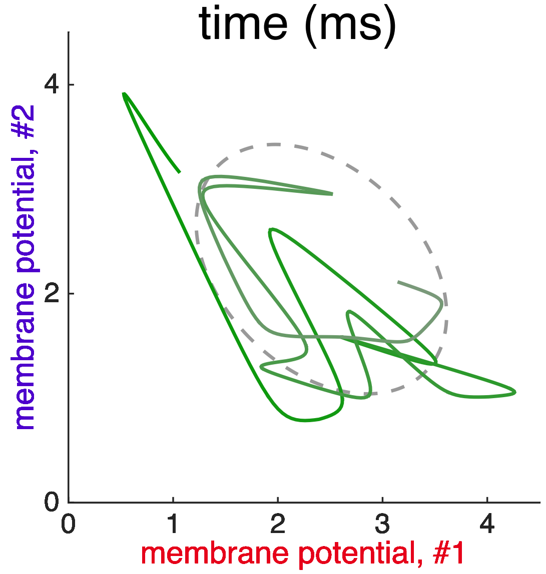
Approximate inference

probability distributions need to be represented



stochastic sampling ⇝ variable membrane potential



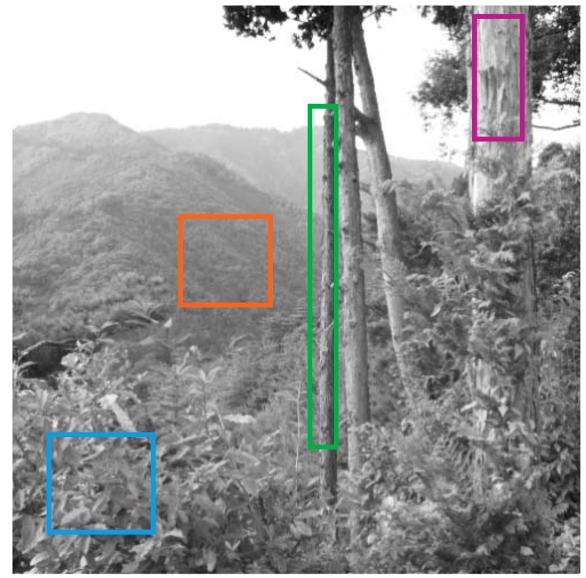


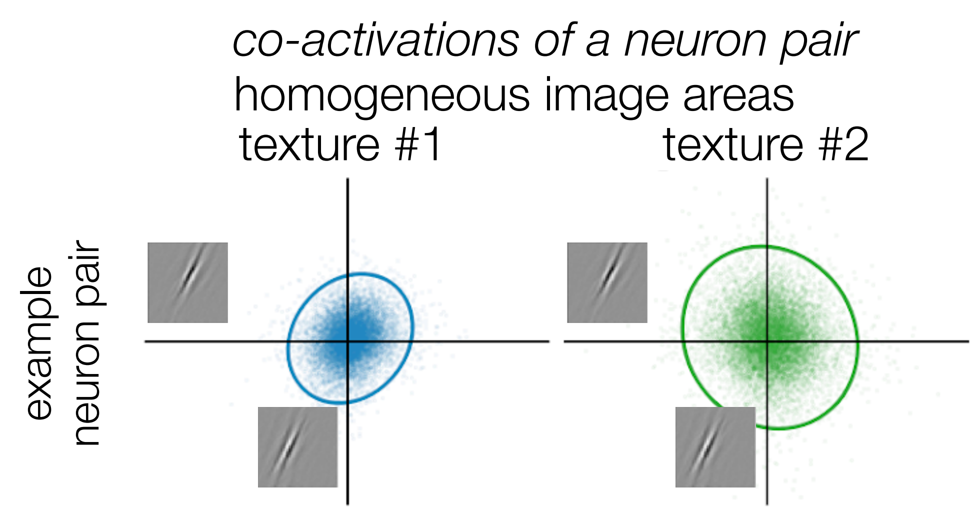
response variability ⇔ subjective uncertainty

Orbán et al (2016) Neuron

Contextual prior

learning exploits regularities, regularities change from context-to context



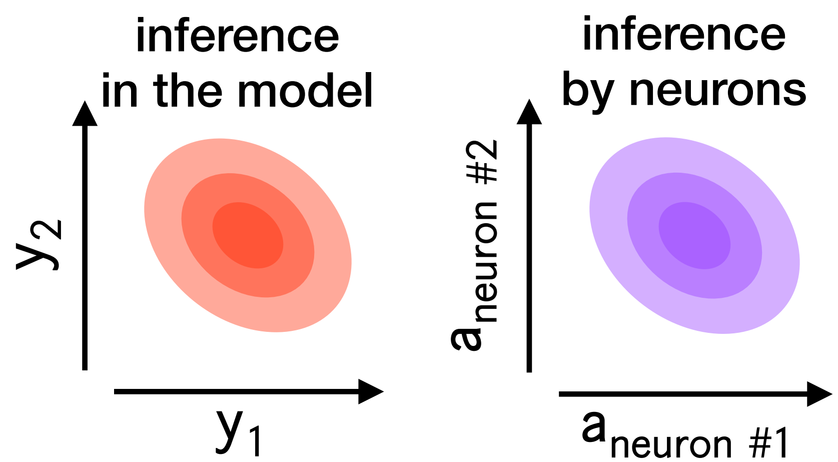


Bányai et al (2019) PNAS

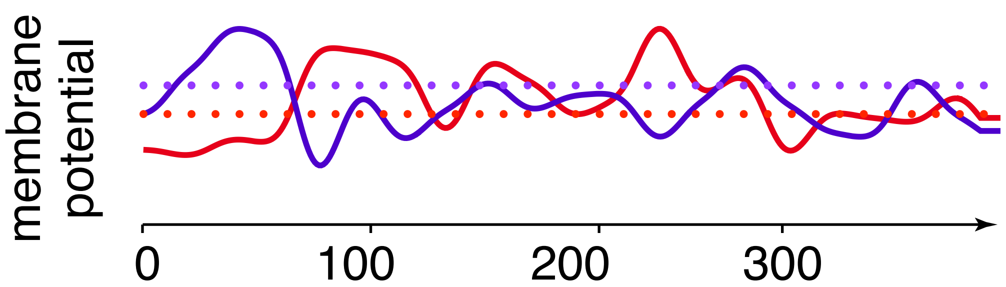
Interpreting the neural code: ML insights

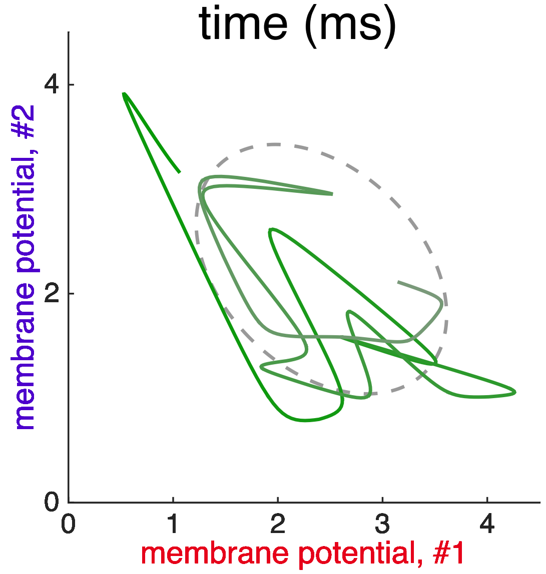
Approximate inference

probability distributions need to be represented



stochastic sampling ⤳ variable membrane potential



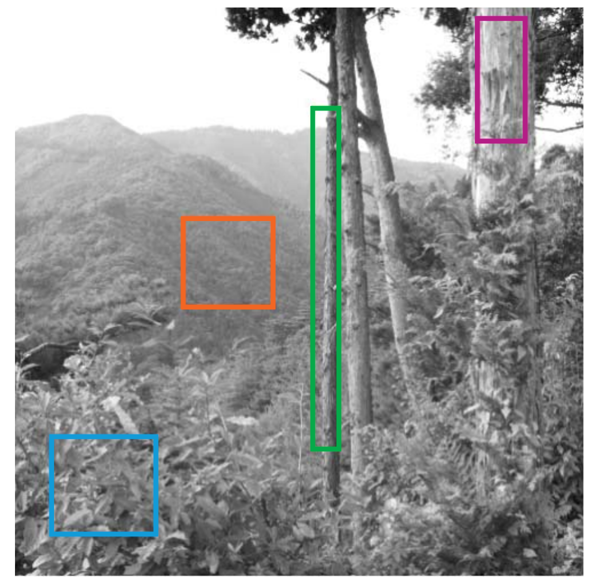


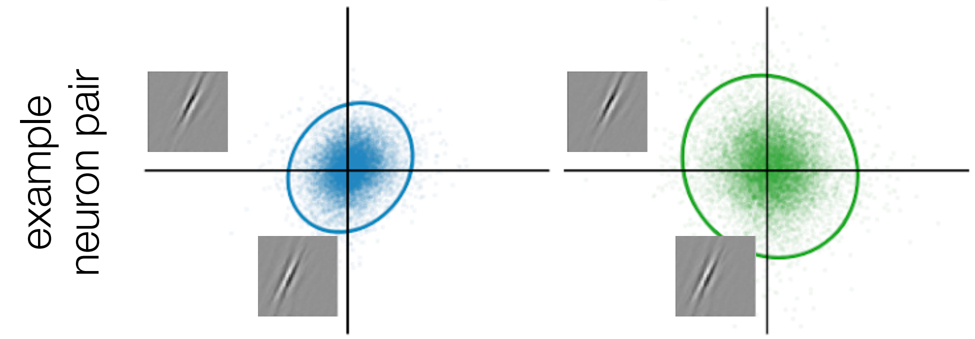
response variability ⇔ subjective uncertainty

Orbán et al (2016) Neuron

Contextual prior

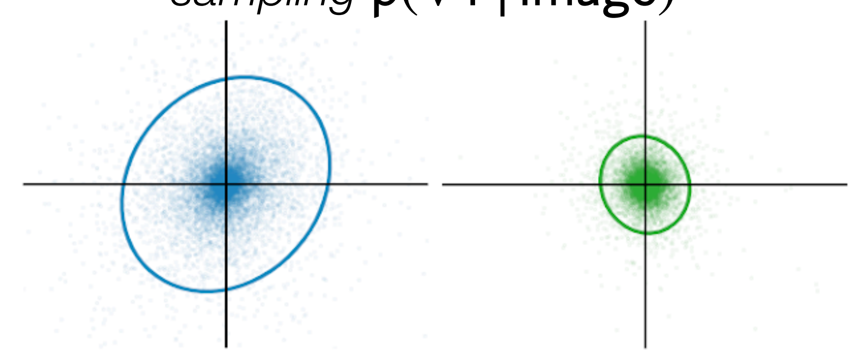
learning exploits regularities, regularities change from context-to context





neural ensembles represent distributions
⤳ context-dependent correlations in priors

sampling p(V1 | image)

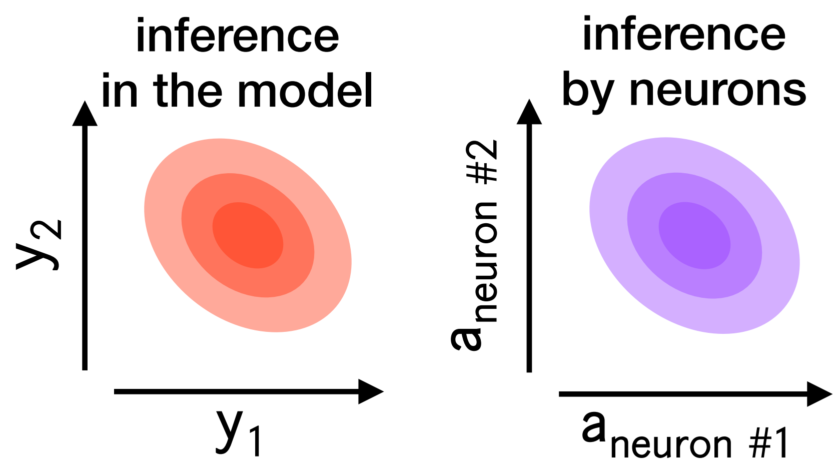


Bányai et al (2019) PNAS

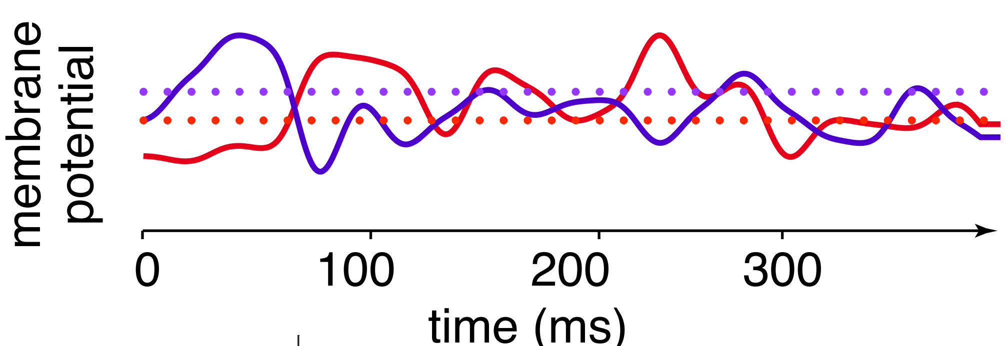
Interpreting the neural code: ML insights

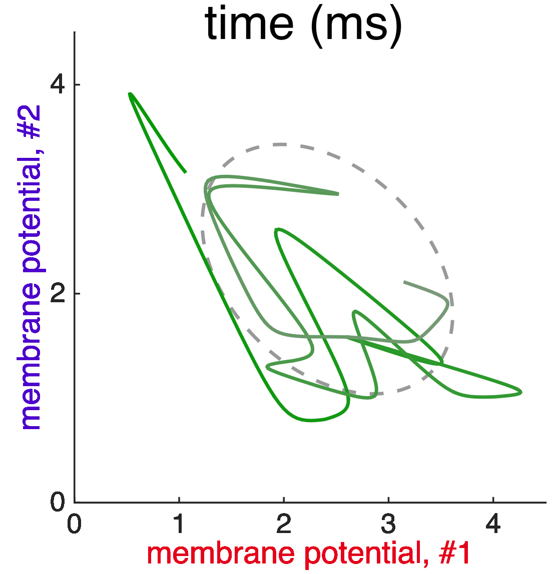
Approximate inference

probability distributions need to be represented



stochastic sampling ⇝ variable membrane potential



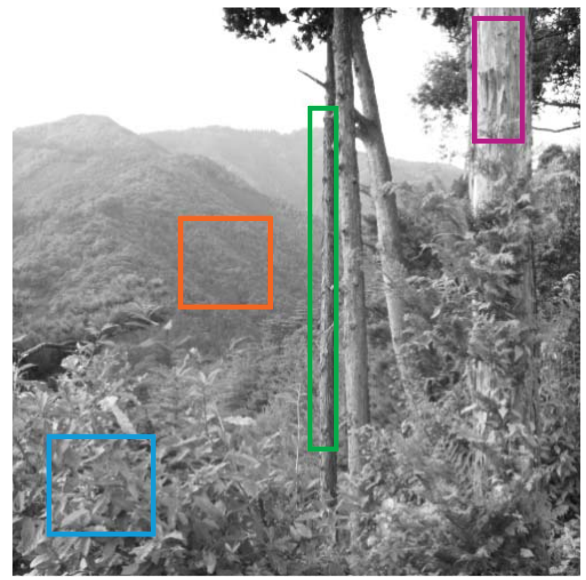


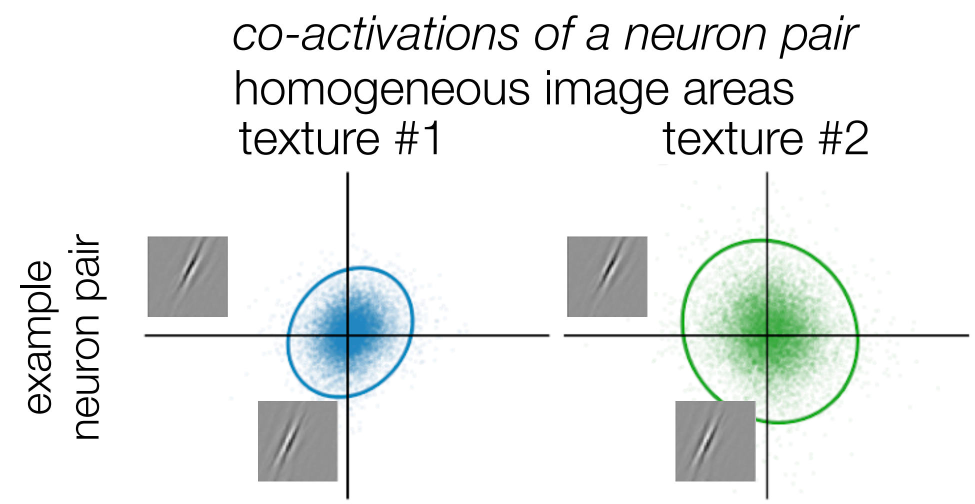
response variability ⇔ subjective uncertainty

Orbán et al (2016) Neuron

Contextual prior

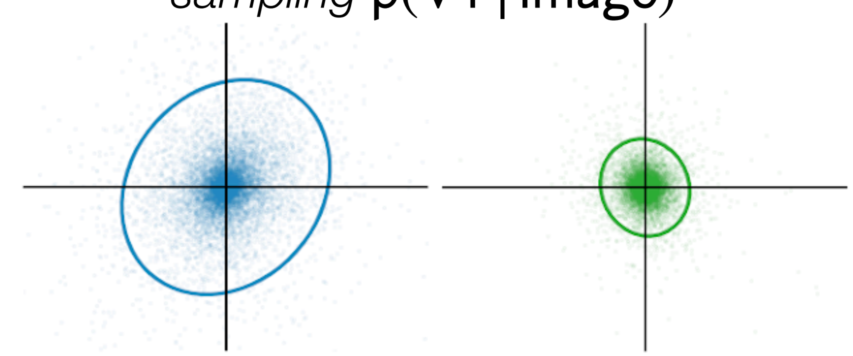
learning exploits regularities, regularities change from context-to context





neural ensembles represent distributions
⇝ context-dependent correlations in priors

sampling p(V1 | image)



response correlations ⇔ correlations in priors

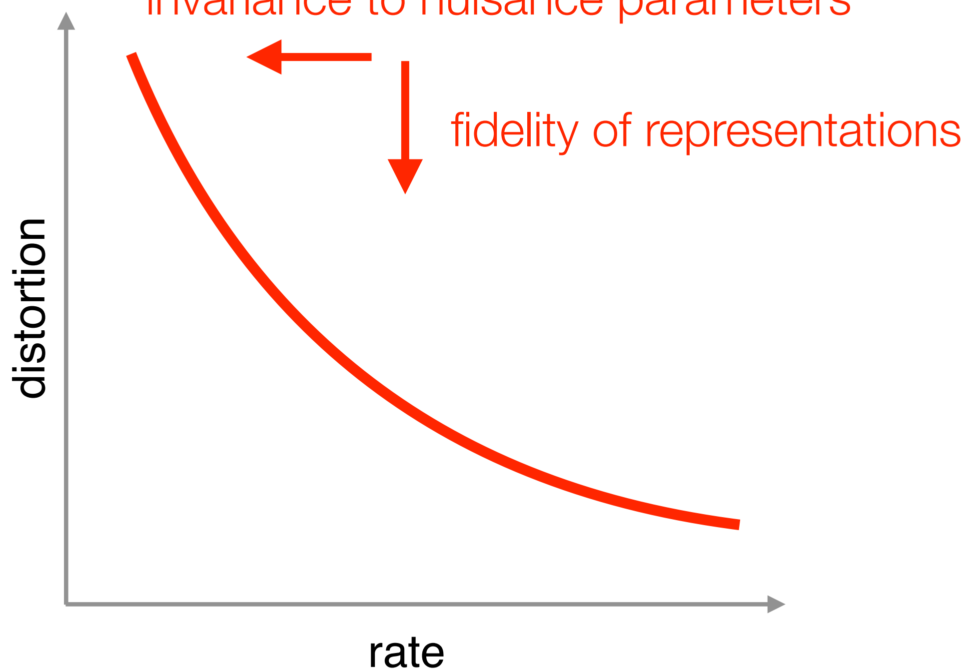
Bányai et al (2019) PNAS

Interpreting the neural code: ML insights

Hierarchical inference

learning captures invariances increasingly complex invariances
⤳ gradual compression

theory of lossy compression
invariance to nuisance parameters



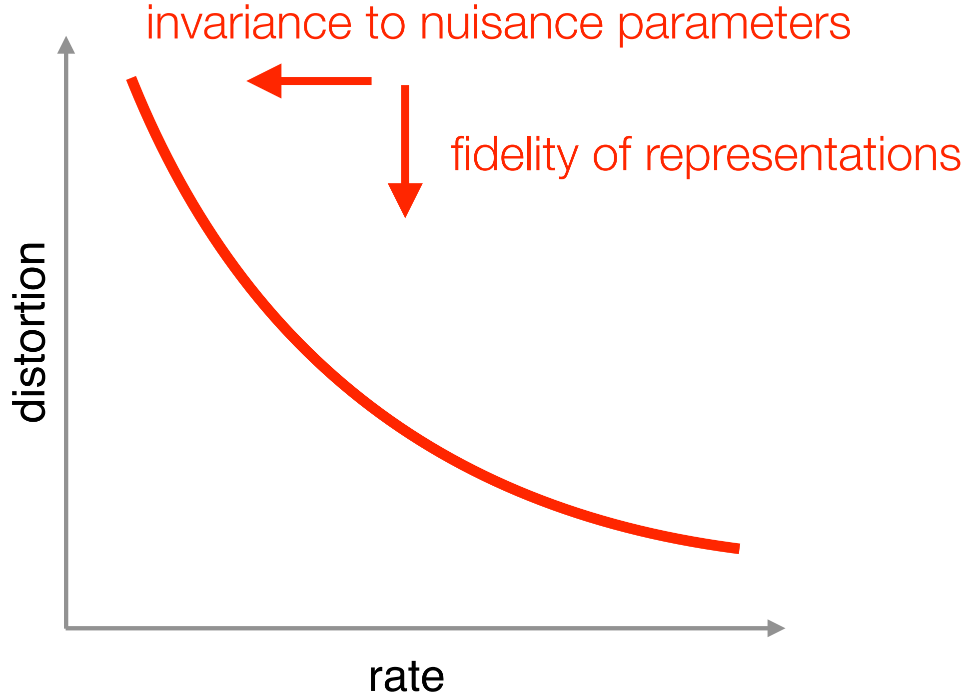
Nagy et et al (2020) PLoS CB
Meszéna et et al (2022) NeurIPS
Csikor et et al, in prep

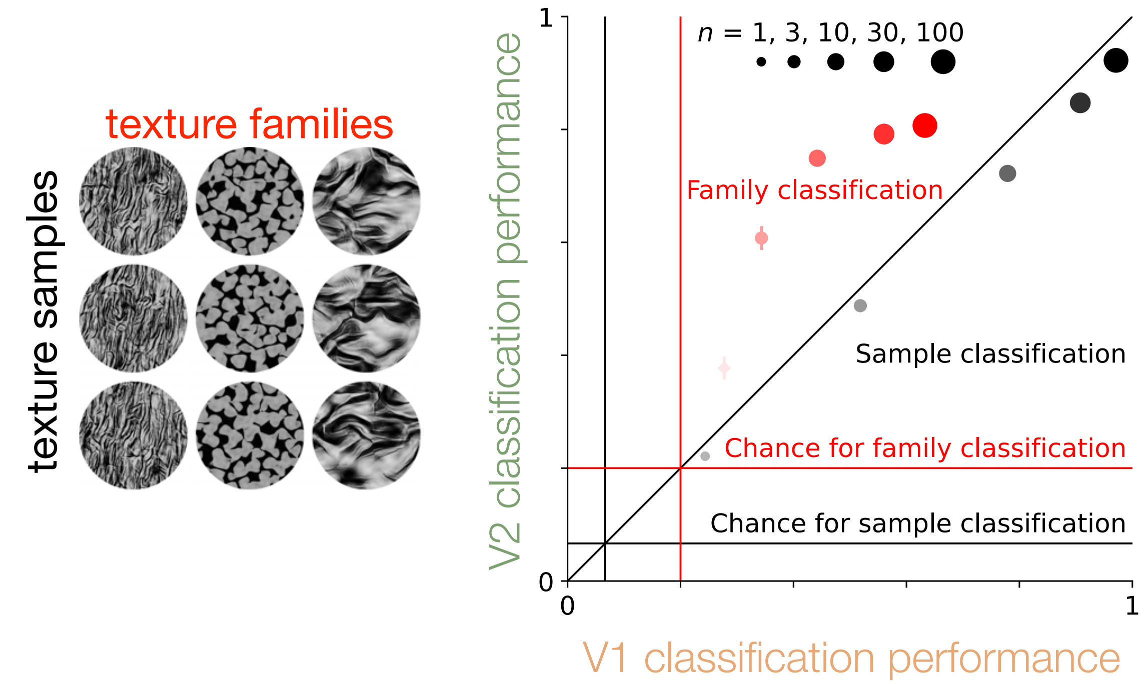
Interpreting the neural code: ML insights

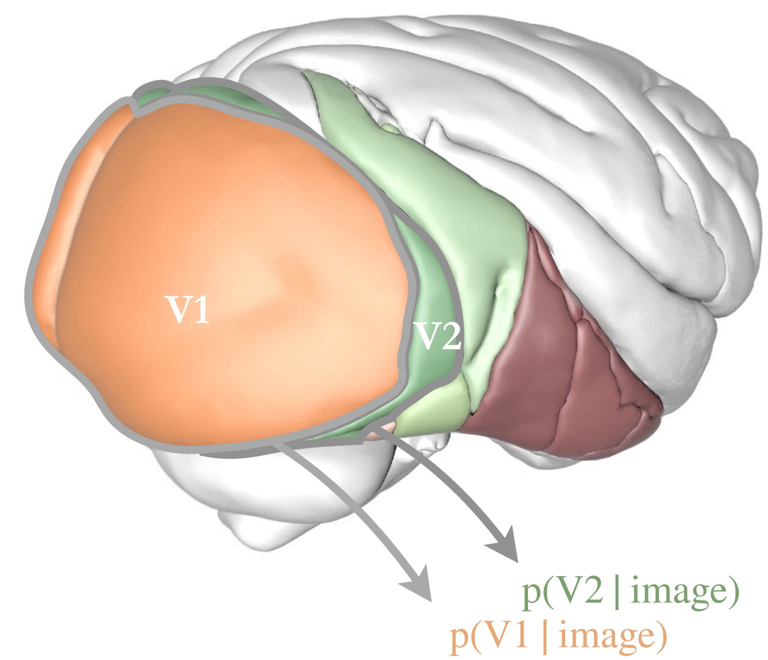
Hierarchical inference

learning captures invariances increasingly complex invariances
⤳ gradual compression

theory of lossy compression
invariance to nuisance parameters







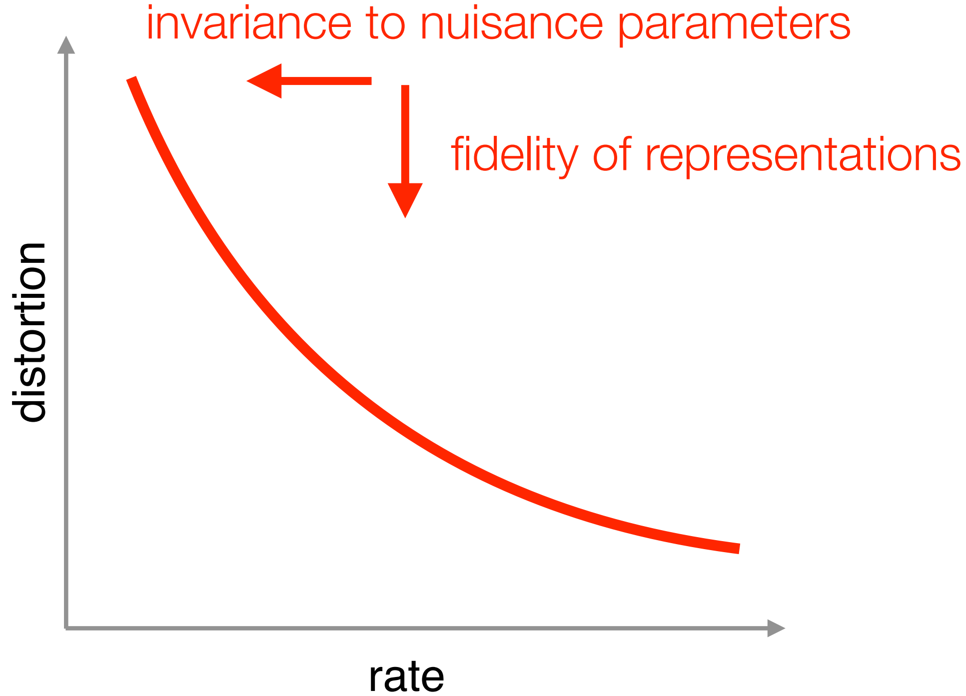
Nagy et et al (2020) PLoS CB
Meszéna et et al (2022) NeurIPS
Csikor et et al, in prep

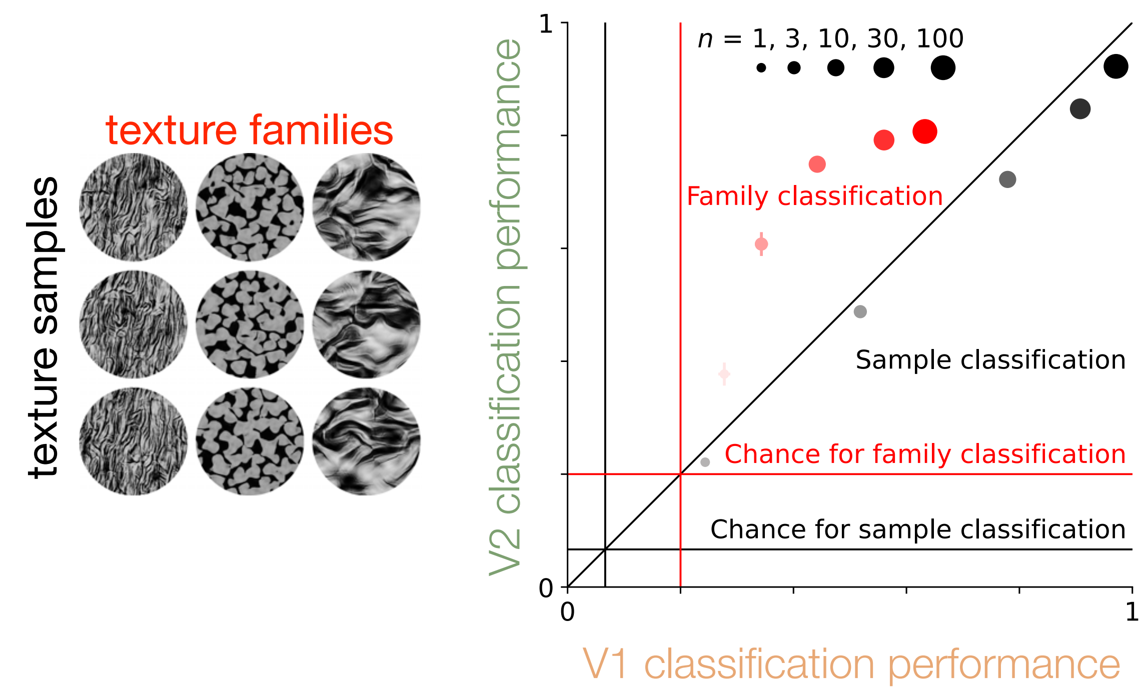
Interpreting the neural code: ML insights

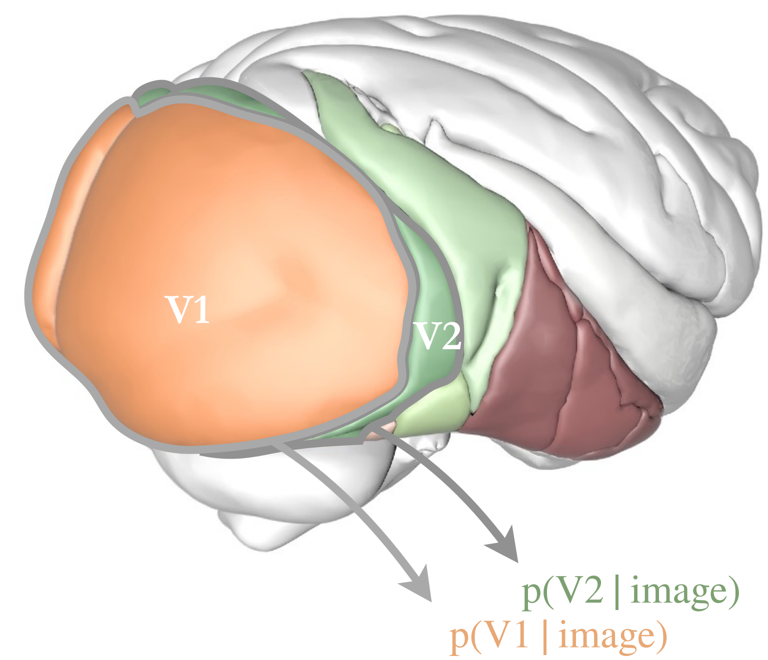
Hierarchical inference

learning captures invariances increasingly complex invariances
⤳ gradual compression

theory of lossy compression
invariance to nuisance parameters







optimally losing information during processing

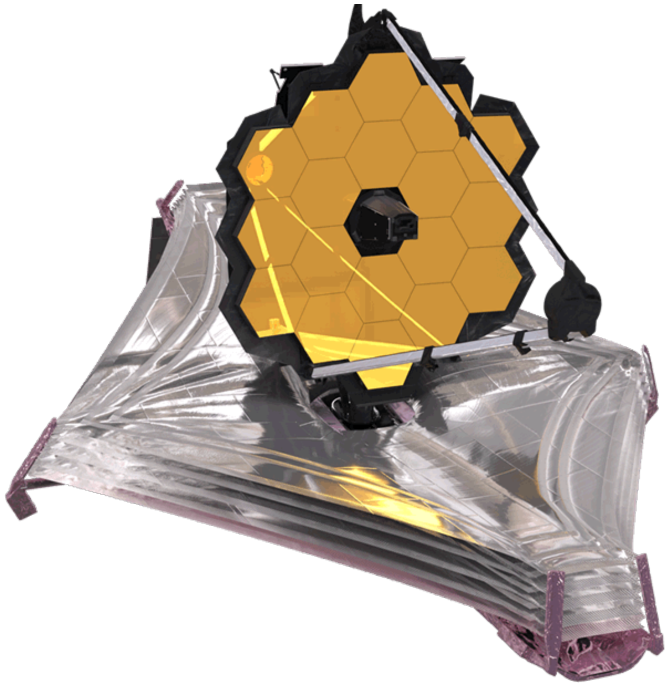
Nagy et et al (2020) PLoS CB
Meszéna et et al (2022) NeurIPS
Csikor et et al, in prep

Outlook: A testbed for theories

$$ELBO = \mathbb{E}_{q(V1\,|\,\mathrm{image},V2)}[p(\mathrm{image}\,|\,V1)] + \mathrm{KL}\,[q(V2\,|\,\mathrm{image})\,||\,p(V2)] + \mathbb{E}_{q(V2\,|\,\mathrm{image})}\left[\mathrm{KL}\,[q(V1\,|\,\mathrm{image},V2)\,||\,p(V1\,|\,V2)]\right]$$

Outlook: A testbed for theories

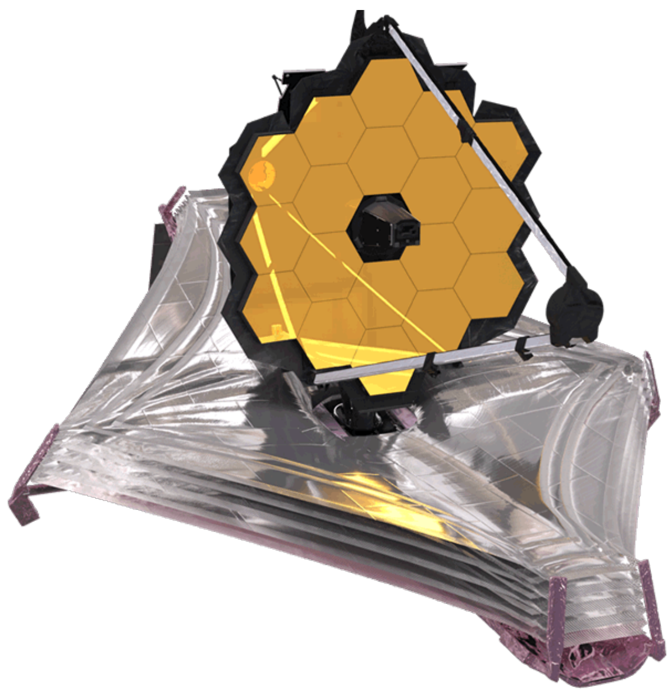
$$ELBO = \mathbb{E}_{q(V1\,|\,\text{image},V2)}[p(\text{image}\,|\,V1)] + \mathrm{KL}\,[q(V2\,|\,\text{image})\,||\,p(V2)] + \mathbb{E}_{q(V2\,|\,\text{image})}\big[\mathrm{KL}\,[q(V1\,|\,\text{image},V2)\,||\,p(V1\,|\,V2)]\big]$$



Space Telescope Science Institute Office of Public Outreach

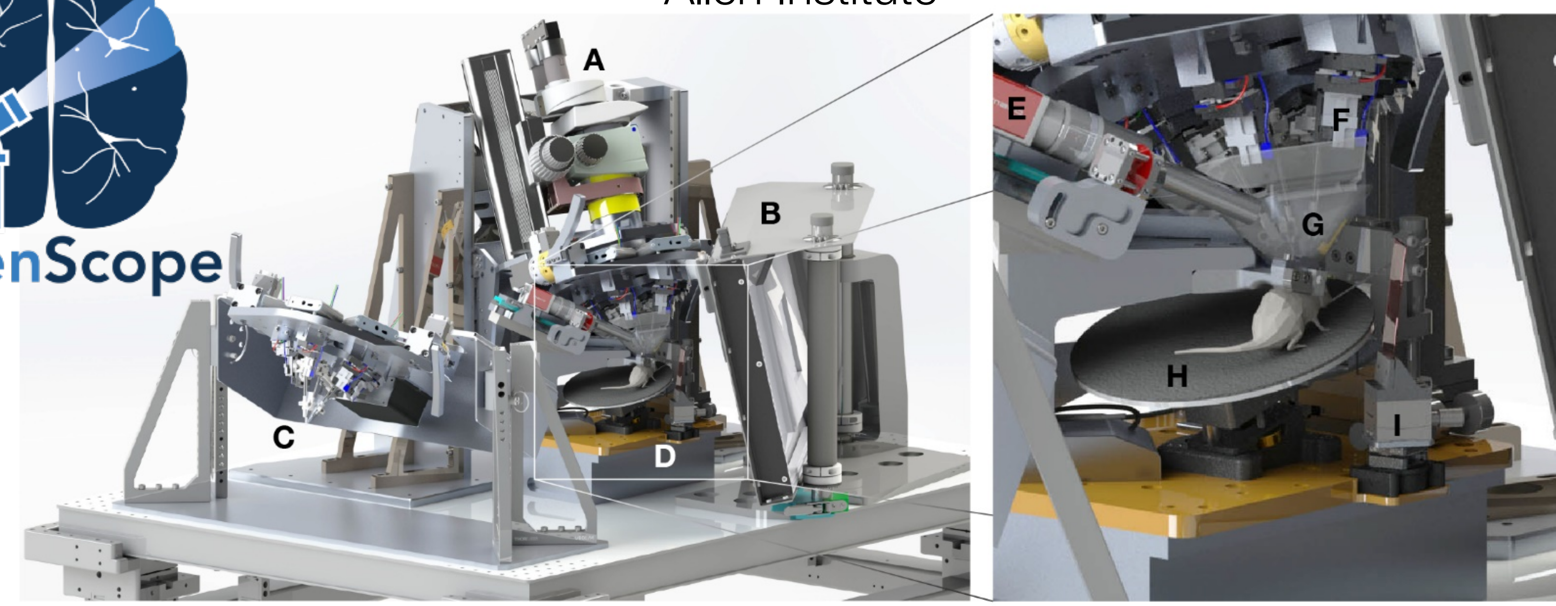
Outlook: A testbed for theories

$$ELBO = \mathbb{E}_{q(V1\,|\,\text{image},V2)}[p(\text{image}\,|\,V1)] + \text{KL}\,[q(V2\,|\,\text{image})\,||\,p(V2)] + \mathbb{E}_{q(V2\,|\,\text{image})}\big[\text{KL}\,[q(V1\,|\,\text{image},V2)\,||\,p(V1\,|\,V2)]\big]$$



OpenScope

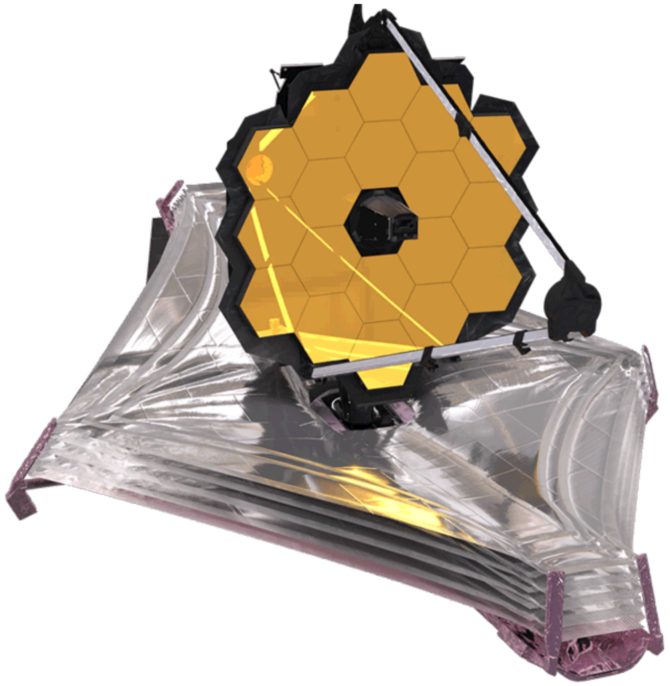
Allen Institute



Space Telescope Science Institute Office of Public Outreach

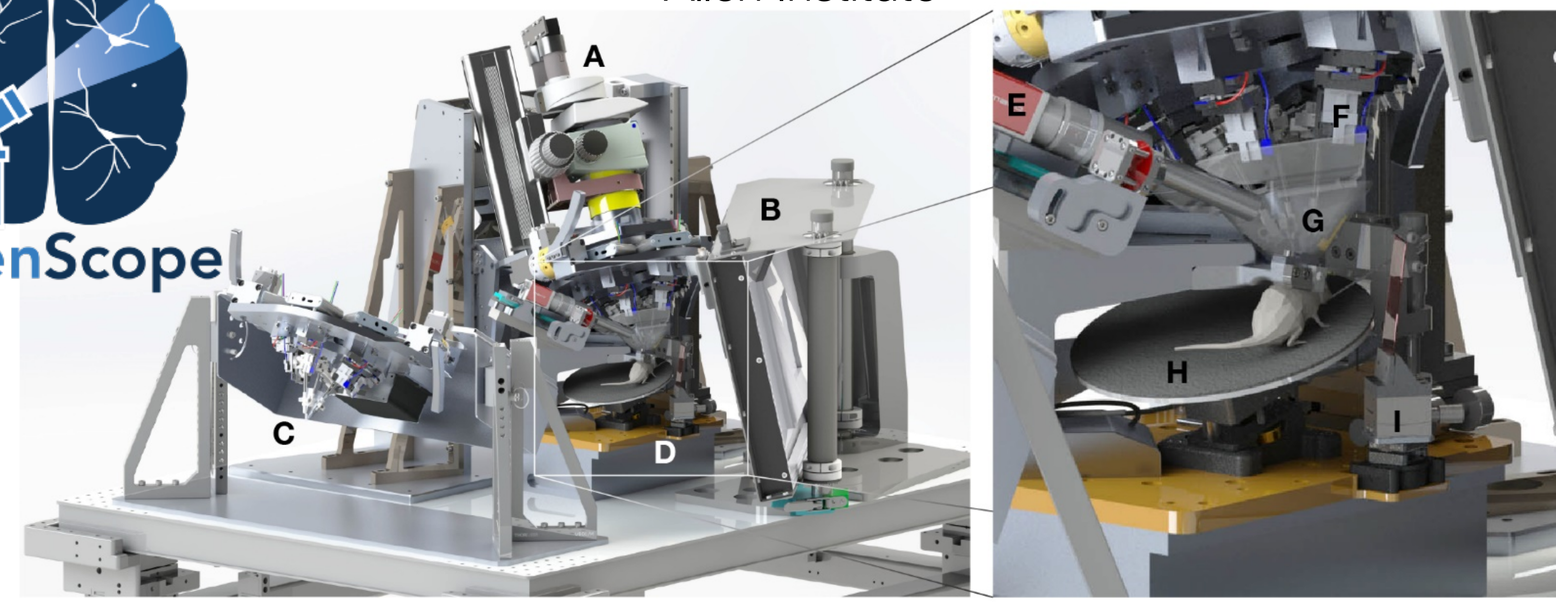
Outlook: A testbed for theories

$$ELBO = \mathbb{E}_{q(V1\,|\,\text{image},V2)}[p(\text{image}\,|\,V1)] + \mathrm{KL}\,[q(V2\,|\,\text{image})\,||\,p(V2)] + \mathbb{E}_{q(V2\,|\,\text{image})}\big[\mathrm{KL}\,[q(V1\,|\,\text{image},V2)\,||\,p(V1\,|\,V2)]\big]$$

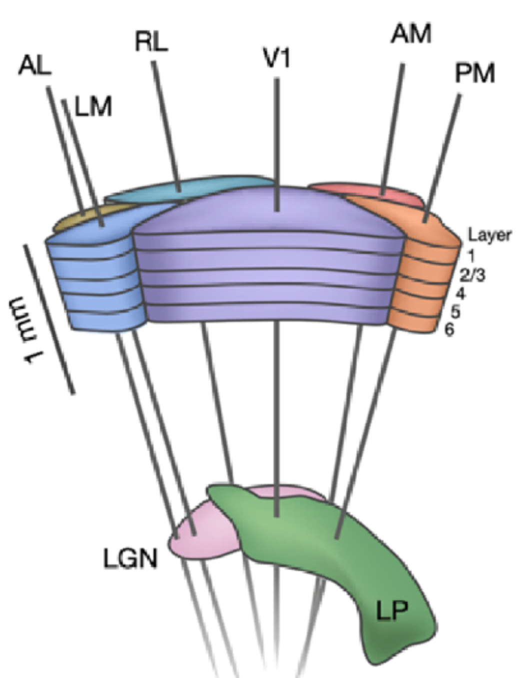


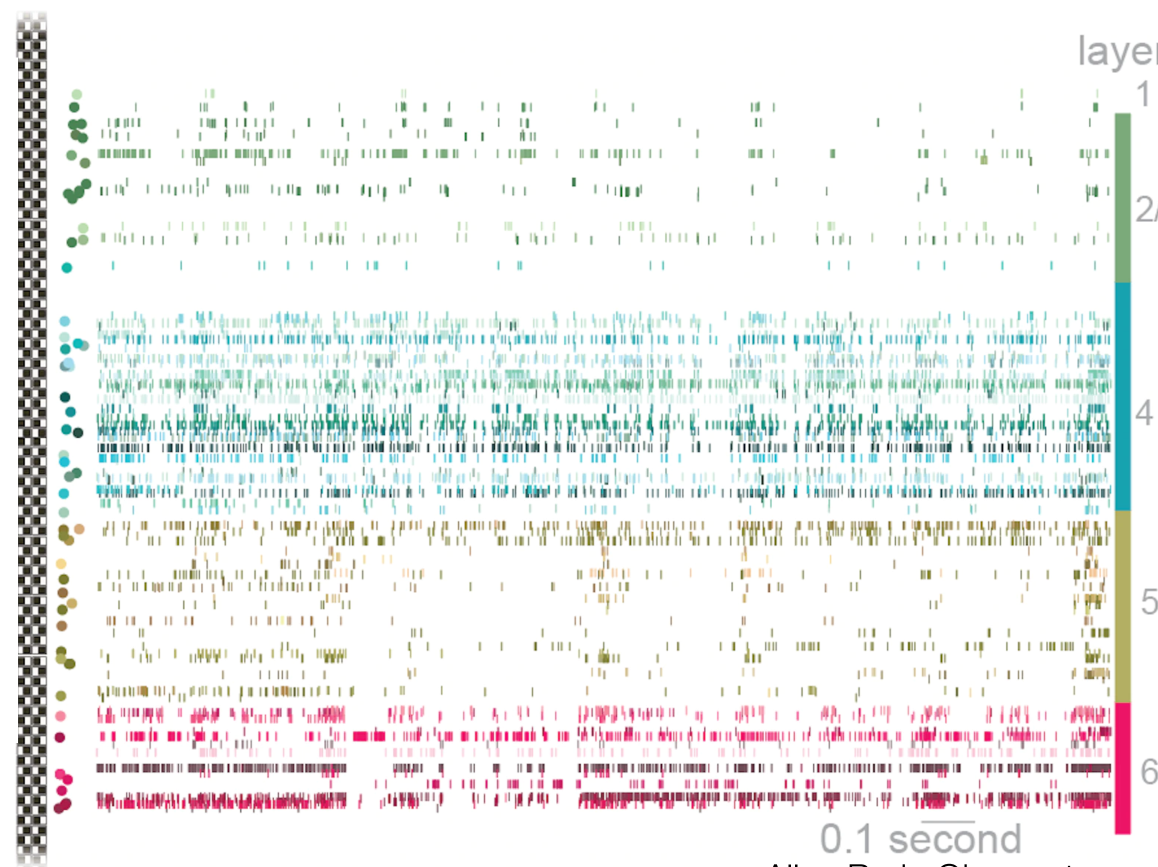
OpenScope

Allen Institute



Space Telescope Science Institute Office of Public Outreach





source: Allen Brain Observatory

Acknowledgements

- Ferenc Csikor (postdoc)
- Márton Hajnal (PhD student)
- Balázs Meszéna (postdoc)
- Anna Székely (PhD student, BME)
- Zsombor Ungvárszki (PhD student, Lyon)
- Zsombor Szabó (research associate)
- Martos Domonkos (research associate)

Students

- Laci Freund (BME)
- Virág Horváth (ELTE)

Funding

HFSP

Human Frontiers Science Program

ARTIFICIAL INTELLIGENCE National Laboratory

HUNGARIAN BRAIN RESEARCH PROGRAM
NEMZETI AGYKUTATÁSI PROGRAM

NATIONAL RESEARCH, DEVELOPMENT AND INNOVATION OFFICE

image credits

Line Icons Pro from Noun Project (CC BY 3.0)
Teewara soontorn from Noun Project (CC BY 3.0)
istockphoto/Getty images