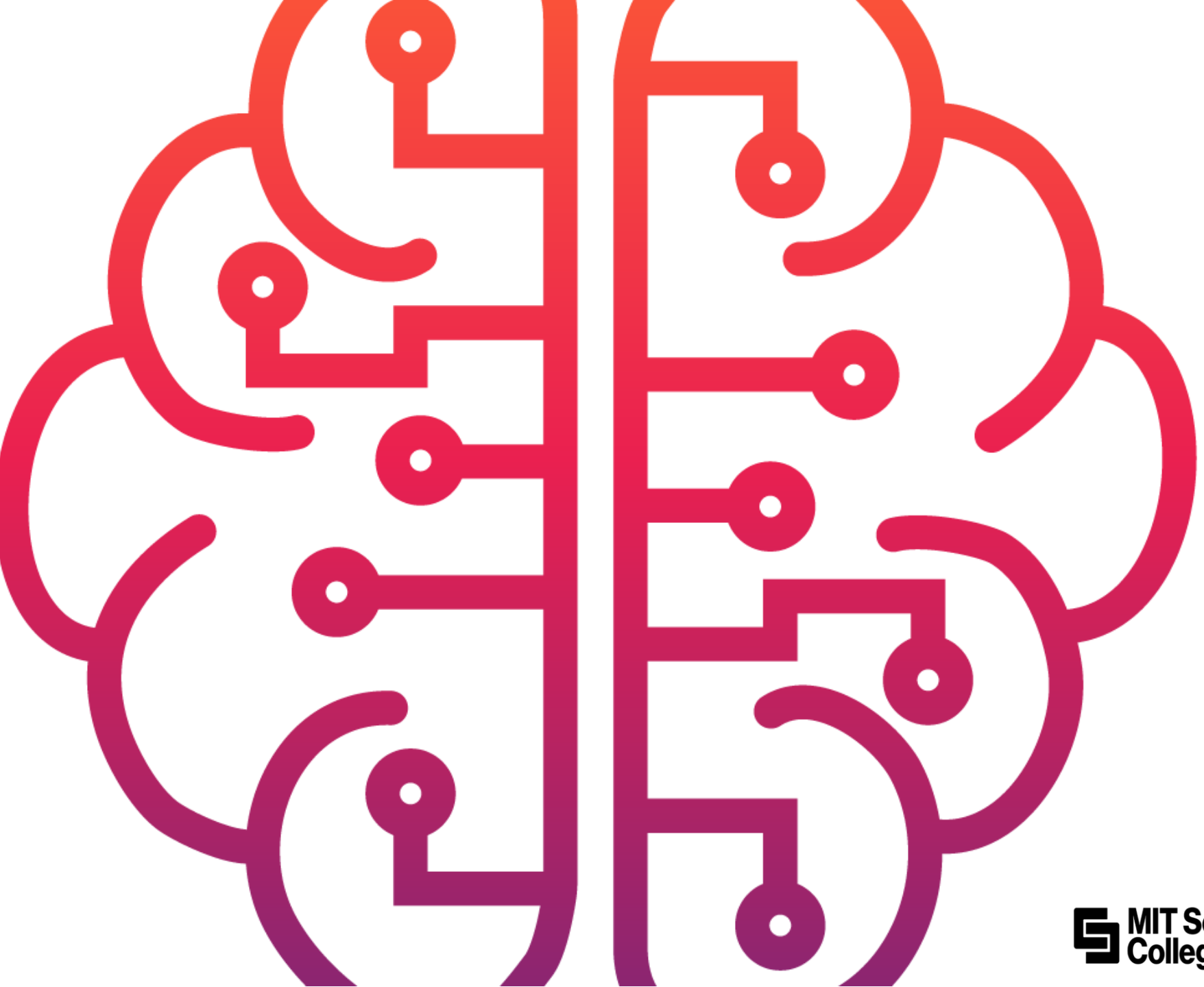


AI@MIT Panel Discussion

Martin Schrimpf (Quest)
Katherine Fairchild (Quest)
Ila Fiete (BCS)

MIT Schwarzman College of Computing

MIT Quest for Intelligence

Advancing Integrative Models of Human Intelligence with Brain-Score

Martin Schrimpf

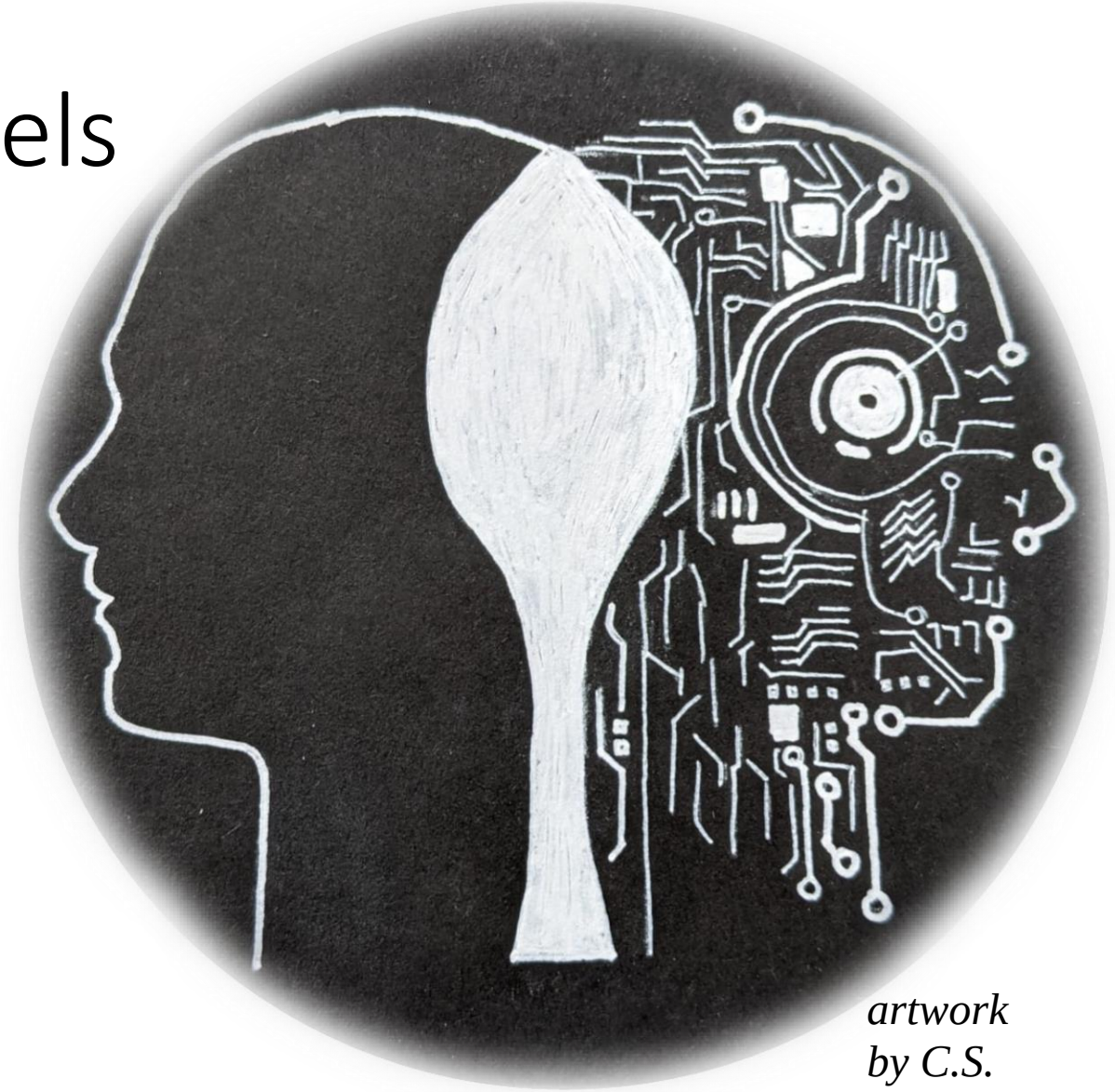
research scientist, MIT

tenure-track assistant professor, EPFL

msch@mit.edu

@martin_schrimpf

mastodon.social/
@mschrimpf



artwork by C.S.

MIT Quest for Intelligence

MIT → EPFL


Jonas Kubilius


Tiago Marques


Kohitij Kar


Franziska Geiger


Joel Dapello


Paul McGrath


Mike Ferguson


Jim DiCarlo


Greta Tuckute


Carina Kauf


Idan Blank

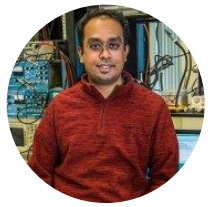

Corey Ziemba


Tony Movshon


Pouya Bashivan


Kailyn Schmidt


Jon Prescott-Roy


Ratan Murty

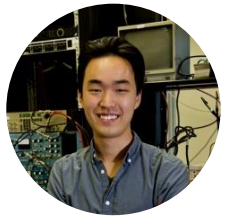

Ev Fedorenko


Chengxu Zhuang


Eshed Margalit


Rishi Rajalingham


Dan Yamins


Michael Lee


Anna Ivanova

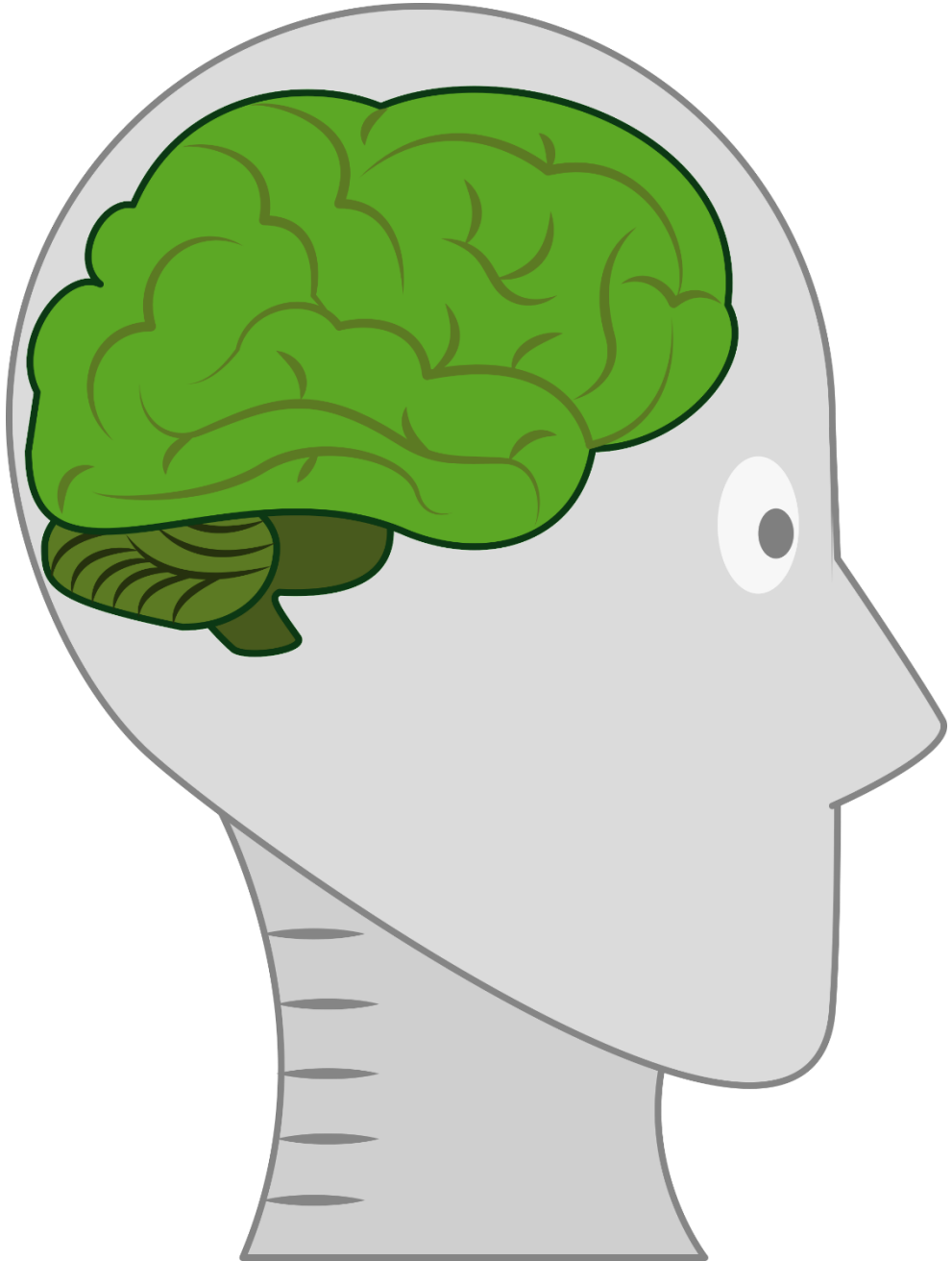

Eghbal Hosseini


Robert Ajemian


Nancy Kanwisher


Josh Tenenbaum

Funding: brain+cognitive sciences, McGovern Institute for Brain Research, Takeda, C-BRIC Center for Brain-Inspired Computing, MIT Quest for Intelligence



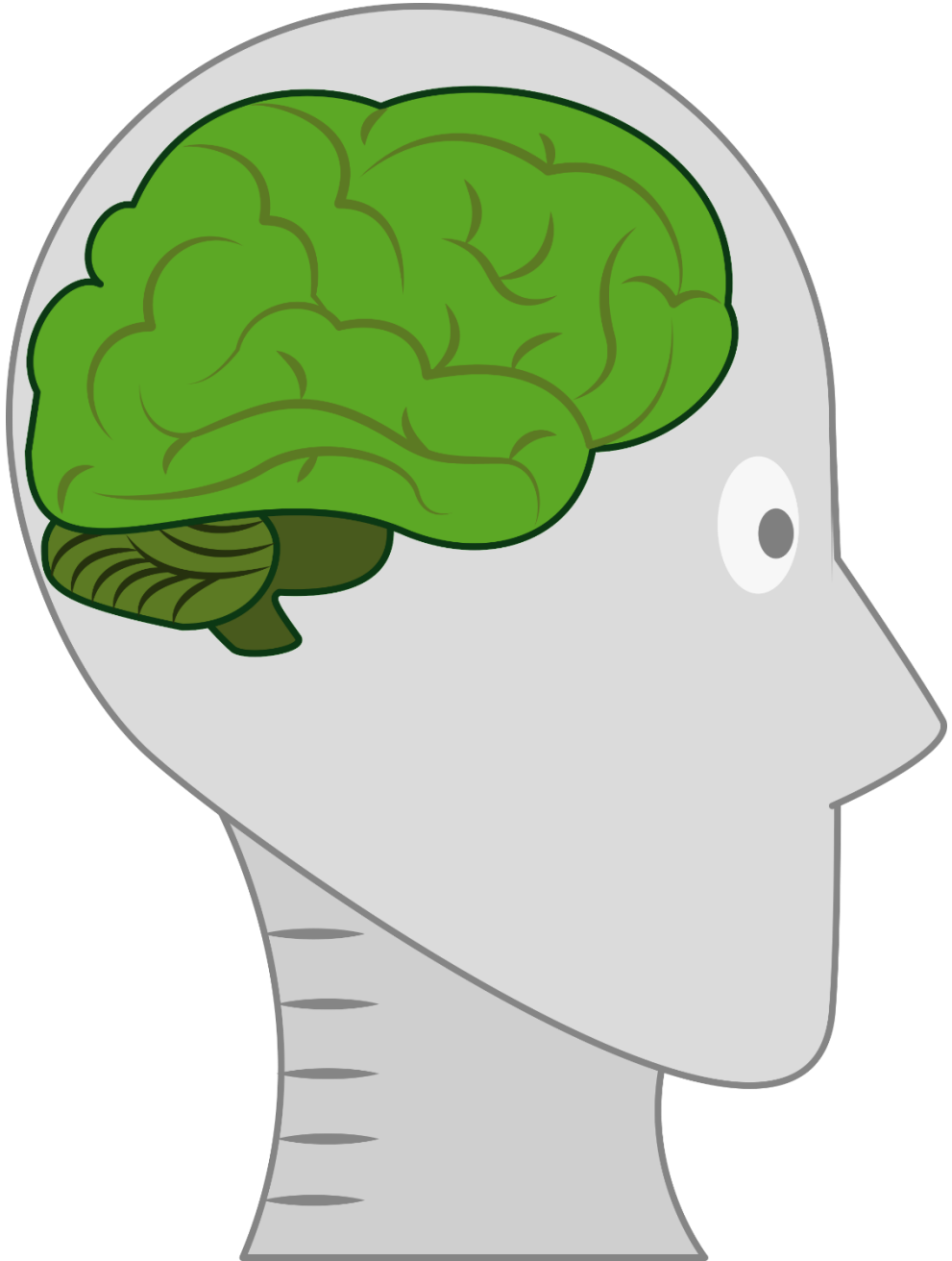
A core goal of our field:

model human intelligence and underlying neural mechanisms

Computational understanding of natural intelligence

Next-generation AI algorithms

Future clinical applications



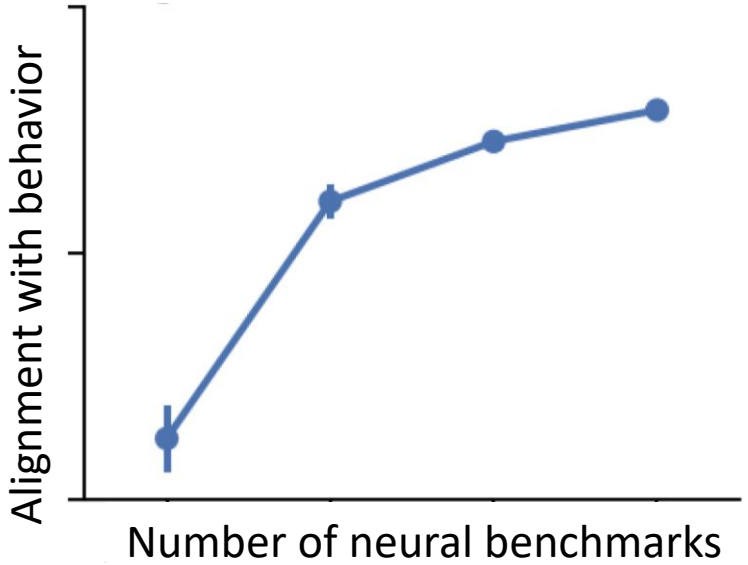
A core goal of our field:
model human intelligence and underlying neural mechanisms

How can we even tell we're making progress?

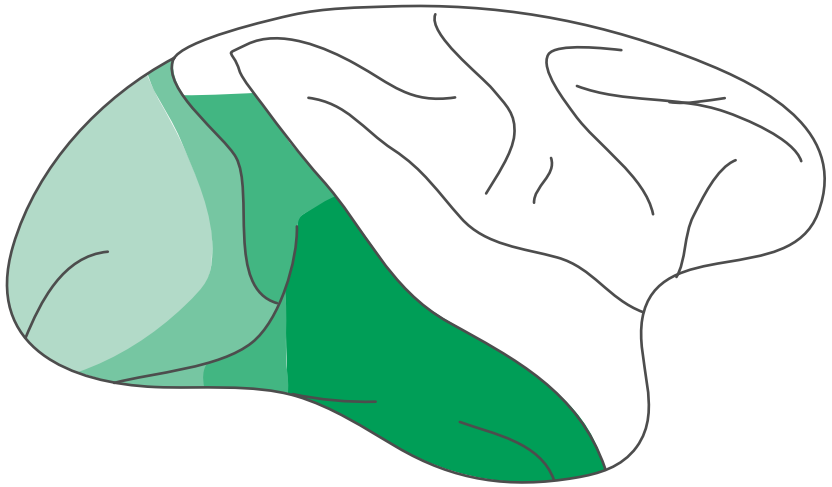
This talk: a proposal for <u>integrative model testing</u> in vision & language

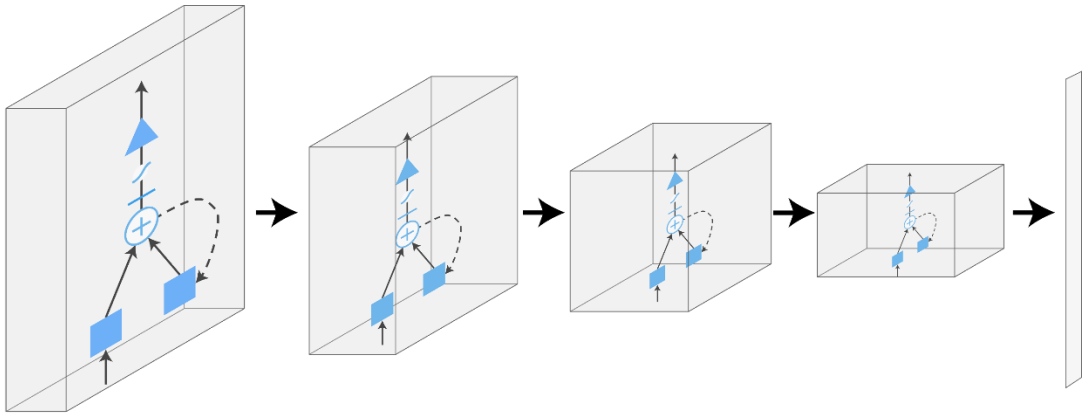
Modeling Primate Visual Intelligence

Evaluate model **alignment to <u>all</u> experiments** in the domain of visual intelligence.



More benchmarks => more likely to do well on the next experiment





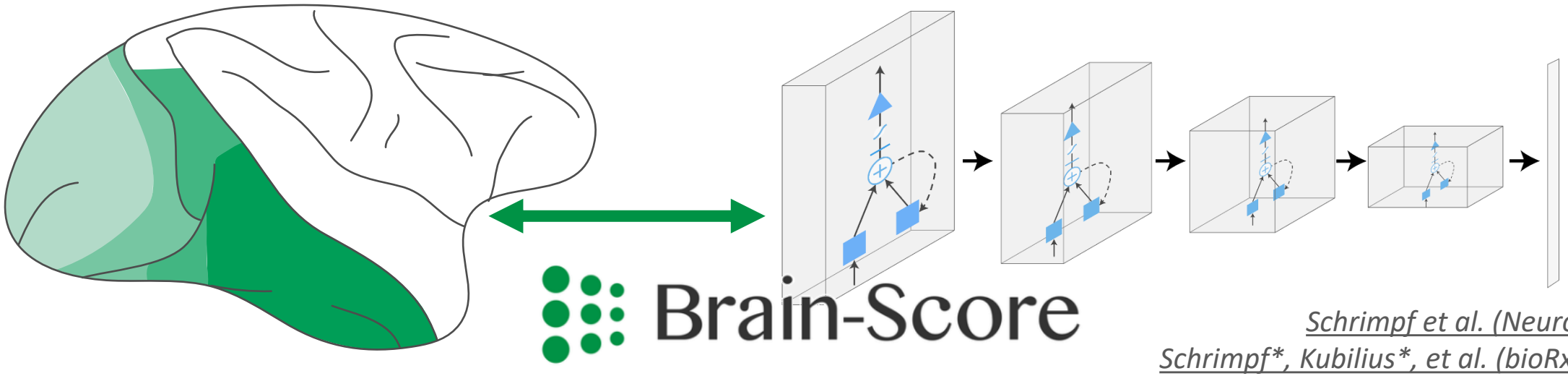
Schrimpf et al. (Neuron 2020)
Schrimpf, Kubilius*, et al. (bioRxiv 2018)*

Modeling Primate Visual Intelligence

Evaluate model **alignment to all experiments** in the domain of visual intelligence.

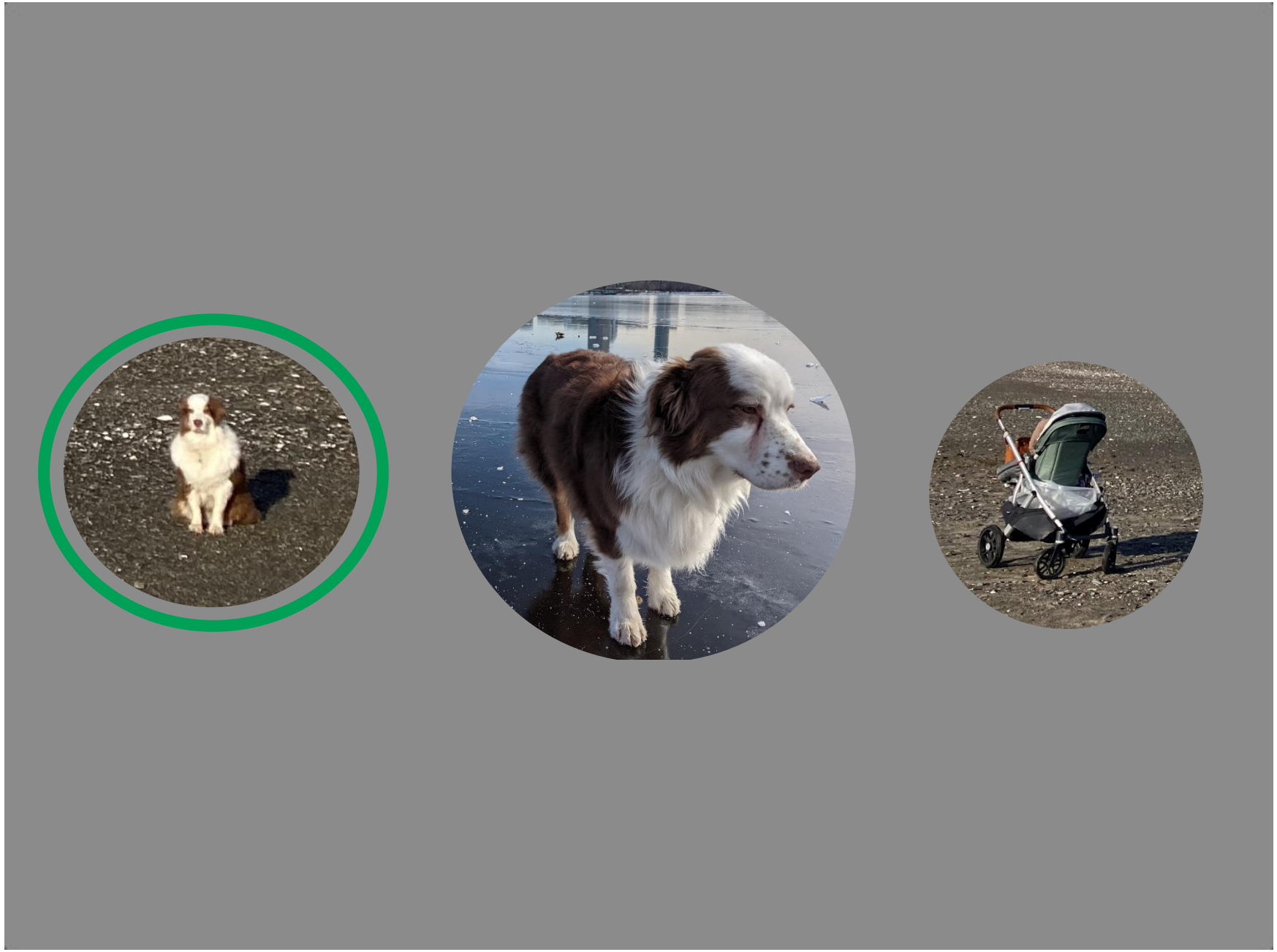
- Test <u>behavioral alignment</u> by showing the **same images** to humans and models and comparing their outputs

- Test <u>neural alignment</u> at the level of **spike rates** since they linearly predict visual behaviors *(*e.g. *Majaj*, Hong*, et al. 2015; I Gusti Bagus et al. 2022).* Include **non-human primates** since their visual behaviors and visual cortex are highly similar to humans (e.g. *Kriegeskorte et al. 2008; Rajalingham et al. 2018*)



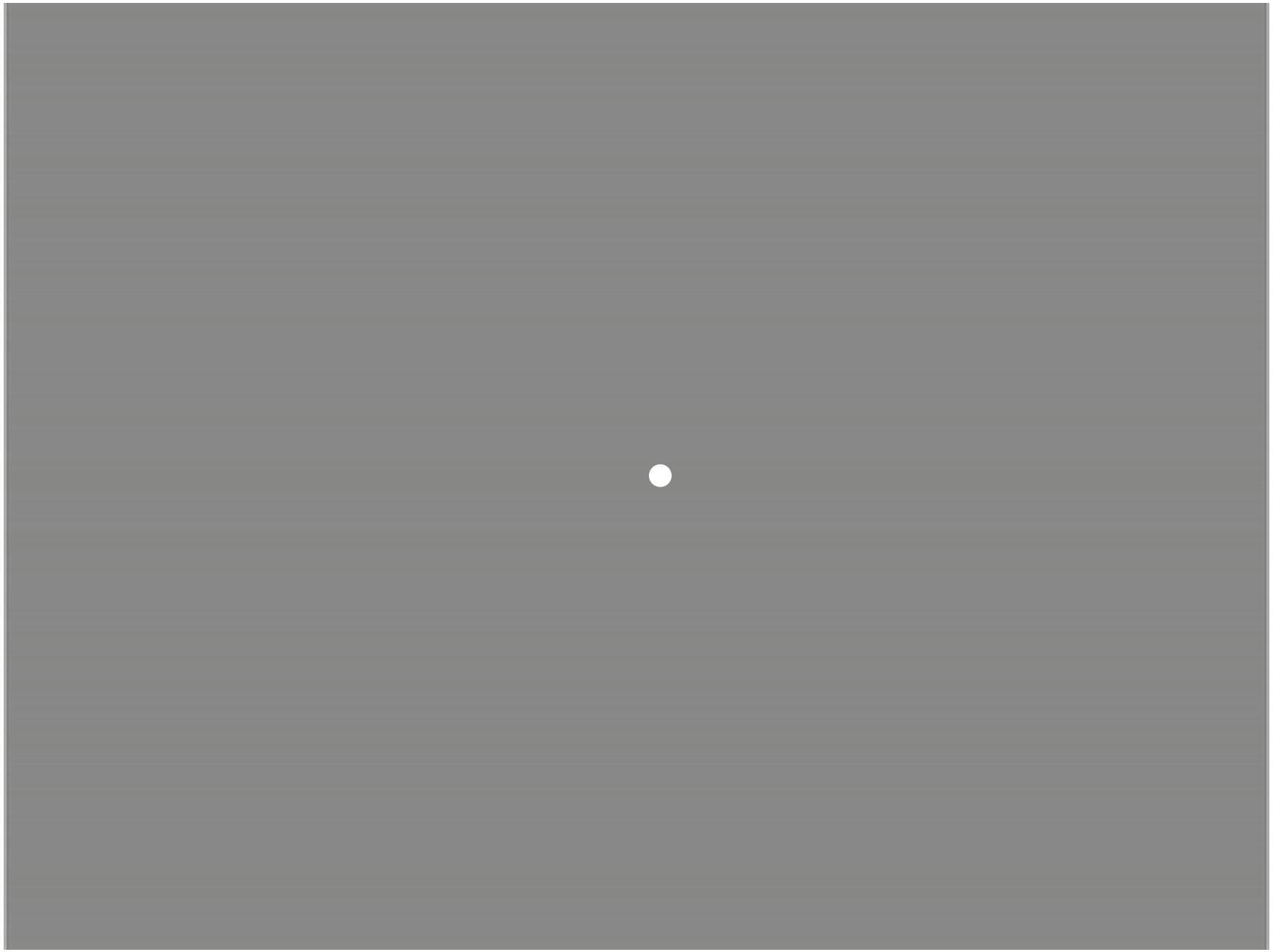
Schrimpf et al. (Neuron 2020)
Schrimpf, Kubilius*, et al. (bioRxiv 2018)*

Behavioral benchmark



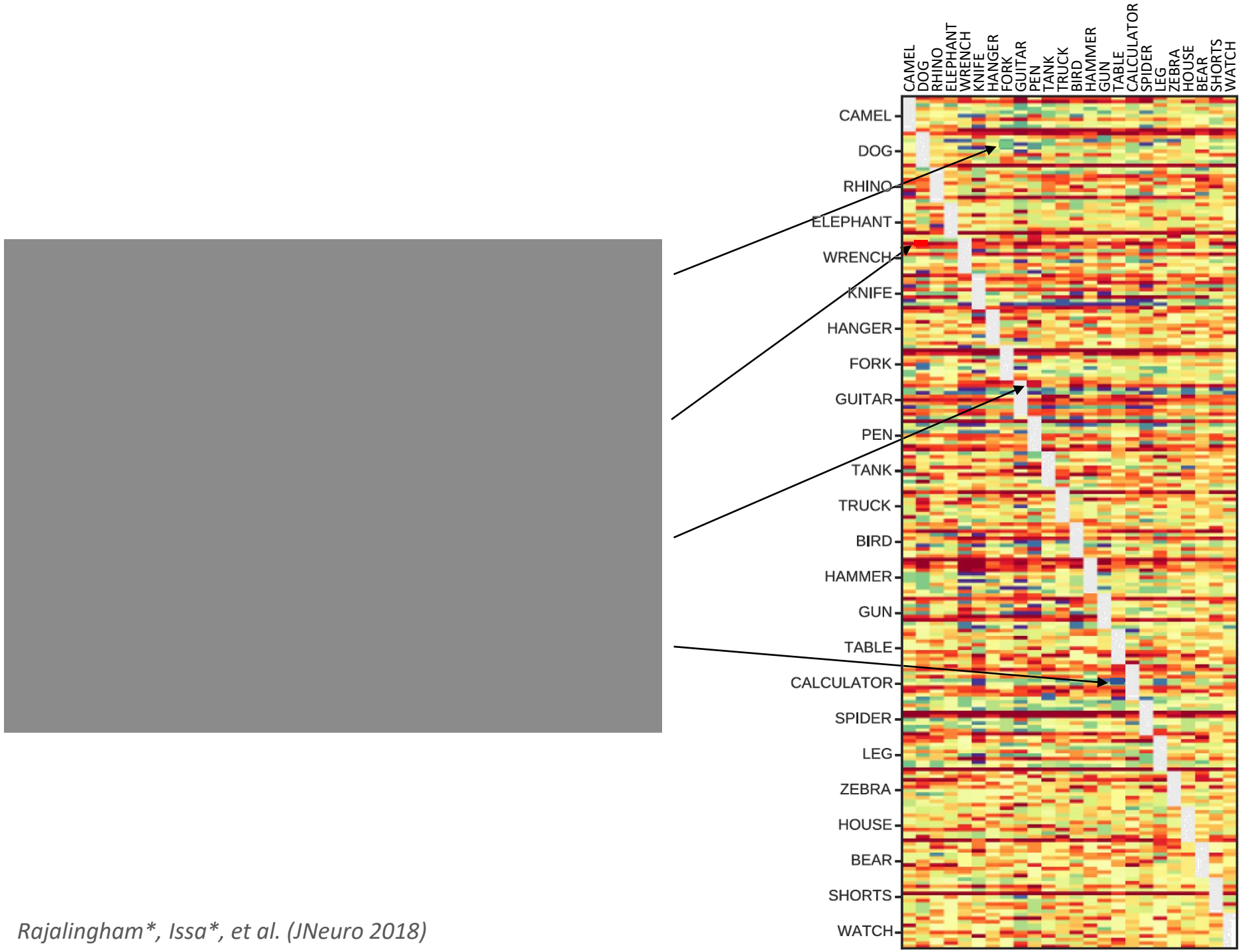
Rajalingham, Issa*, et al. (JNeuro 2018)*

Behavioral benchmark



Rajalingham, Issa*, et al. (JNeuro 2018)*

Behavioral benchmark



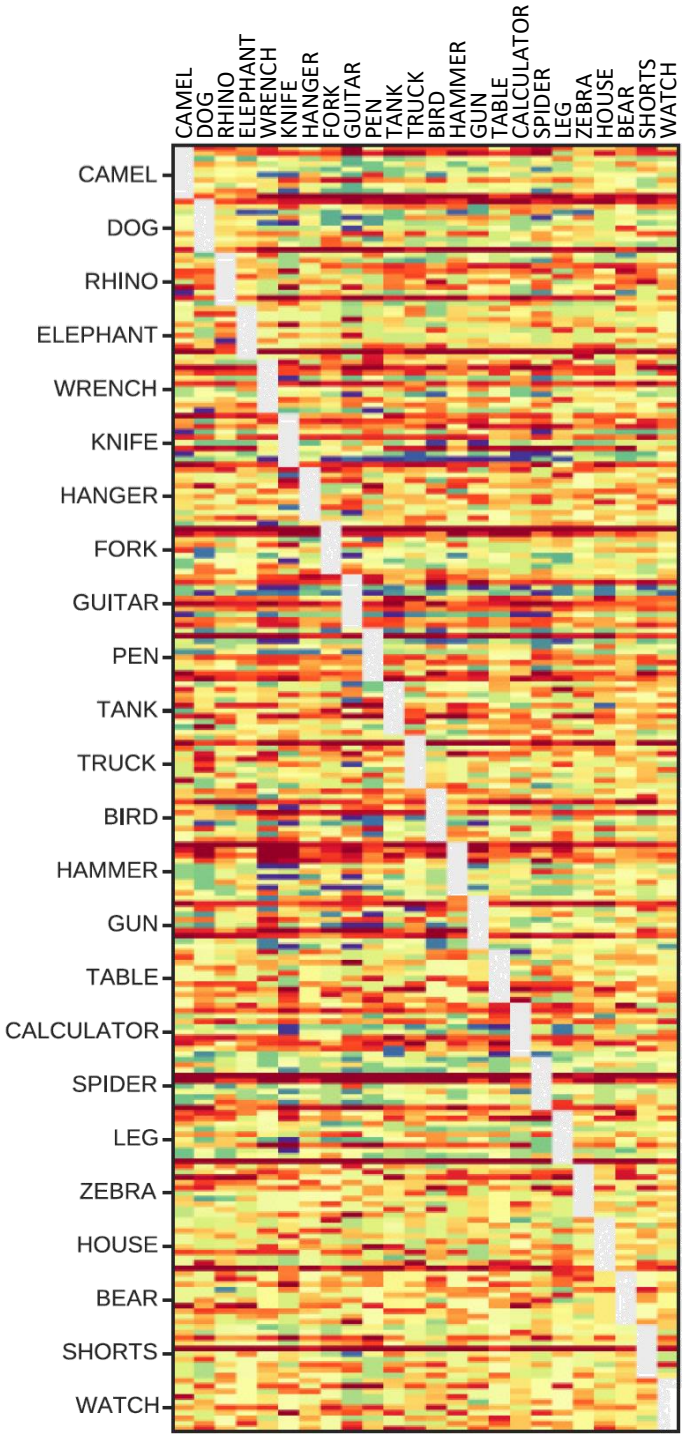
performance per binary categorization

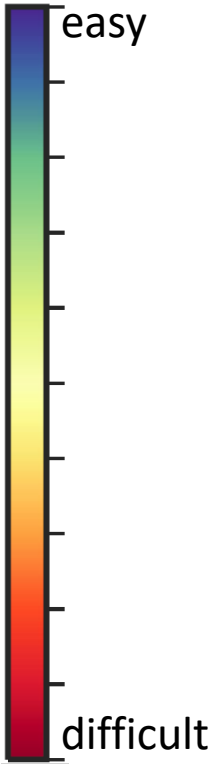
easy

difficult

Rajalingham, Issa*, et al. (JNeuro 2018)*

Behavioral benchmark

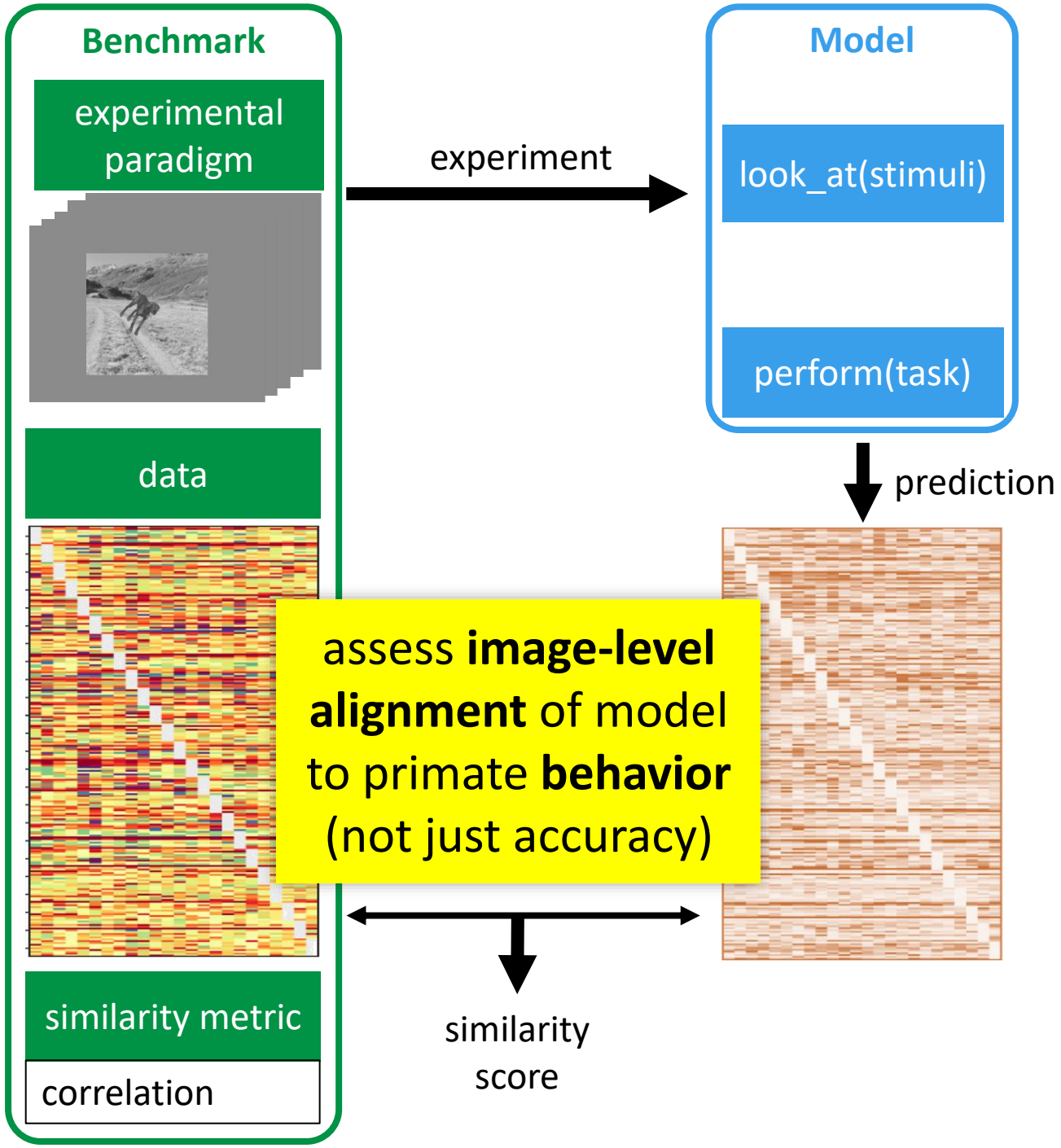




performance per binary categorization

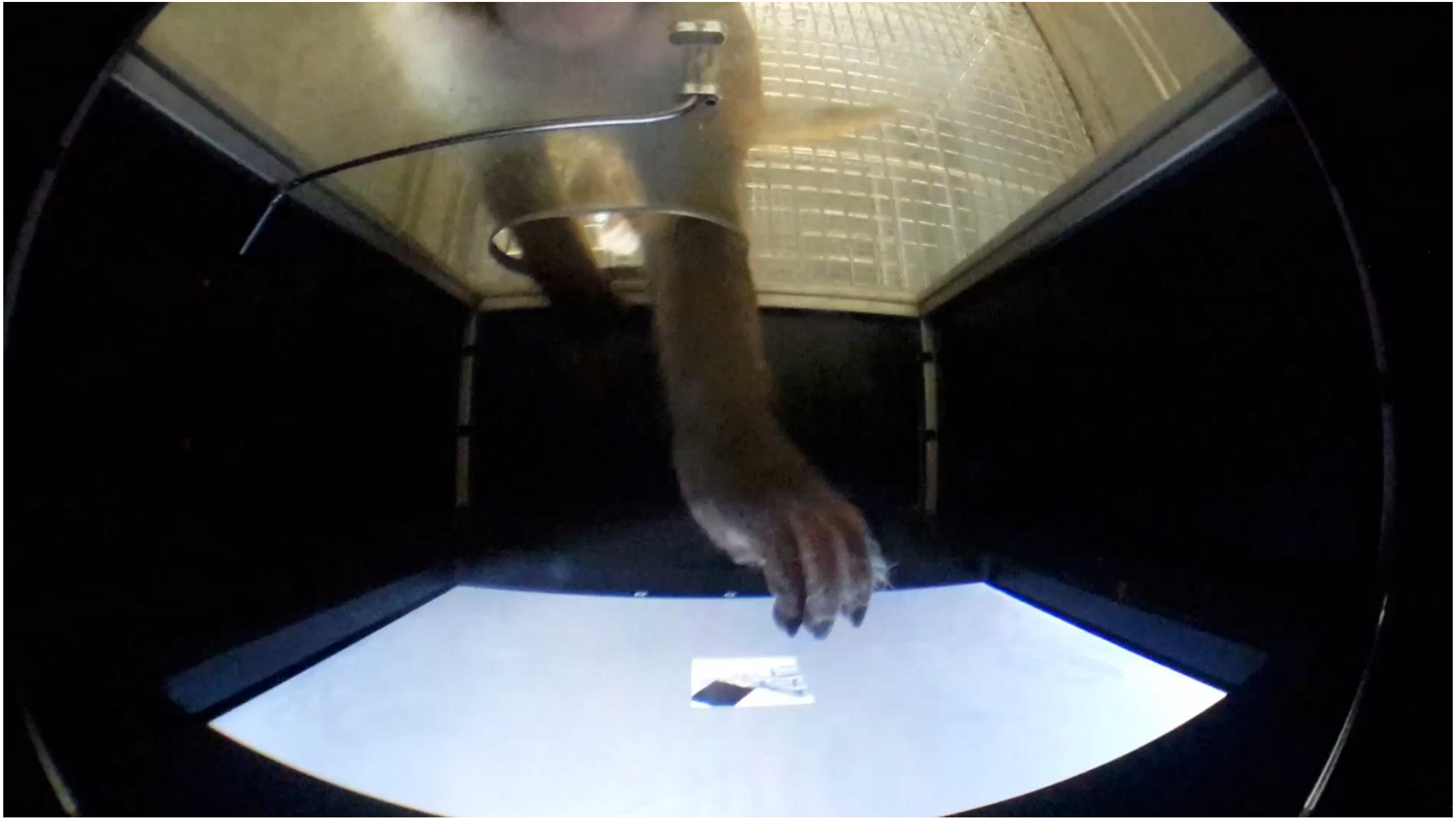
Rajalingham, Issa*, et al. (JNeuro 2018)*

Behavioral benchmark



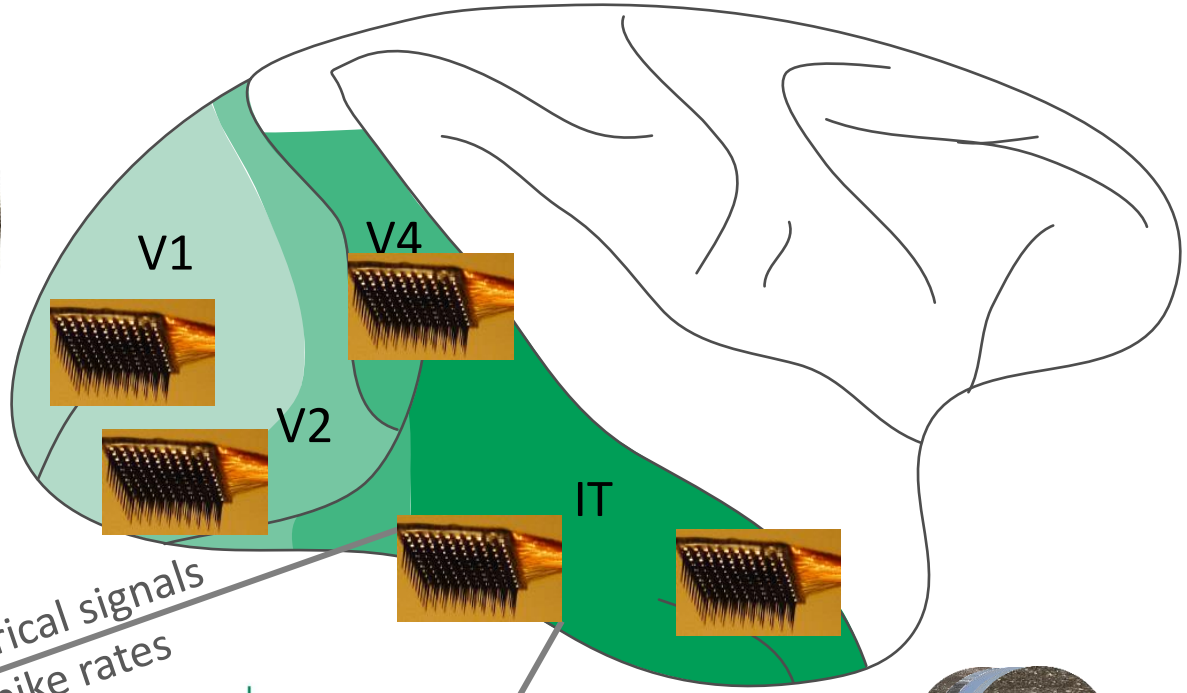
Neural benchmarks

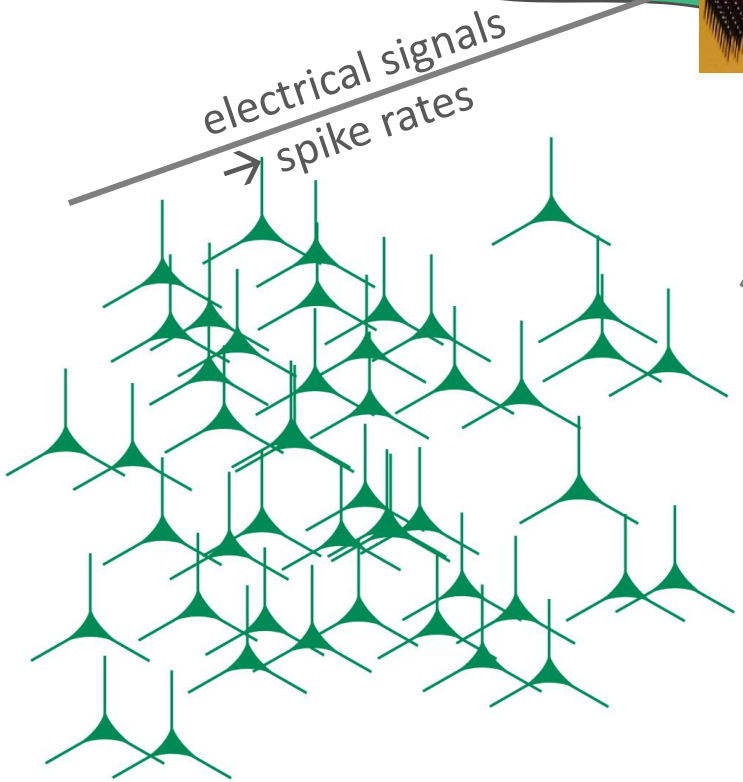
Neural benchmarks

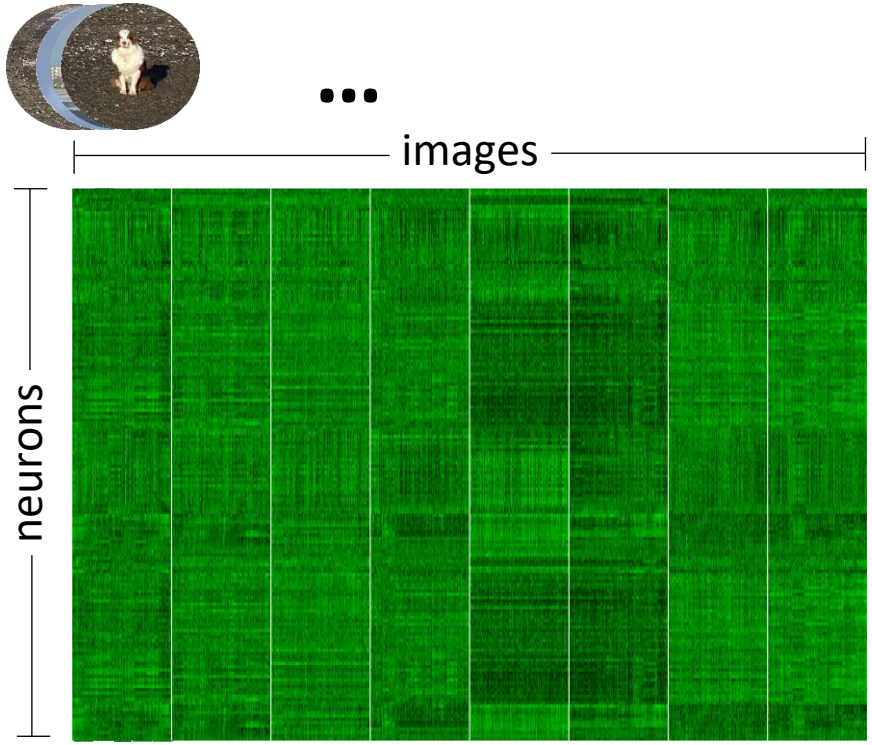


video courtesy of Kailyn Schmidt

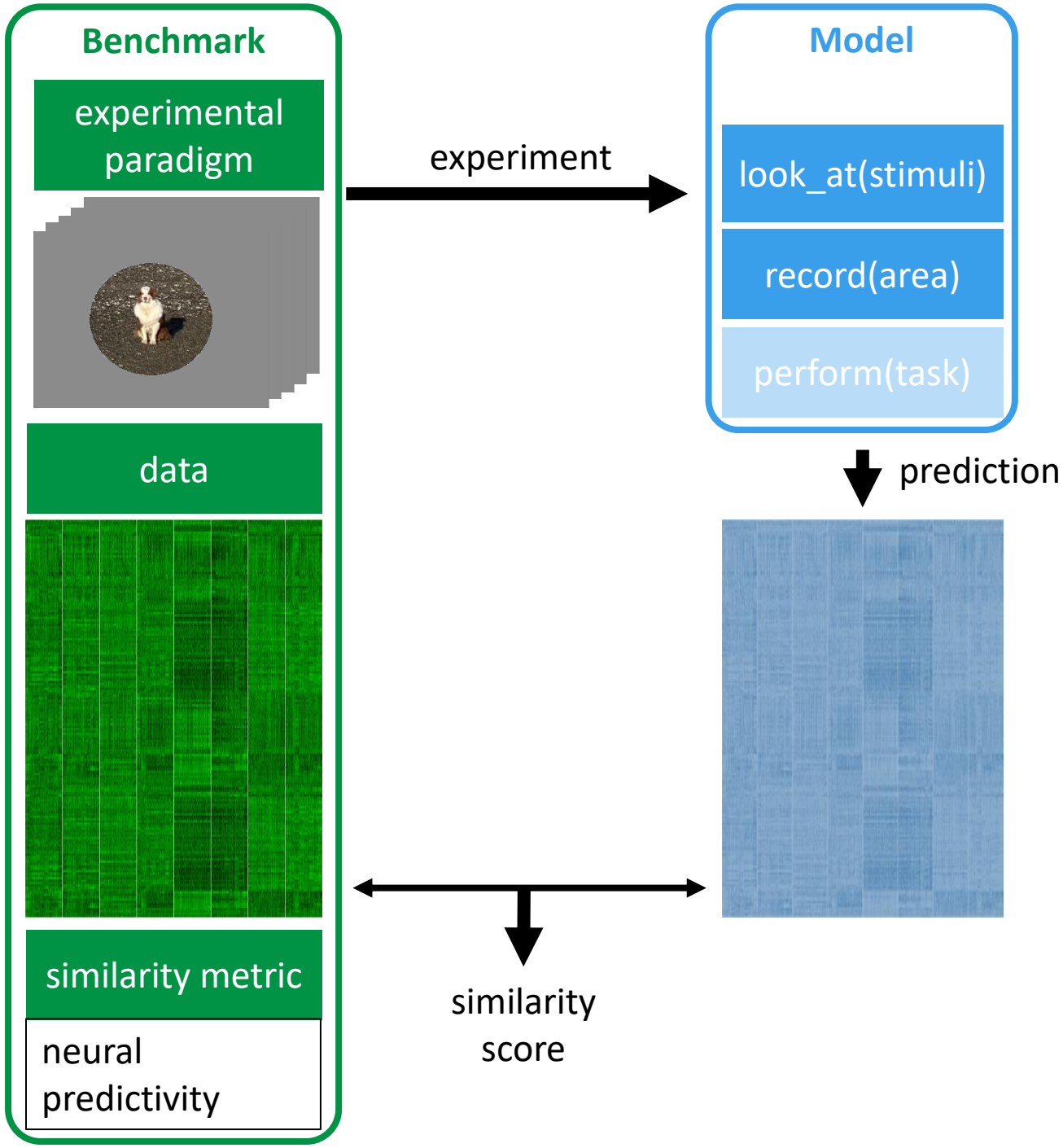
Neural benchmarks



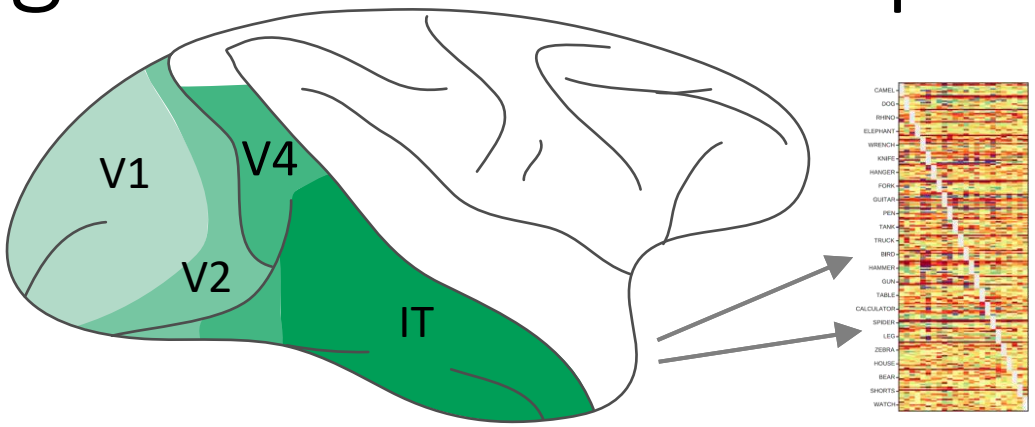


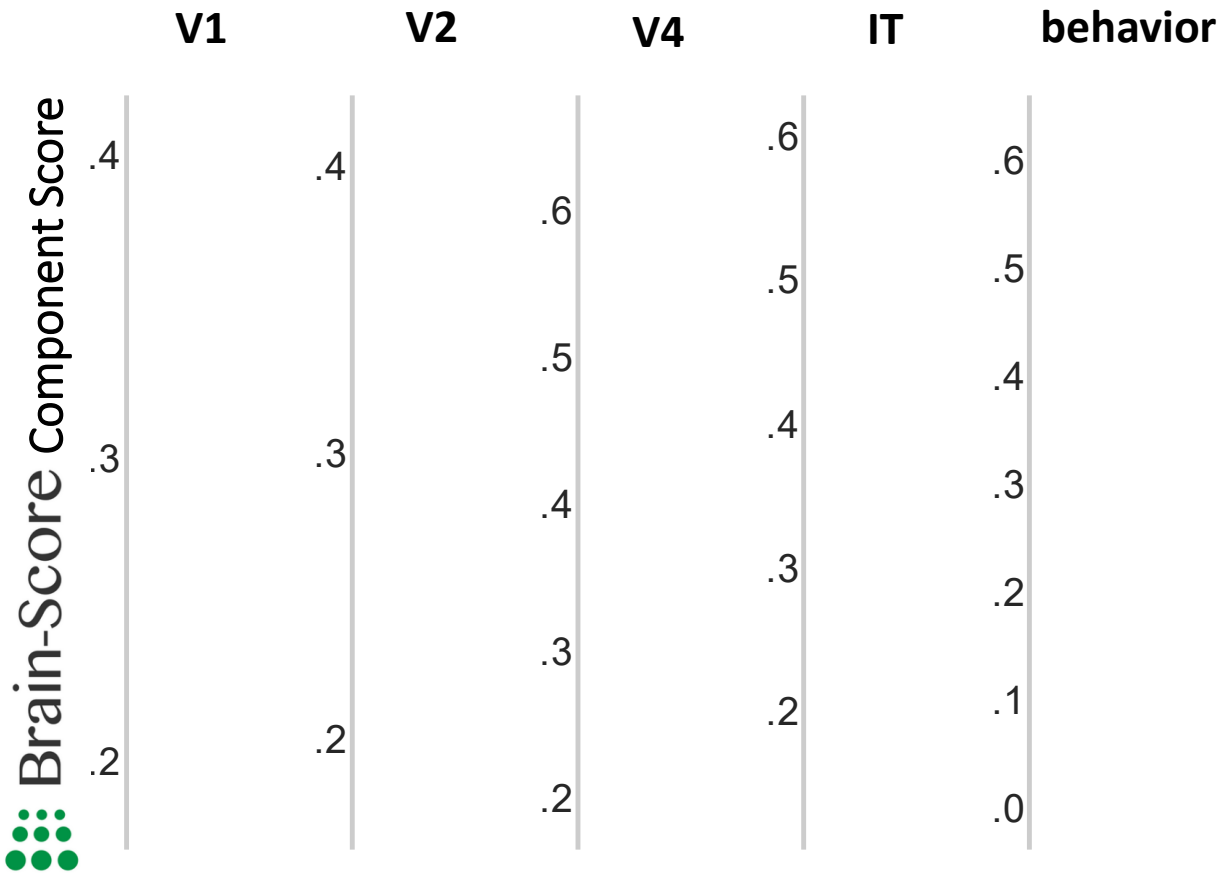


Neural benchmarks



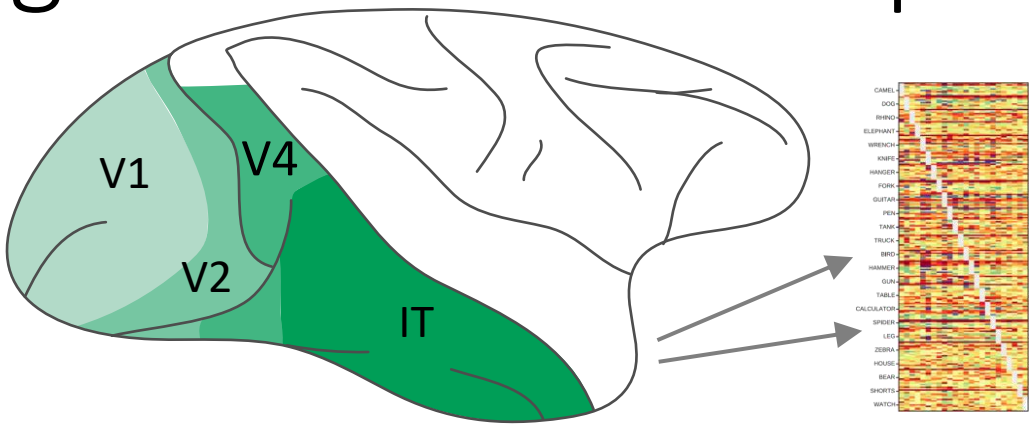
Integrative model comparison on Brain-Score

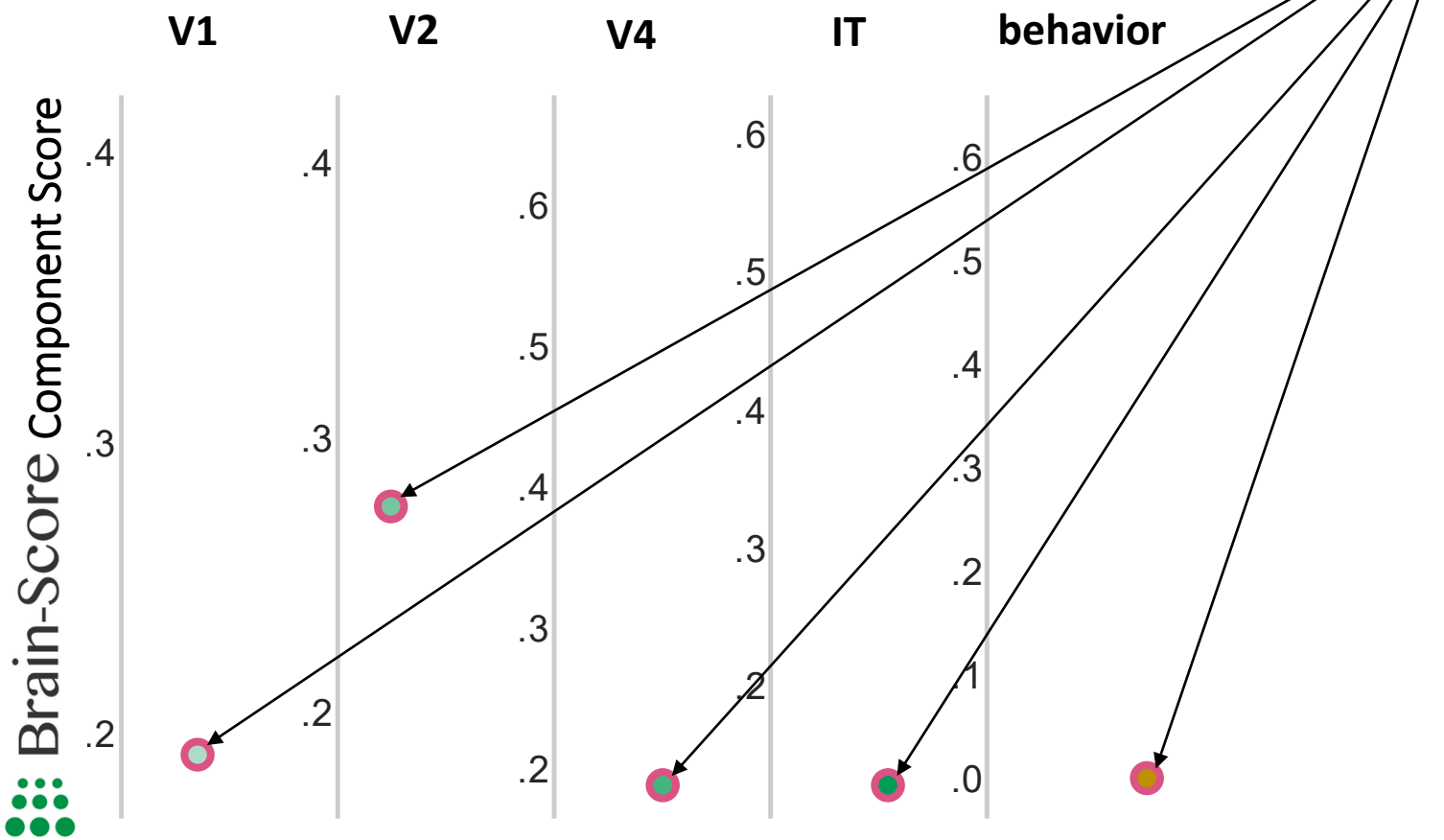




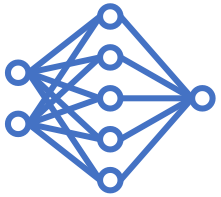
Schrimpf, Kubilius*, et al. (bioRxiv 2018)*
V1, V2 data: Freeman, Ziemba*, et al. (NatNeuro 2013)*
V4, IT data: Majaj, Hong*, et al. (JNeuro 2015)*
behavioral data: Rajalingham, Issa*, et al. (JNeuro 2018)*

Integrative model comparison on Brain-Score





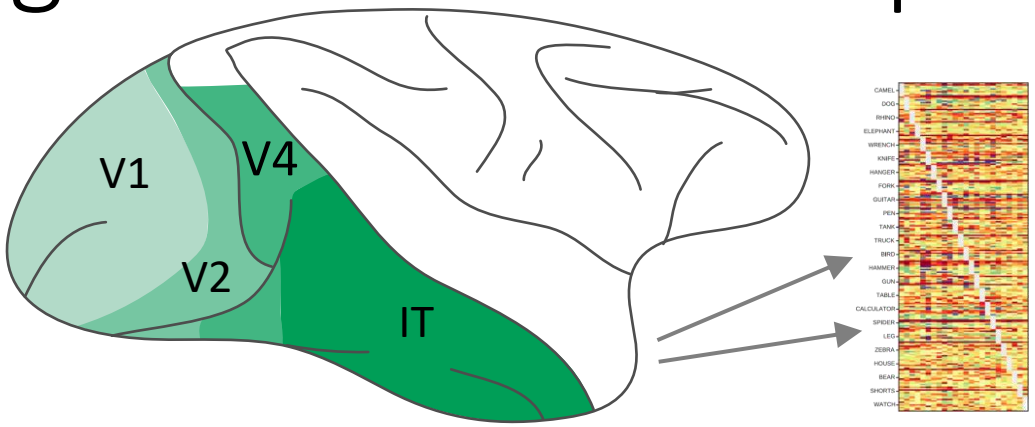
Model candidates tested:

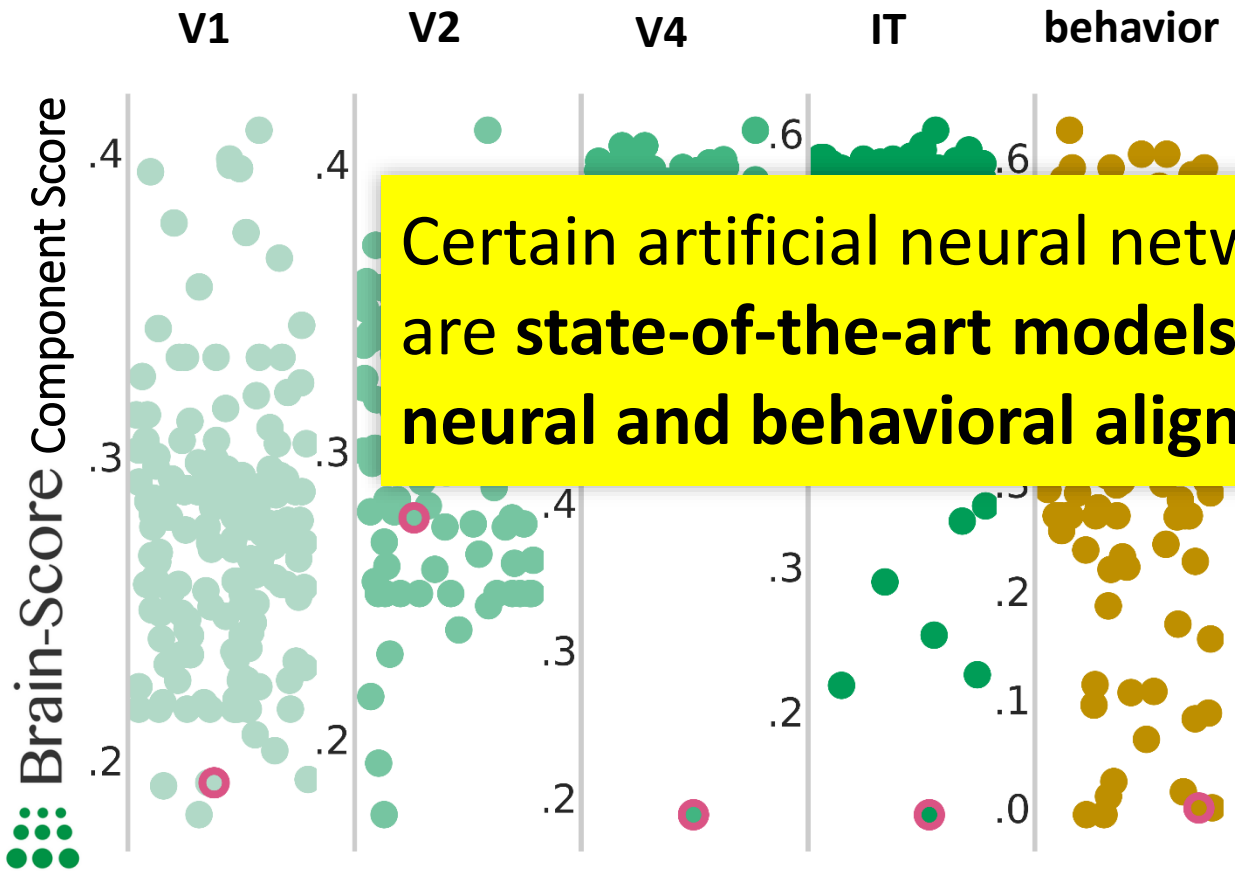


hmax *classic neuroscience model*

Schrimpf, Kubilius*, et al. (bioRxiv 2018)*
V1, V2 data: Freeman, Ziemba*, et al. (NatNeuro 2013)*
V4, IT data: Majaj, Hong*, et al. (JNeuro 2015)*
behavioral data: Rajalingham, Issa*, et al. (JNeuro 2018)*

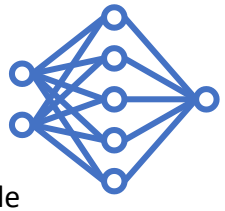
Integrative model comparison on Brain-Score





Certain artificial neural networks are state-of-the-art models of neural and behavioral alignment

Model candidates tested:



hmax		nasnet_large
vgg-16		nasnet_mobile
vgg-19		pnasnet_large
densenet-121		resnet-101_v1
densenet-169		resnet-101_v2
densenet-201		resnet-152_v1
inception_resnet_v2		resnet-152_v2
inception_v1		resnet-18
inception_v2		resnet-34
inception_v3	*ML models*	resnet-50_v1
inception_v4		resnet-50_v2
mobilenet_v1_0.25_128		squeezenet1_0
mobilenet_v1_0.25_160		squeezenet1_1
mobilenet_...		xception
mobilenet_v2_1.3_224		...
mobilenet_v2_1.4_224		

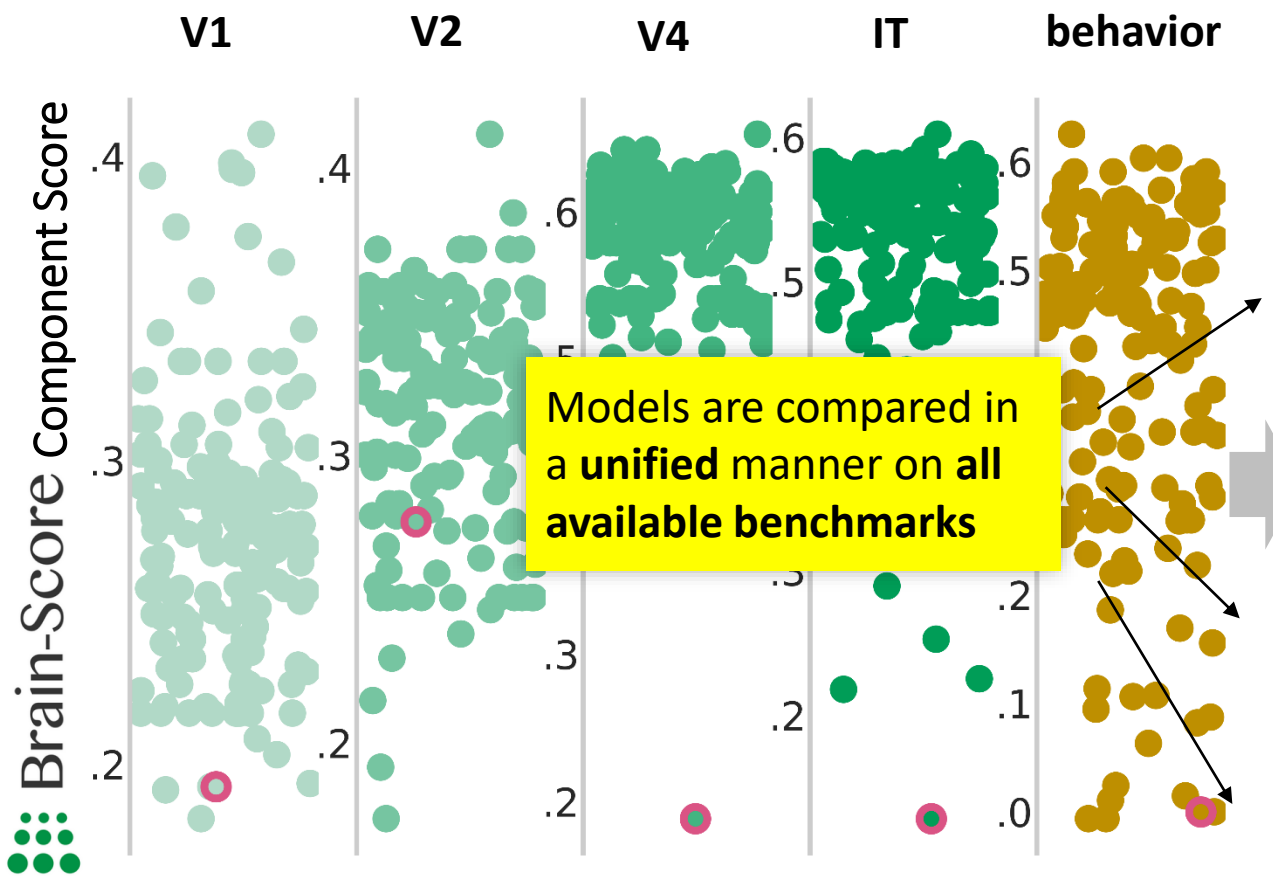
*translated into System Models:
- assign layers to regions
- assign pixels to visual degrees

Schrimpf, Kubilius*, et al. (bioRxiv 2018)*
V1, V2 data: Freeman, Ziemba*, et al. (NatNeuro 2013)*
V4, IT data: Majaj, Hong*, et al. (JNeuro 2015)*
behavioral data: Rajalingham, Issa*, et al. (JNeuro 2018)*

Brain-Score: Integrative Benchmarking

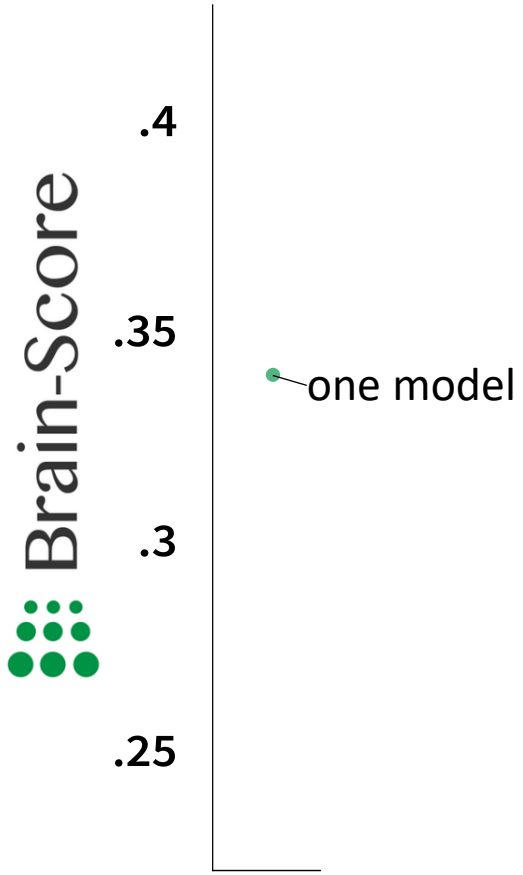
Brain-Score

www.Brain-Score.org



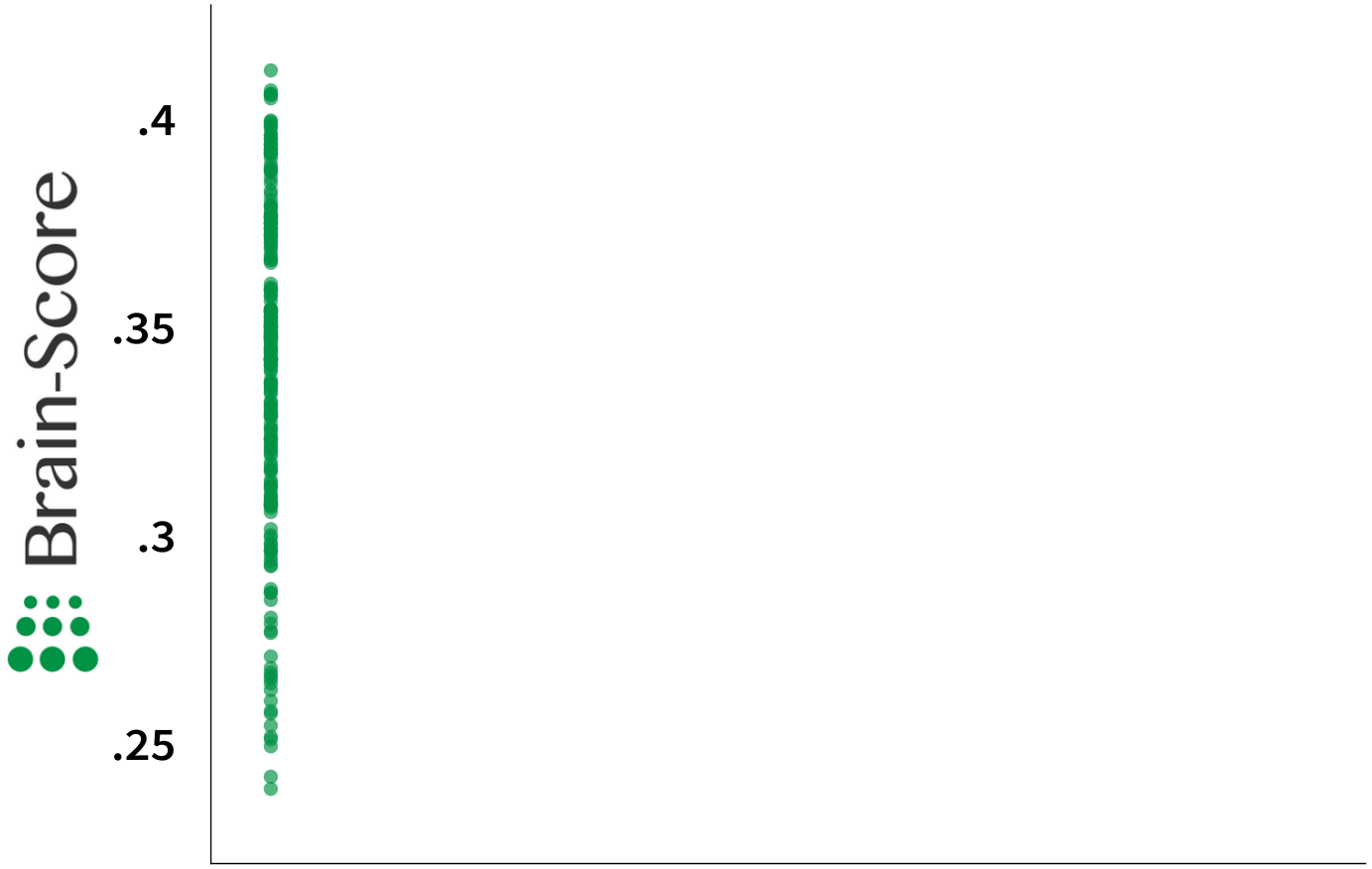
Models are compared in a **unified** manner on **all available benchmarks**

Our community is adding **more and more benchmarks** that evaluate all available models – and **guard against overfitting** (currently **51 benchmarks**)



Schrimpf, Kubilius*, et al. (bioRxiv 2018)*

What explains the model differences?

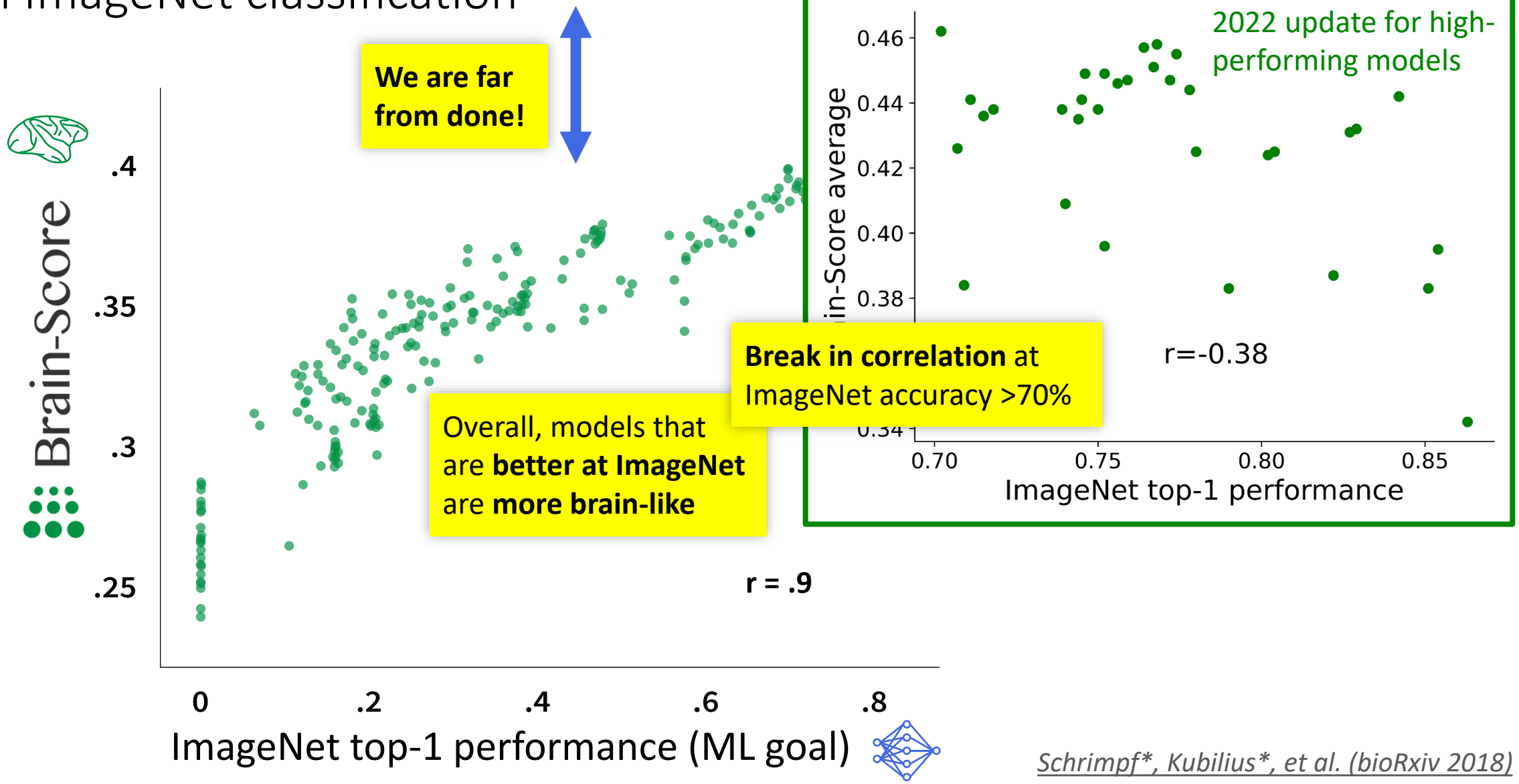


Normative variable

cf. Yamins, Hong*, et al. (PNAS 2014)*
Schrimpf, Kubilius*, et al. (bioRxiv 2018)*

Task performance correlates with Brain-Score

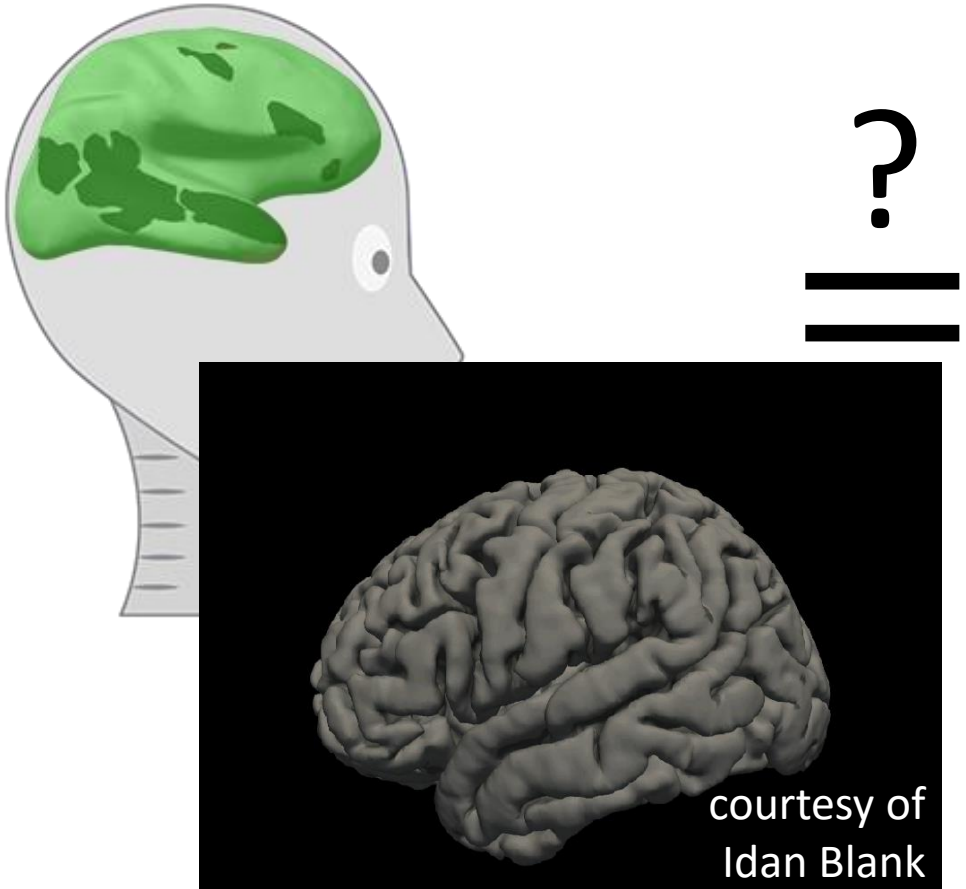
on ImageNet classification



Schrimpf, Kubilius*, et al. (bioRxiv 2018)*

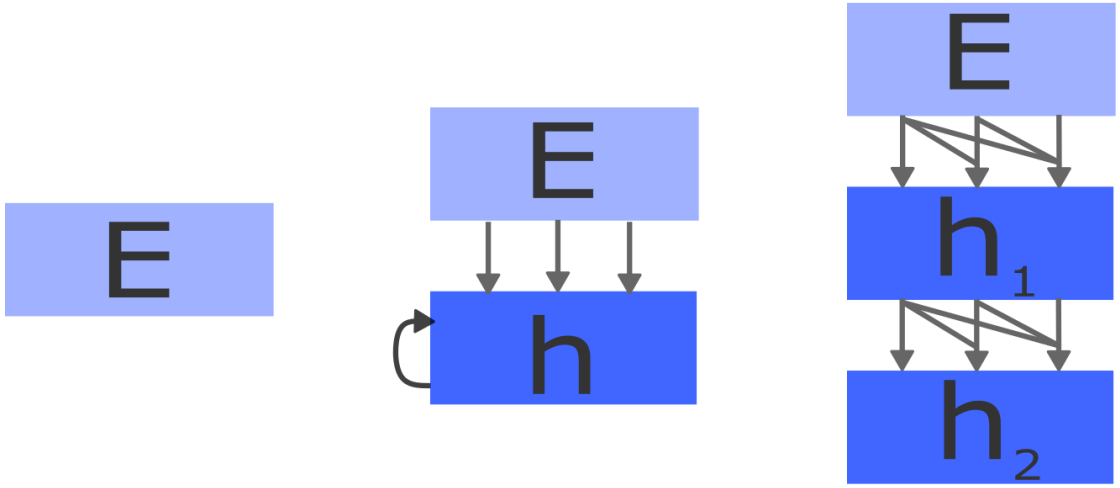
Integrative testing yields insights across domains of intelligence such as language

Humans

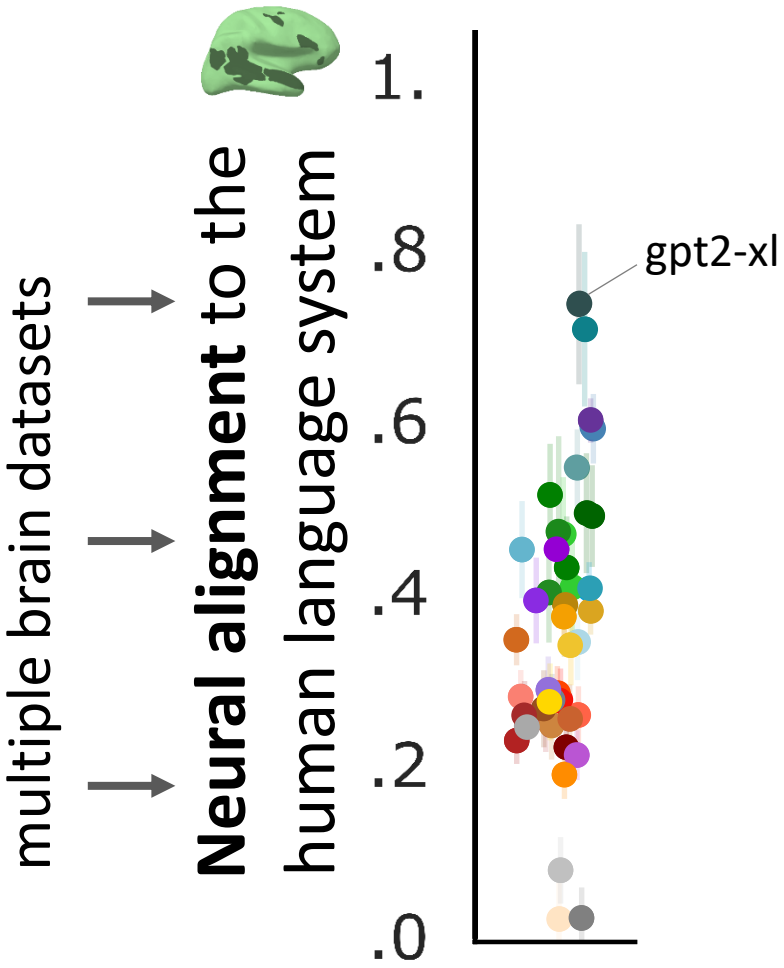


?

Models

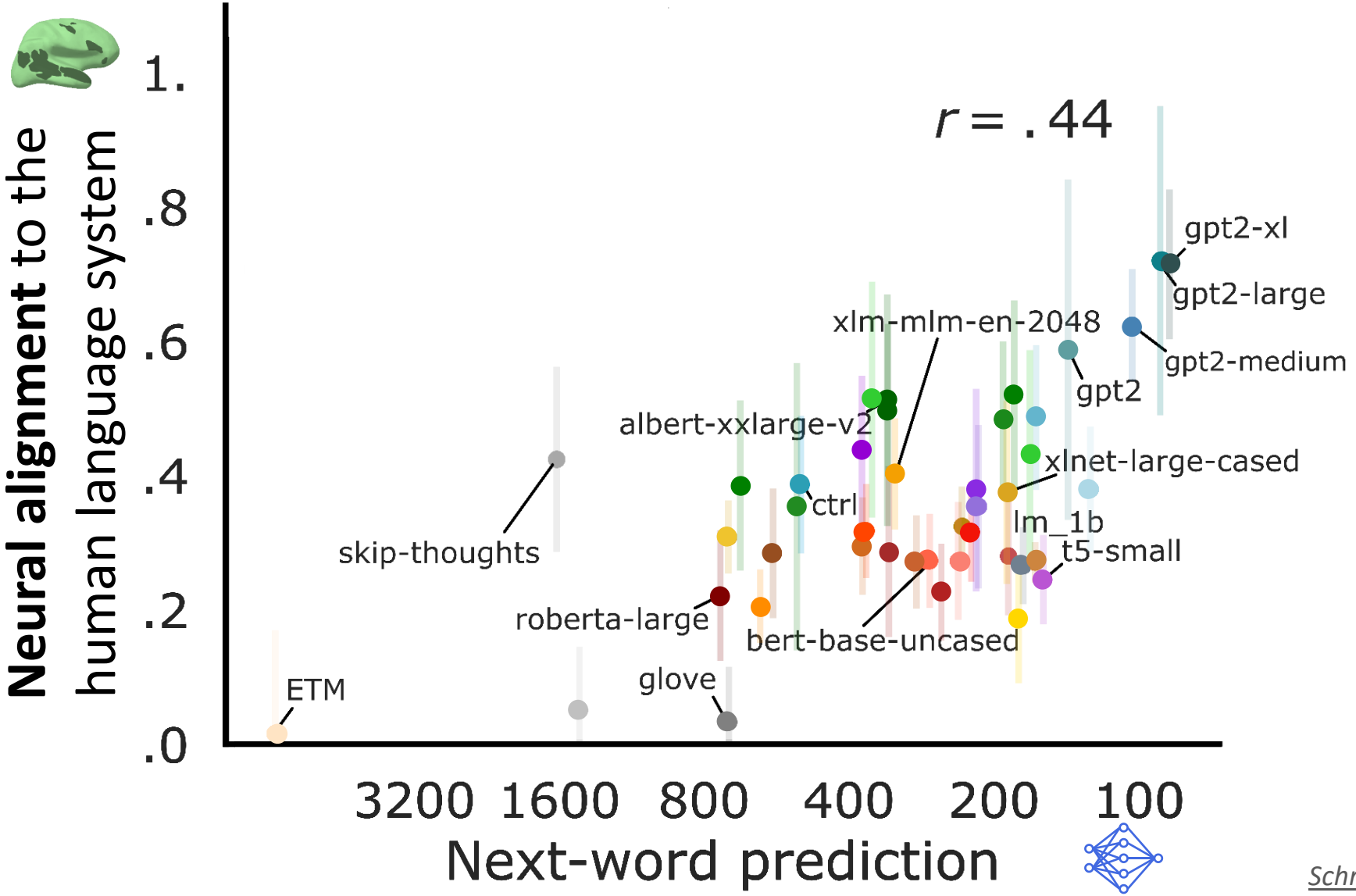


Particular ML language models predict the human language system



Schrimpf et al. (PNAS 2021)

The better models can predict the next word, the more brain-like they are

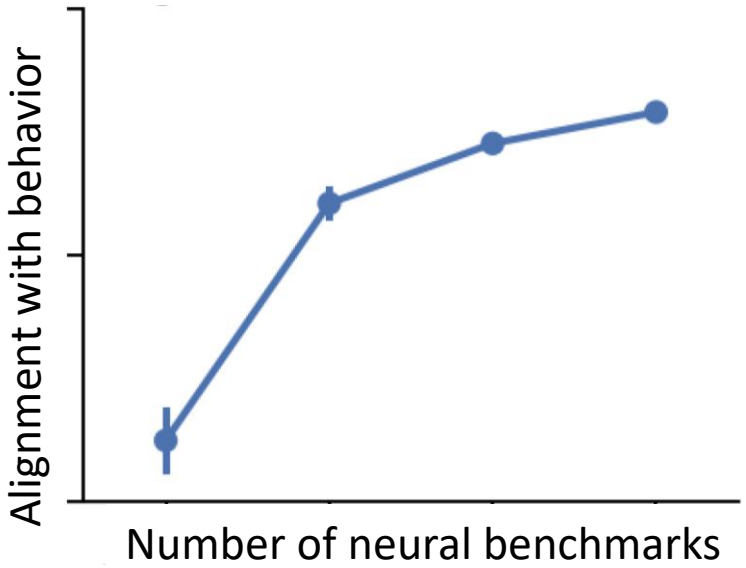


Schrimpf et al. (PNAS 2021)

Contributions

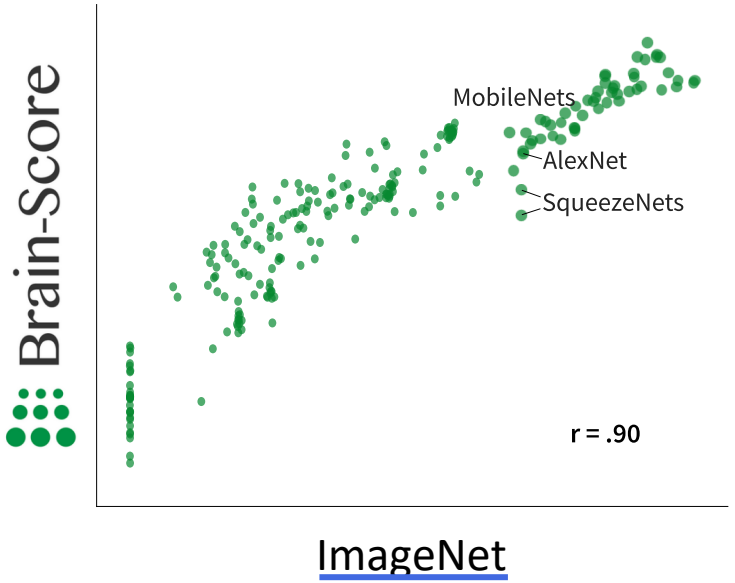
msch@mit.edu
mastodon.social/
@mschrimpf
@martin_schrimpf

To meaningfully **make progress** in modeling human intelligence (behavior + underlying neural activity), we must *integrate experimental results*.



Brain-Score is an implementation of this approach, currently in the domains of vision and language.

- *Identify* most brain-like models
- Provide *empirical constraints* for developing new models
- Discover key *relationships* between neural function, behavior, and computation



We could use your help

Brain-Score.org github.com/brain-score