Task-Optimized Models of the Brain

Aran Nayebi
McGovern Institute, MIT

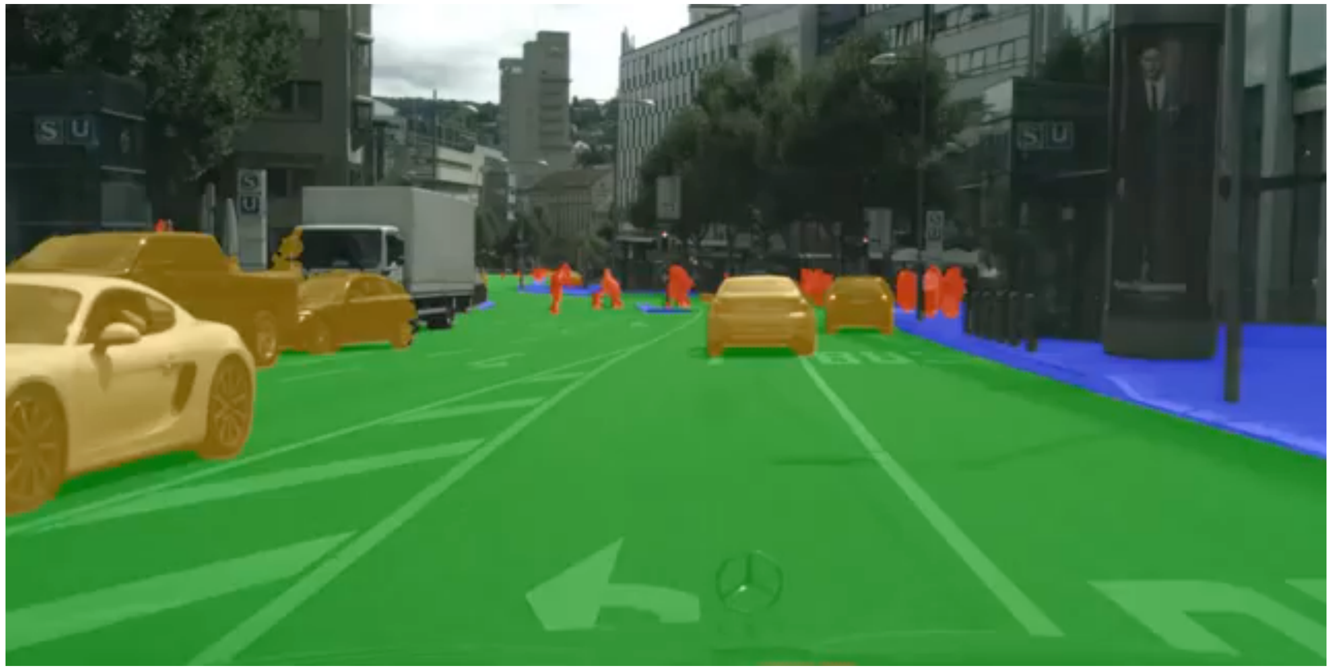
Carnegie Mellon University
Machine Learning Department,
School of Computer Science
2024.04.16

MCGOVERN INSTITUTE

K. LISA YANG ICoN CENTER
Integrative Computational Neuroscience

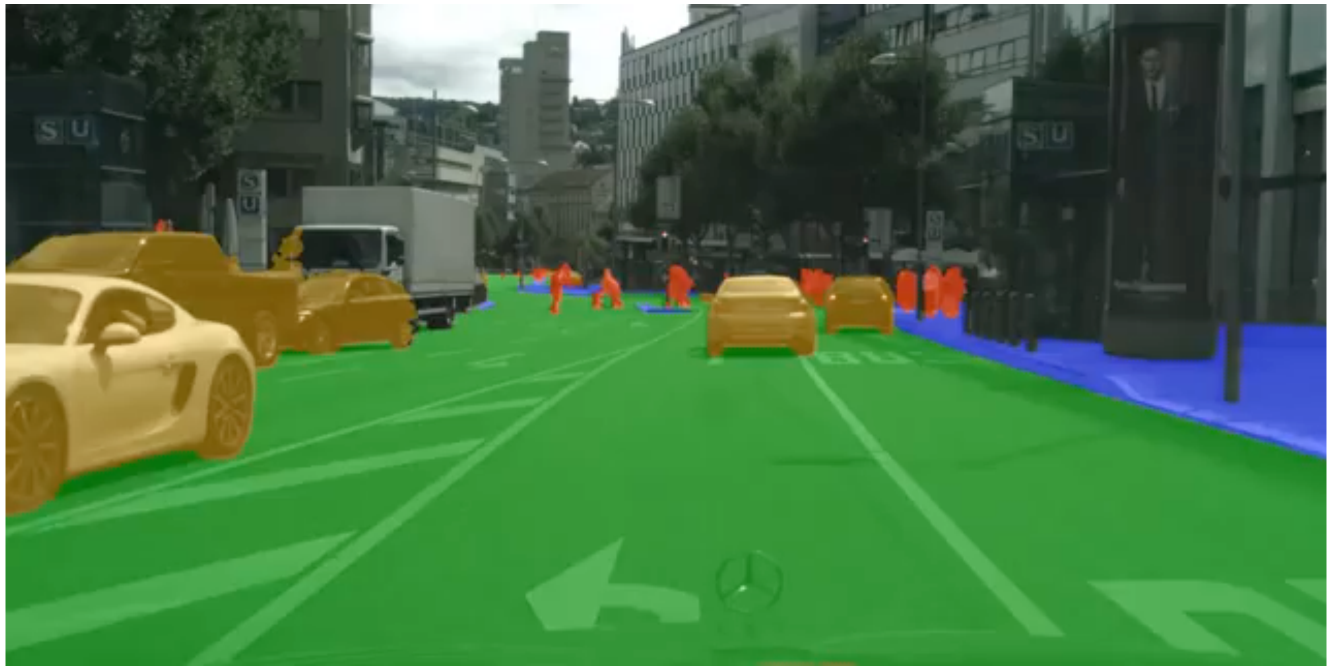
From Neurons to Behavior

Scene Understanding



From Neurons to Behavior

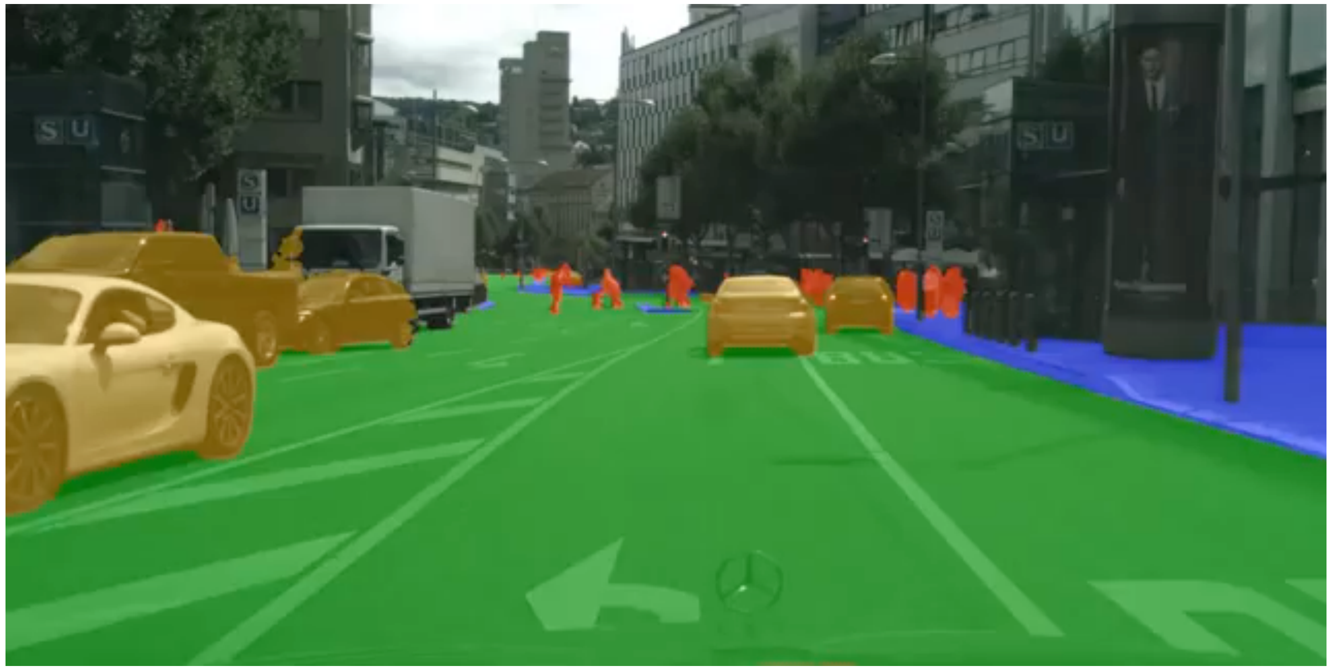
Scene Understanding



Multi-Step Planning

From Neurons to Behavior

Scene Understanding

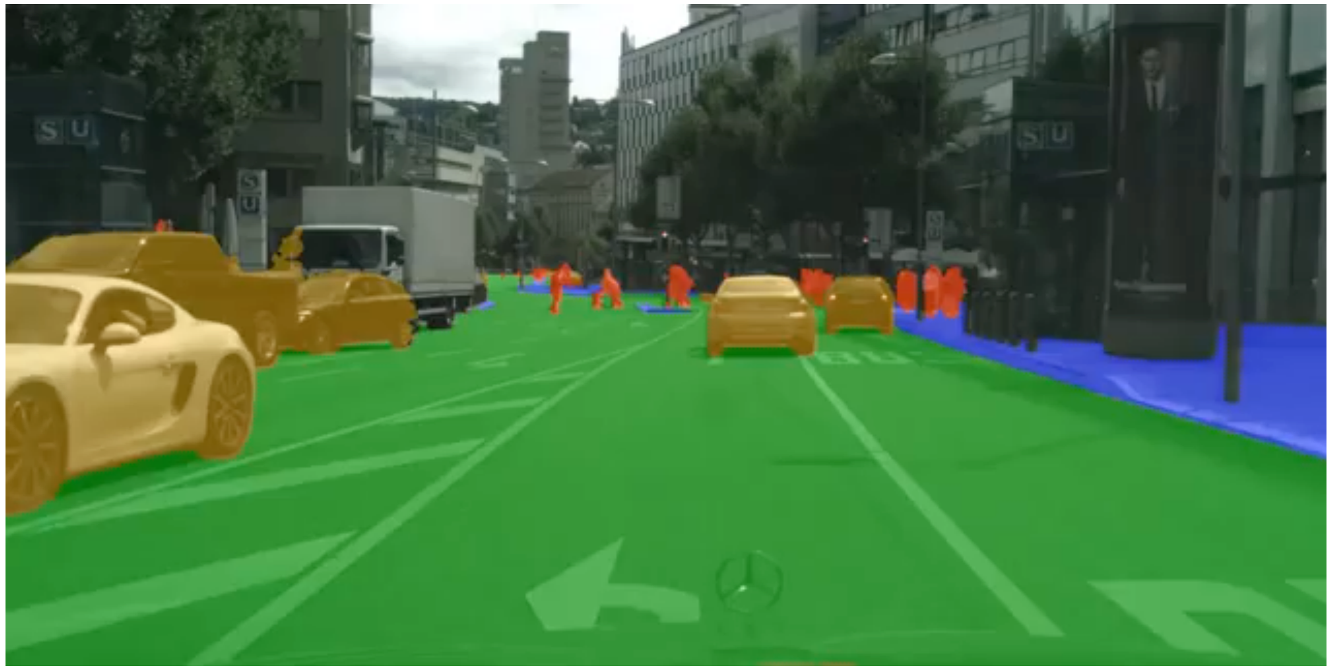


Multi-Step Planning

Navigation

From Neurons to Behavior

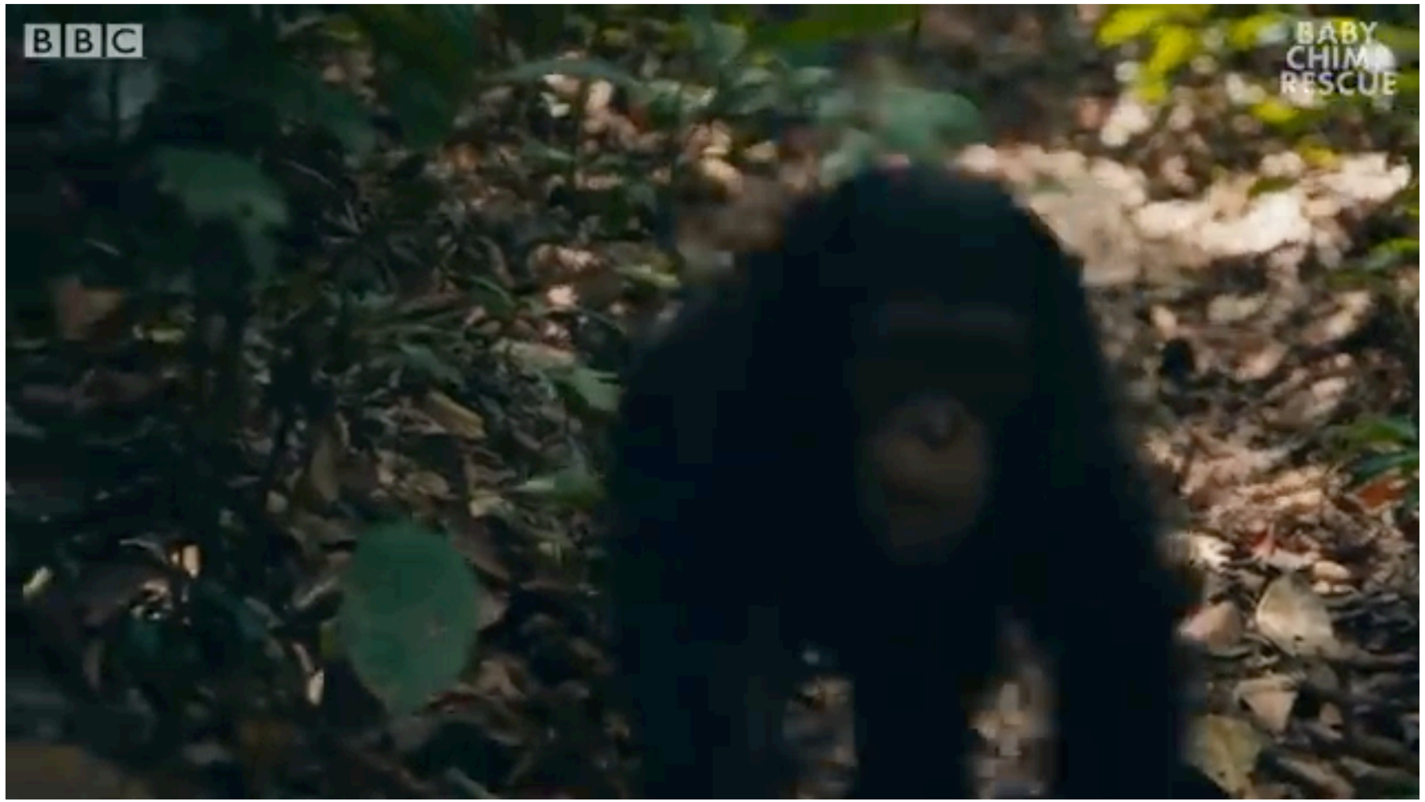
Scene Understanding



Multi-Step Planning

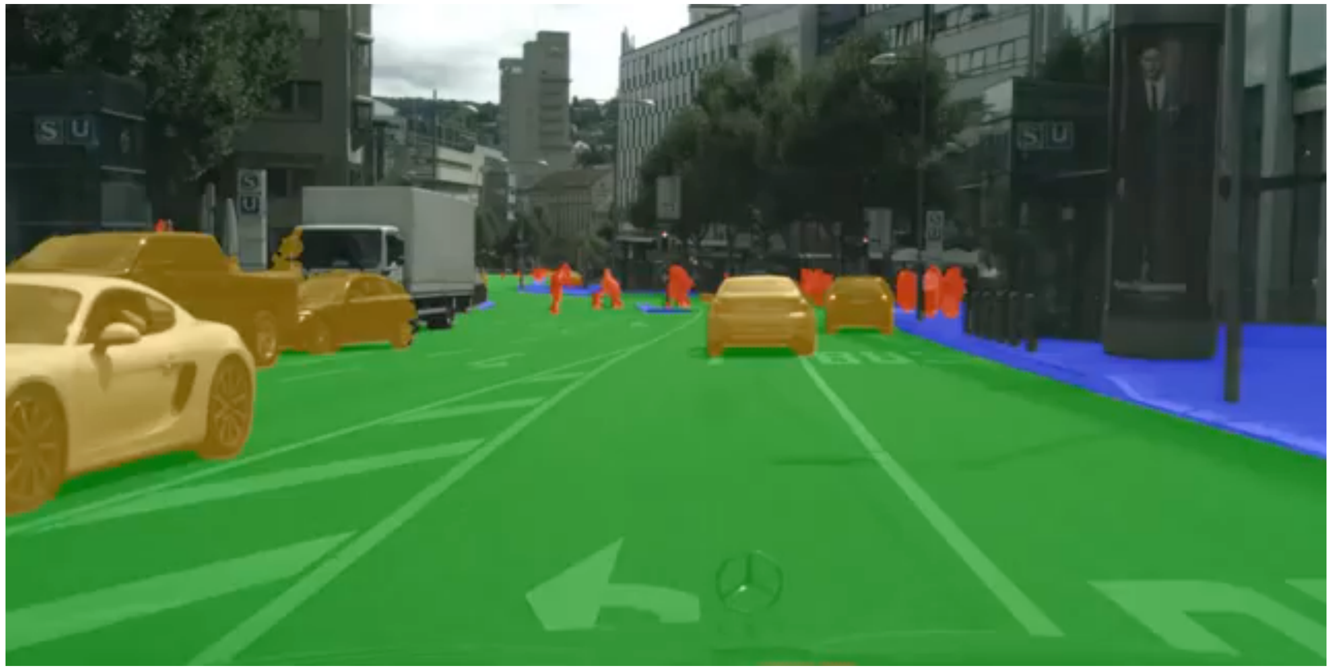
Navigation

Flexible Embodiment



From Neurons to Behavior

Scene Understanding

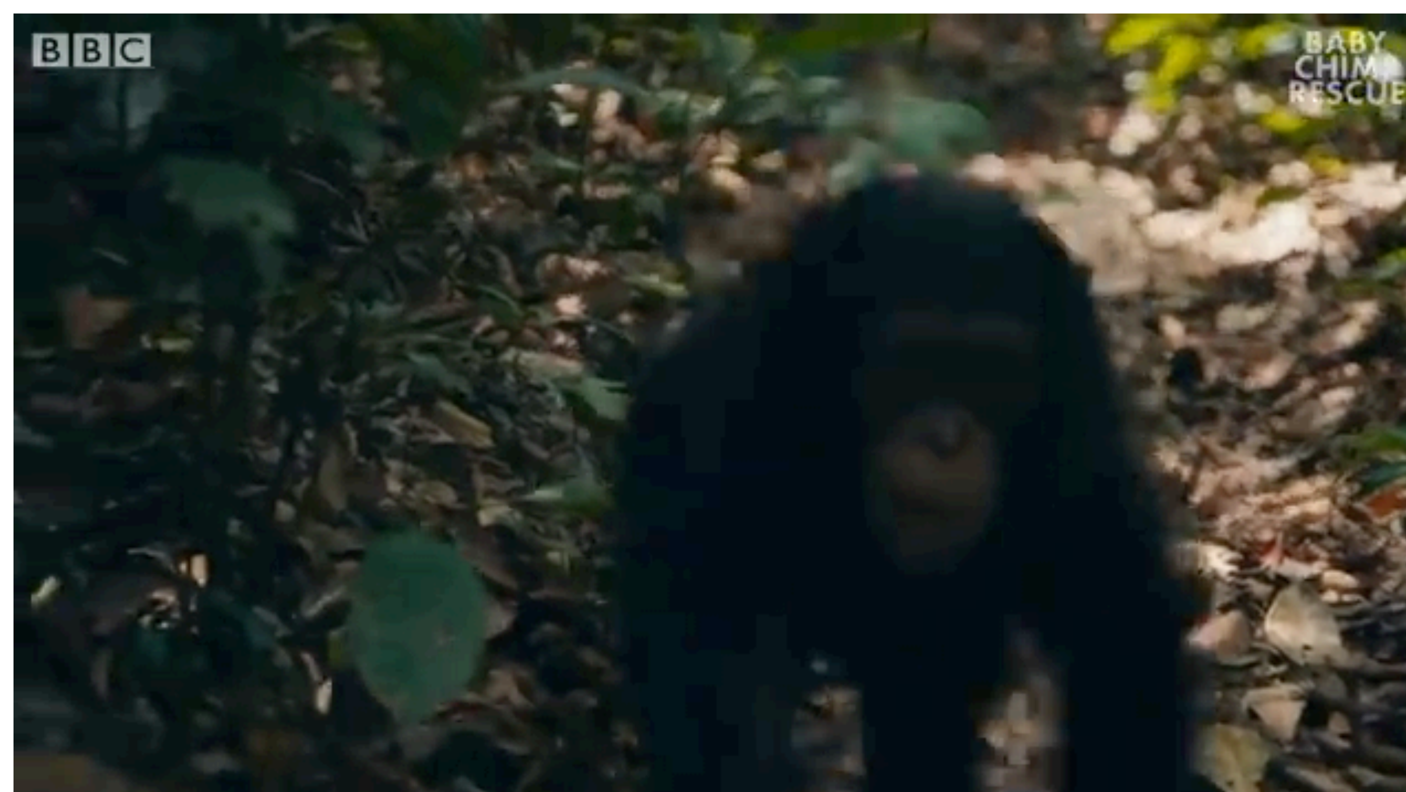


Multi-Step Planning

What are the core design principles that give rise to these abilities?

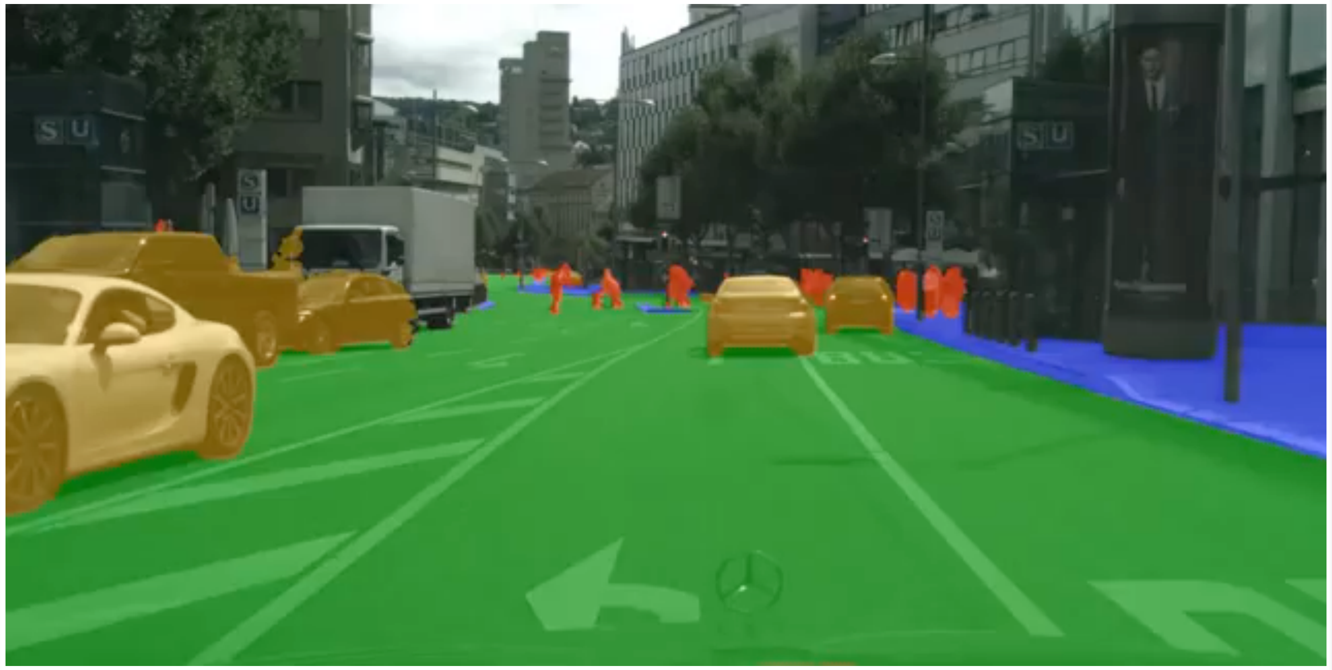
Navigation

Flexible Embodiment



From Neurons to Behavior

Scene Understanding

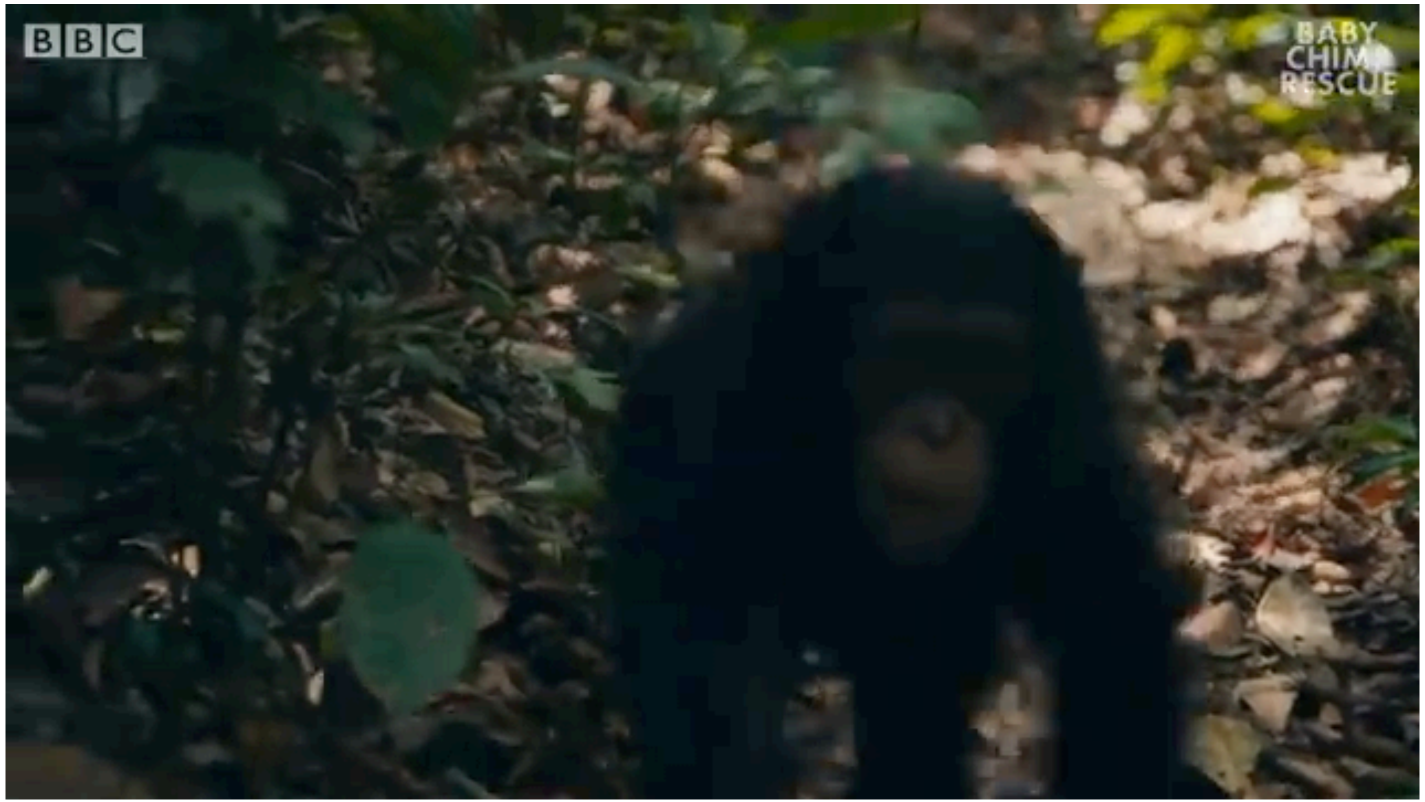


Multi-Step Planning

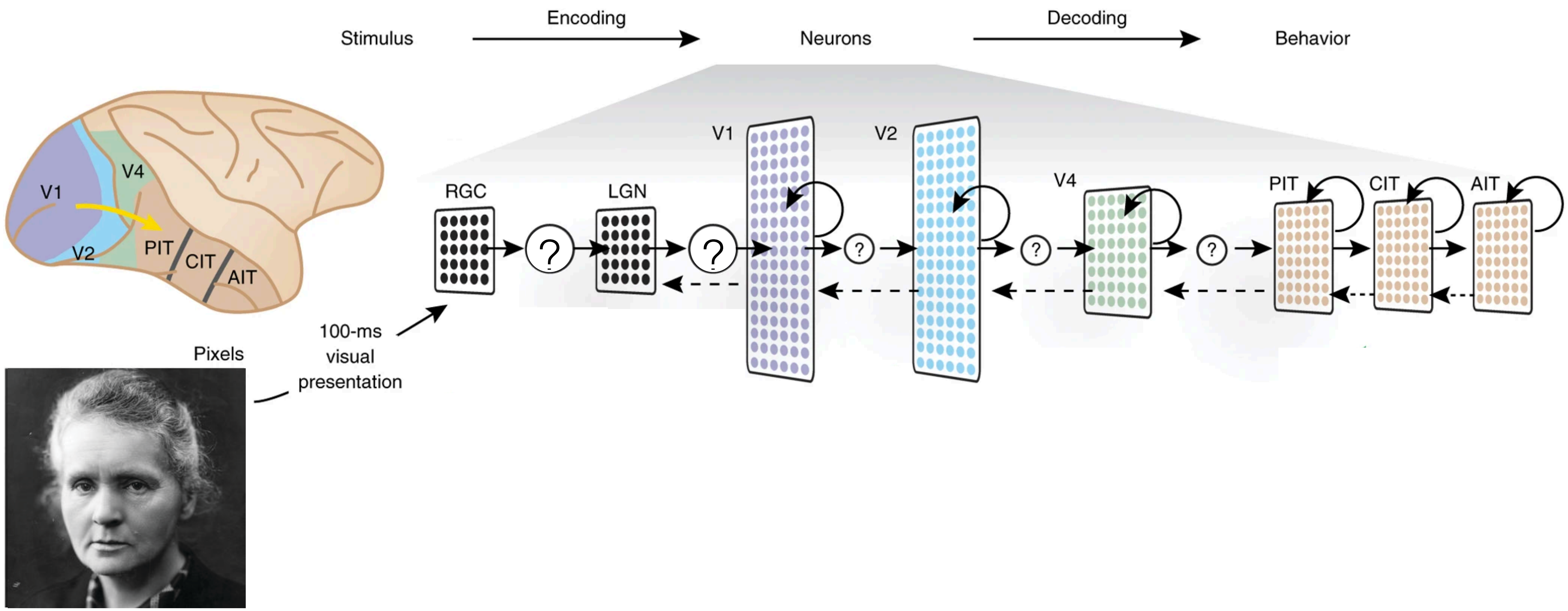
How do we bridge the gap from neurons to behavior?

Navigation

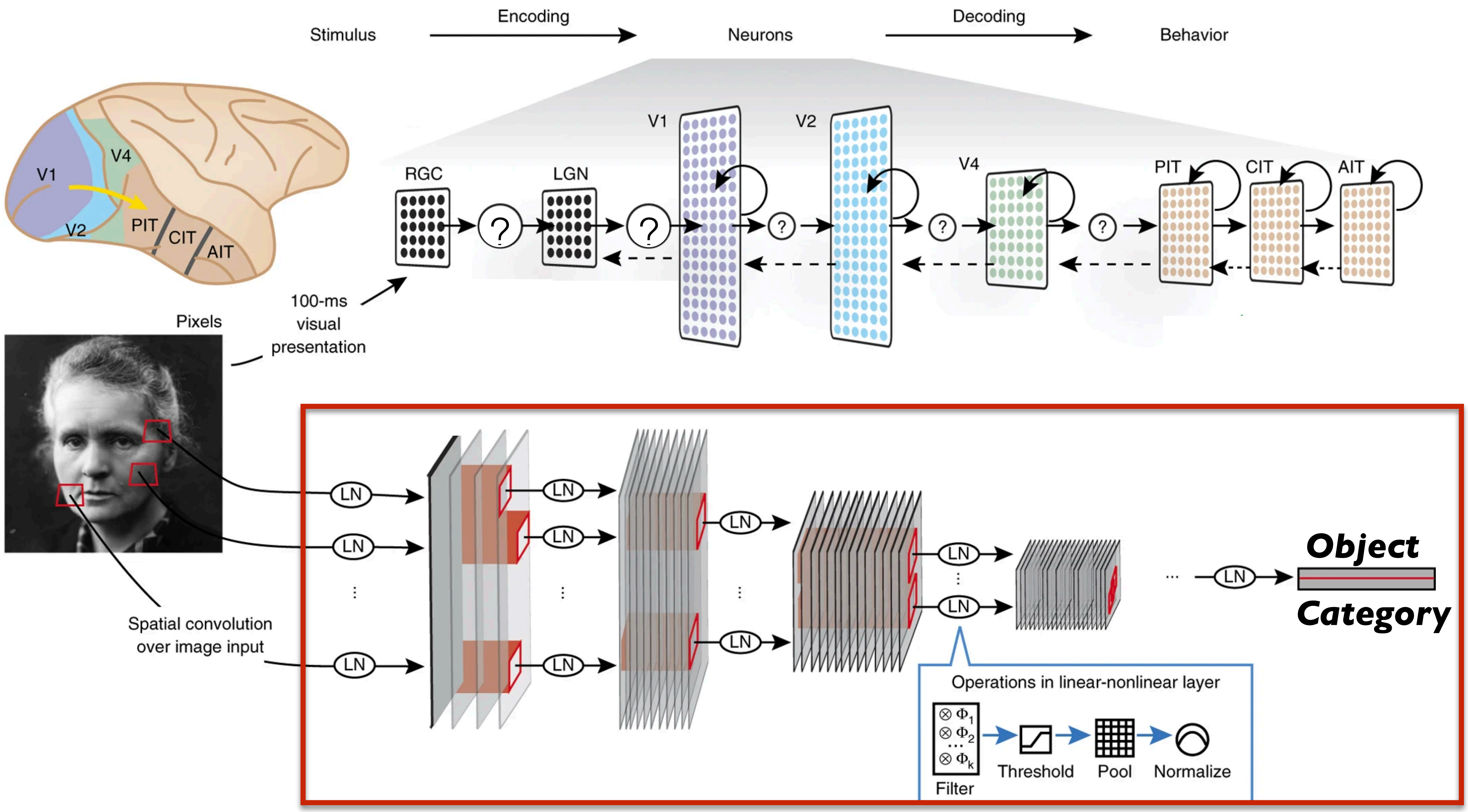
Flexible Embodiment



Primate Ventral Stream Implements Object Recognition



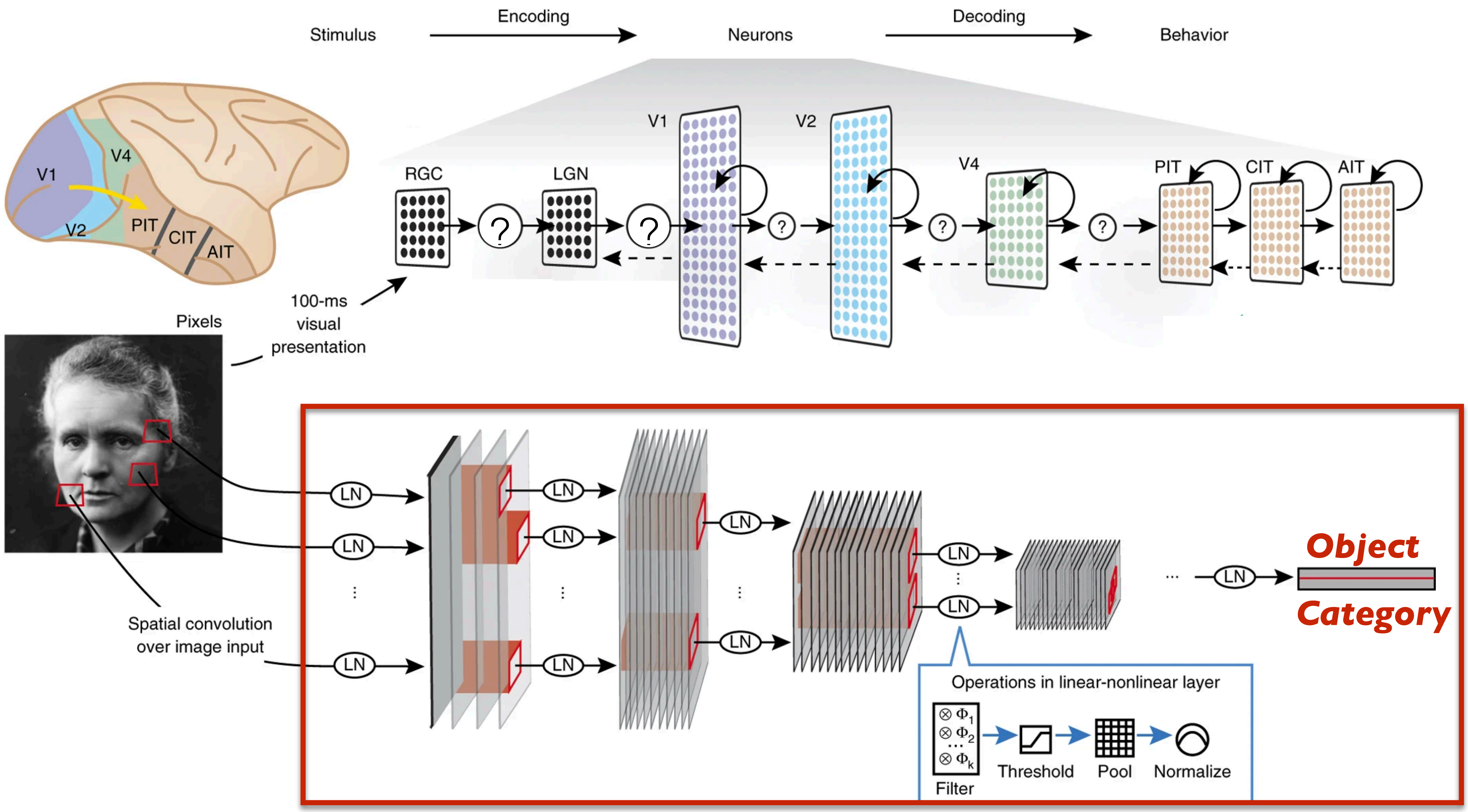
CNNs as Models of Primate Object Recognition



CNNs are inspired by visual neuroscience:

1) **hierarchy**
2) **retinotopy** (spatially tiled)

CNNs as Models of Primate Object Recognition



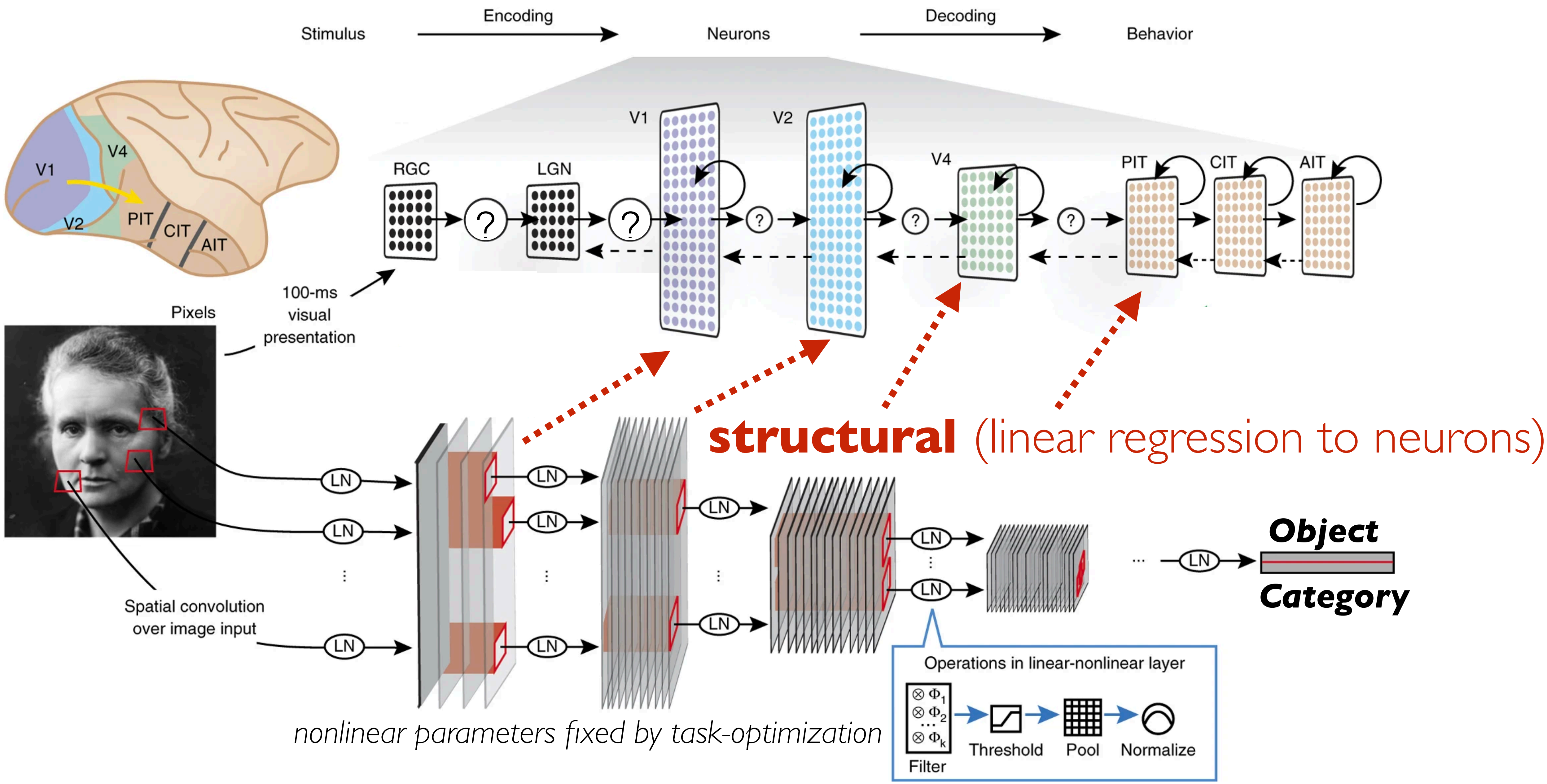
CNNs are inspired by visual neuroscience:

1) **hierarchy**

2) **retinotopy** (spatially tiled)

functional (performs behavior)

CNNs as Models of Primate Object Recognition



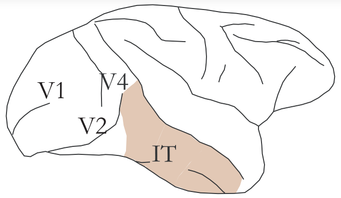
CNNs are inspired by visual neuroscience:

1) **hierarchy**

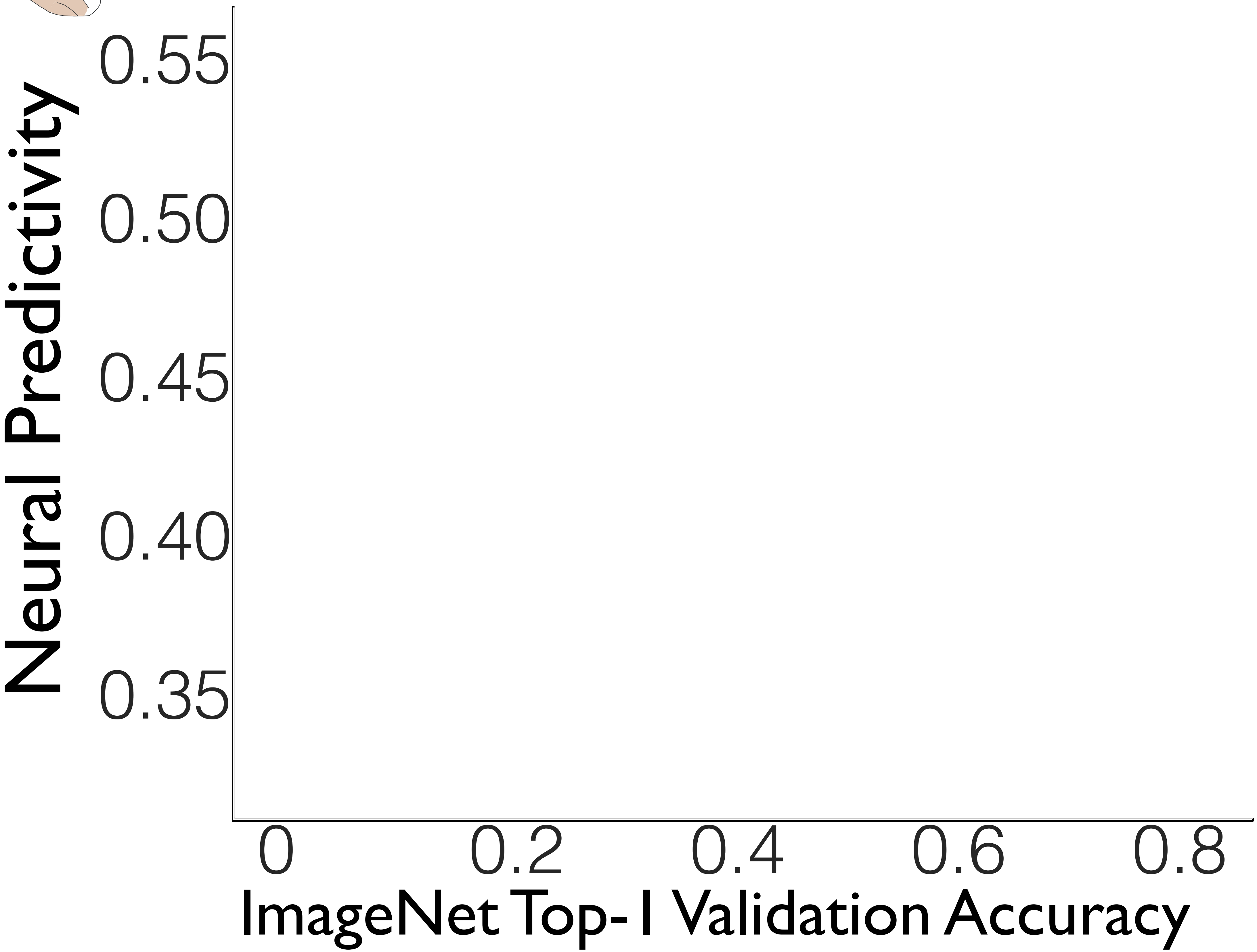
2) **retinotopy** (spatially tiled)

functional (performs behavior)

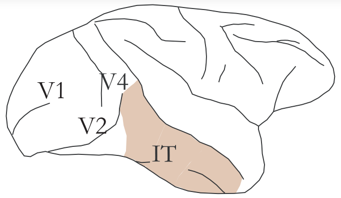
Task Performance Correlated with Neural Predictivity



Schrimpf, Kubilius* et al. 2018*

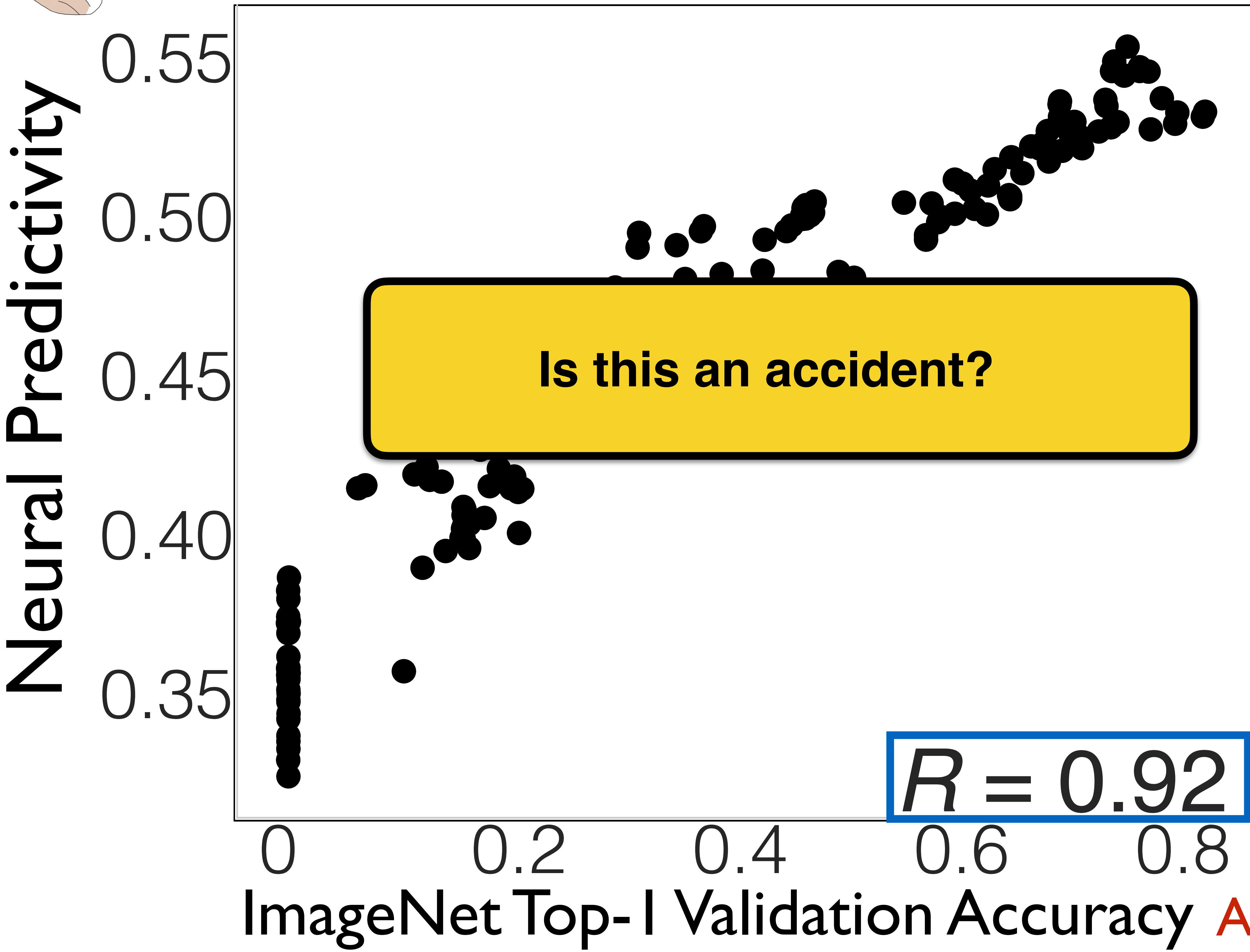


Task Performance Correlated with Neural Predictivity



A Neuroscience Goal

Schrimpf, Kubilius* et al. 2018*



An AI Goal

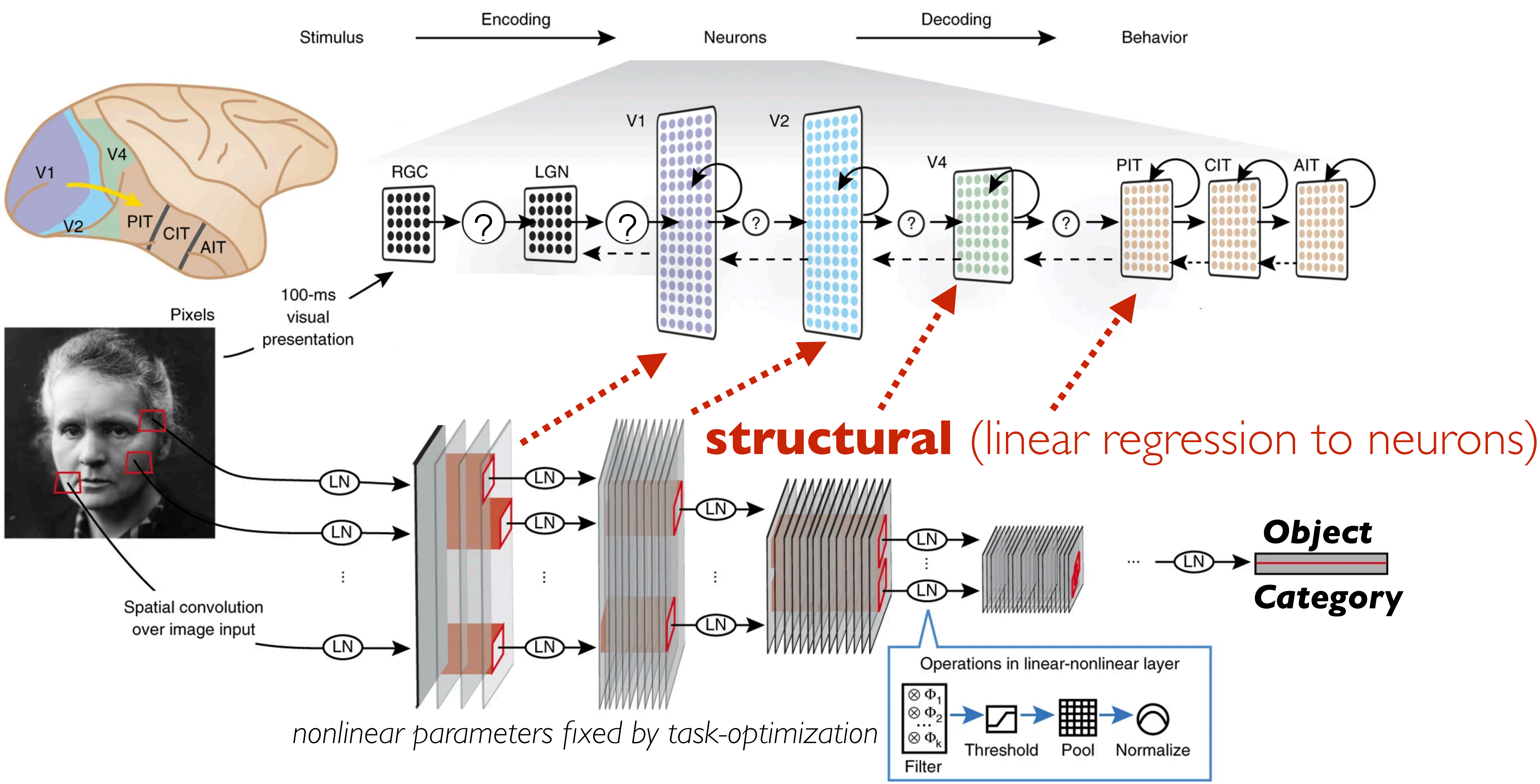
CNNs as Models of Primate Object Recognition

L = *learning rule*

Backpropagation

T = *task loss*

Categorization



ImageNet

functional (performs behavior)

CNNs

D = *data stream*

A = *architecture class*

Task-Optimized Modeling: Four Components

Task-Optimization (ML)

1.

A = *architecture class*

2.

T = *task loss*

3.

D = *dataset*

4.

L = *learning rule*

Task-Optimized Modeling: Four Components

Task-Optimization (ML)

Neurobiology

1.

A = *architecture class*

2.

T = *task loss*

3.

D = *dataset*

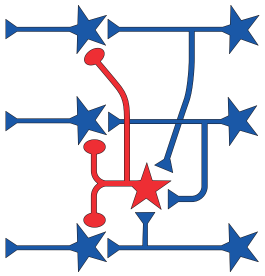
4.

L = *learning rule*

Task-Optimized Modeling: Four Components

<u>Task-Optimization (ML)</u>

<u>Neurobiology</u>



1.

A = *architecture class* **= *circuit neuroanatomy***

2.

T = *task loss*

3.

D = *dataset*

4.

L = *learning rule*

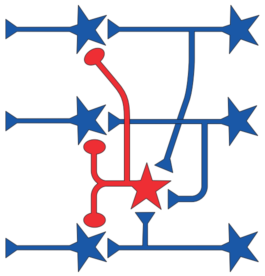
Task-Optimized Modeling: Four Components

Task-Optimization (ML)

Neurobiology

1.

A = *architecture class* **= *circuit neuroanatomy***



2.

T = *task loss* **= *ecological niche/behavior***

3.

D = *dataset*

4.

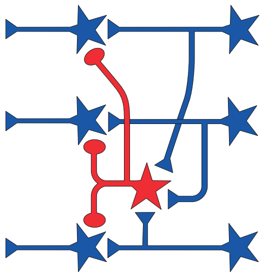
L = *learning rule*

Task-Optimized Modeling: Four Components

Task-Optimization (ML) | **Neurobiology**

1.

A = *architecture class* **= *circuit neuroanatomy***

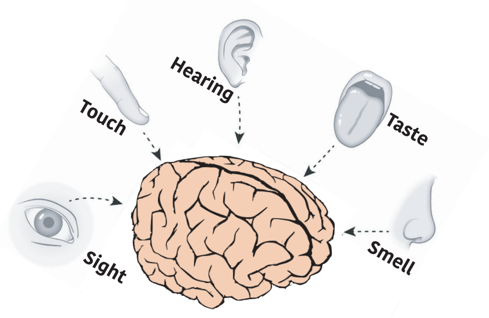


2.

T = *task loss* **= *ecological niche/behavior***

3.

D = *dataset* **= *environment***



4.

L = *learning rule*

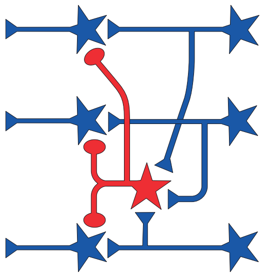
Task-Optimized Modeling: Four Components

Task-Optimization (ML)

Neurobiology

1.

A = *architecture class* **= *circuit neuroanatomy***

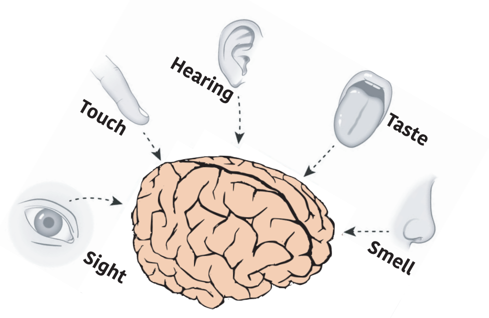


2.

T = *task loss* **= *ecological niche/behavior***

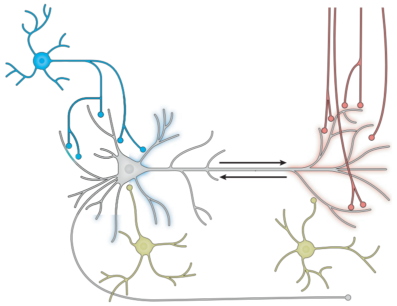
3.

D = *dataset* **= *environment***



4.

L = *learning rule* **= *natural selection + synaptic plasticity***



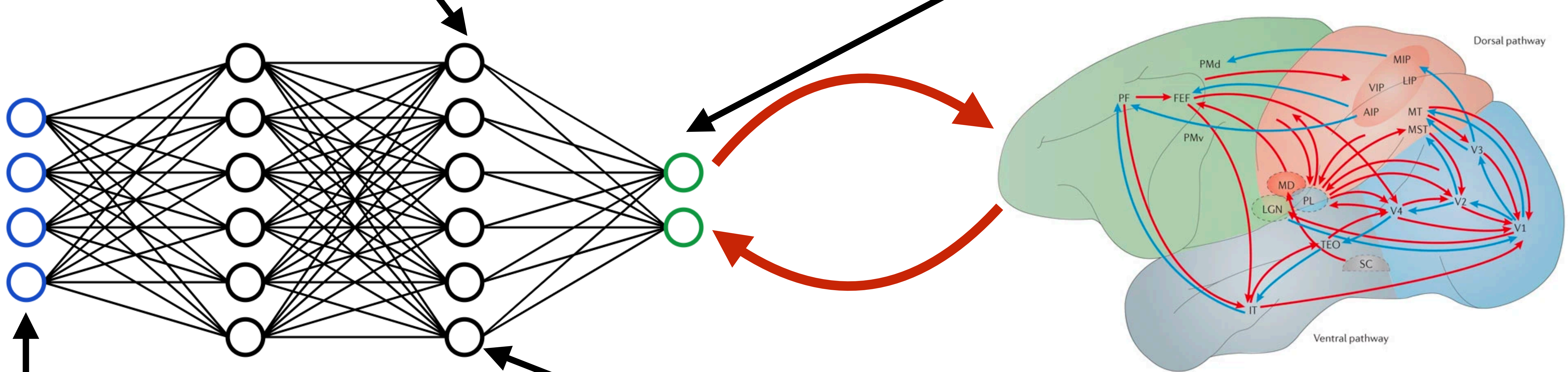
Task-Optimized Modeling: Four Components

L = *learning rule*

"Natural selection + plasticity"

T = *task loss*

"Ecological niche/ behavior"



"Environment"

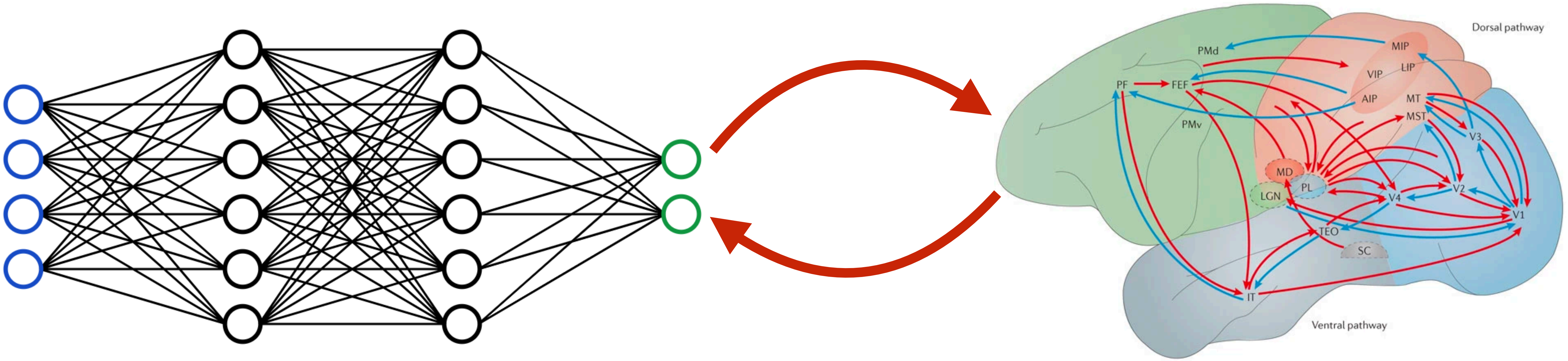
D = *data stream*

"Circuit"

A = *architecture class*

Task-Optimized Modeling

Design ML Algorithms Optimized to Perform Organism's Behavior under Organism's Constraints



Yields:

Quantitatively Accurate & Practically Useful Brain Models

AND

Principles of *Why* Neural Responses Are As They Are

Outline

- Role of Recurrent Processing During Object Recognition
- Visually-Grounded Mental Simulation
- Vision and Navigation in Rodents

- Future Directions

Outline

- Role of Recurrent Processing During Object Recognition
- Visually-Grounded Mental Simulation
- Vision and Navigation in Rodents

- Future Directions

Role of Recurrent Processing During Object Recognition

L = *learning rule*

"Natural selection + plasticity"

Backpropagation

T = *task loss*

"Ecological niche/ behavior"

Categorization

A. Nayebi*, D. Bear*, J. Kubilius*, *et al.*
Task-Driven Convolutional Recurrent Models of the Visual System. *NeurIPS 2018*

A. Nayebi, *et al.*
Recurrent Connections in the Primate Ventral Visual Stream Mediate a Tradeoff Between Task Performance and Network Size During Core Object Recognition. *Neural Computation 2022*

Daniel Yamins

Daniel Bear

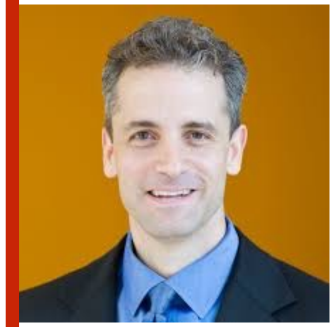
Jonas Kubilius

Kohitij Kar

Surya Ganguli

Javier Sagastuy

David Sussillo

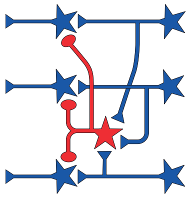


Jim DiCarlo

ImageNet

"Environment"

D = *data stream*

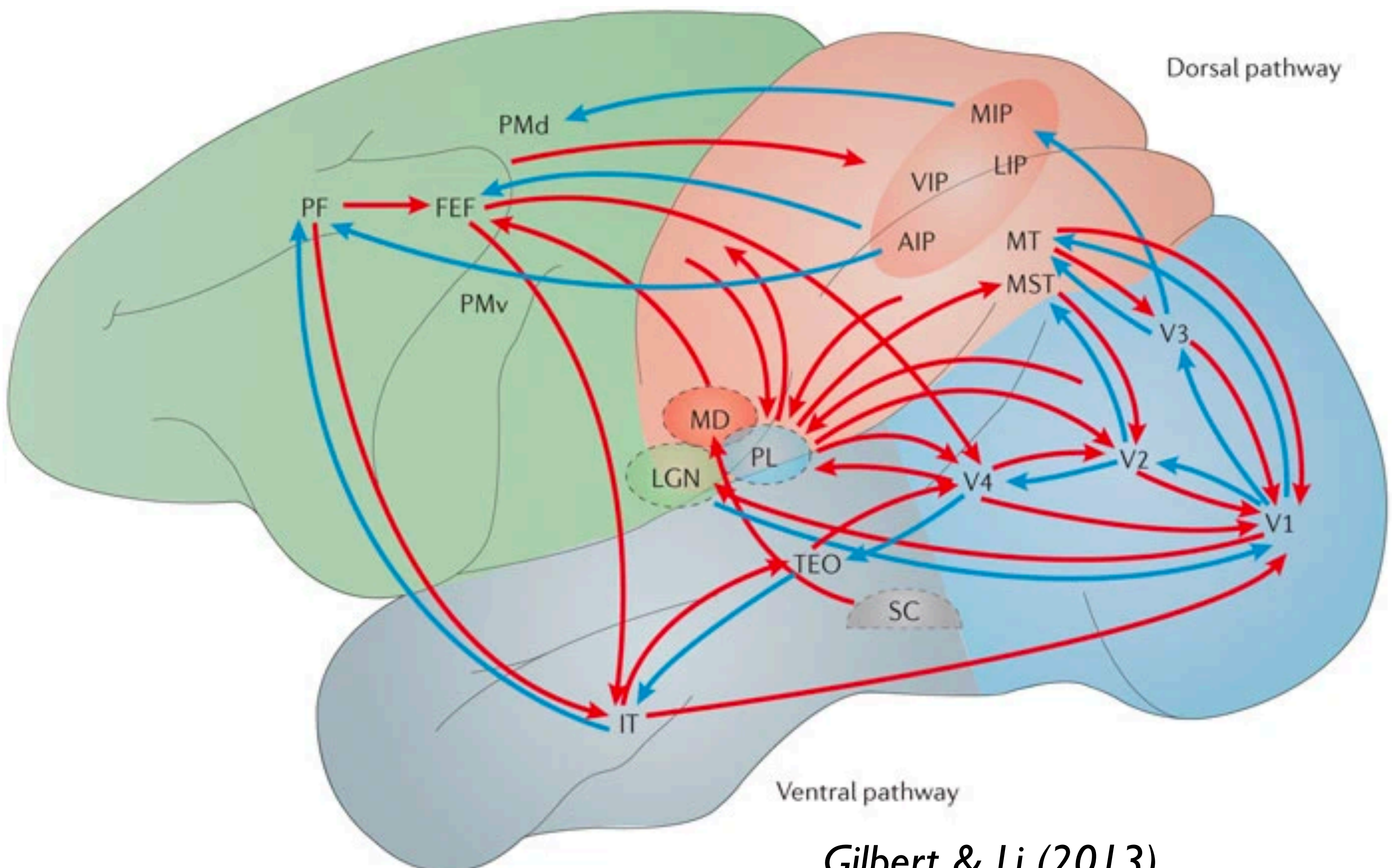


"Circuit"

A = *architecture class*

Recurrent Connections Are Ubiquitous

Recurrent connections are everywhere anatomically:



Gilbert & Li (2013)

. . . but what role do they play in behavior (if any)?

CNNs as Models of Primate Object Recognition

L = *learning rule*

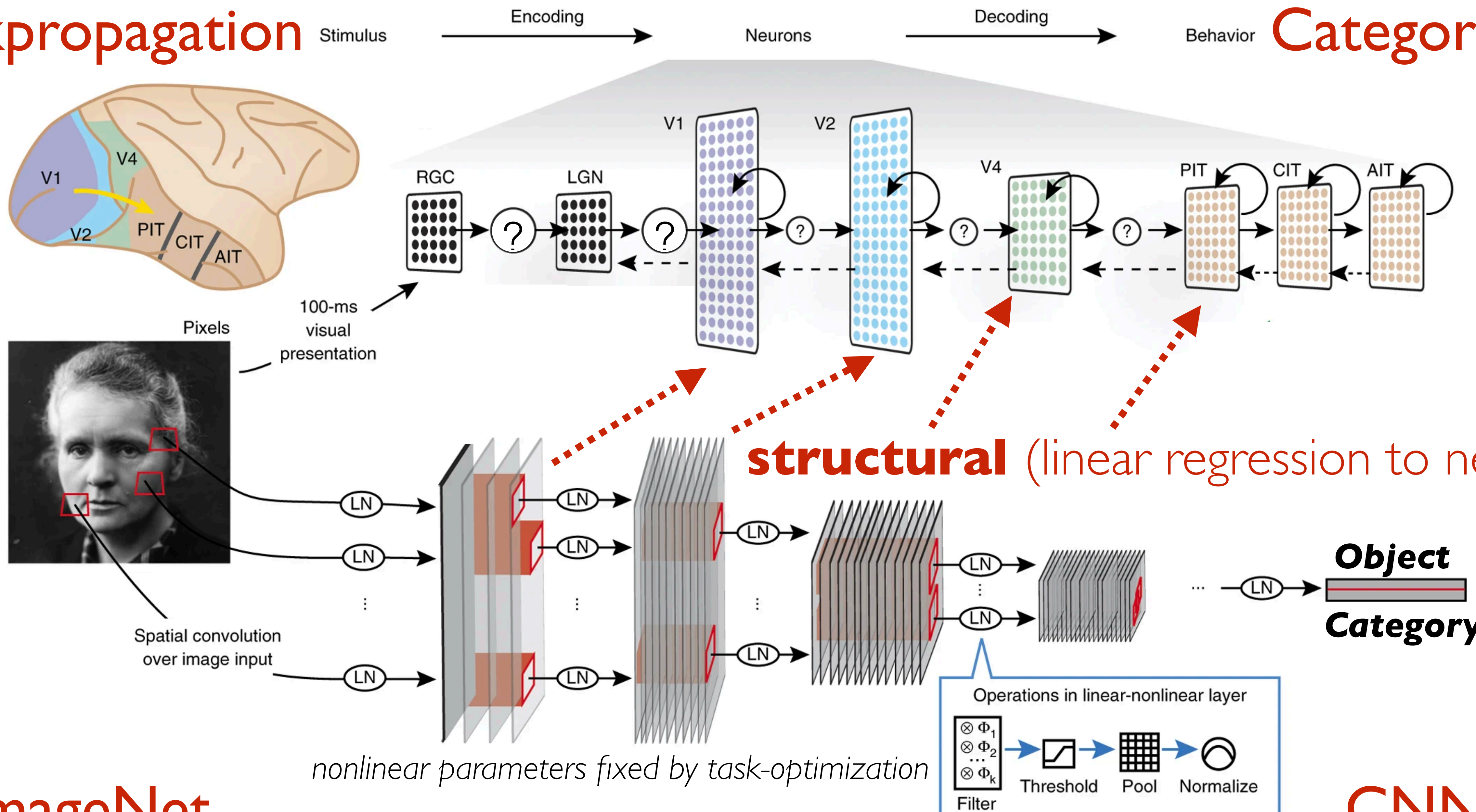
"Natural selection + plasticity"

Backpropagation

T = *task loss*

"Ecological niche/ behavior"

Categorization



ImageNet

"Environment"

functional (performs behavior)

CNNs

"Circuit"

D = *data stream*

A = *architecture class*

CNNs as Models of Primate Object Recognition

L = *learning rule*

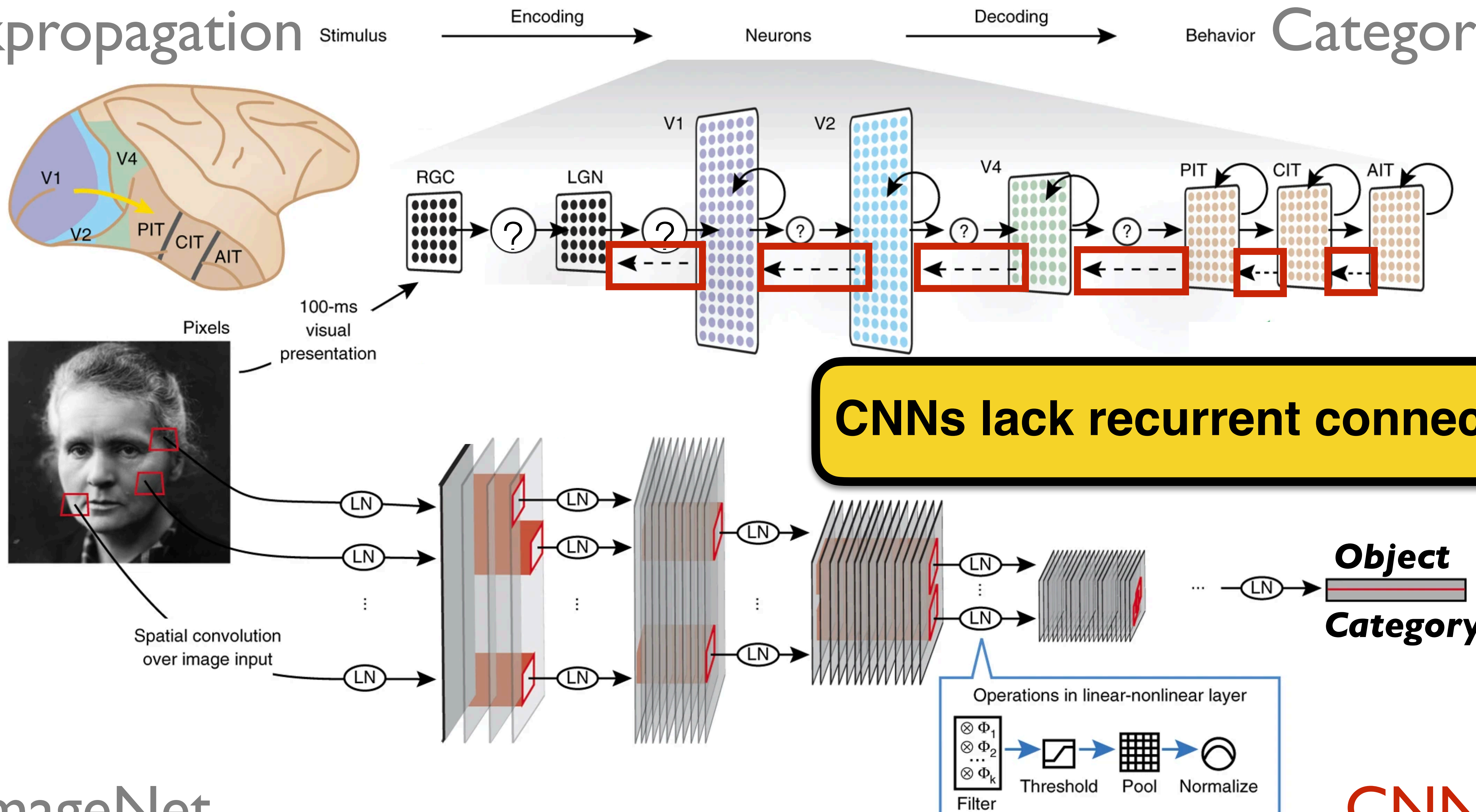
"Natural selection + plasticity"

Backpropagation

T = *task loss*

"Ecological niche/ behavior"

Categorization



CNNs lack recurrent connections!

ImageNet

"Environment"

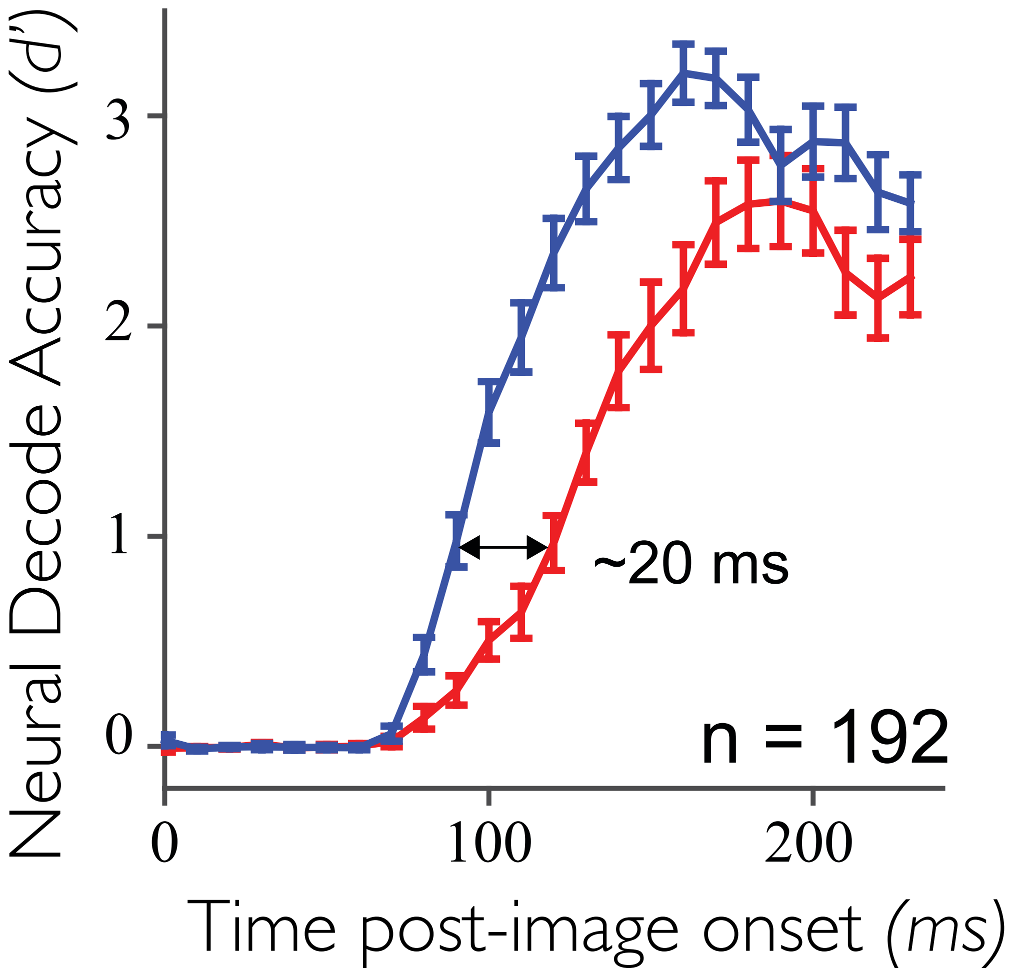
D = *data stream*

CNNs

"Circuit"

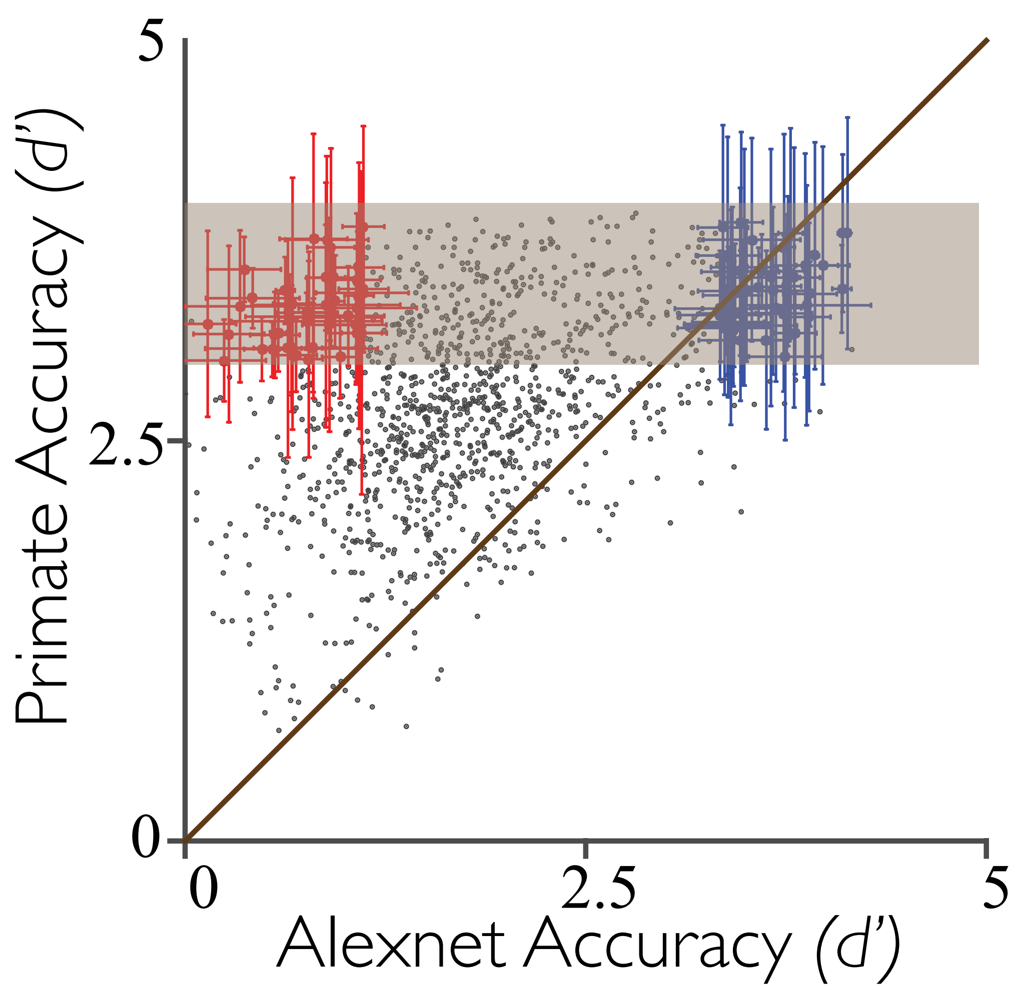
A = *architecture class*

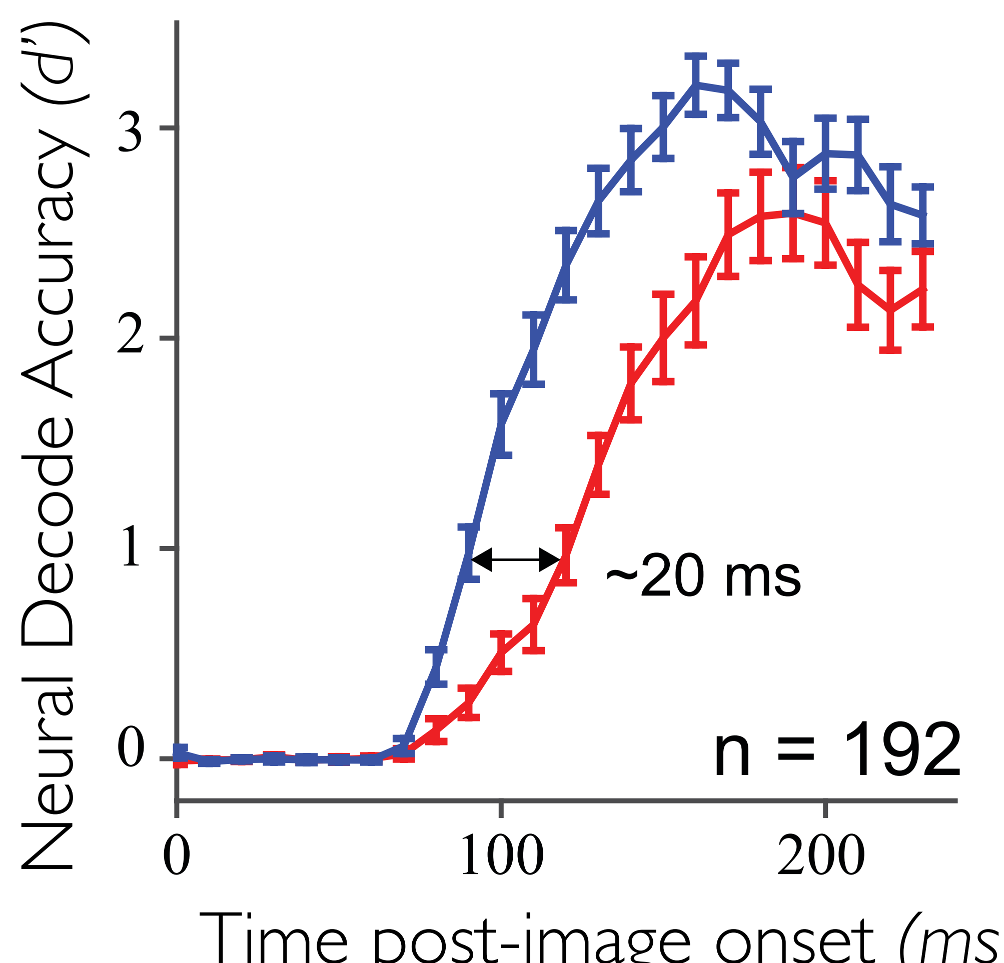
Evidence of Functional Relevance During Object Recognition



Kar et. al. (2019)

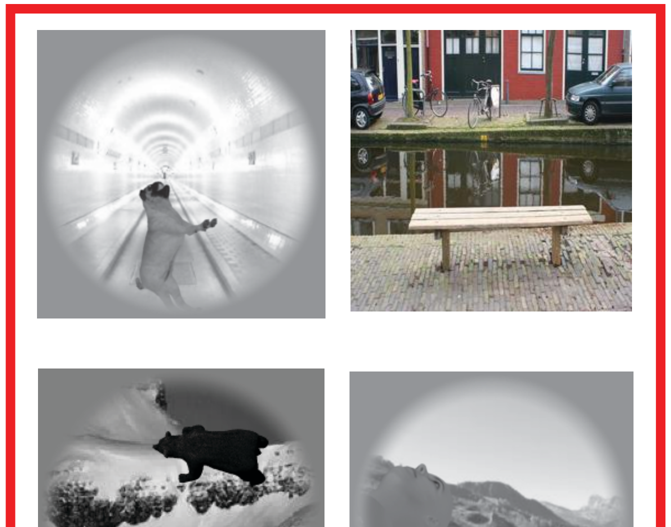
Evidence of Functional Relevance During Object Recognition

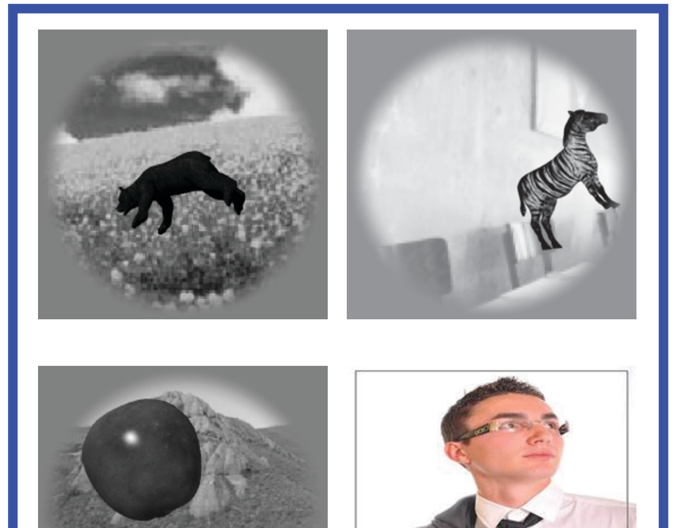




Kar et. al. (2019)

Evidence of Functional Relevance During Object Recognition

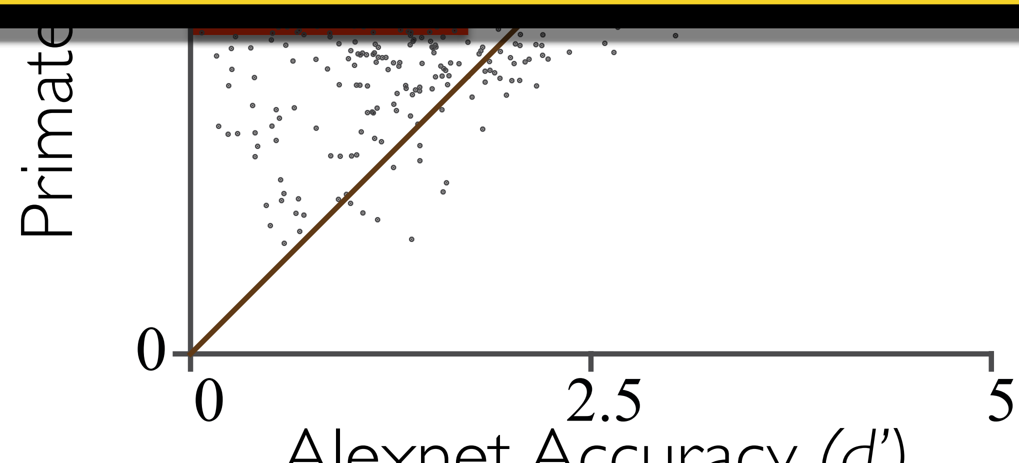


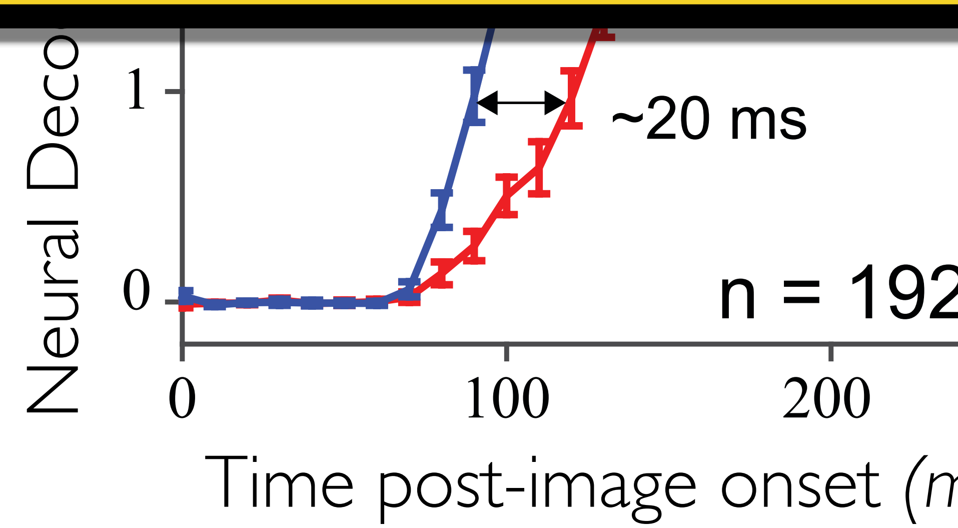


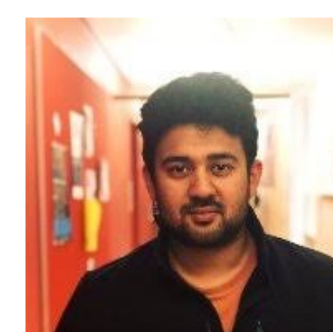
CNN-not-solved images ***are*** solved by the primate ventral stream later in time!

Neurobiological Puzzle:

What is the role of recurrent processing in the primate ventral stream during object recognition?



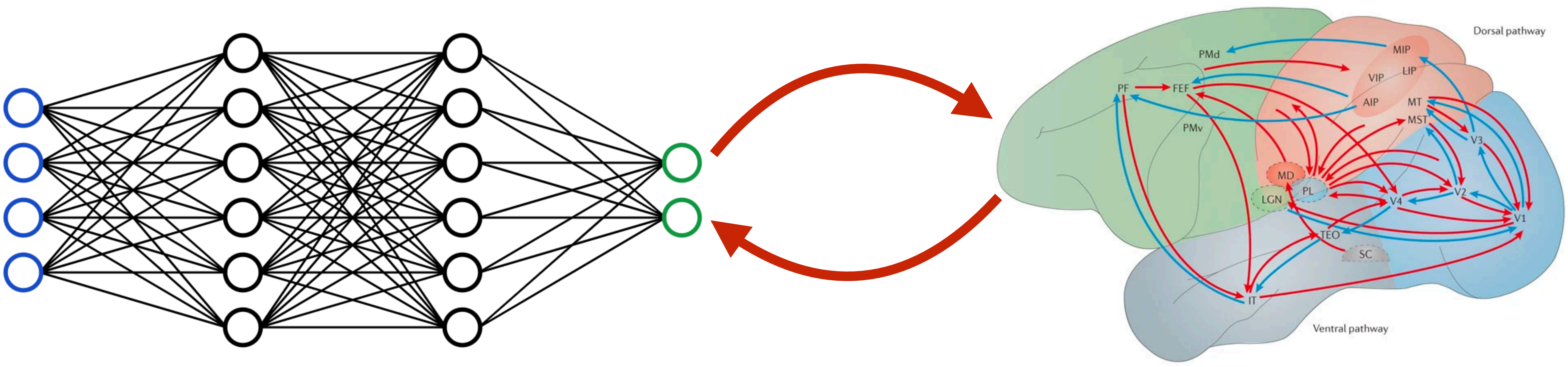




Kar et. al. (2019)

Task-Optimized Modeling

Design ML Algorithms Optimized to Perform Organism's Behavior under Organism's Constraints



Yields:

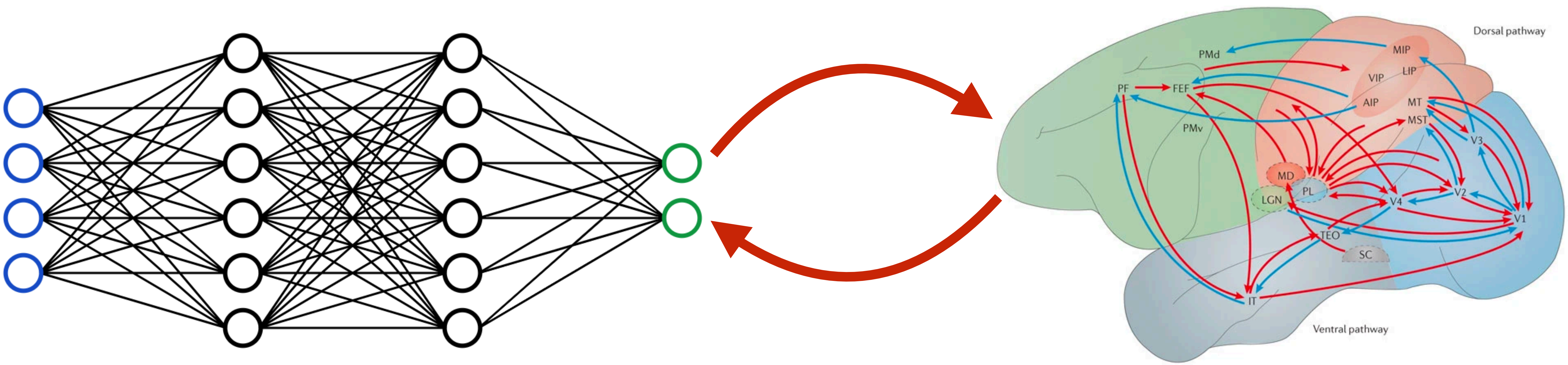
Quantitatively Accurate & Practically Useful Brain Models

AND

Principles of *Why* Neural Responses Are As They Are

Task-Optimized Modeling

Design ML Algorithms Optimized to Perform Organism's Behavior under Organism's Constraints



Yields:

Quantitatively Accurate & Practically Useful Brain Models

AND

Principles of *Why* Neural Responses Are As They Are

CNNs as Models of Primate Object Recognition

L = *learning rule*

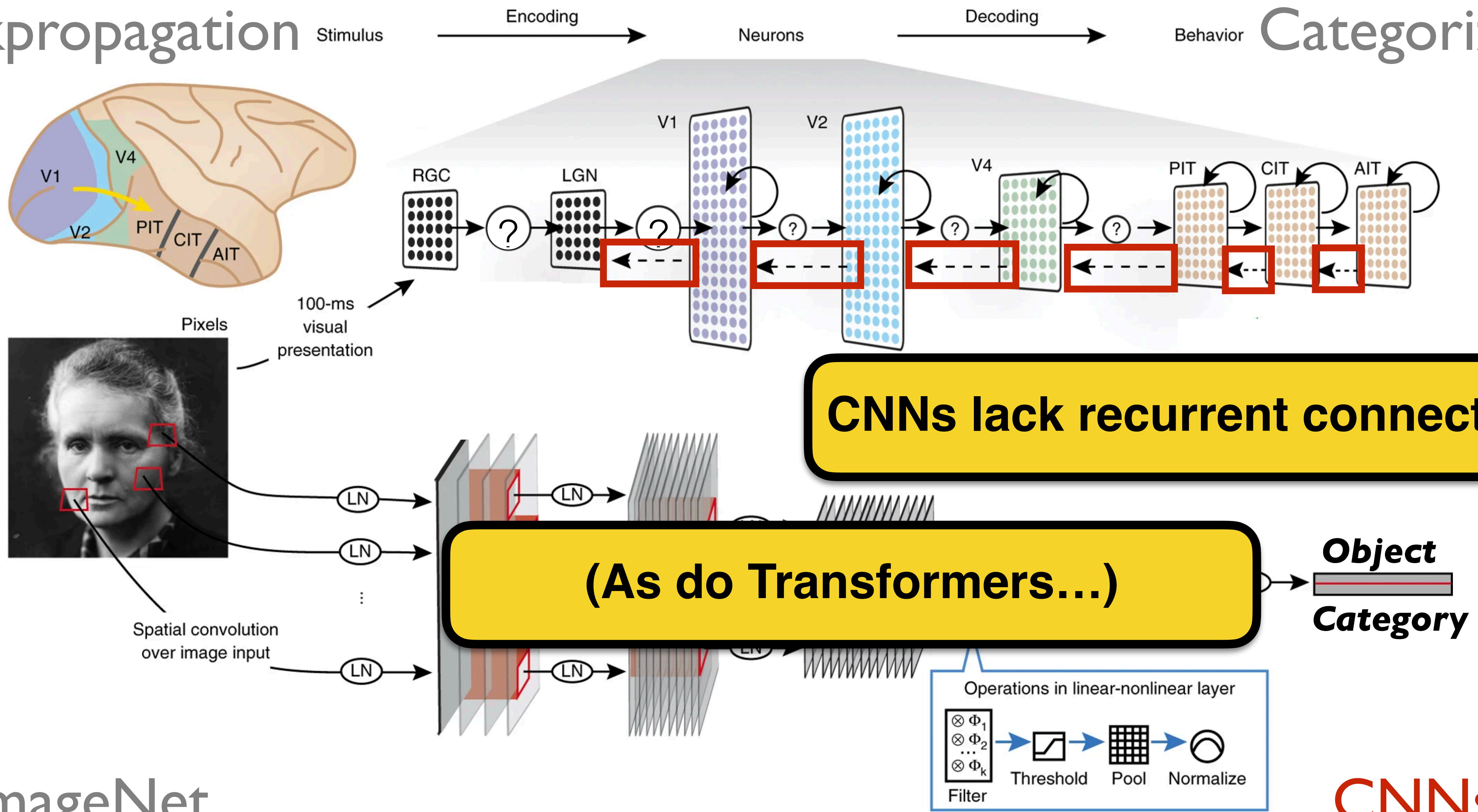
"Natural selection + plasticity"

Backpropagation

T = *task loss*

"Ecological niche/ behavior"

Categorization



CNNs lack recurrent connections!

(As do Transformers…)

ImageNet

"Environment"

D = *data stream*

CNNs

"Circuit"

A = *architecture class*

Convolutional Recurrent Networks (ConvRNNs)

L = *learning rule*

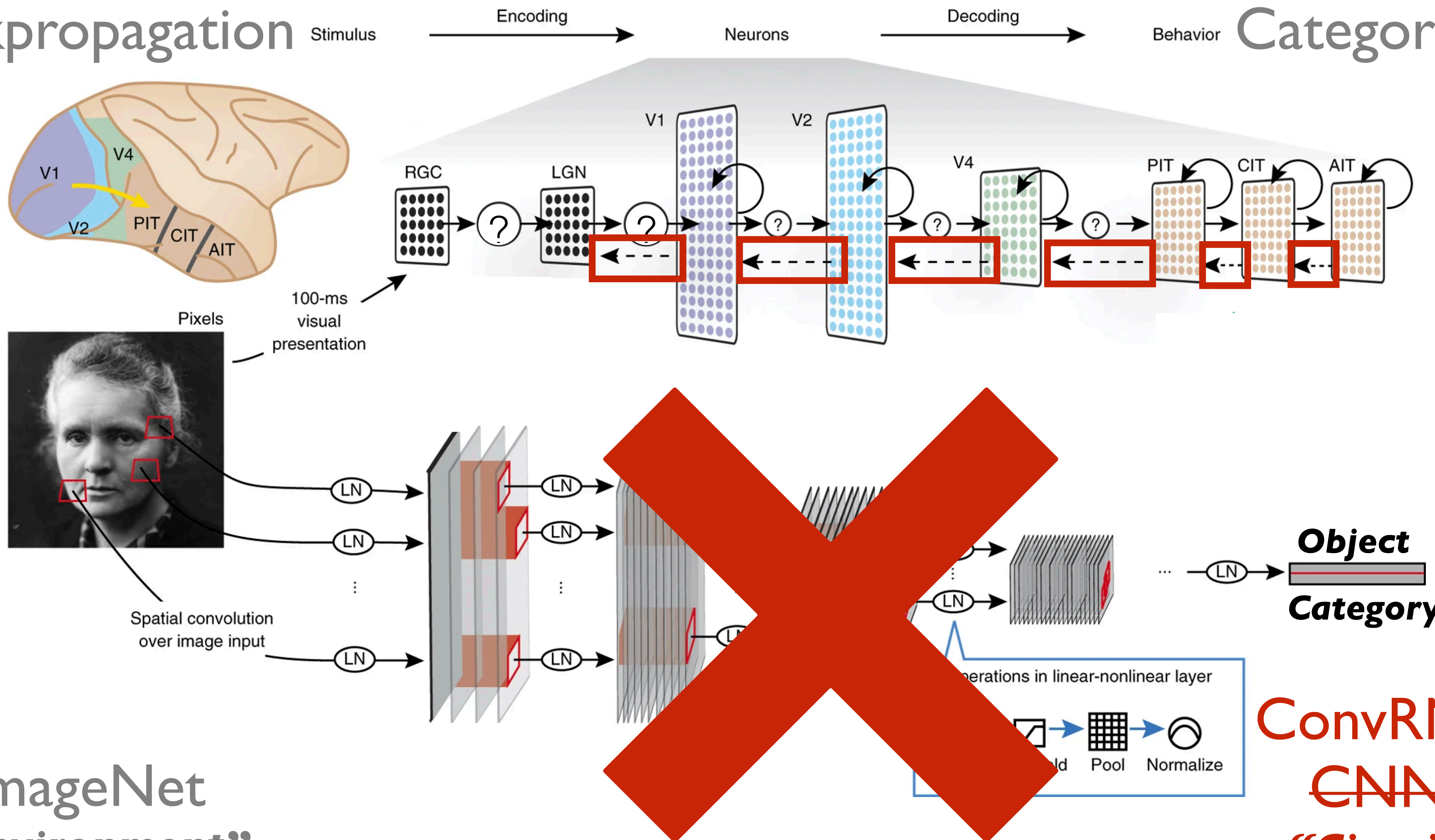
"Natural selection + plasticity"

Backpropagation

T = *task loss*

"Ecological niche/ behavior"

Categorization



ImageNet

"Environment"

D = *data stream*

ConvRNNs

~~CNNs~~

"Circuit"

A = *architecture class*

Convolutional Recurrent Networks (ConvRNNs)

L = *learning rule*

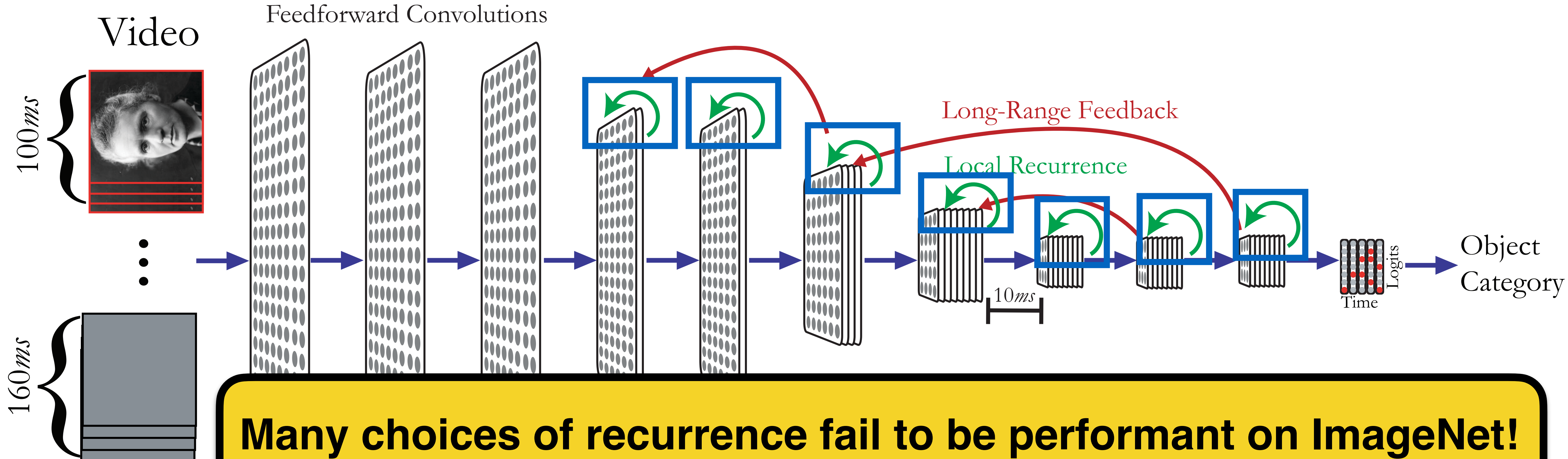
"Natural selection + plasticity"

Backpropagation

T = *task loss*

"Ecological niche/ behavior"

Categorization



Many choices of recurrence fail to be performant on ImageNet!

ImageNet

"Environment"

Each time-step (10 *ms*) is treated equally — including feedforward steps

ConvRNNs

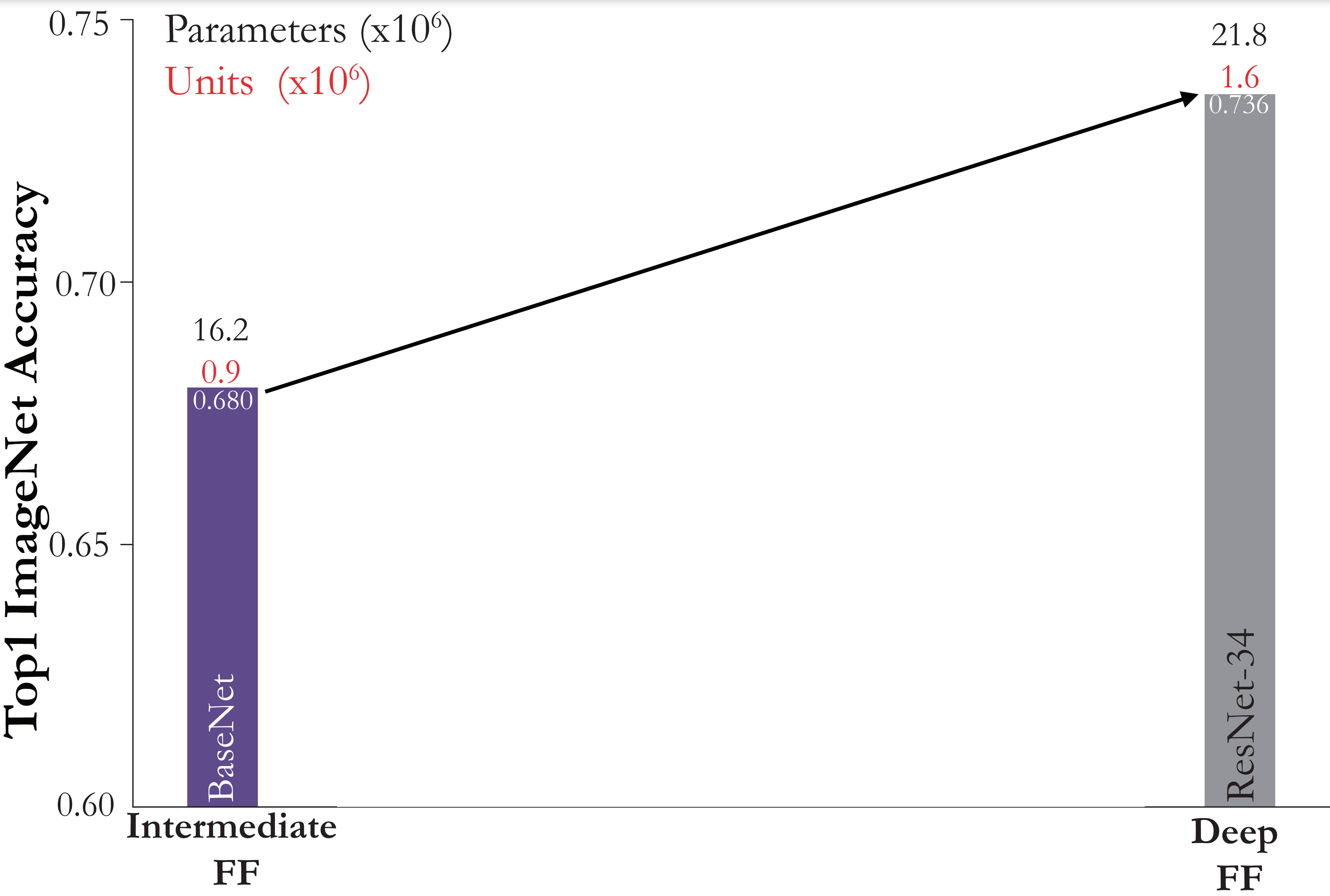
~~CNNs~~

"Circuit"

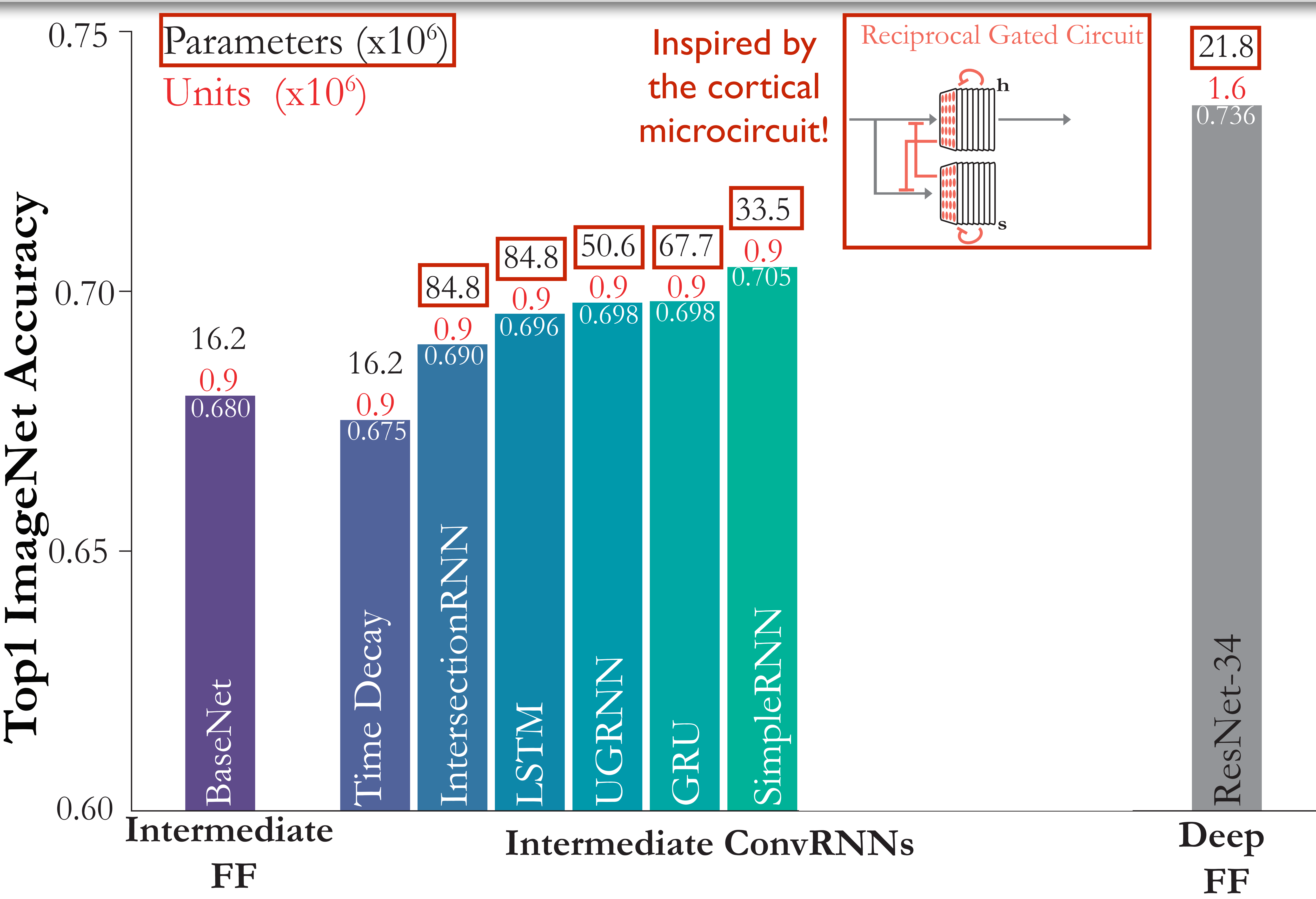
D = *data stream*

A = *architecture class*

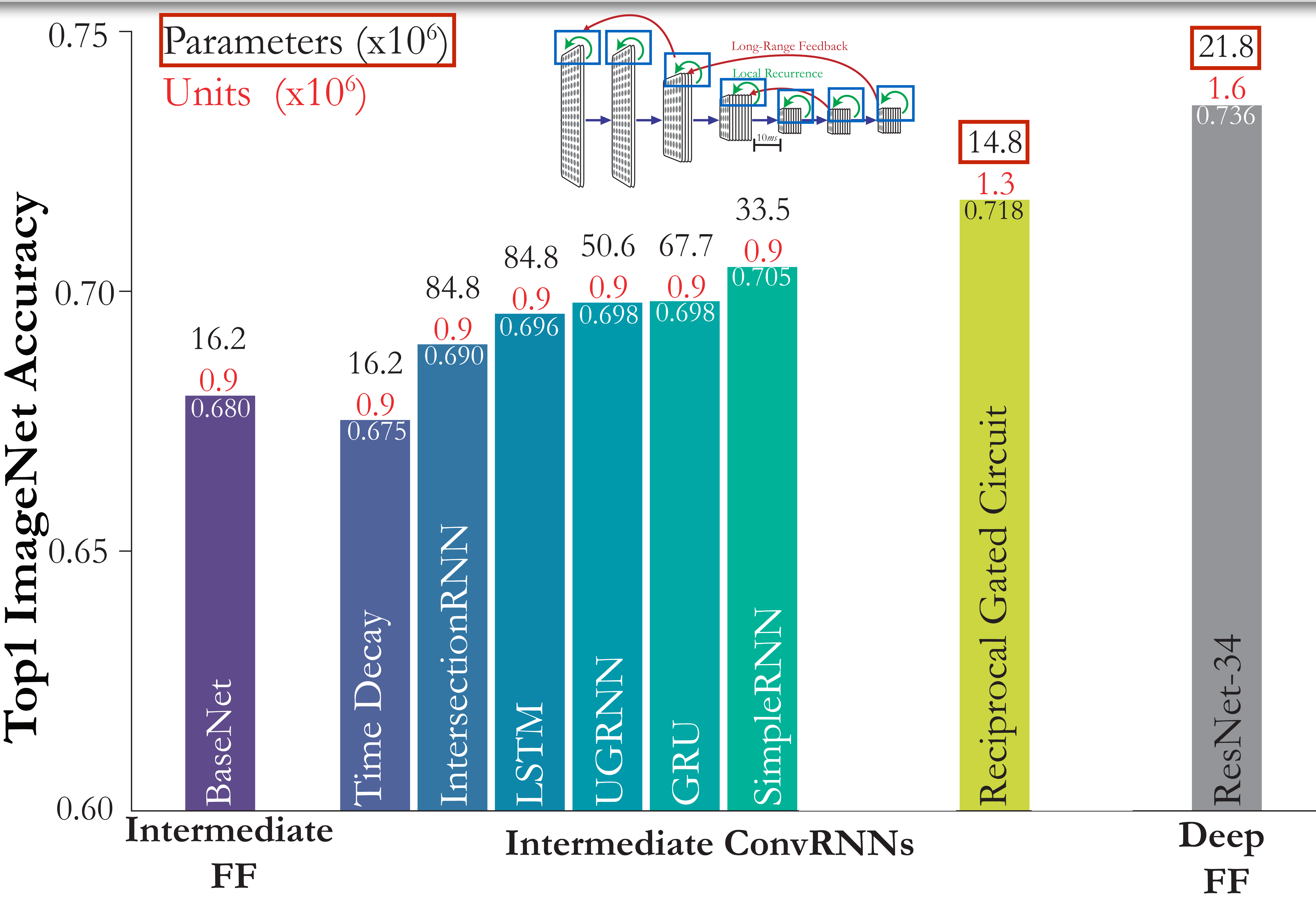
Implanting Local Recurrence into Feedforward CNNs



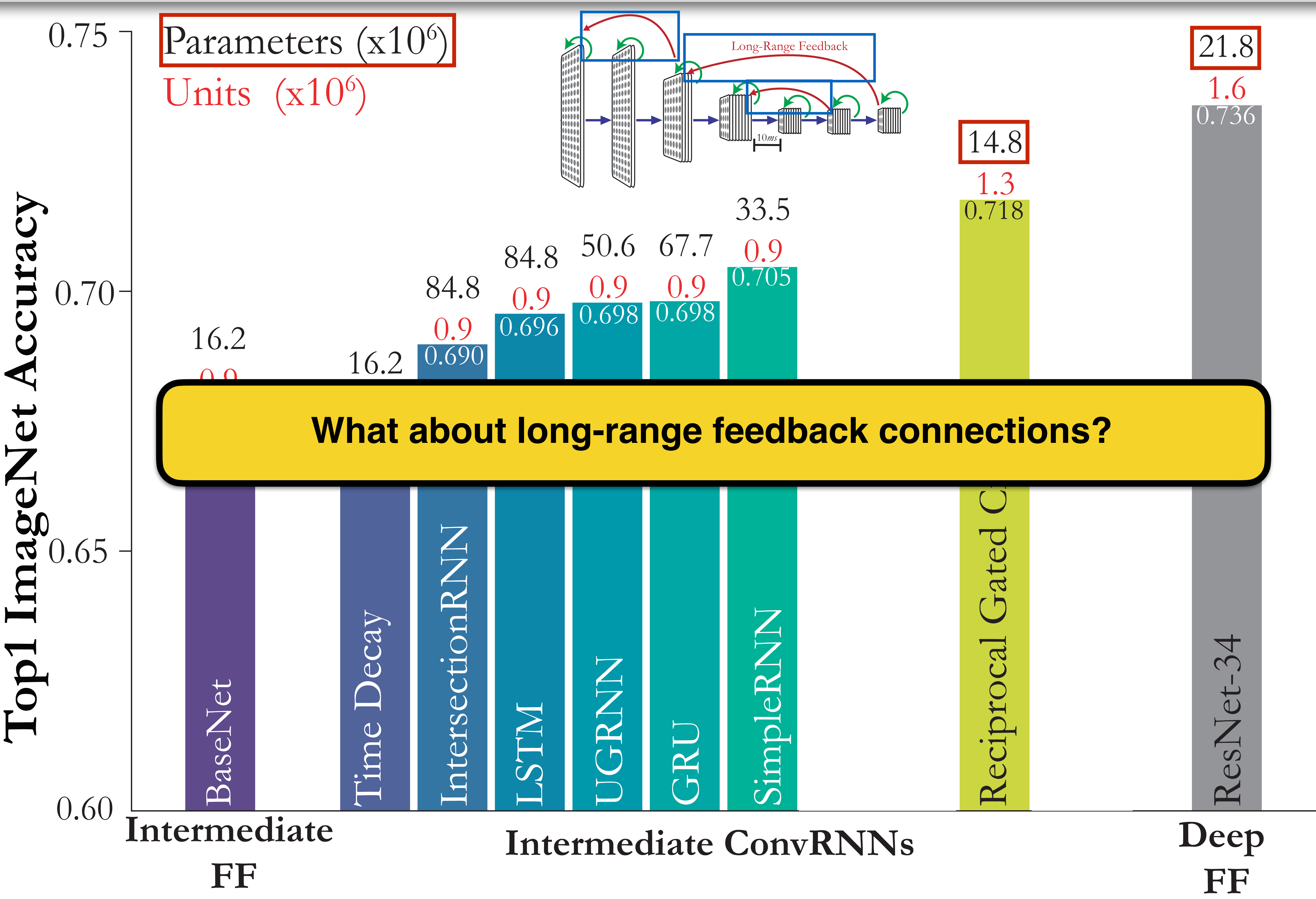
Adding Standard RNNs Helps Incrementally, but Add **Lots** of Parameters!



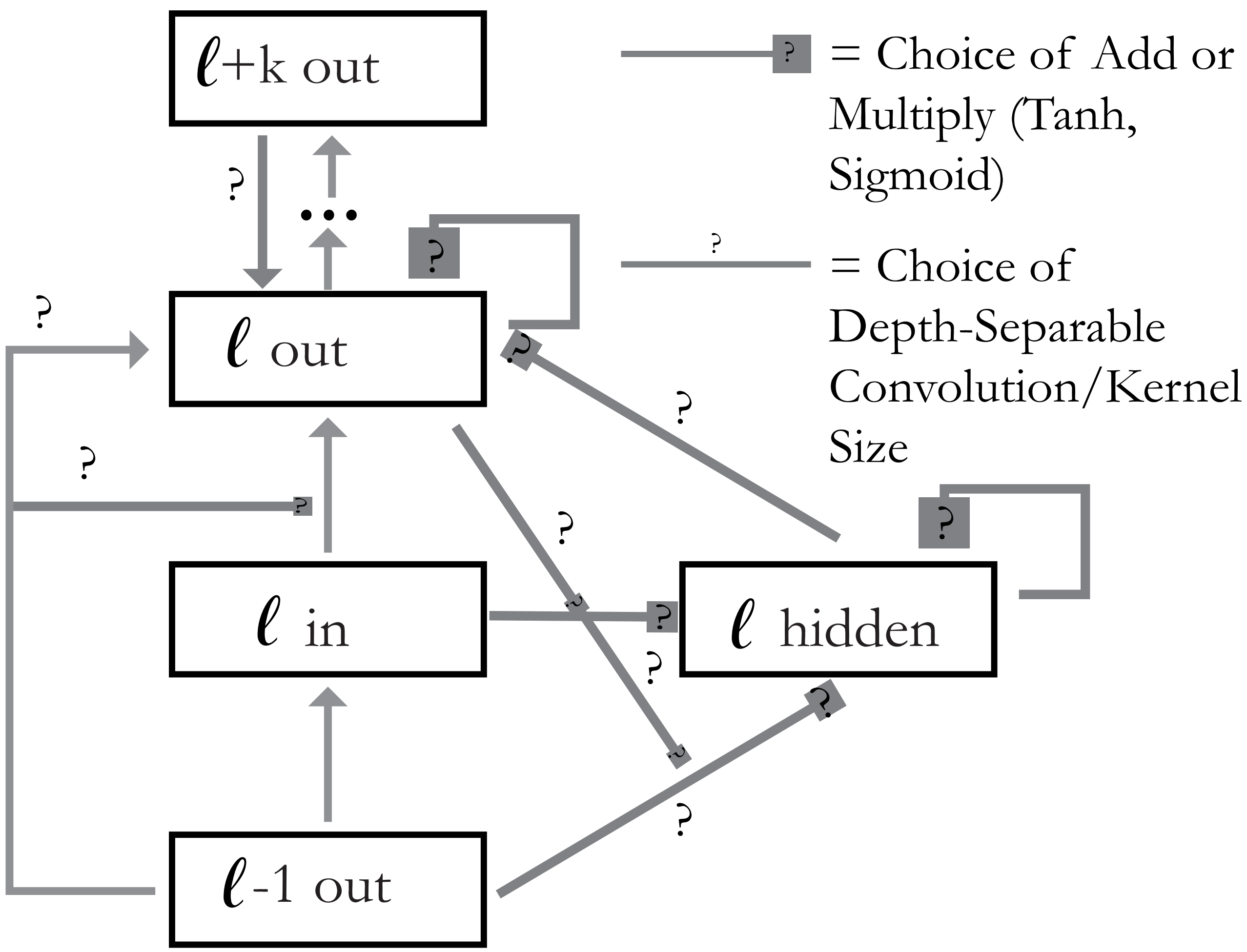
Novel Recurrent Cells Yield Improved ImageNet Performance

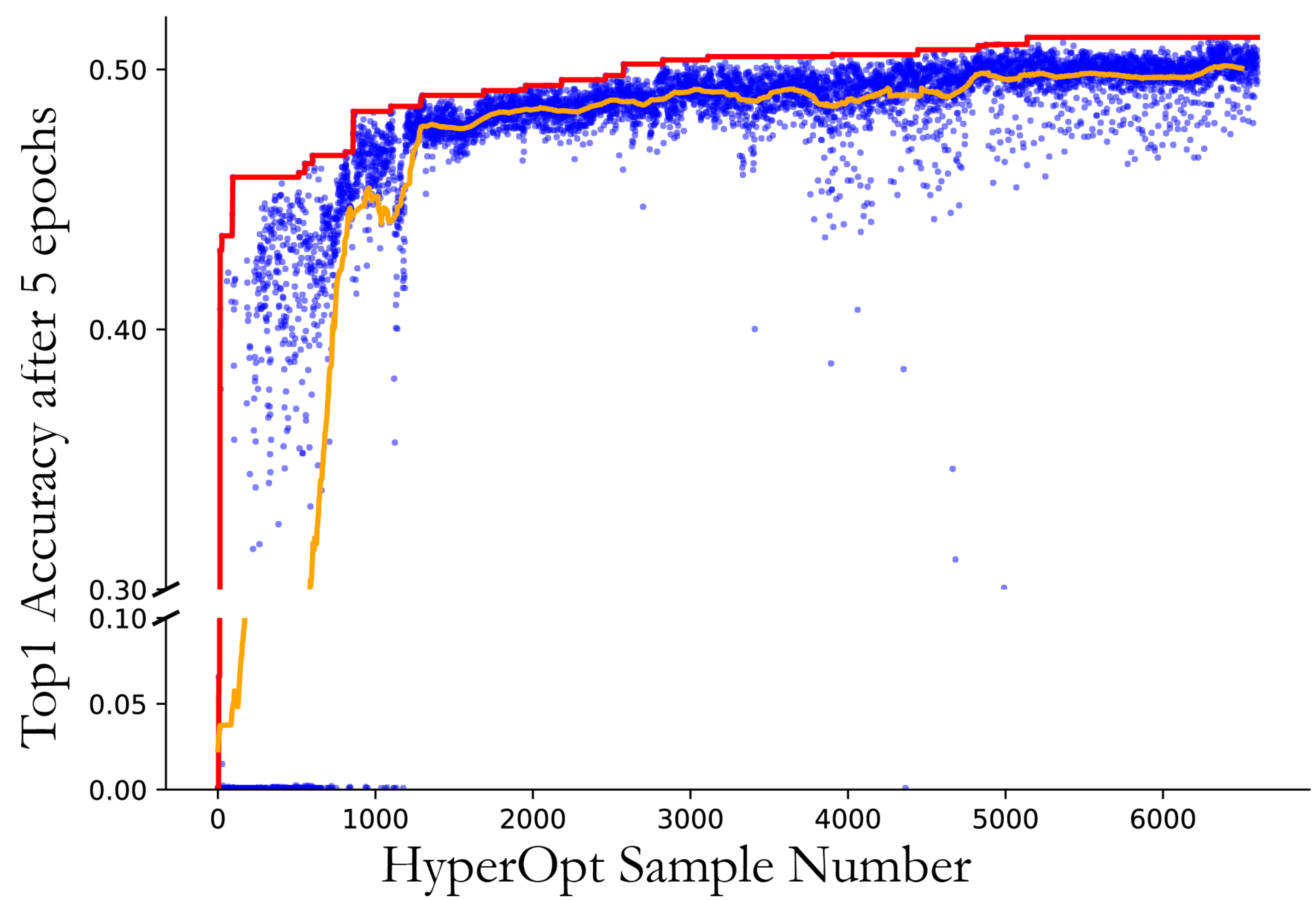


Novel Recurrent Cells Yield Improved ImageNet Performance

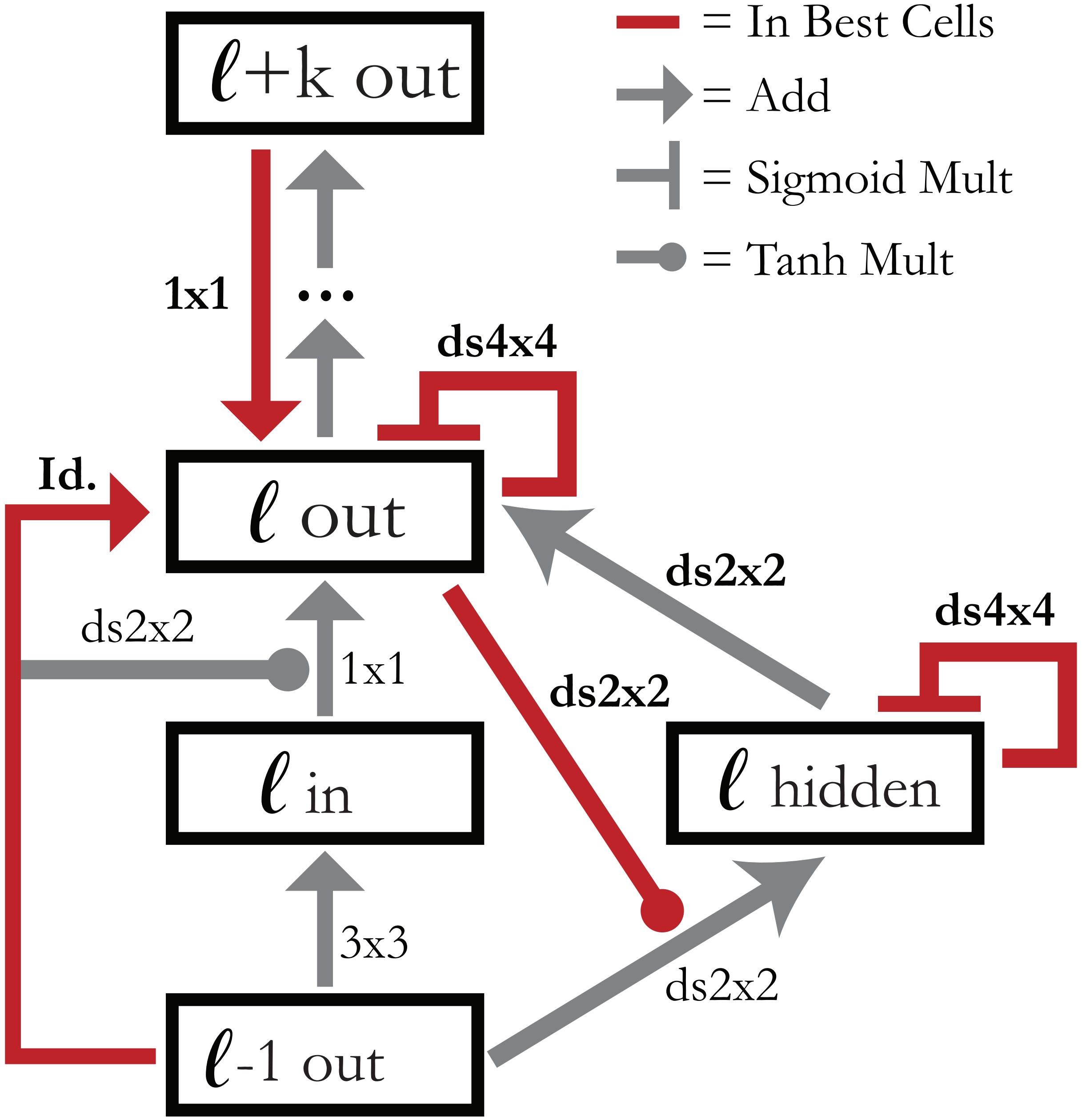


Large-Scale Search Over Long-Range Feedback Connections



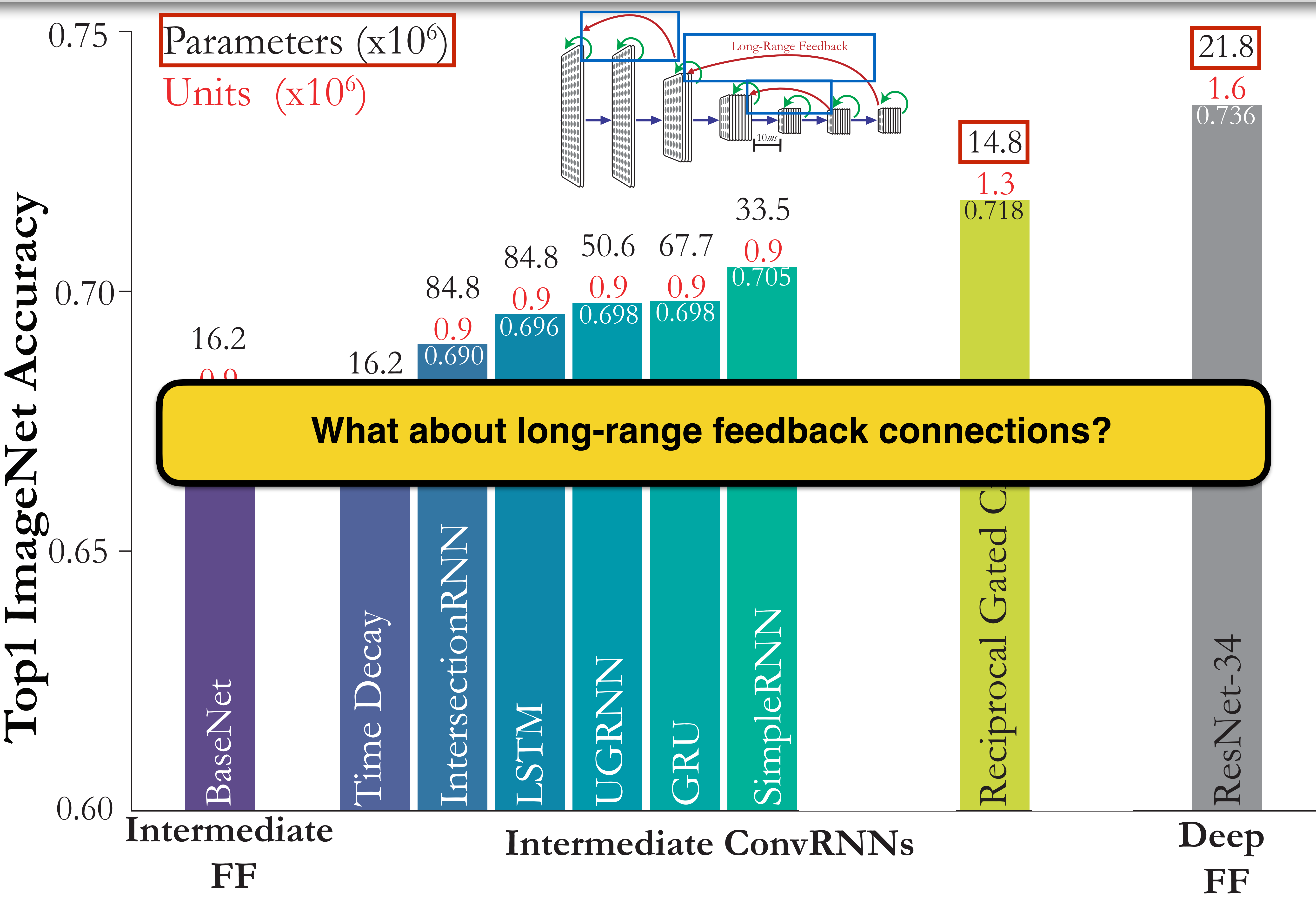


Emergent Global Connectivity Patterns

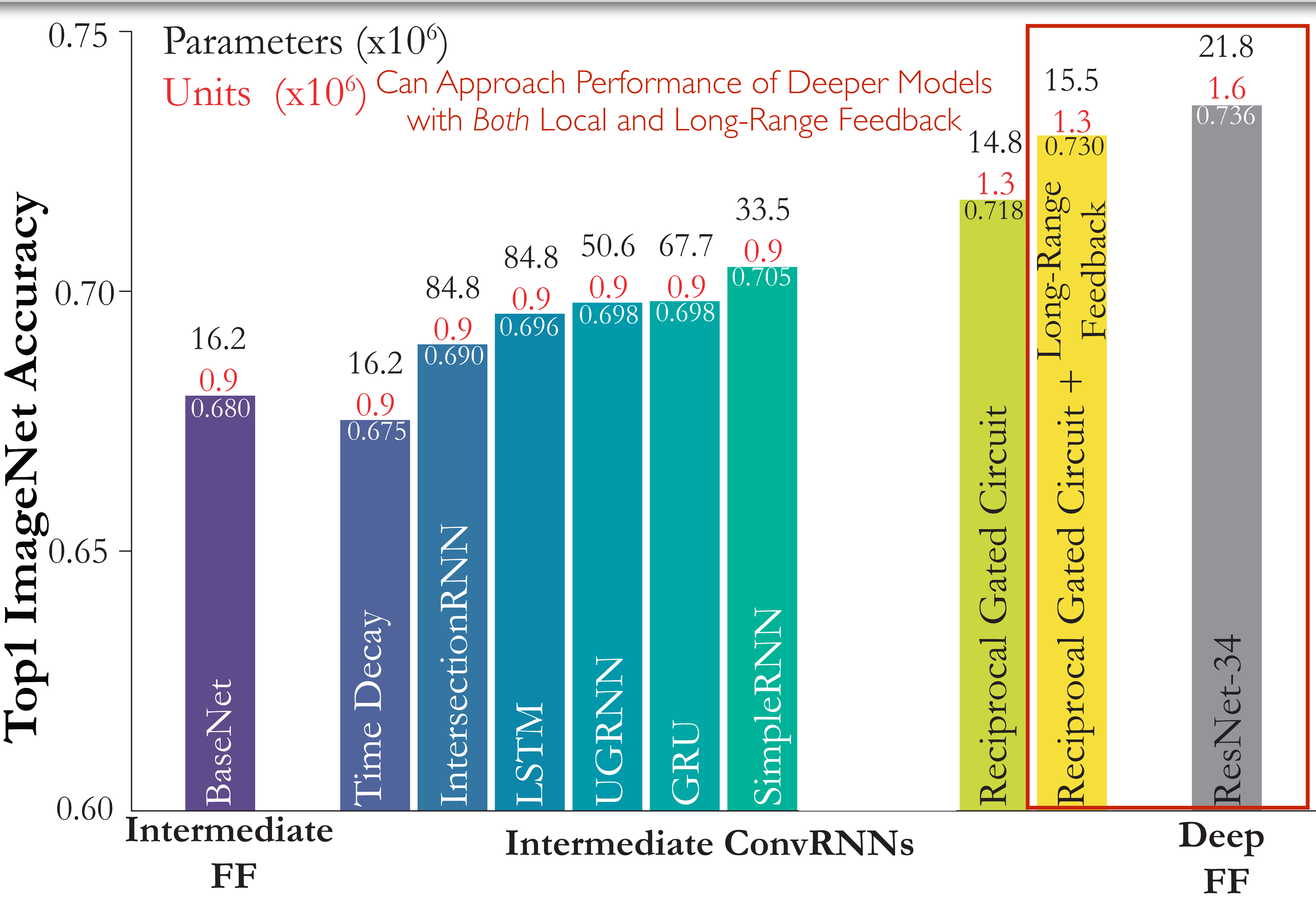


Conservation on parameter count as a byproduct of evolutionary optimization

Novel Recurrent Cells Yield Improved ImageNet Performance

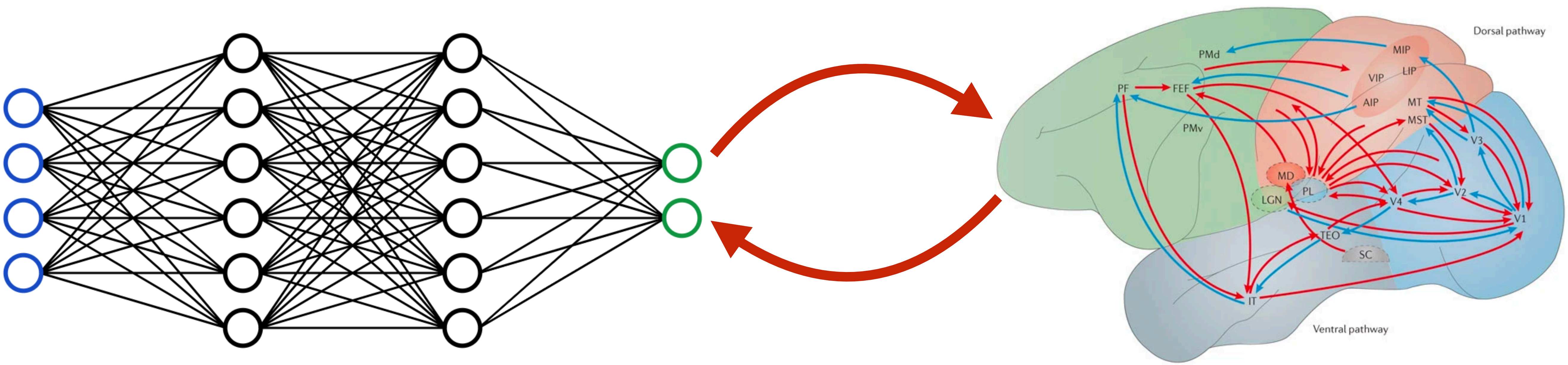


Long-Range Feedback Connections Matter



Task-Optimized Modeling

Design ML Algorithms Optimized to Perform Organism's Behavior under Organism's Constraints



Yields:

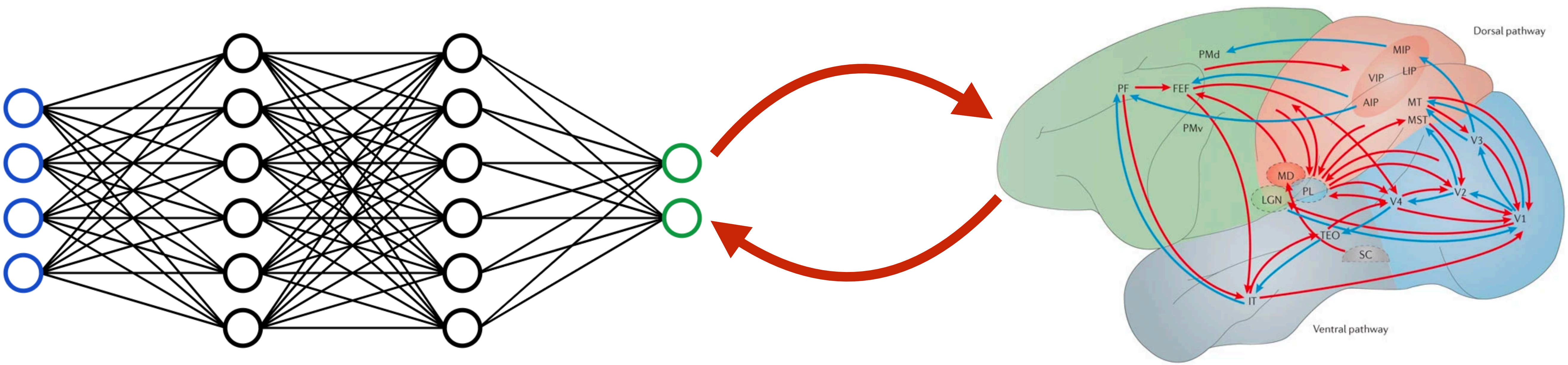
Quantitatively Accurate & Practically Useful Brain Models

AND

Principles of *Why* Neural Responses Are As They Are

Task-Optimized Modeling

Design ML Algorithms Optimized to Perform Organism's Behavior under Organism's Constraints



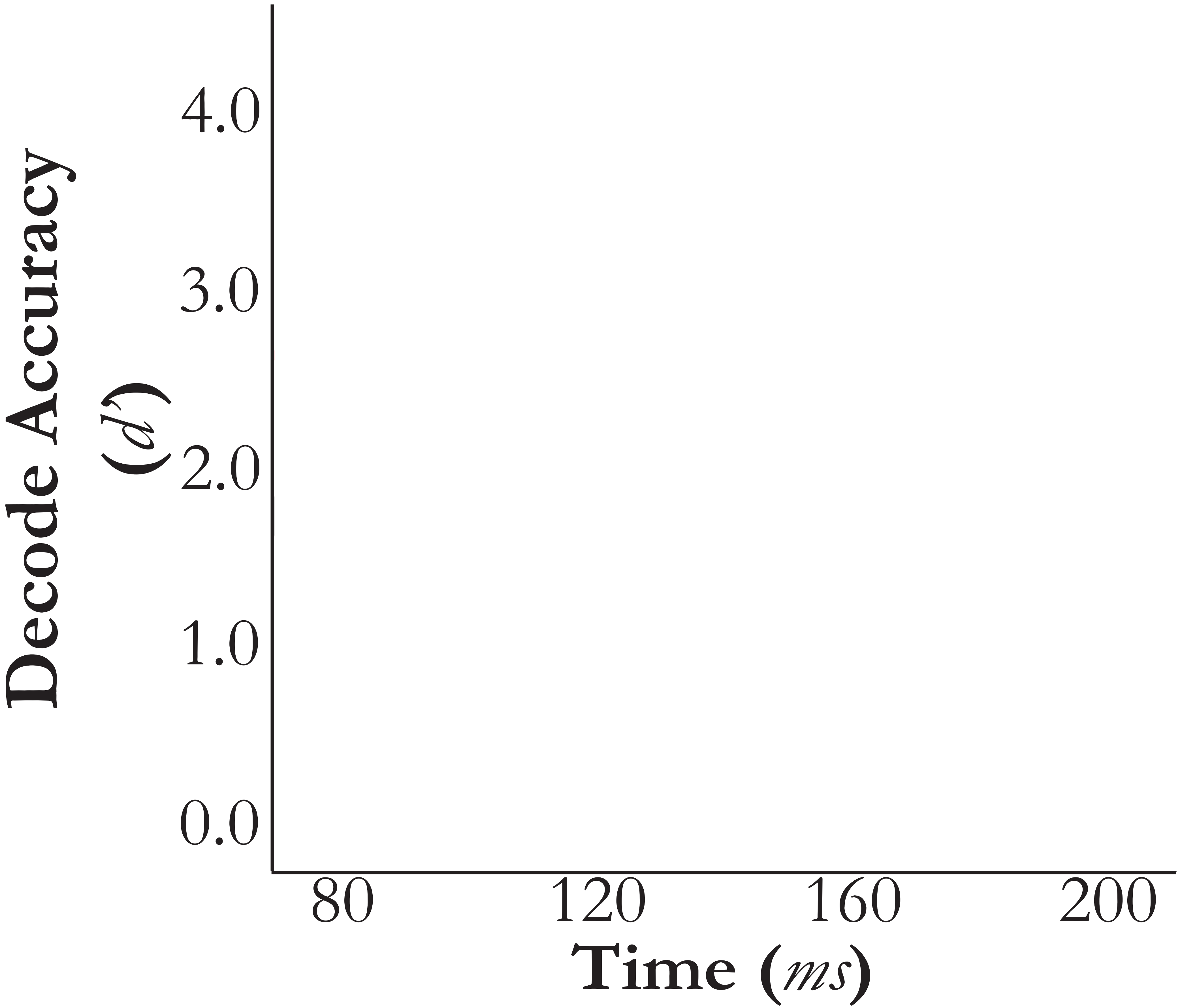
Yields:

Quantitatively Accurate & Practically Useful Brain Models

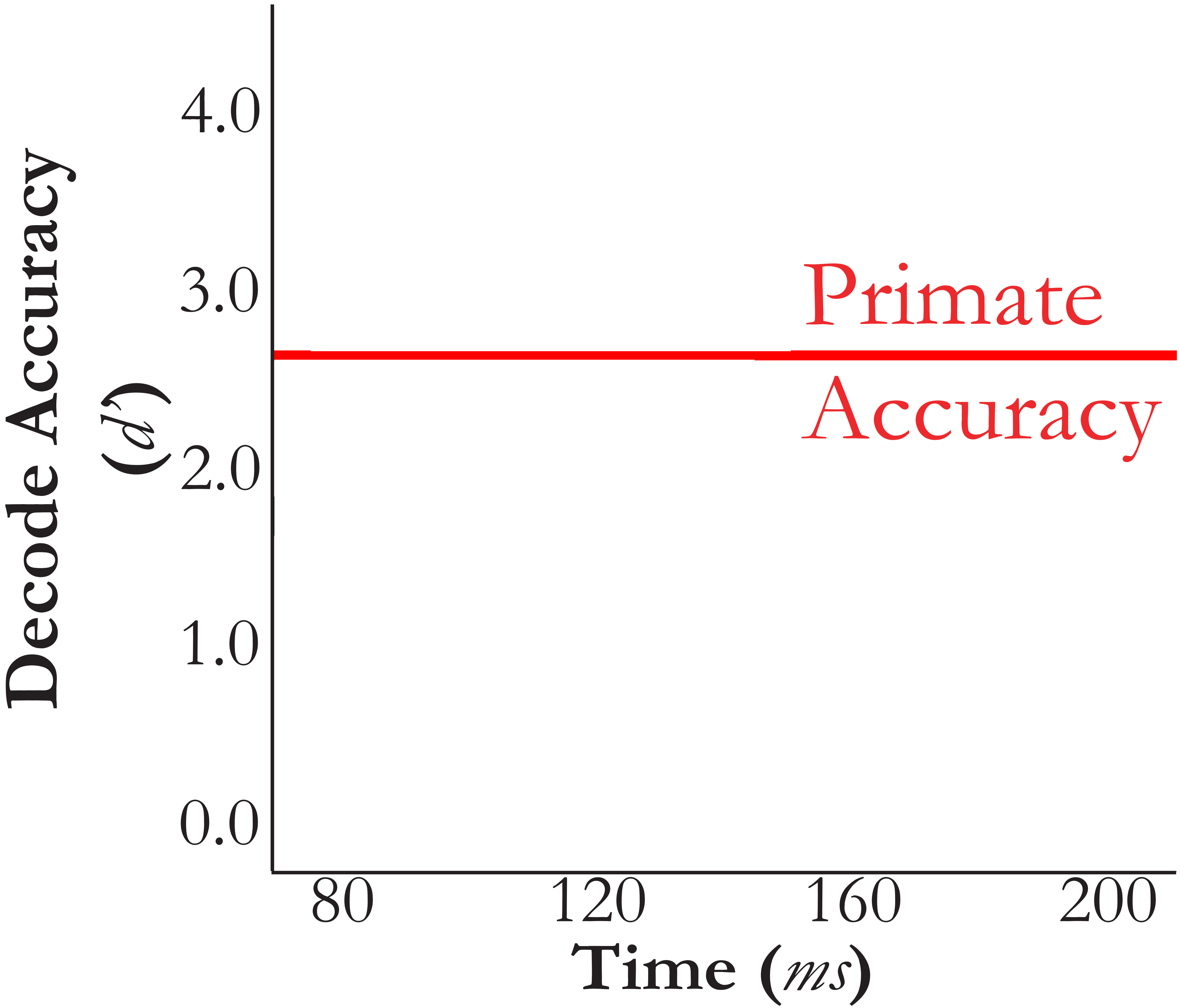
AND

Principles of *Why* Neural Responses Are As They Are

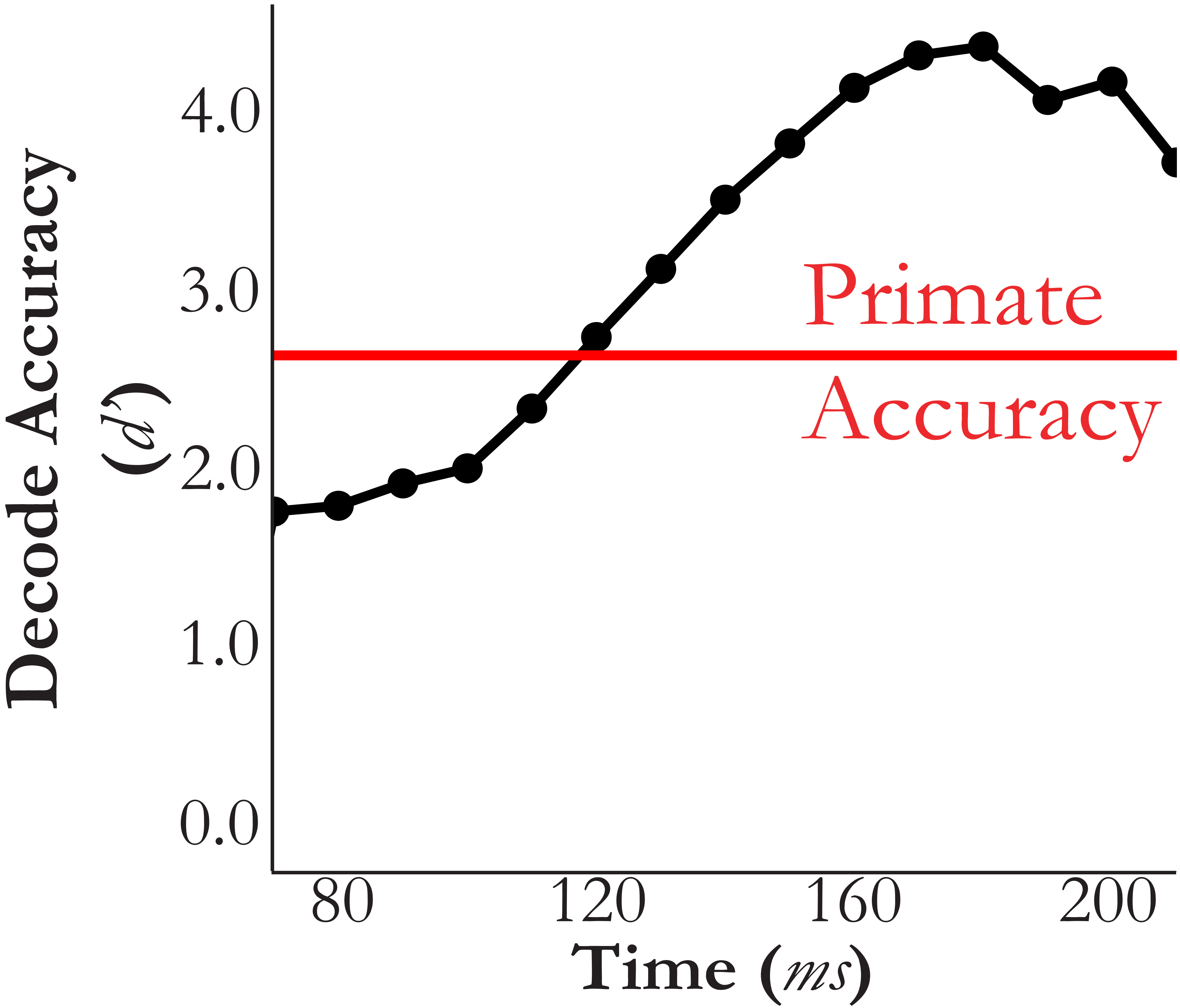
Comparing to Primate Object Solution Times (OSTs)



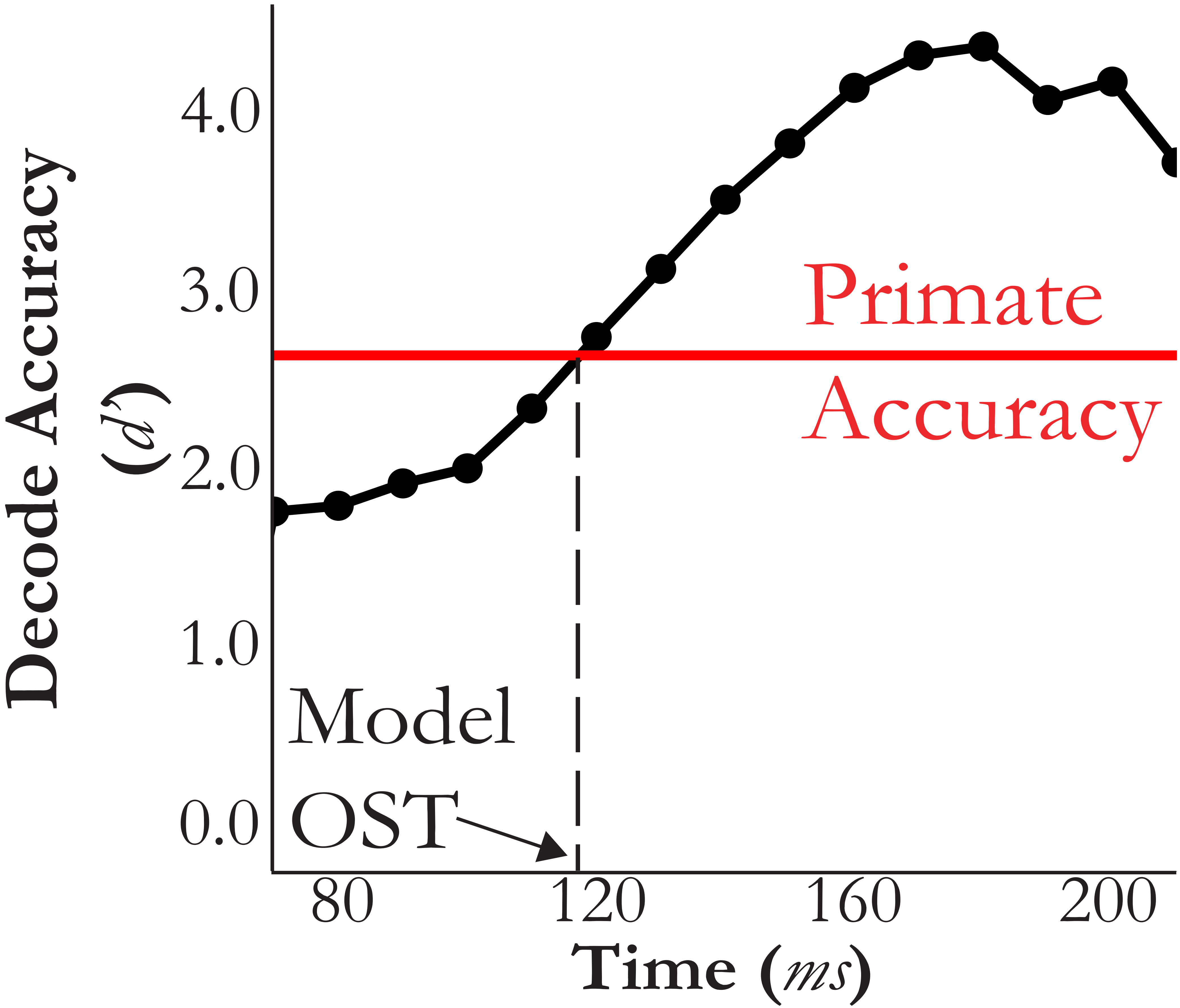
Comparing to Primate Object Solution Times (OSTs)



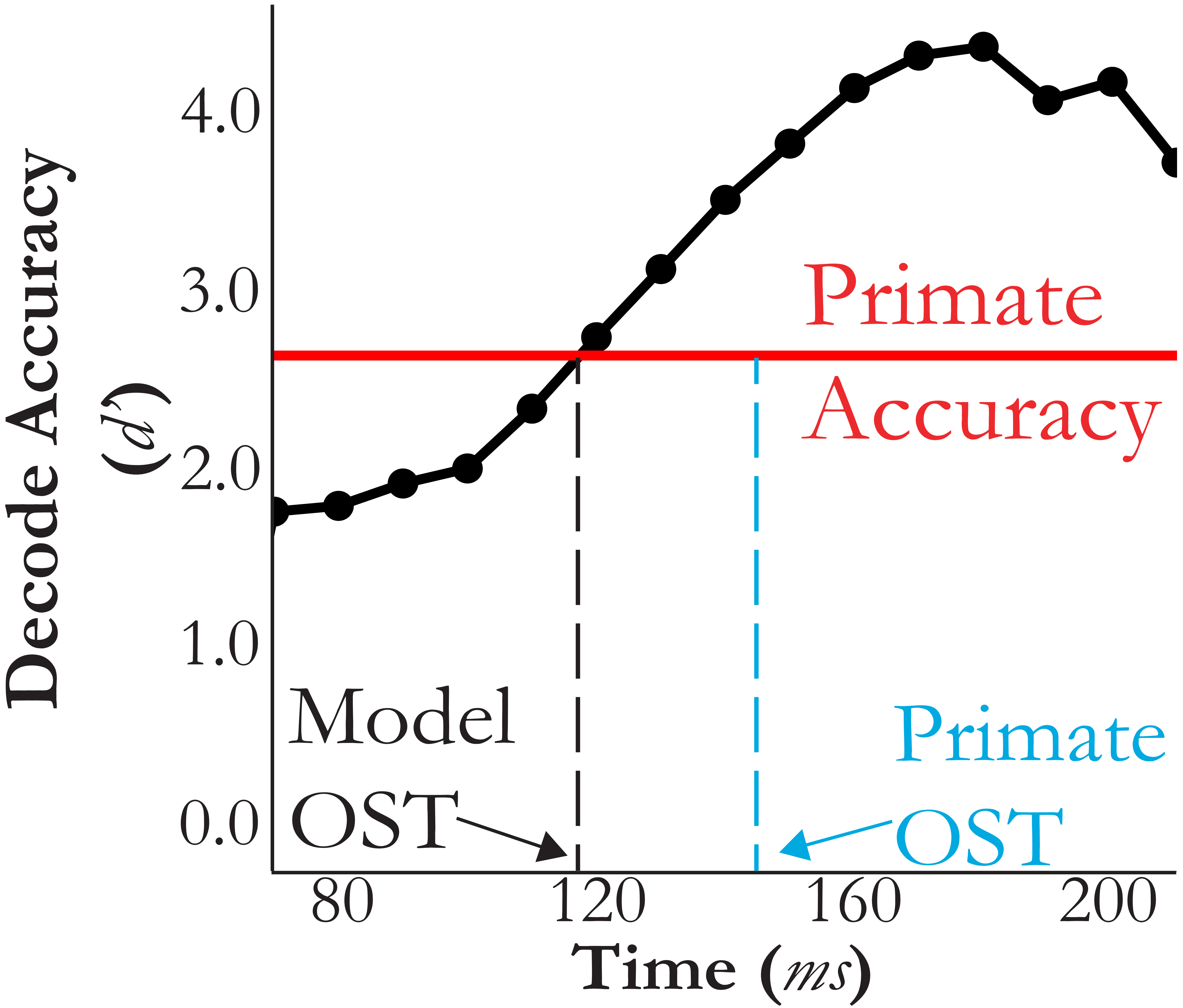
Comparing to Primate Object Solution Times (OSTs)



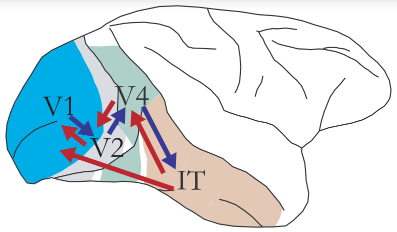
Comparing to Primate Object Solution Times (OSTs)

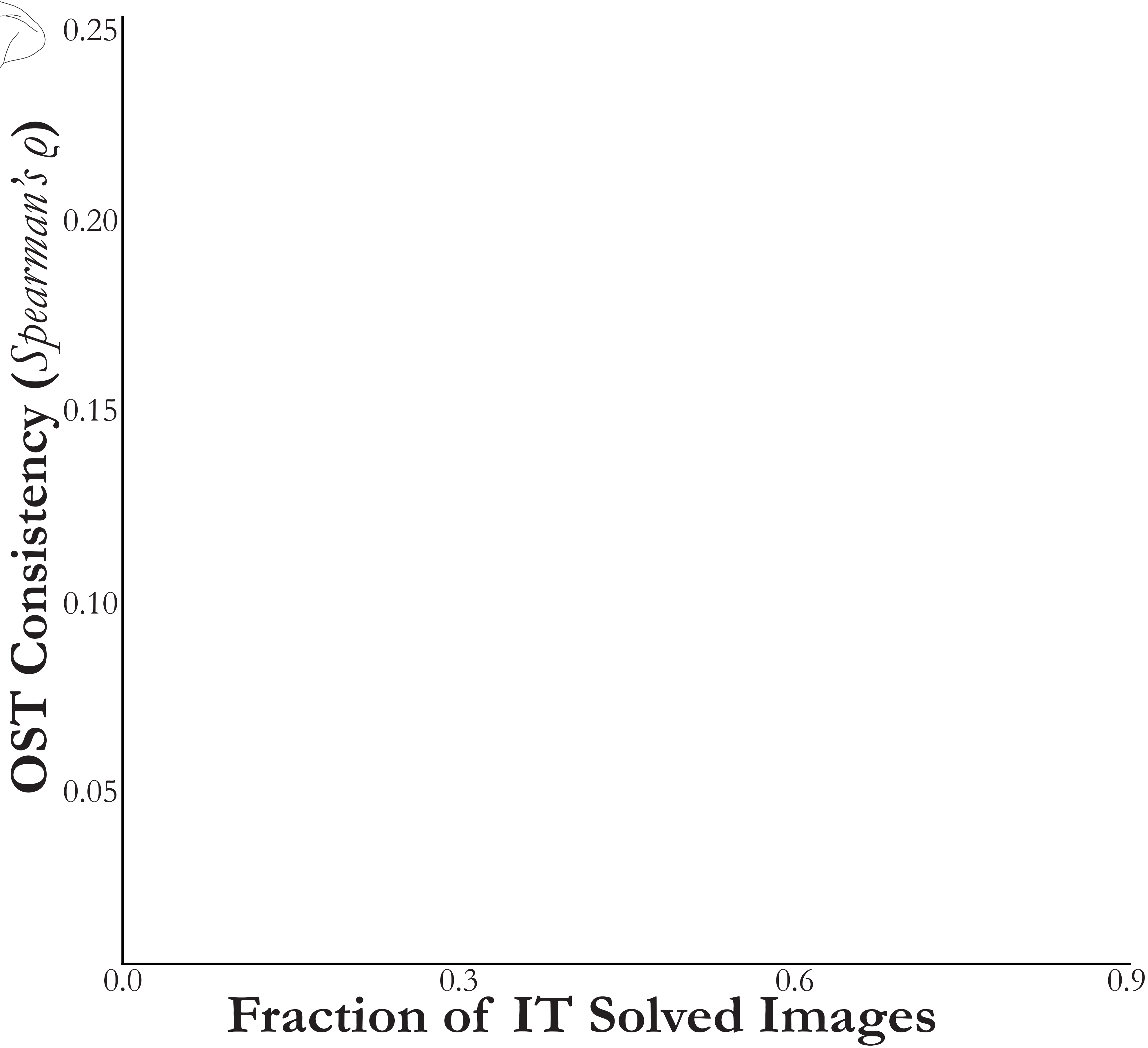


Comparing to Primate Object Solution Times (OSTs)

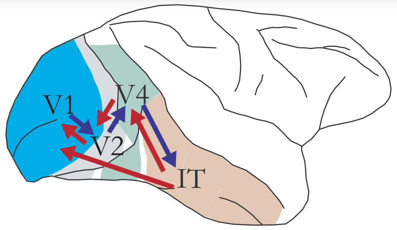


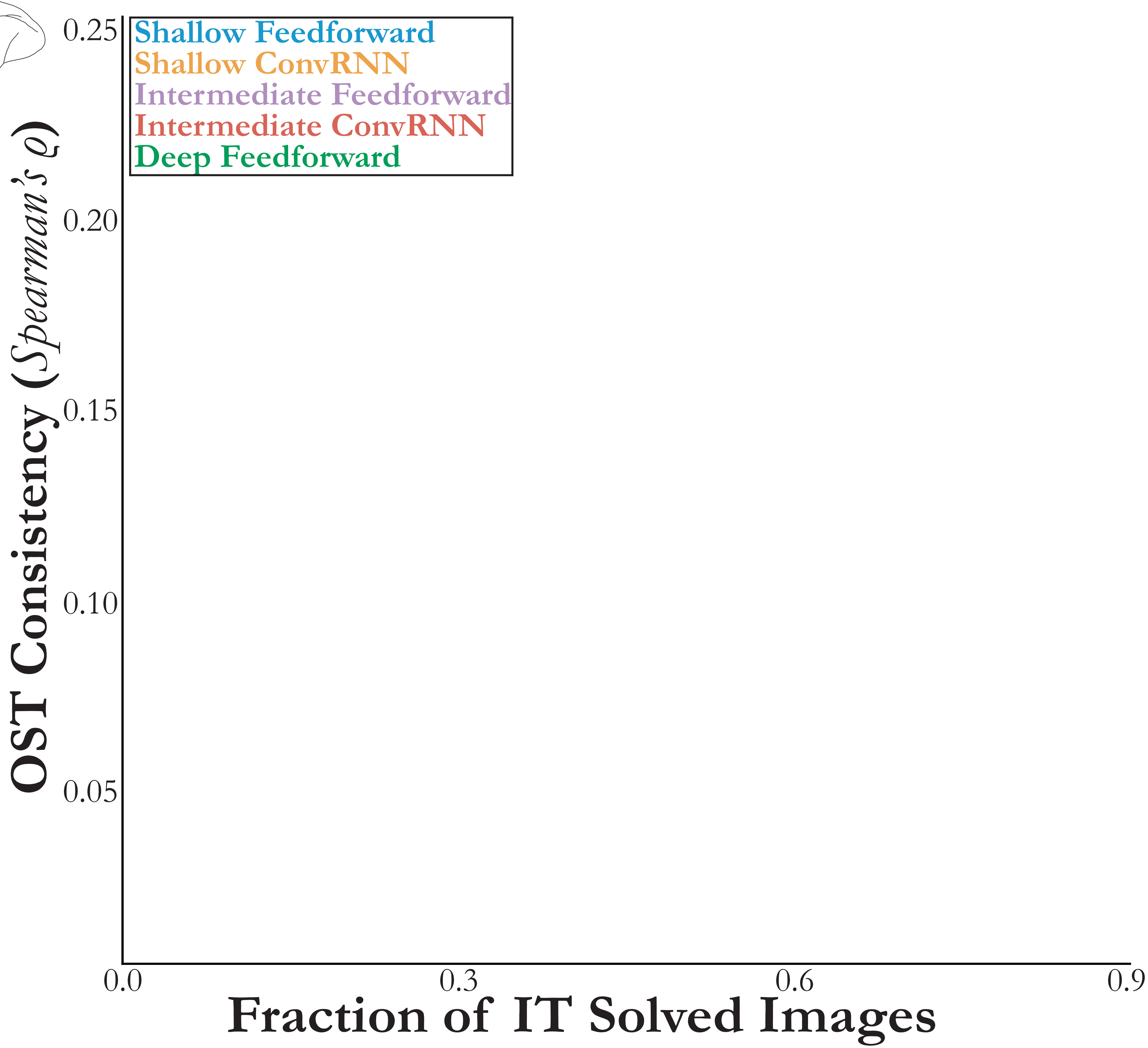
Comparing to Primate Object Solution Times (OSTs)



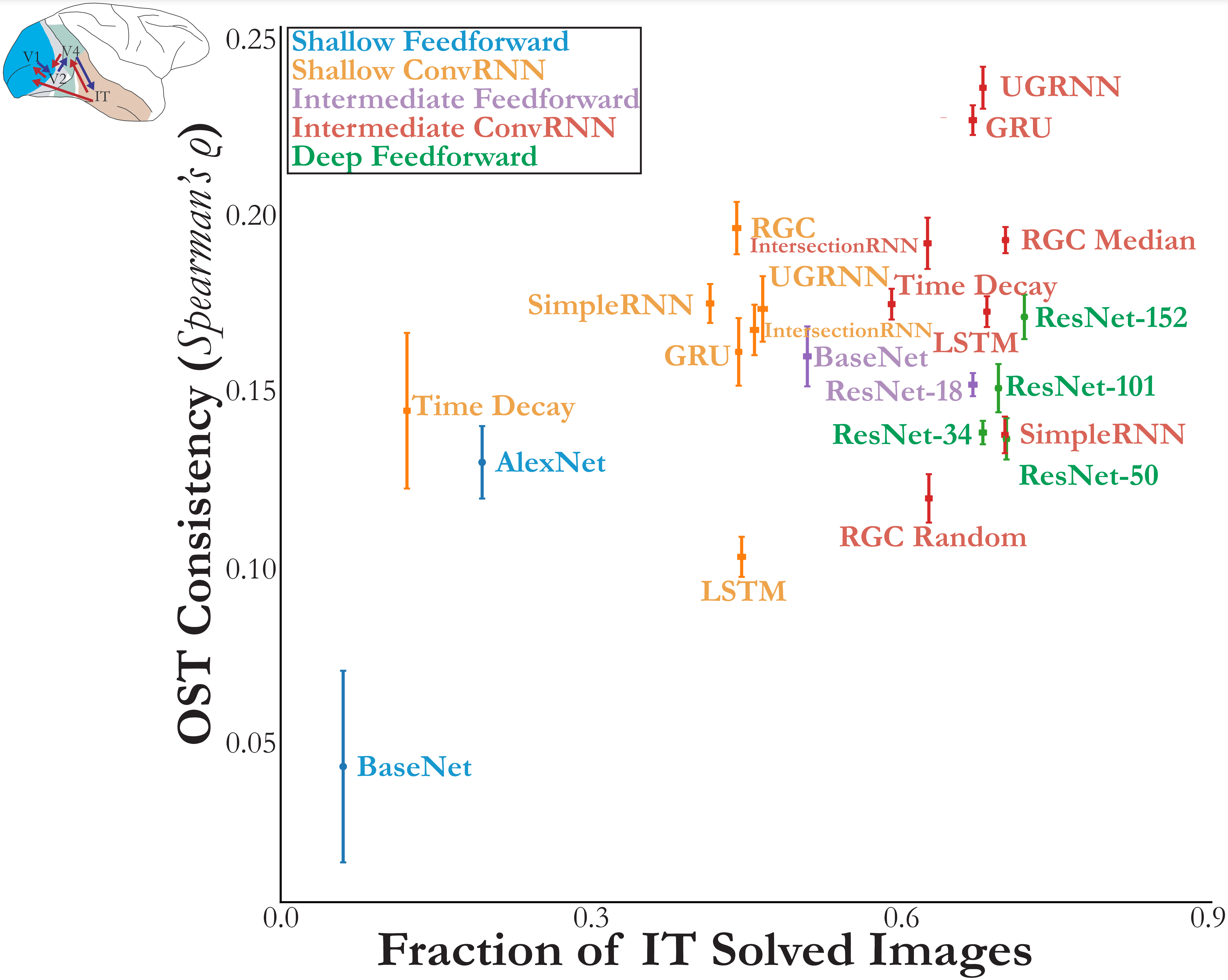


Comparing to Primate Object Solution Times (OSTs)

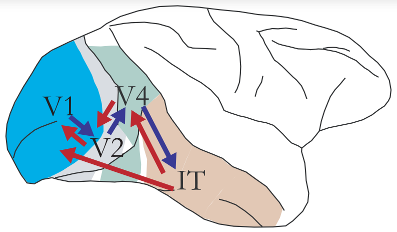


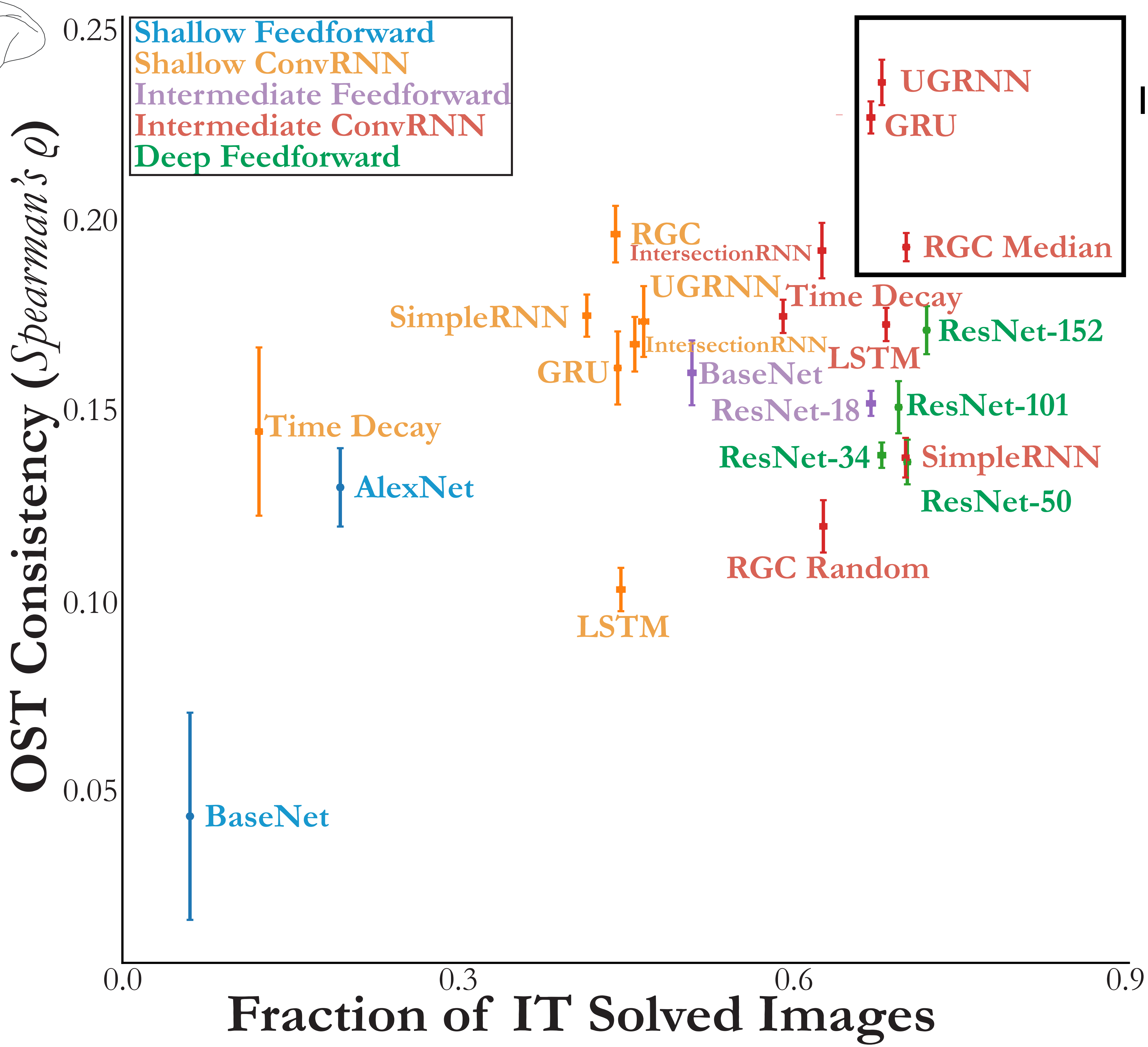


Comparing to Primate Object Solution Times (OSTs)



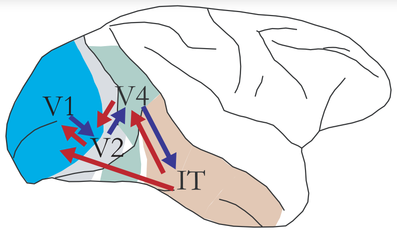
Intermediate Depth ConvRNNs best match OSTs

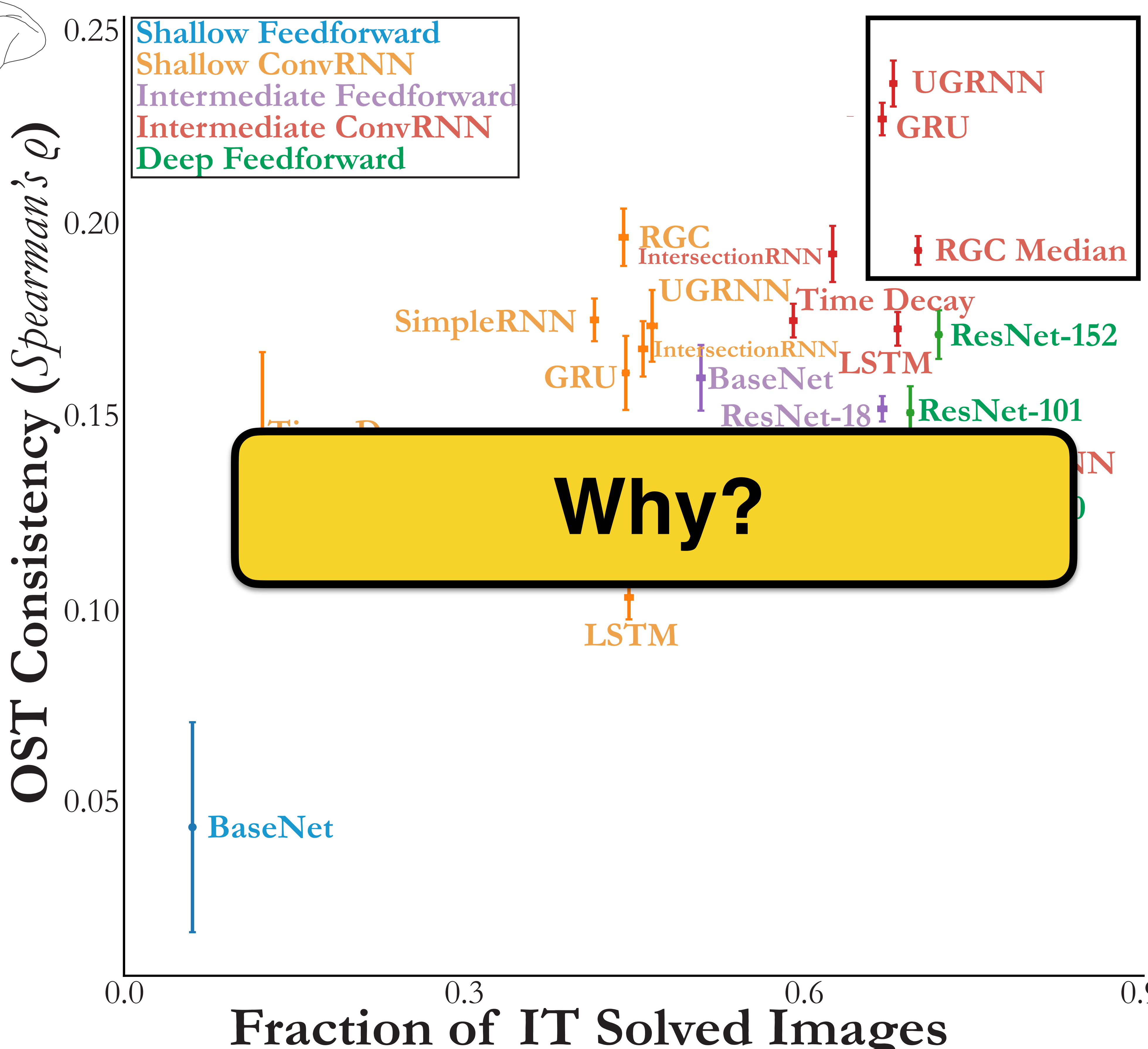




Intermediate Depth ConvRNNs

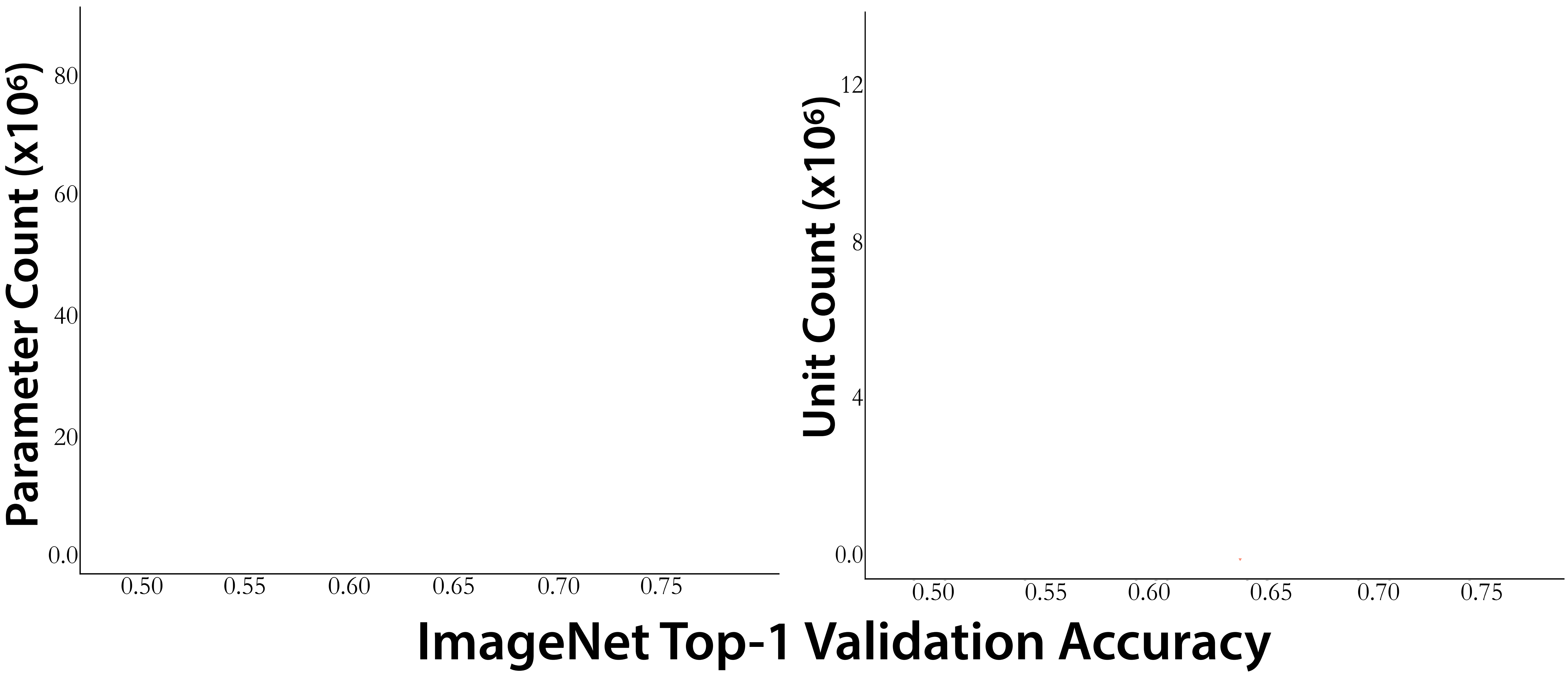
Intermediate Depth ConvRNNs best match OSTs



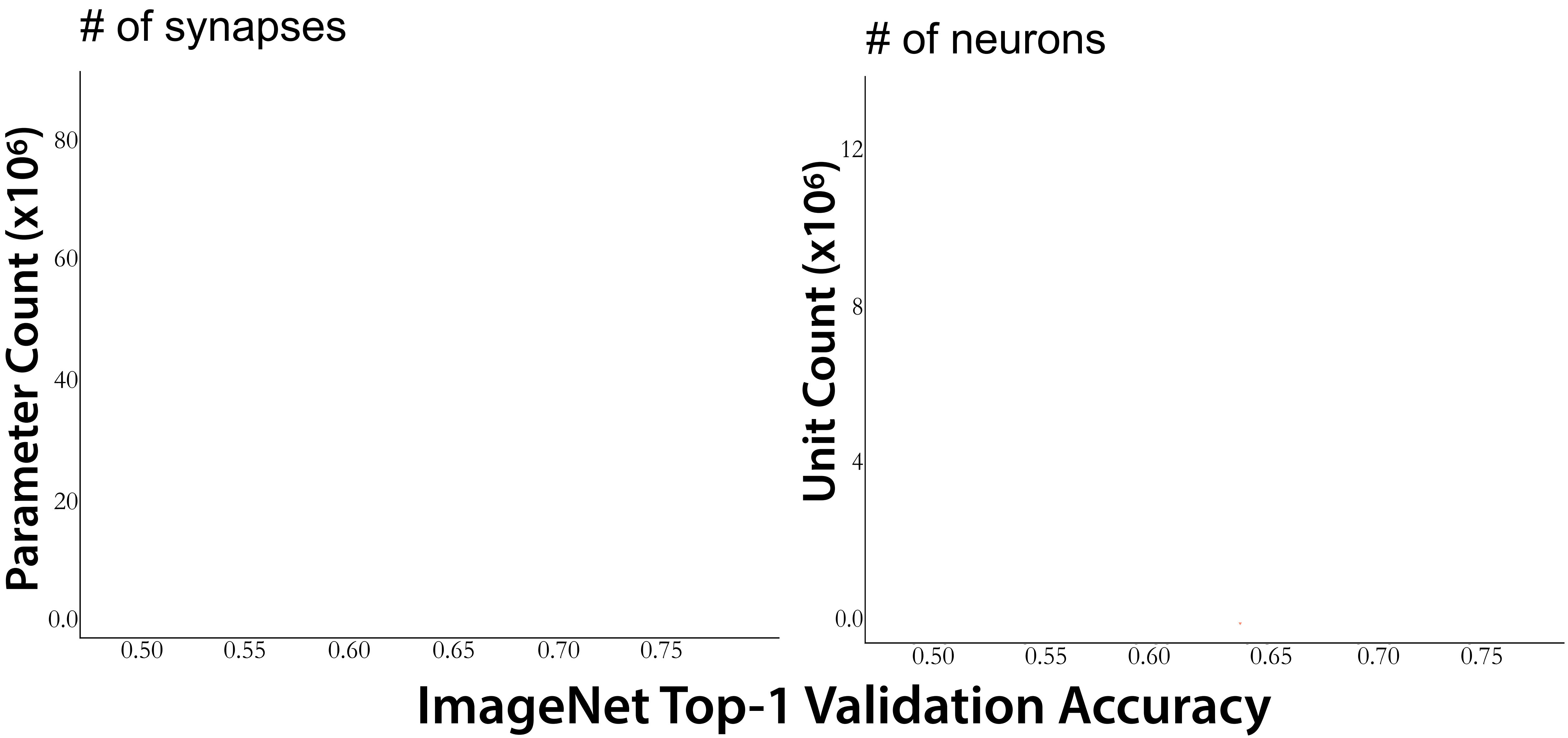


Intermediate Depth ConvRNNs

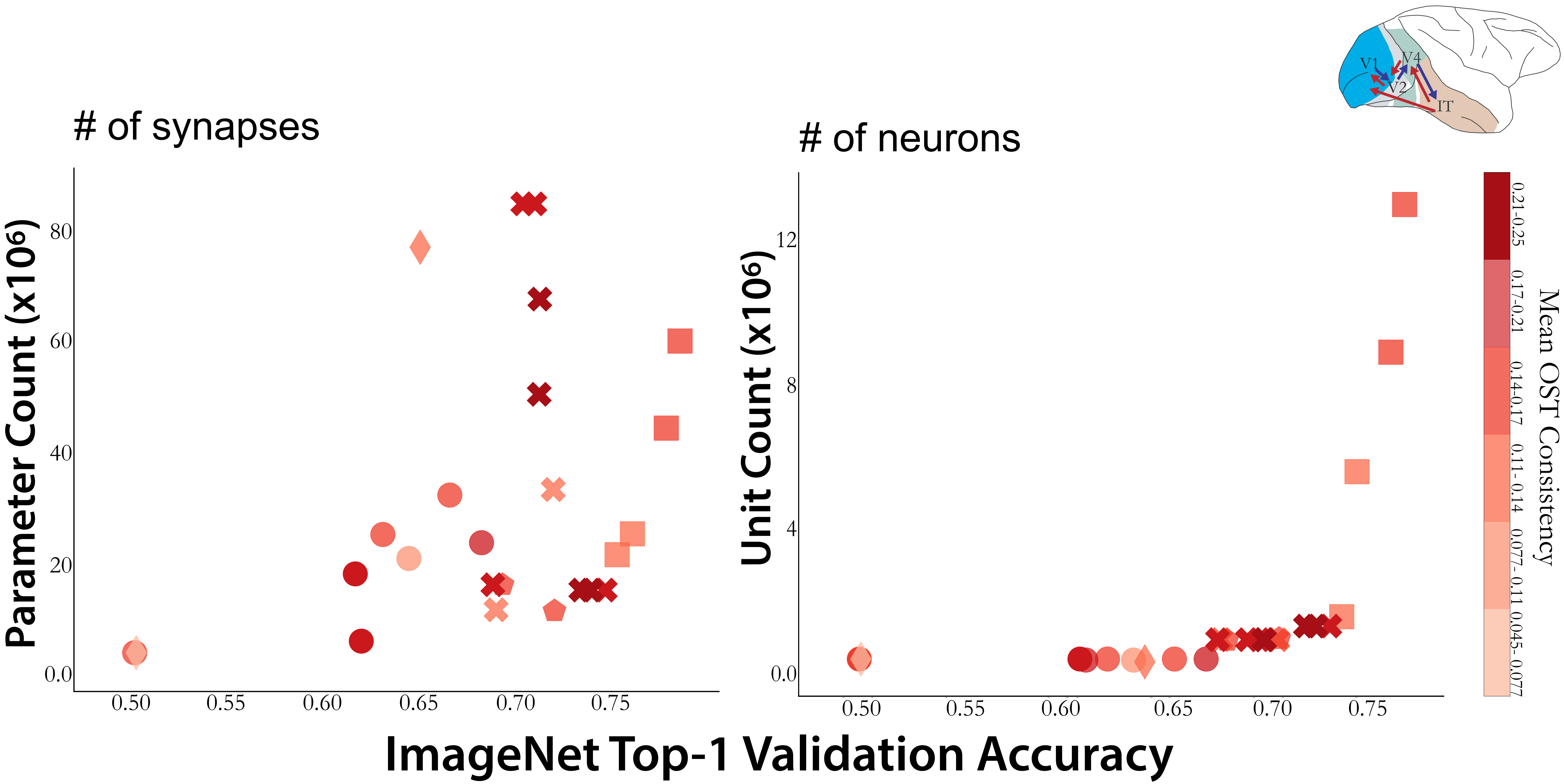
Why Are ConvRNNs the Most Brain-Like?



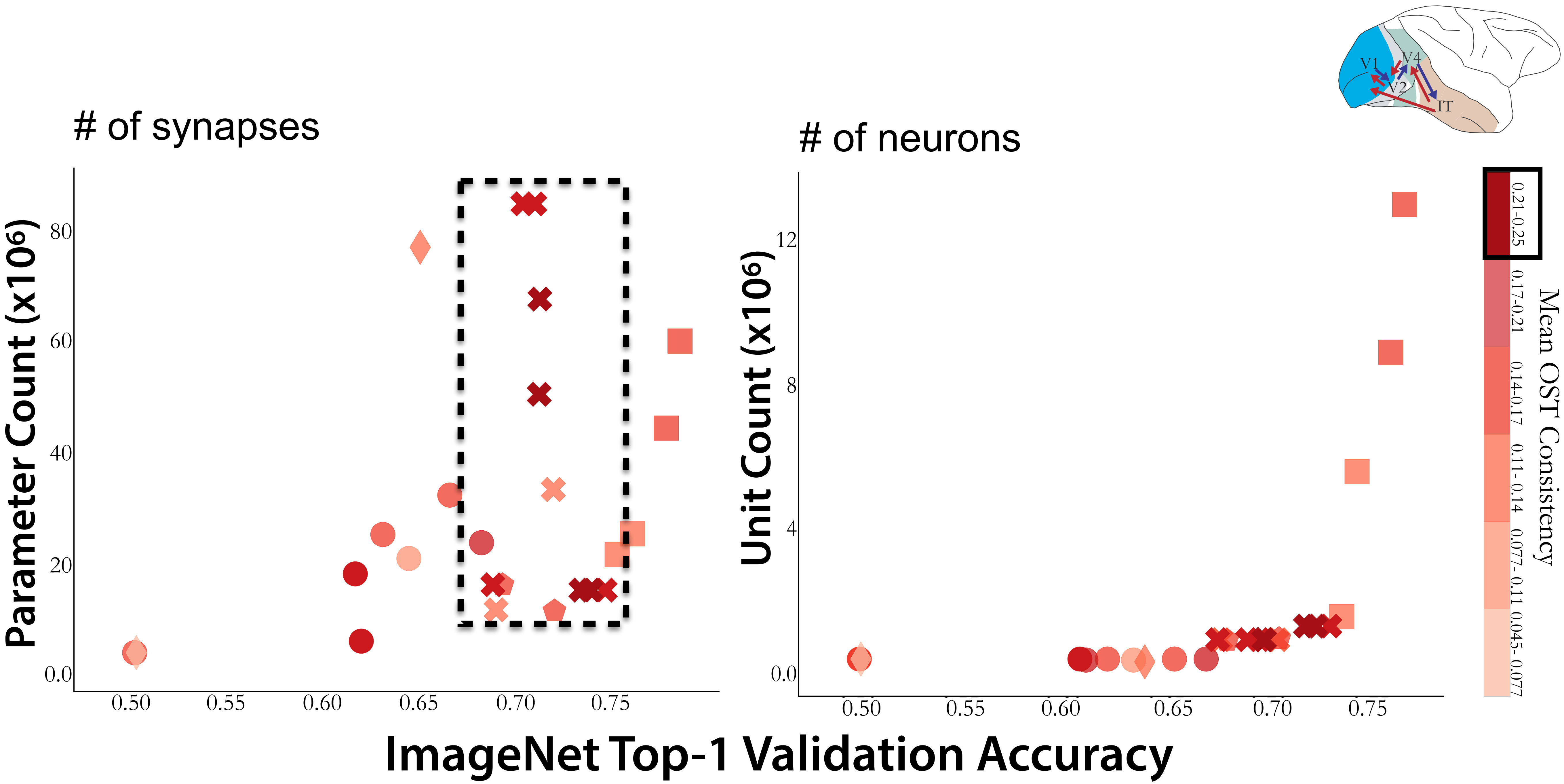
Why Are ConvRNNs the Most Brain-Like?



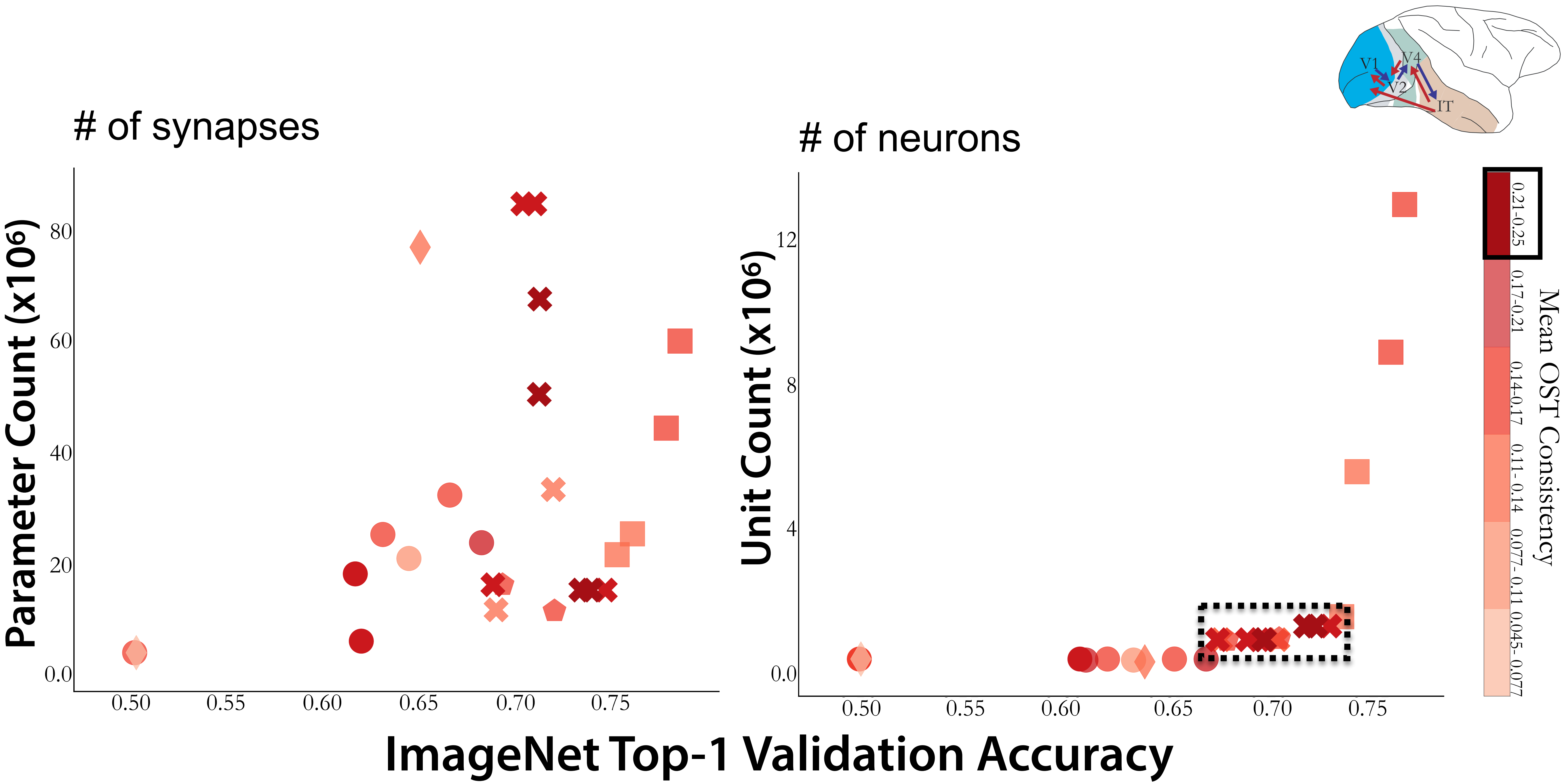
Why Are ConvRNNs the Most Brain-Like?



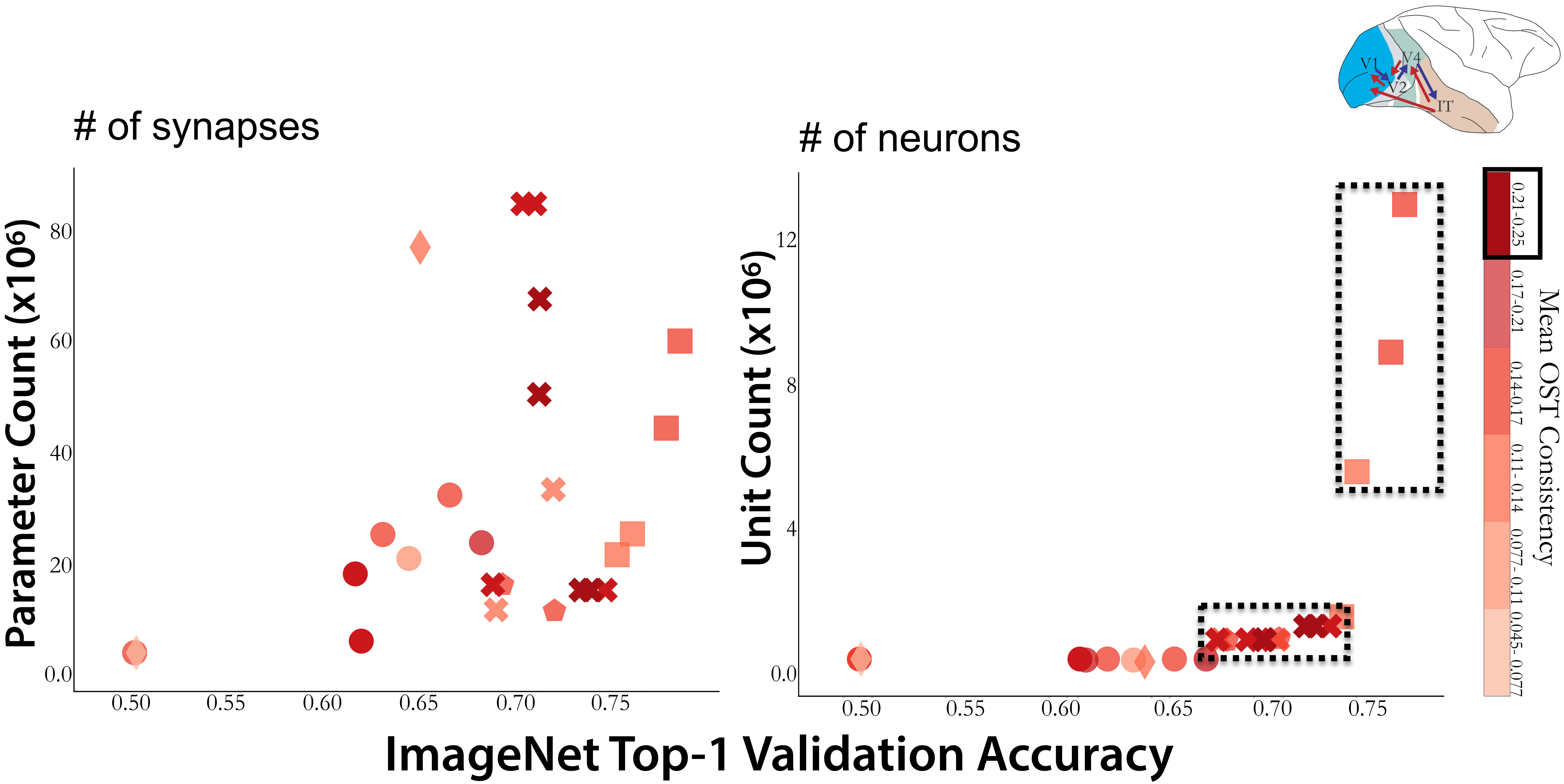
Why Are ConvRNNs the Most Brain-Like?



Smaller Networks that are Still *Performant* are More Brain-Like!



Smaller Networks that are Still *Performant* are More Brain-Like!



Takeaways

L = *learning rule*

"Natural selection + plasticity"

Backpropagation

T = *task loss*

"Ecological niche/ behavior"

Categorization

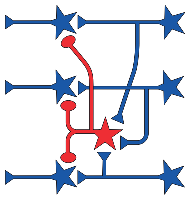
Neurobiological Puzzle:

What is the role of recurrent processing in the primate ventral stream during object recognition?

ImageNet

"Environment"

D = *data stream*



"Circuit"

A = *architecture class*

Takeaways

L *= learning rule*

"Natural selection + plasticity"

Backpropagation

T *= task loss*

"Ecological niche/ behavior"

Categorization

Neurobiological Puzzle:

What is the role of recurrent processing in the primate ventral stream during object recognition?

ImageNet

"Environment"

D *= data stream*

ConvRNNs

~~CNNs~~

"Circuit"

A *= architecture class*

Takeaways

Neurobiological Puzzle:

What is the role of recurrent processing in the primate ventral stream during object recognition?

Findings:

Enables the primate ventral stream to attain high object recognition ability under a physical size constraint,

through temporal rather than spatial complexity,

specifically by conserving on number of neurons rather than synapses.

Outline

- Role of Recurrent Processing During Object Recognition
- Visually-Grounded Mental Simulation
- Vision and Navigation in Rodents

- Future Directions

Visually-Grounded Mental Simulation

L = *learning rule*

"Natural selection + plasticity"

Backpropagation

T = *task loss*

"Ecological niche/ behavior"

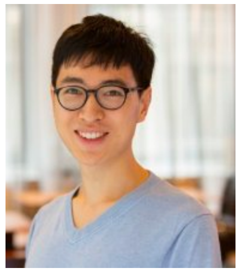
A. Nayebi, R. Rajalingham, M. Jazayeri, G.R. Yang

Neural foundations of mental simulation: future prediction of latent representations on dynamic scenes.

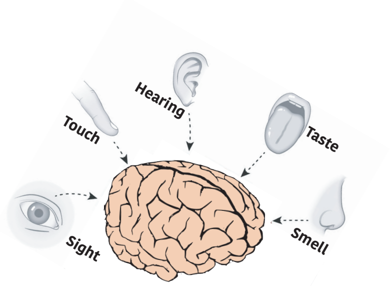
NeurIPS 2023 (spotlight)

Rishi Rajalingham

Mehrdad Jazayeri

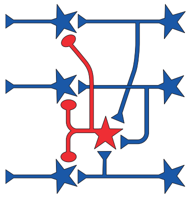


Guangyu Robert Yang



"Environment"

D = *data stream*



"Circuit"

A = *architecture class*

Motivation

Motivation

Infer:
Has this ice block been out longer?

Motivation

Infer:
Has this ice block been out longer?

Motivation

Infer:
Has this ice block been out longer?

Predict:
Will this box support me?

Motivation

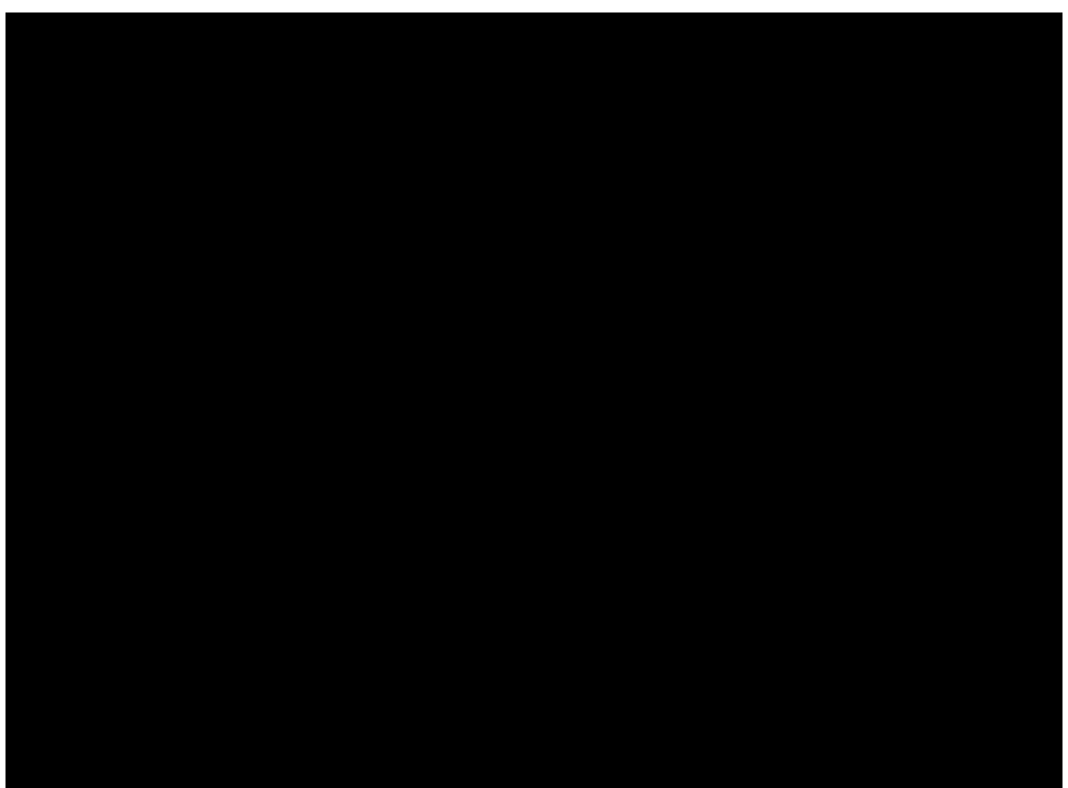
Infer:
Has this ice block been out longer?

Plan:
How would I take these hats off the rack?

Predict:
Will this box support me?

Motivation

<u>Infer:</u>
Has this ice block been out longer?

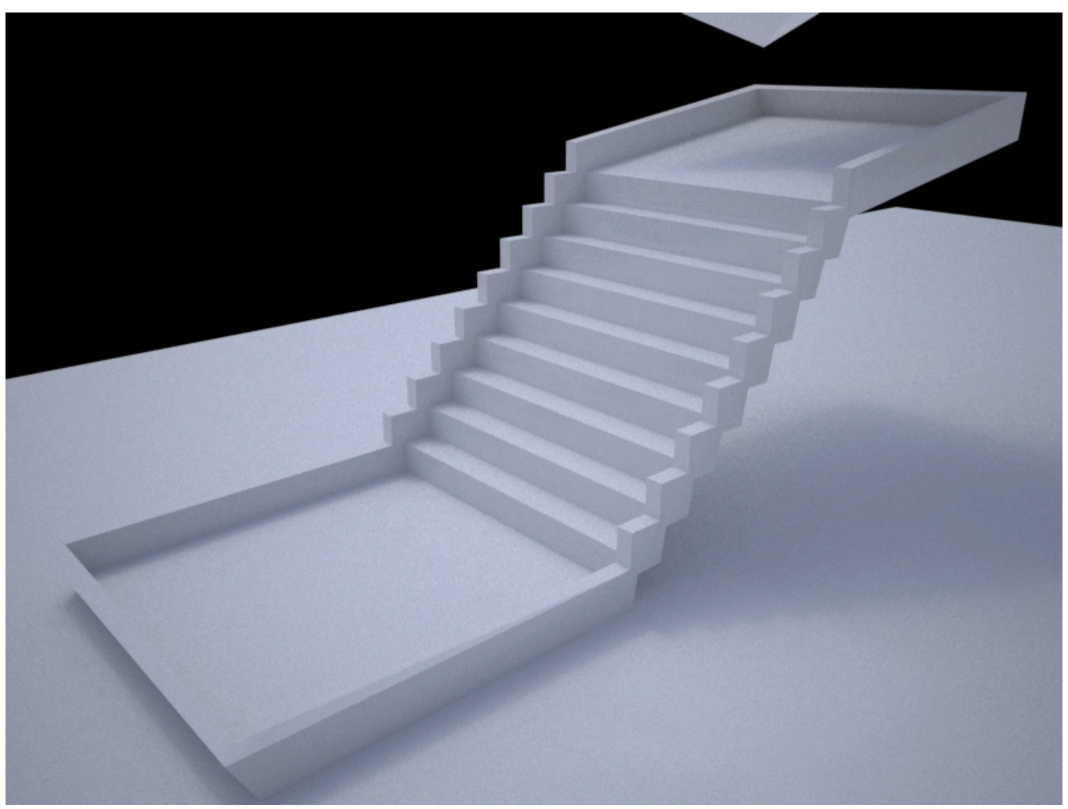


<u>Plan:</u>
How would I take these hats off the rack?

<u>Predict:</u>
Will this box support me?

Motivation

Infer:
Has this ice block been out longer?



Plan:
How would I take these hats off the rack?

Predict:
Will this box support me?

The Mental Simulation Hypothesis

The Nature of Explanation

My hypothesis then is that thought models, or parallels, reality – that its essential feature is not 'the mind', 'the self', 'sense-data', nor propositions but symbolism, and that this symbolism is largely of the same kind as that which is familiar to us in mechanical devices which aid thought and calculation. . .

If the organism carries a 'small-scale model' of external reality and of its own possible actions within its head, it is able to try out various alternatives, conclude which is the best of them, react to future situations before they arise, utilize the knowledge of past events in dealing with the present and future, and in every way to react in a much fuller, safer, and more competent manner to the emergencies which face it.

Craik (1943): The brain builds **mental models** of the external physical world, that support physical inferences via **mental simulations.**

Kenneth Craik

The Mental Simulation Hypothesis

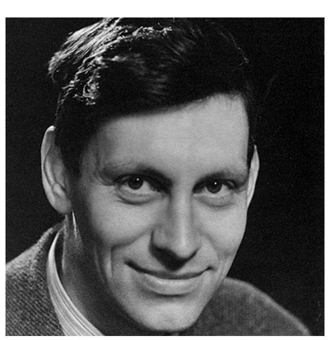
The Nature of Explanation

Pre-dates the modern computer!

My hypothesis then is that thought models, or parallels, reality – that its essential feature is not 'the mind', 'the self', 'sense-data', nor propositions but symbolism, and that this symbolism is largely of the same kind as that which is familiar to us in mechanical devices which aid thought and calculation. . .

If the organism carries a 'small-scale model' of external reality and of its own possible actions within its head, it is able to try out various alternatives, conclude which is the best of them, react to future situations before they arise, utilize the knowledge of past events in dealing with the present and future, and in every way to react in a much fuller, safer, and more competent manner to the emergencies which face it.

Craik (1943): The brain builds **mental models** of the external physical world, that support physical inferences via **mental simulations.**



Kenneth Craik

The Mental Simulation Hypothesis: Behavioral Evidence

The Nature of Explanation

My hypothesis then is that thought models, or parallels, reality – that its essential feature is not 'the mind', 'the self', 'sense-data', nor propositions but symbolism, and that this symbolism is largely of the same kind as that which is familiar to us in mechanical devices which aid thought and calculation. . .

If the organism carries a 'small-scale model' of external reality and of its own possible actions within its head, it is able to try out various alternatives, conclude which is the best of them, react to future situations before they arise, utilize the knowledge of past events in dealing with the present and future, and in every way to react in a much fuller, safer, and more competent manner to the emergencies which face it.

Craik (1943): The brain builds **mental models** of the external physical world, that support physical inferences via **mental simulations.**

Focus on *physical* simulation

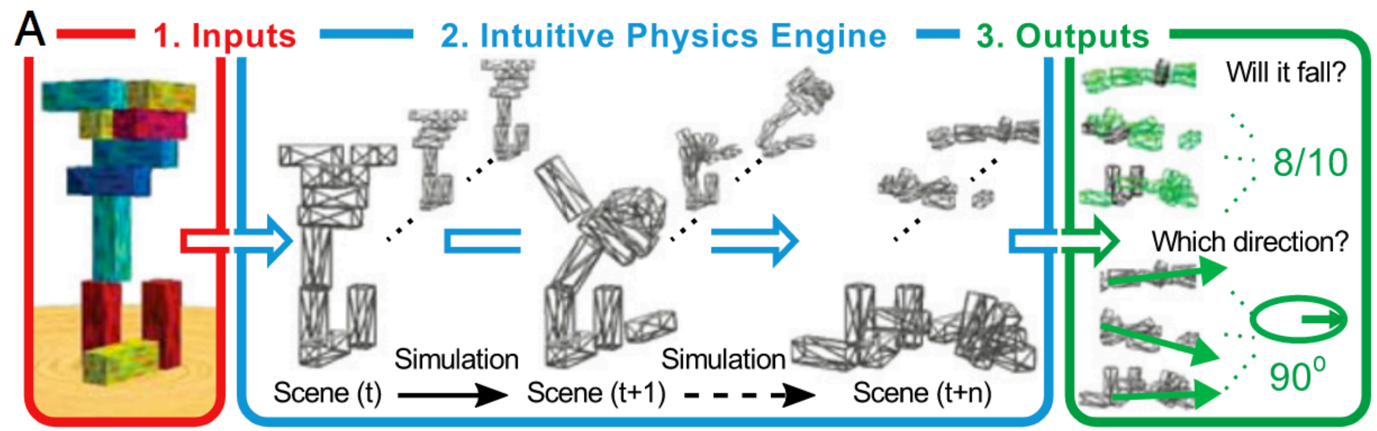
The Mental Simulation Hypothesis: Behavioral Evidence

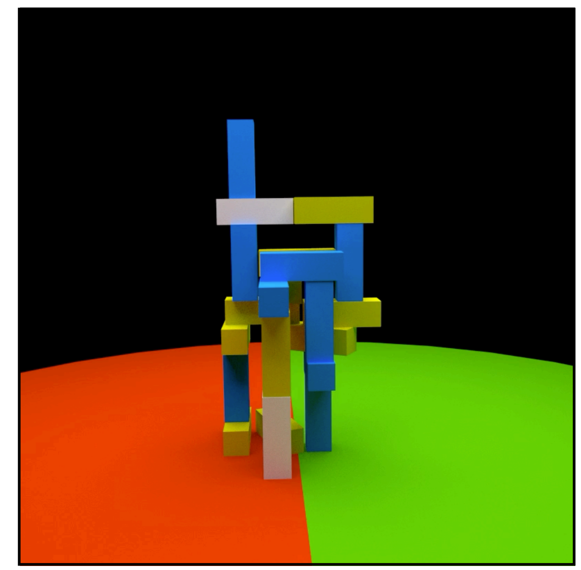
The Nature of Explanation

My hypothesis then is that thought models, or parallels, reality – that its essential feature is not 'the mind', 'the self', 'sense-data', nor propositions but symbolism, and that this symbolism is largely of the same kind as that which is familiar to us in mechanical devices which aid thought and calculation. . .

If the organism carries a 'small-scale model' of external reality and of its own possible actions within its head, it is able to try out various alternatives, conclude which is the best of them, react to future situations before they arise, utilize the knowledge of past events in dealing with the present and future, and in every way to react in a much fuller, safer, and more competent manner to the emergencies which face it.

Craik (1943): The brain builds **mental models** of the external physical world, that support physical inferences via **mental simulations.**





Ullman et al. 2017 *Battaglia, Hamrick, Tenenbaum 2013*

Intuitive Physics Engine (IPE) can match human physical judgements

Peter Battaglia

Tomer Ullman

Jessica Hamrick

Joshua Tenenbaum

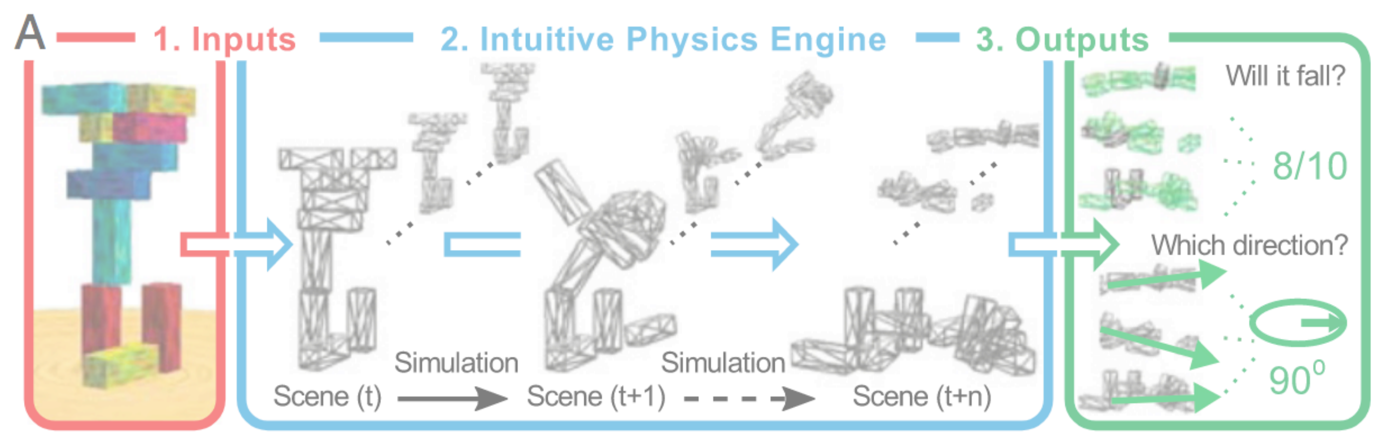
The Mental Simulation Hypothesis: Human Neuroimaging Evidence

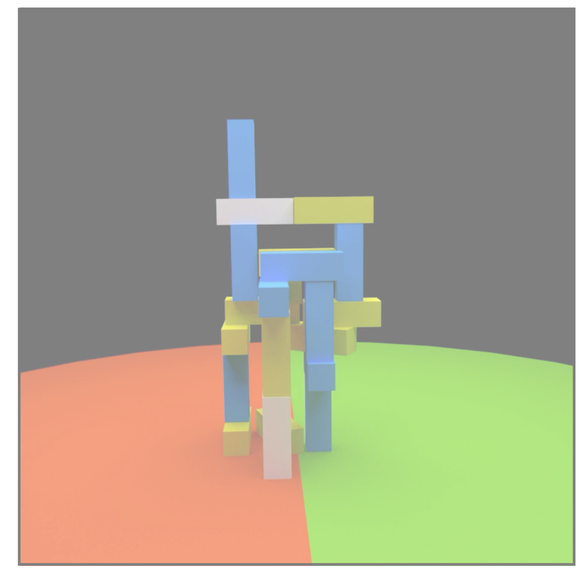
The Nature of Explanation

My hypothesis then is that thought models, or parallels, reality – that its essential feature is not 'the mind', 'the self', 'sense-data', nor propositions but symbolism, and that this symbolism is largely of the same kind as that which is familiar to us in mechanical devices which aid thought and calculation. . .

If the organism carries a 'small-scale model' of external reality and of its own possible actions within its head, it is able to try out various alternatives, conclude which is the best of them, react to future situations before they arise, utilize the knowledge of past events in dealing with the present and future, and in every way to react in a much fuller, safer, and more competent manner to the emergencies which face it.

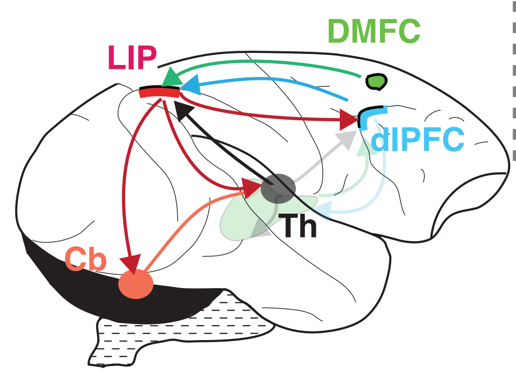
Craik (1943): The brain builds **mental models** of the external physical world, that support physical inferences via **mental simulations.**



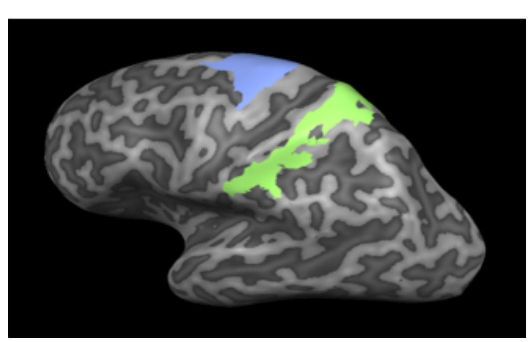


Ullman et al. 2017 *Battaglia, Hamrick, Tenenbaum 2013*

The Brain's "Physics Engine"



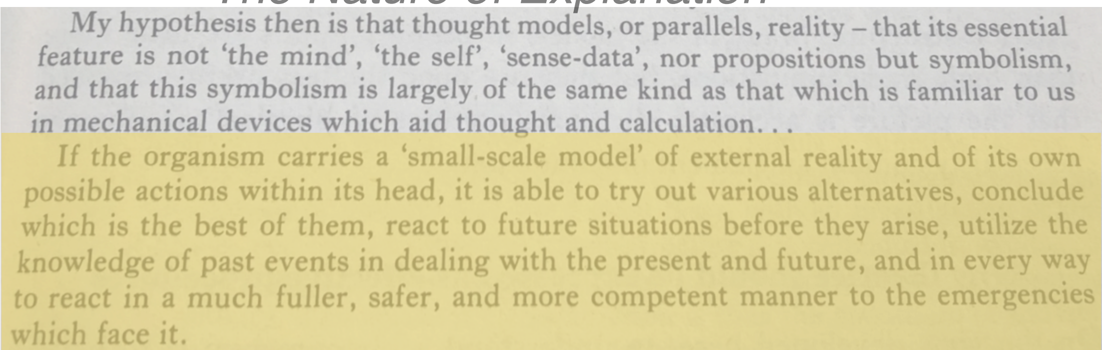
Fronto-Parietal Network



Nancy Kanwisher

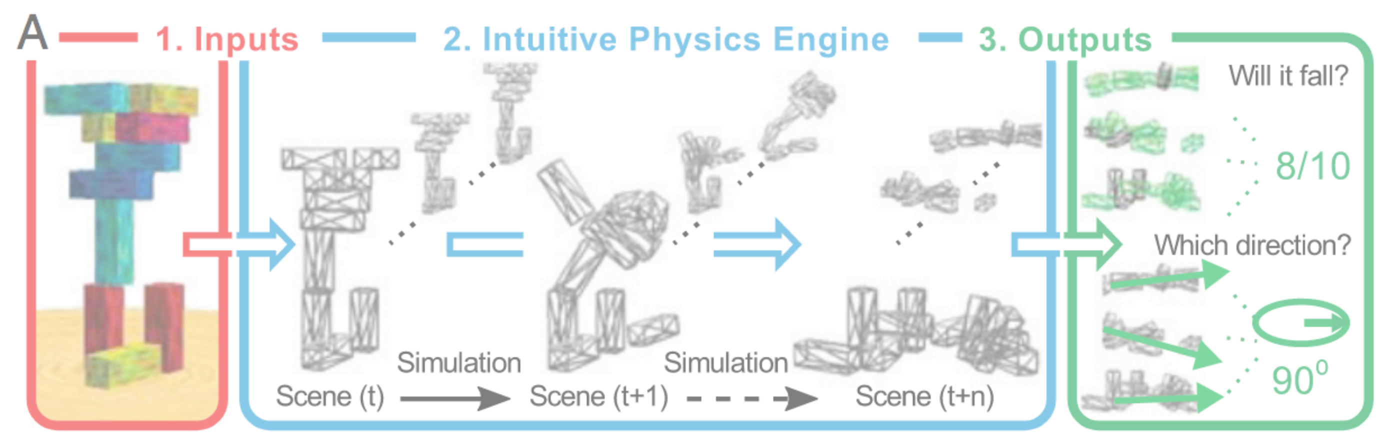
The Mental Simulation Hypothesis: Human Neuroimaging Evidence

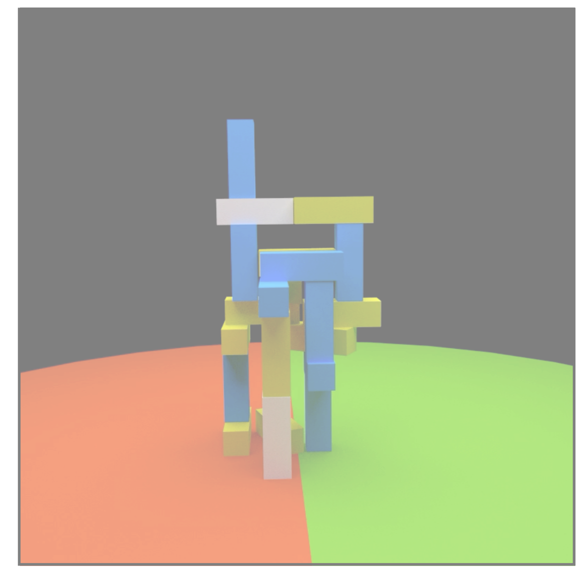
The Nature of Explanation


My hypothesis then is that thought models, or parallels, reality – that its essential feature is not 'the mind', 'the self', 'sense-data', nor propositions but symbolism, and that this symbolism is largely of the same kind as that which is familiar to us in mechanical devices which aid thought and calculation. . .

If the organism carries a 'small-scale model' of external reality and of its own possible actions within its head, it is able to try out various alternatives, conclude which is the best of them, react to future situations before they arise, utilize the knowledge of past events in dealing with the present and future, and in every way to react in a much fuller, safer, and more competent manner to the emergencies which face it.

<u>Craik (1943):</u> The brain builds **mental models** of the external physical world, that support physical inferences via **mental simulations.**

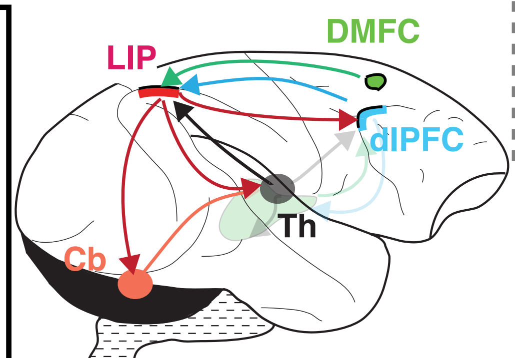




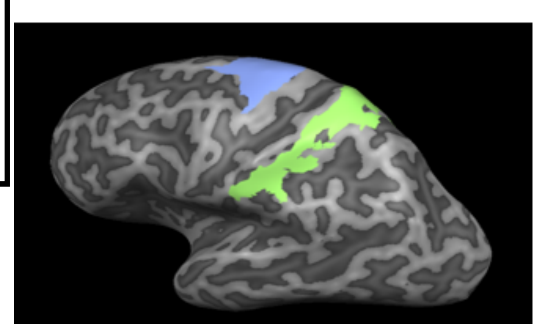
Ullman et al. 2017 *Battaglia, Hamrick, Tenenbaum 2013*

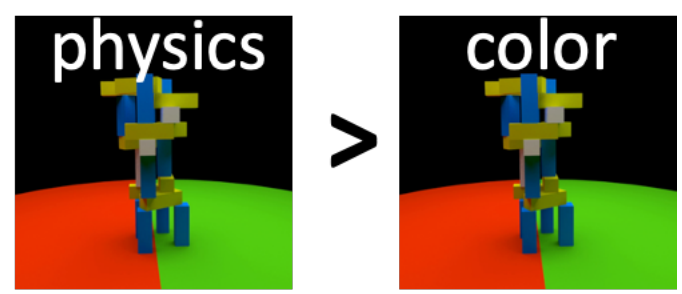
The Brain's "Physics Engine"

- A network of brain regions recruited by physical inferences (*Fischer et al. 2016*)



Fronto-Parietal Network





Fischer et al. 2016

Jason Fischer

Nancy Kanwisher

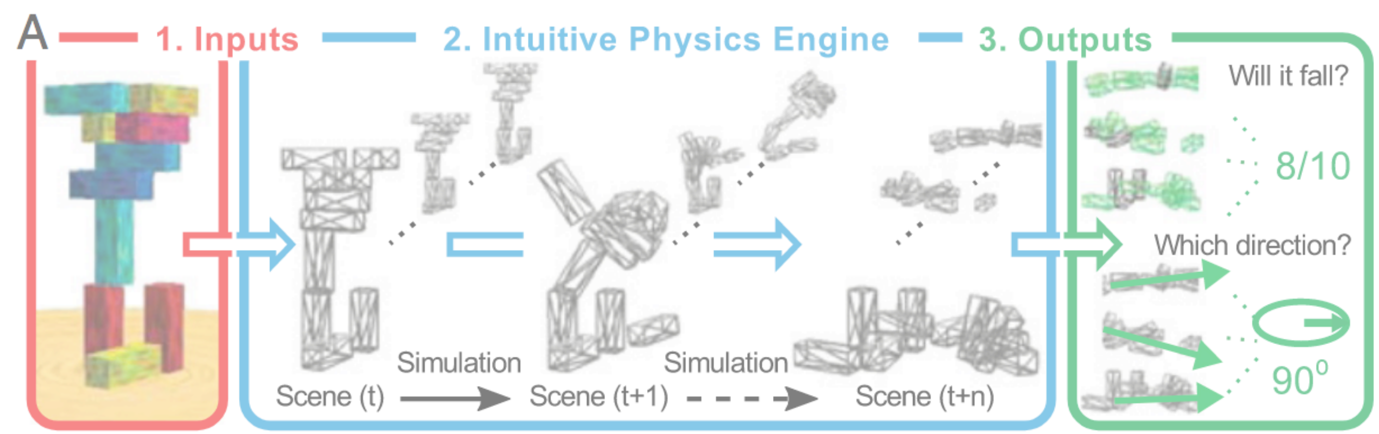
The Mental Simulation Hypothesis: Human Neuroimaging Evidence

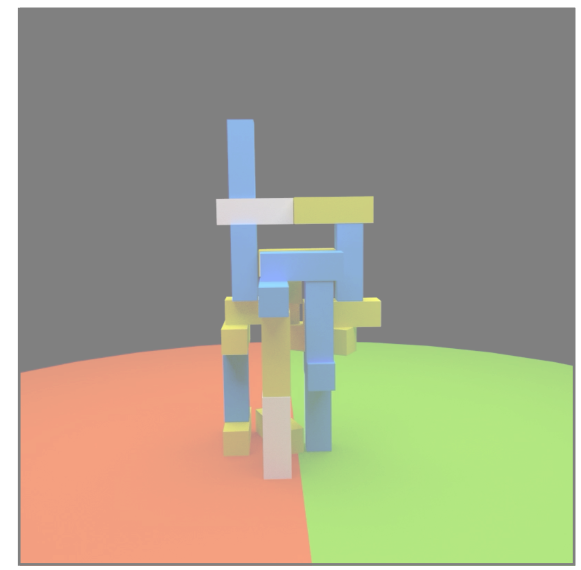
The Nature of Explanation

My hypothesis then is that thought models, or parallels, reality – that its essential feature is not 'the mind', 'the self', 'sense-data', nor propositions but symbolism, and that this symbolism is largely of the same kind as that which is familiar to us in mechanical devices which aid thought and calculation. . .

If the organism carries a 'small-scale model' of external reality and of its own possible actions within its head, it is able to try out various alternatives, conclude which is the best of them, react to future situations before they arise, utilize the knowledge of past events in dealing with the present and future, and in every way to react in a much fuller, safer, and more competent manner to the emergencies which face it.

Craik (1943): The brain builds **mental models** of the external physical world, that support physical inferences via **mental simulations.**

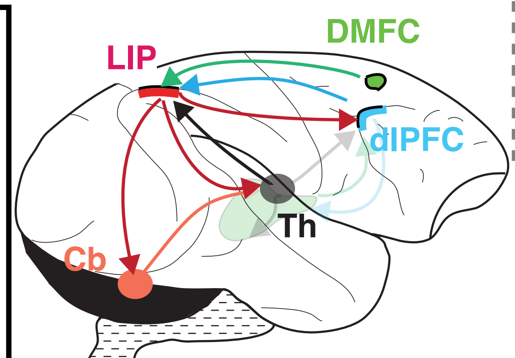




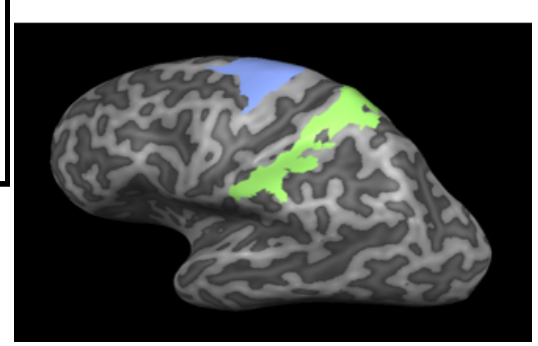
Ullman et al. 2017 *Battaglia, Hamrick, Tenenbaum 2013*

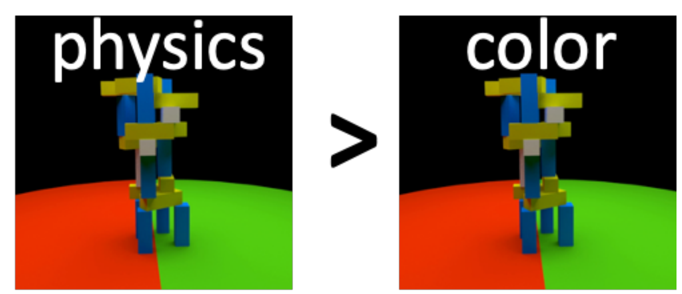
The Brain's "Physics Engine"

- A network of brain regions recruited by physical inferences (*Fischer et al. 2016*)
- Contains information about mass (*Schwettmann et al. 2019*)

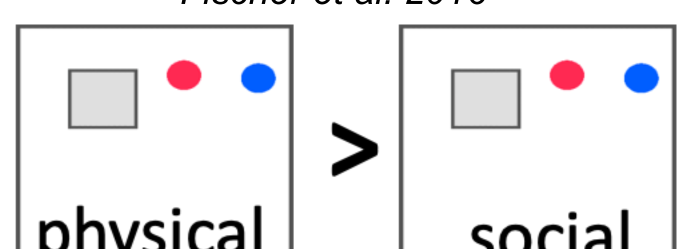


Fronto-Parietal Network





Fischer et al. 2016



Schwettmann et al. 2019

Sarah Schwettmann

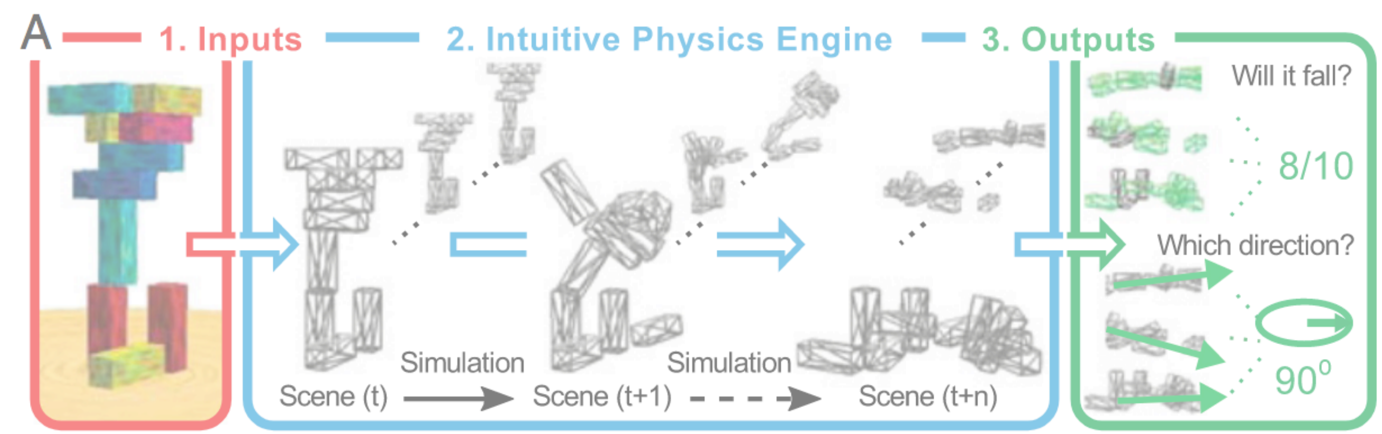
Nancy Kanwisher

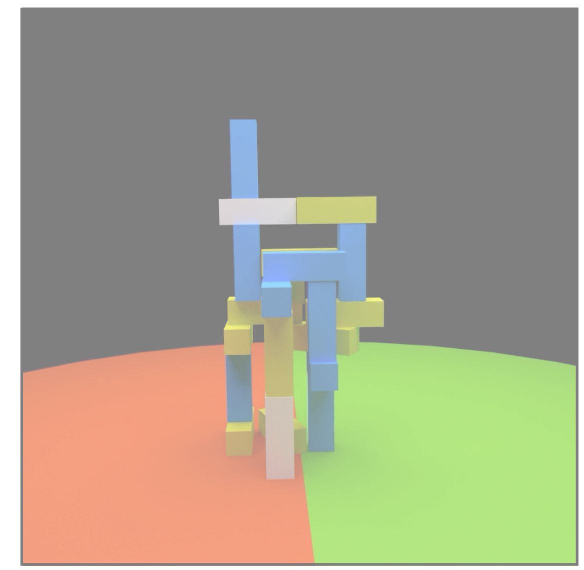
The Mental Simulation Hypothesis: Human Neuroimaging Evidence

The Nature of Explanation

> My hypothesis then is that thought models, or parallels, reality – that its essential feature is not 'the mind', 'the self', 'sense-data', nor propositions but symbolism, and that this symbolism is largely of the same kind as that which is familiar to us in mechanical devices which aid thought and calculation. . .
>
> If the organism carries a 'small-scale model' of external reality and of its own possible actions within its head, it is able to try out various alternatives, conclude which is the best of them, react to future situations before they arise, utilize the knowledge of past events in dealing with the present and future, and in every way to react in a much fuller, safer, and more competent manner to the emergencies which face it.

Craik (1943): The brain builds **mental models** of the external physical world, that support physical inferences via **mental simulations.**

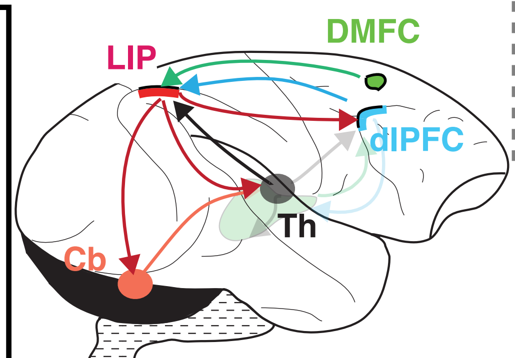




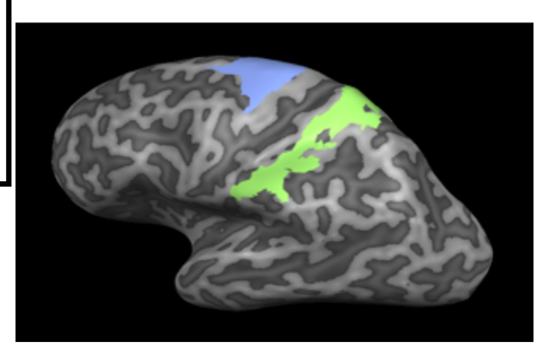
Ullman et al. 2017 Battaglia, Hamrick, Tenenbaum 2013

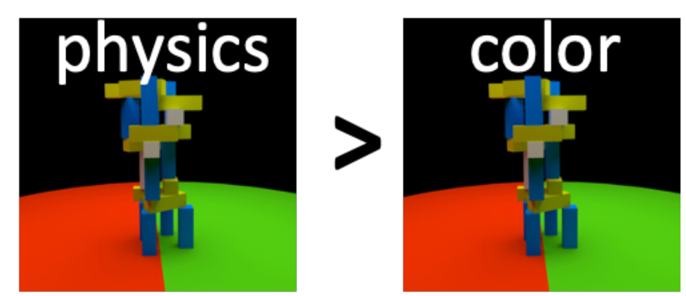
The Brain's "Physics Engine"

- A network of brain regions recruited by physical inferences (*Fischer et al. 2016*)
- Contains information about mass (*Schwettmann et al. 2019*)
- Contains information about physical stability (*Pramod et al. 2022*)

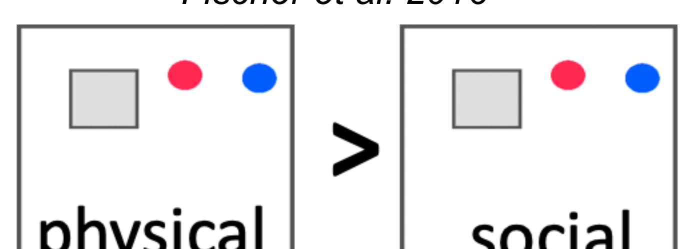


Fronto-Parietal Network





Fischer et al. 2016



Schwettmann et al. 2019

Pramod et al. 2022

RT Pramod Nancy Kanwisher

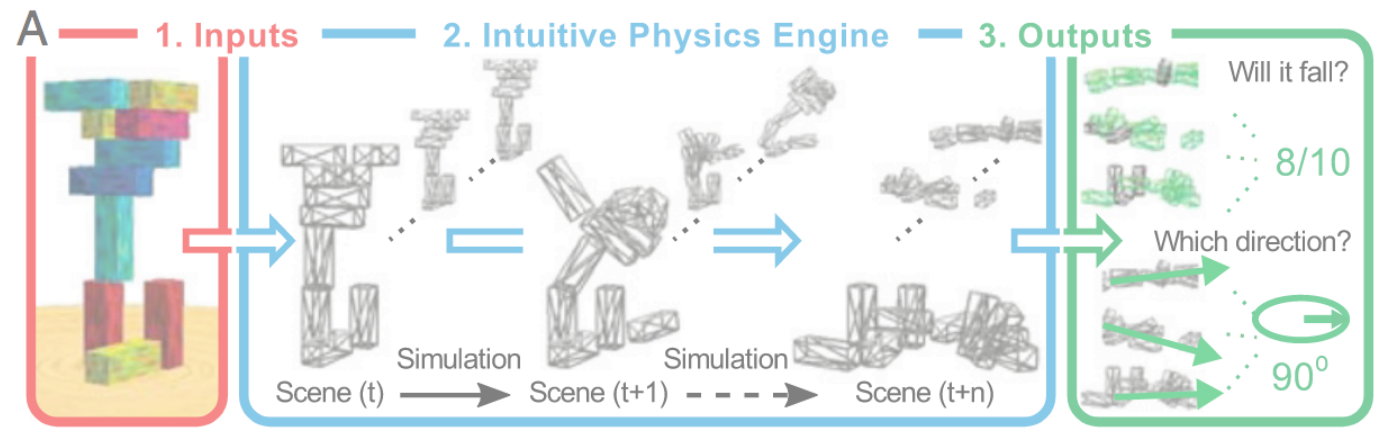
The Mental Simulation Hypothesis: Primate Electrophysiological Evidence

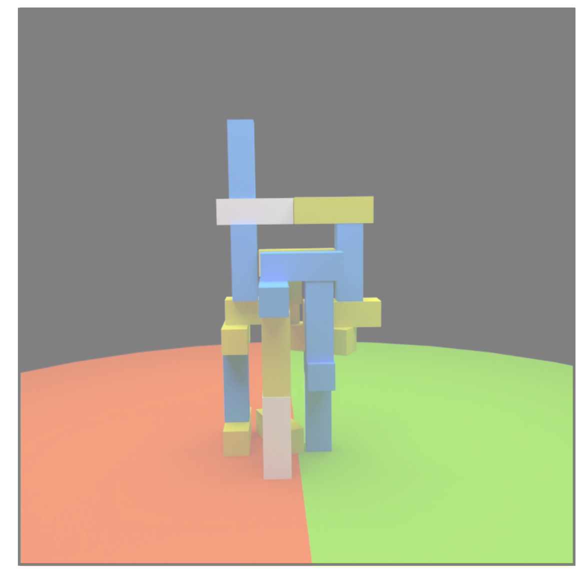
The Nature of Explanation

My hypothesis then is that thought models, or parallels, reality – that its essential feature is not 'the mind', 'the self', 'sense-data', nor propositions but symbolism, and that this symbolism is largely of the same kind as that which is familiar to us in mechanical devices which aid thought and calculation. . .

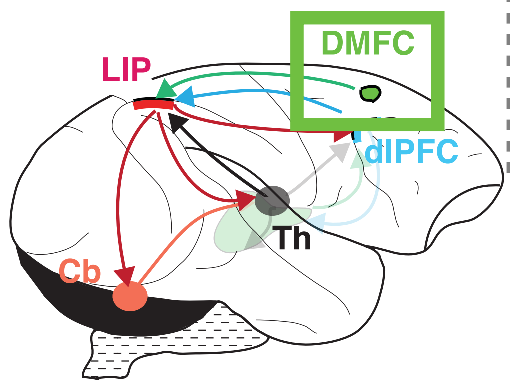
If the organism carries a 'small-scale model' of external reality and of its own possible actions within its head, it is able to try out various alternatives, conclude which is the best of them, react to future situations before they arise, utilize the knowledge of past events in dealing with the present and future, and in every way to react in a much fuller, safer, and more competent manner to the emergencies which face it.

Craik (1943): The brain builds **mental models** of the external physical world, that support physical inferences via **mental simulations.**

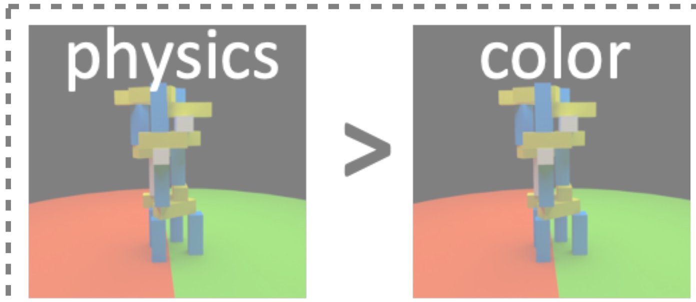




Ullman et al. 2017 Battaglia, Hamrick, Tenenbaum 2013

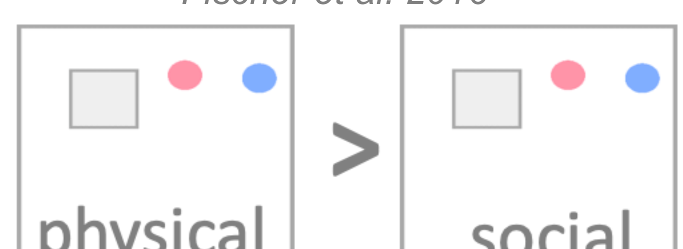


Fronto-Parietal Network

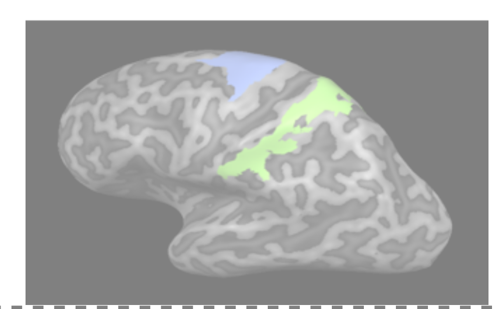


Fischer et al. 2016

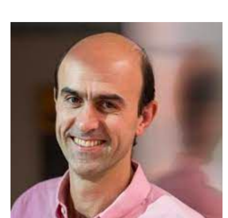
Pramod et al. 2022



Schwettmann et al. 2019

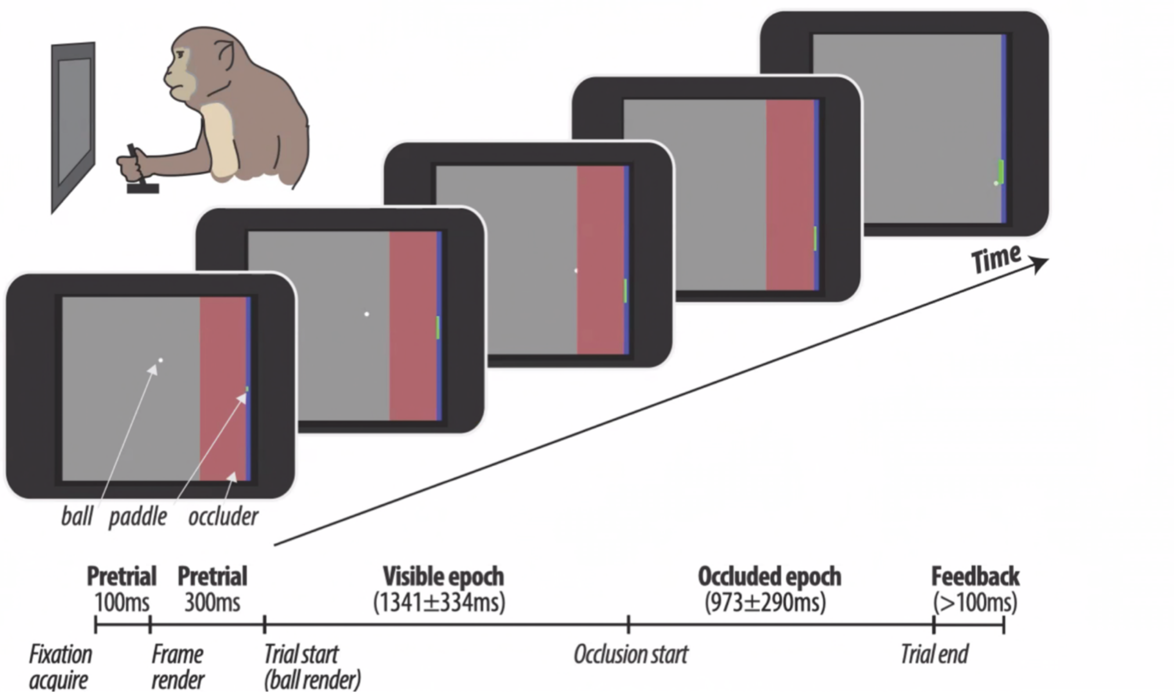


Rishi Rajalingham



Mehrdad Jazayeri

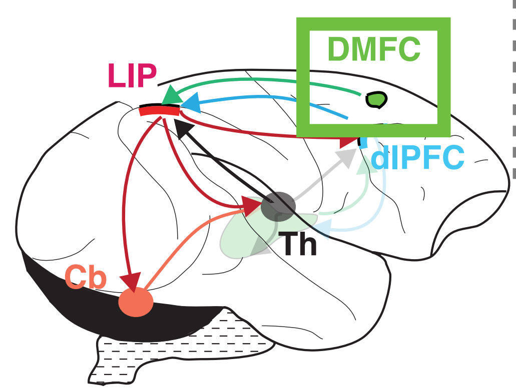
The Mental Simulation Hypothesis: Primate Electrophysiological Evidence



The role of mental simulation in primate physical inference abilities

Rishi Rajalingham, Aida Piccato, Mehrdad Jazayeri

doi: https://doi.org/10.1101/2021.01.14.426741



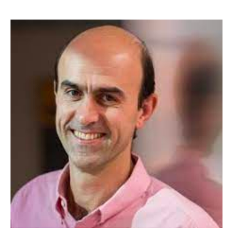
Fronto-Parietal Network

Dynamic tracking of objects in the macaque dorsomedial frontal cortex

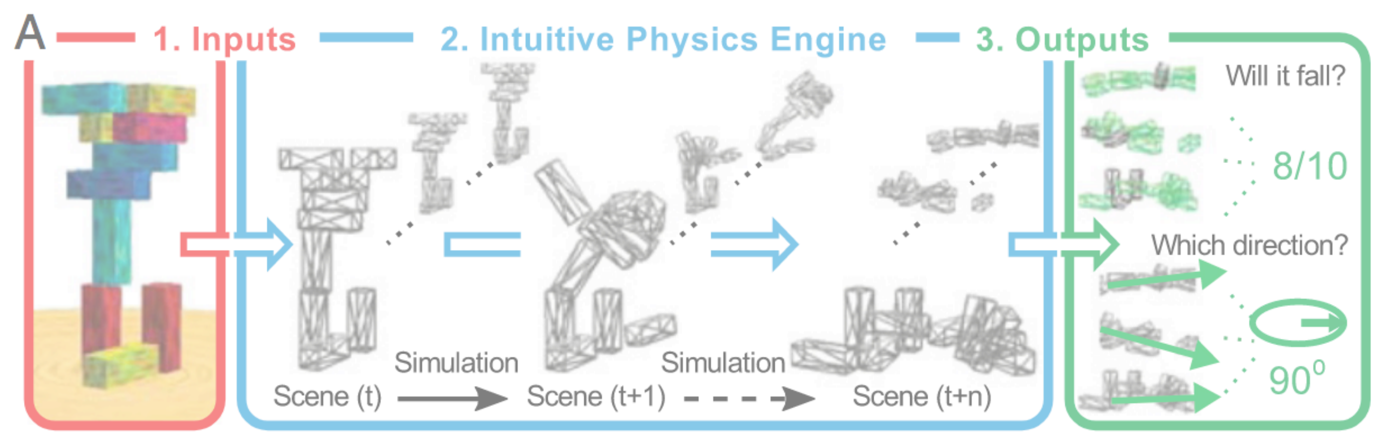
Rishi Rajalingham, Hansem Sohn, Mehrdad Jazayeri

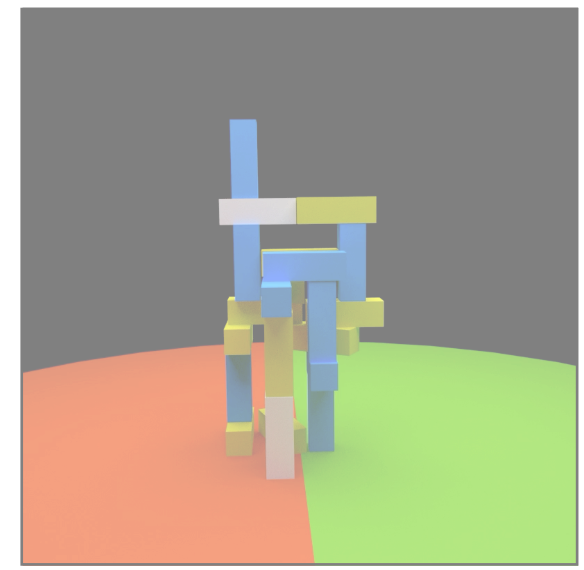
doi: https://doi.org/10.1101/2022.06.24.497529

Rishi Rajalingham



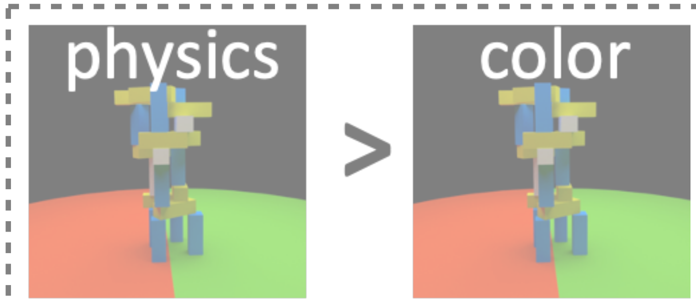
Mehrdad Jazayeri





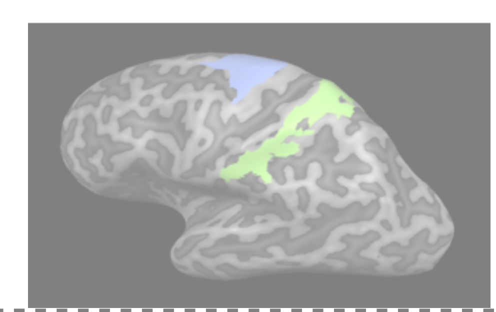
Ullman et al. 2017

Battaglia, Hamrick, Tenenbaum 2013



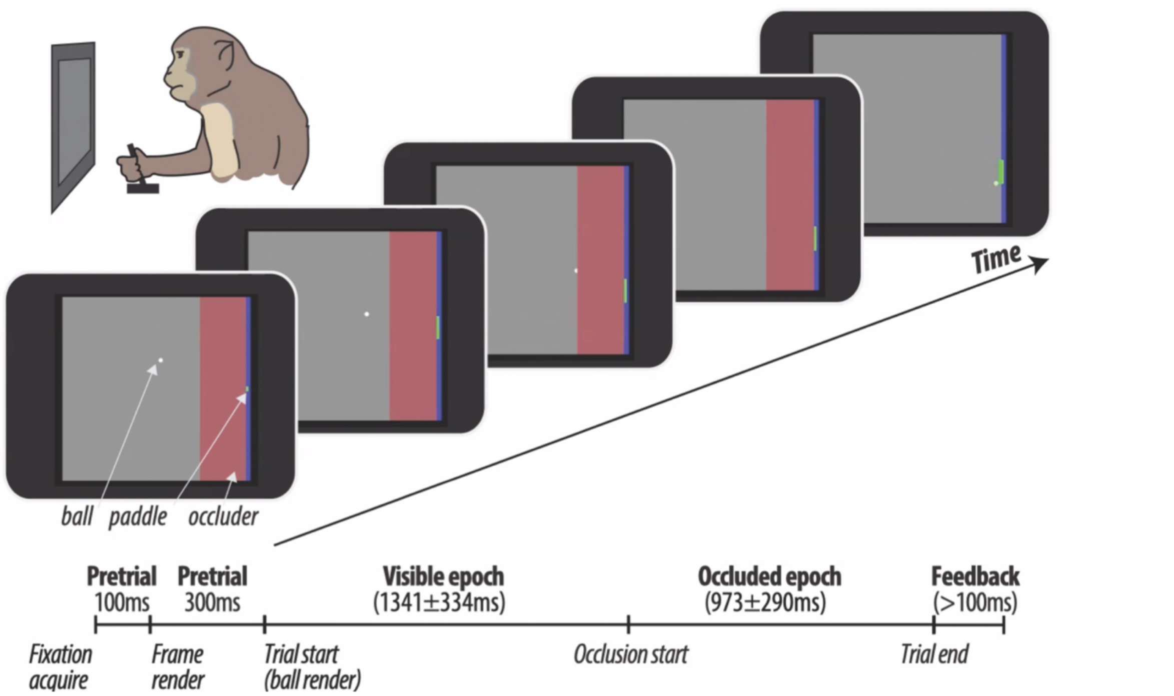
Fischer et al. 2016

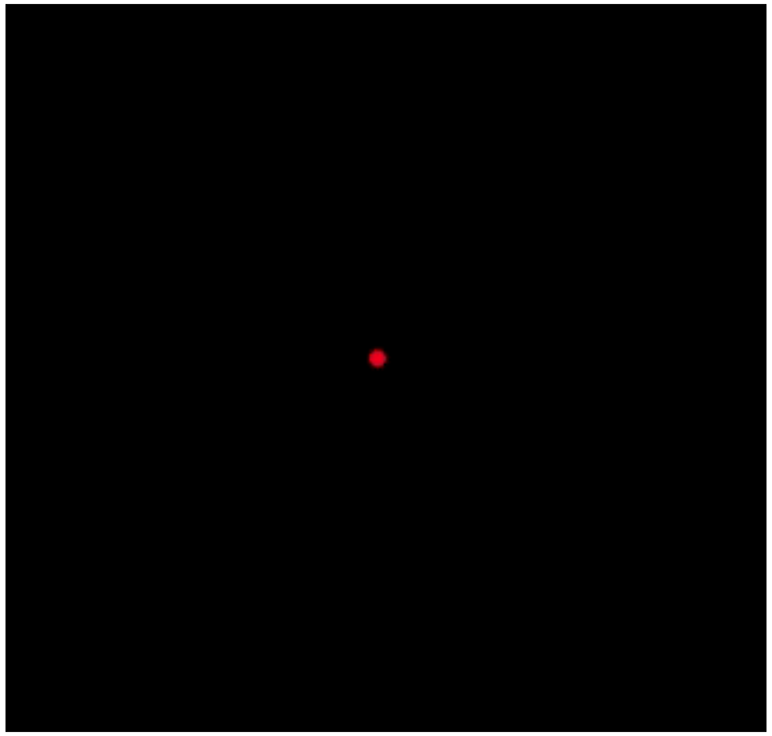
Pramod et al. 2022

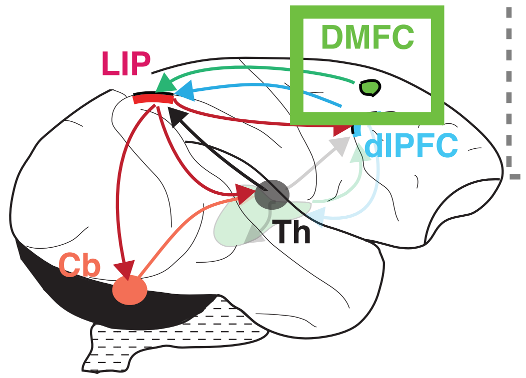


Schwettmann et al. 2019

The Mental Simulation Hypothesis: Primate Electrophysiological Evidence



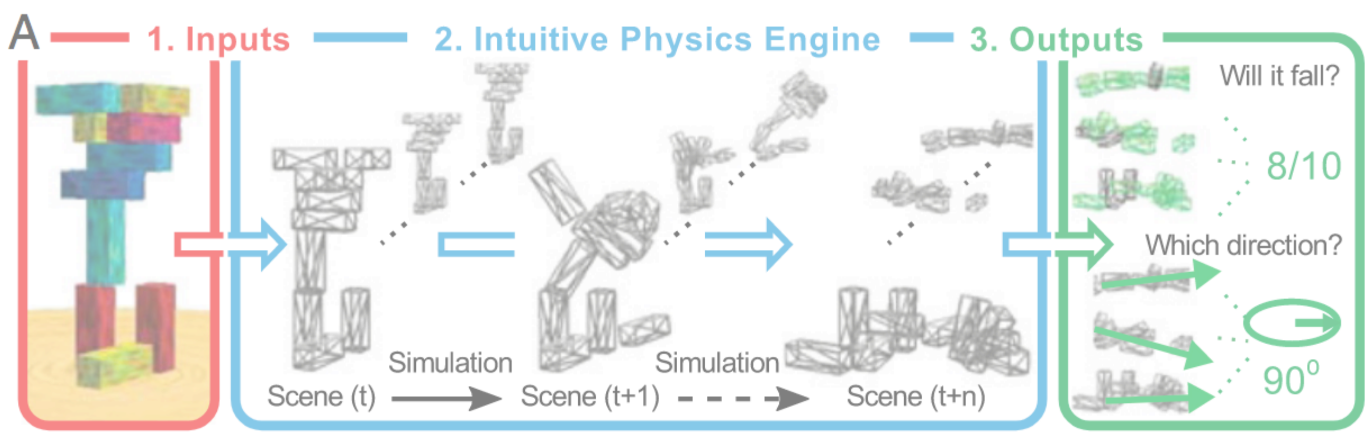


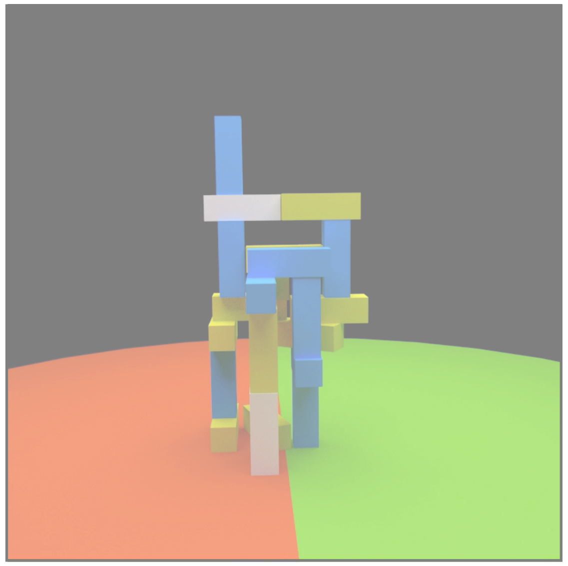


Fronto-Parietal Network

Rishi Rajalingham

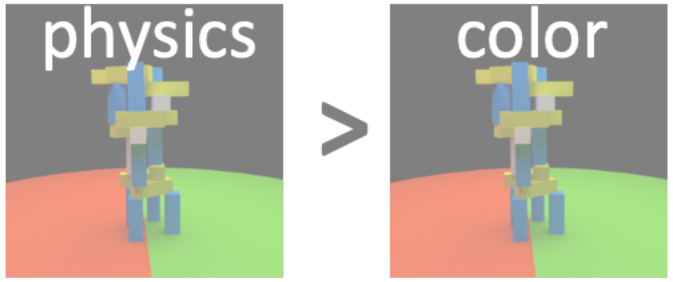
Mehrdad Jazayeri





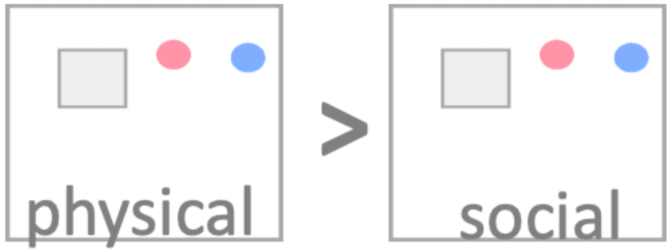
Ullman et al. 2017

Battaglia, Hamrick, Tenenbaum 2013

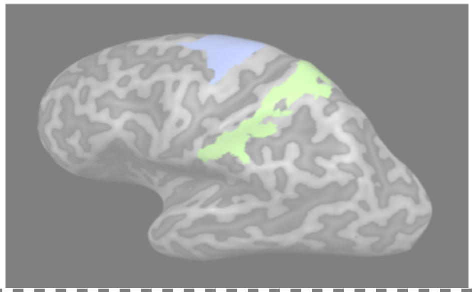


Fischer et al. 2016

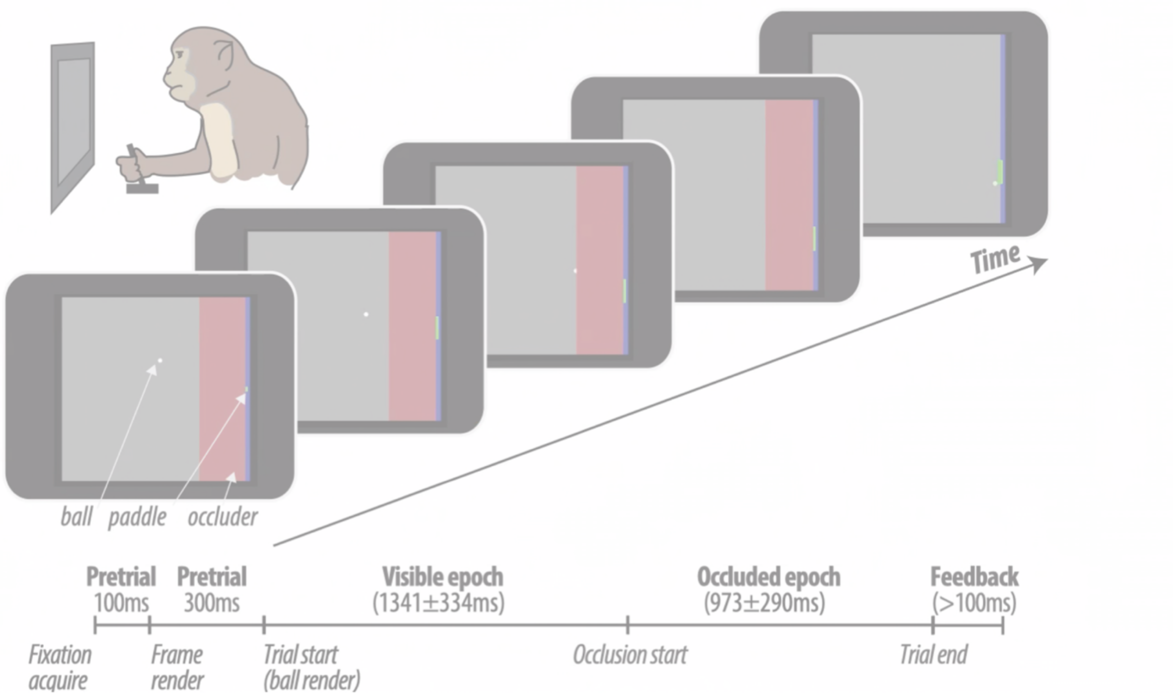
Pramod et al. 2022

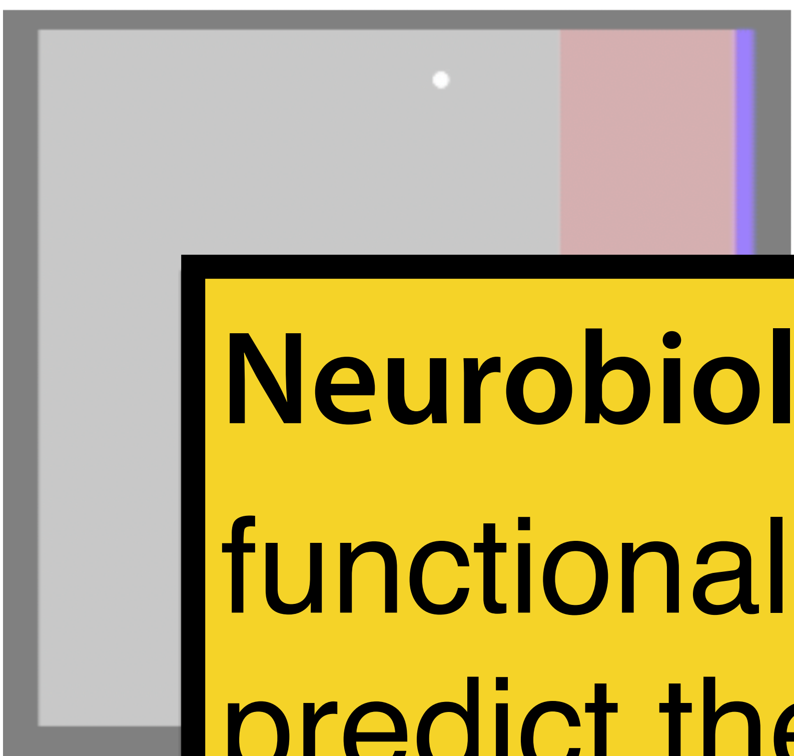


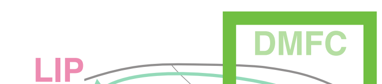
Schwettmann et al. 2019

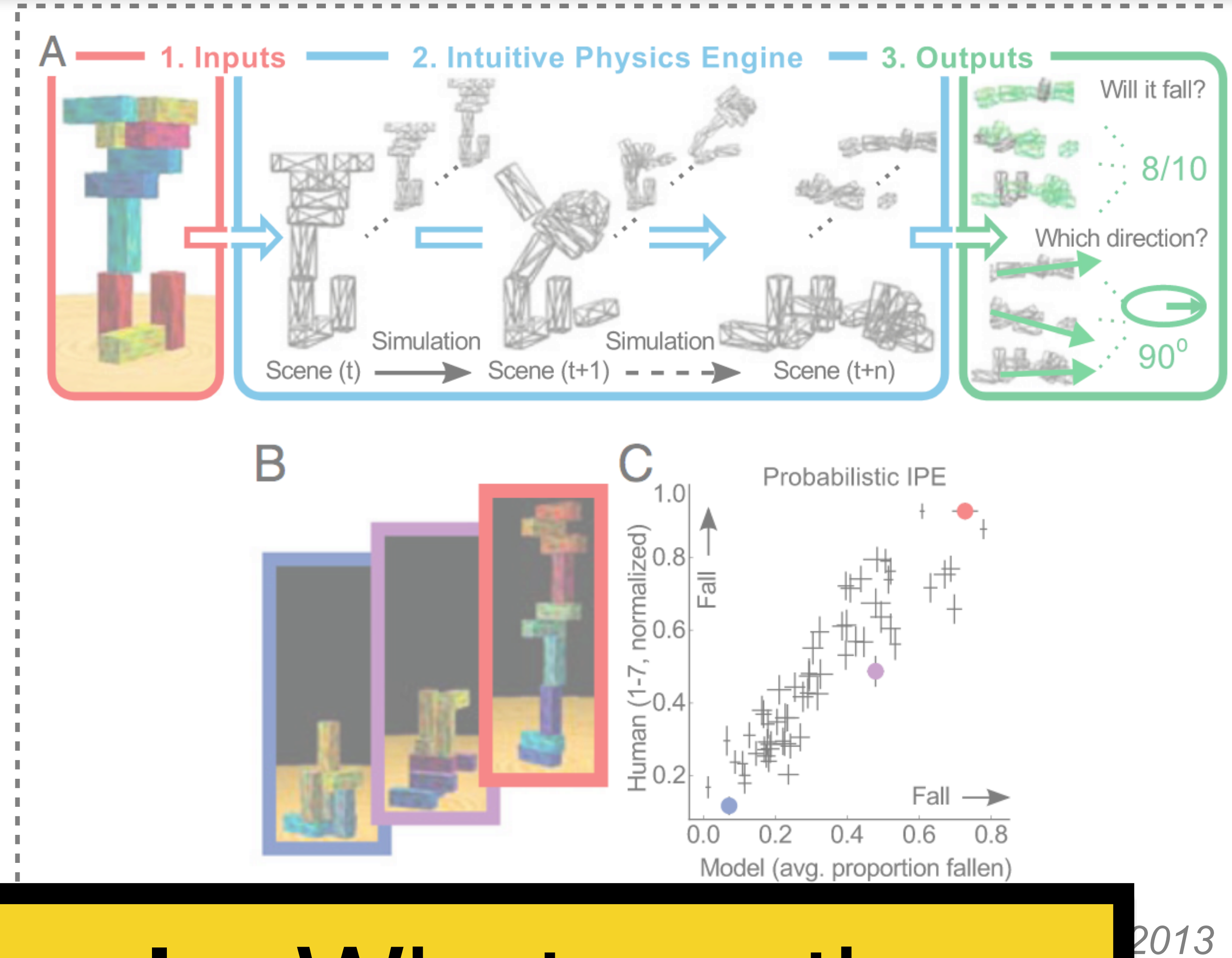


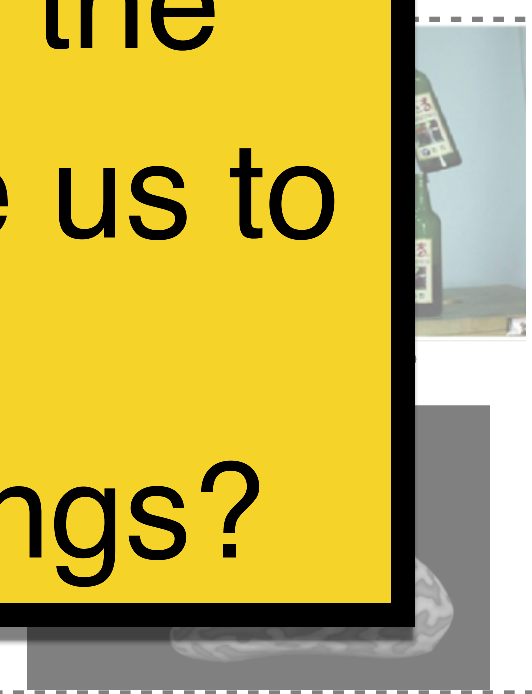
Functional Constraints of Mental Simulation *Across* Environments?











Neurobiological Puzzle: What are the functional constraints that enable us to predict the future state of our environment *across* diverse settings?

Schwettmann et al. 2019

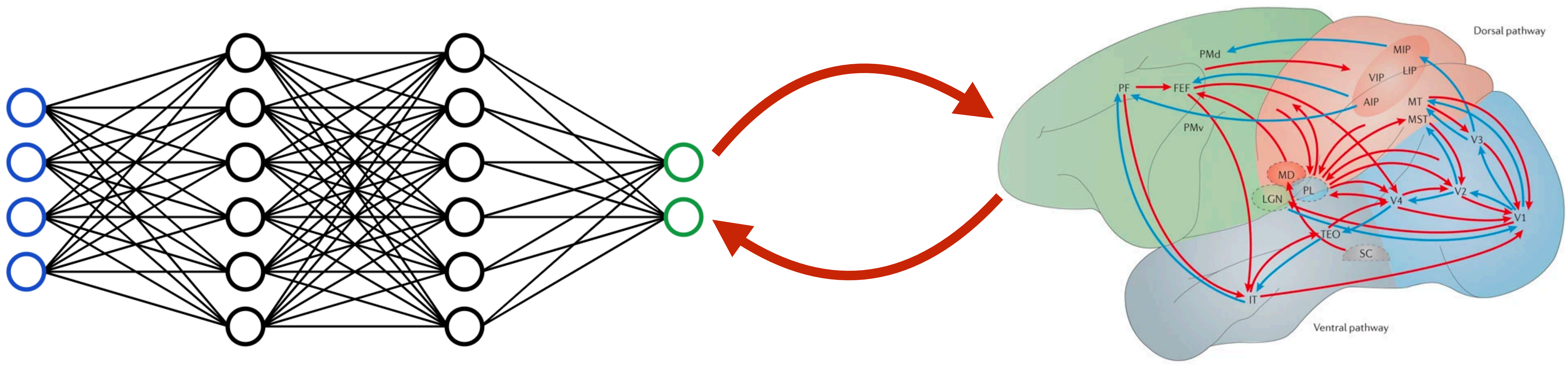
Rishi Rajalingham Mehrdad Jazayeri

Task-Optimized Approach

Neurobiological Puzzle: What are the functional constraints that enable us to predict the future state of our environment *across* diverse settings?

Task-Optimized Approach

Design ML Algorithms Optimized to Perform Organism's Behavior under Organism's Constraints



Yields:

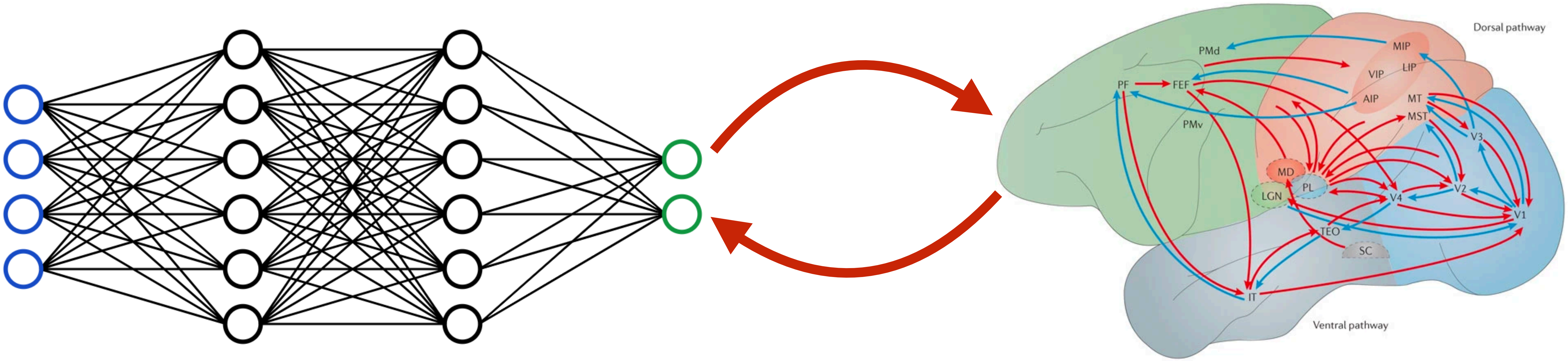
Quantitatively Accurate & Practically Useful Brain Models

AND

Principles of *Why* Neural Responses Are As They Are

Task-Optimized Approach

Design ML Algorithms Optimized to Perform Organism's Behavior under Organism's Constraints



Yields:

Quantitatively Accurate & Practically Useful Brain Models

AND

Principles of *Why* Neural Responses Are As They Are

Task-Optimized Approach

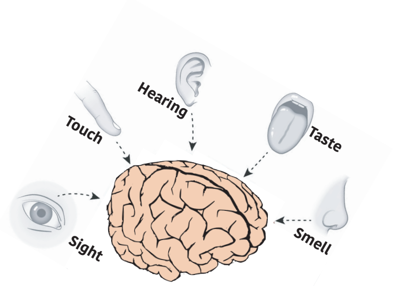
L *= learning rule*

"Natural selection + plasticity"

Backpropagation

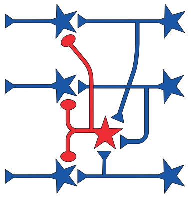
T *= task loss*

"Ecological niche/ behavior"



"Environment"

D *= data stream*



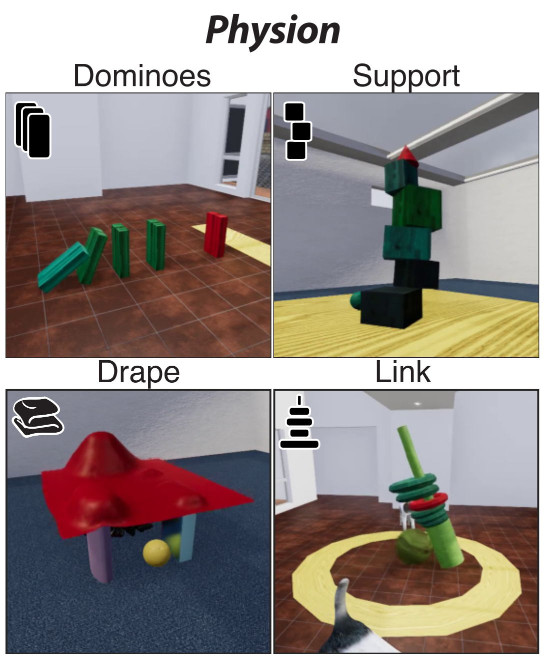
"Circuit"

A *= architecture class*

Task-Optimized Approach

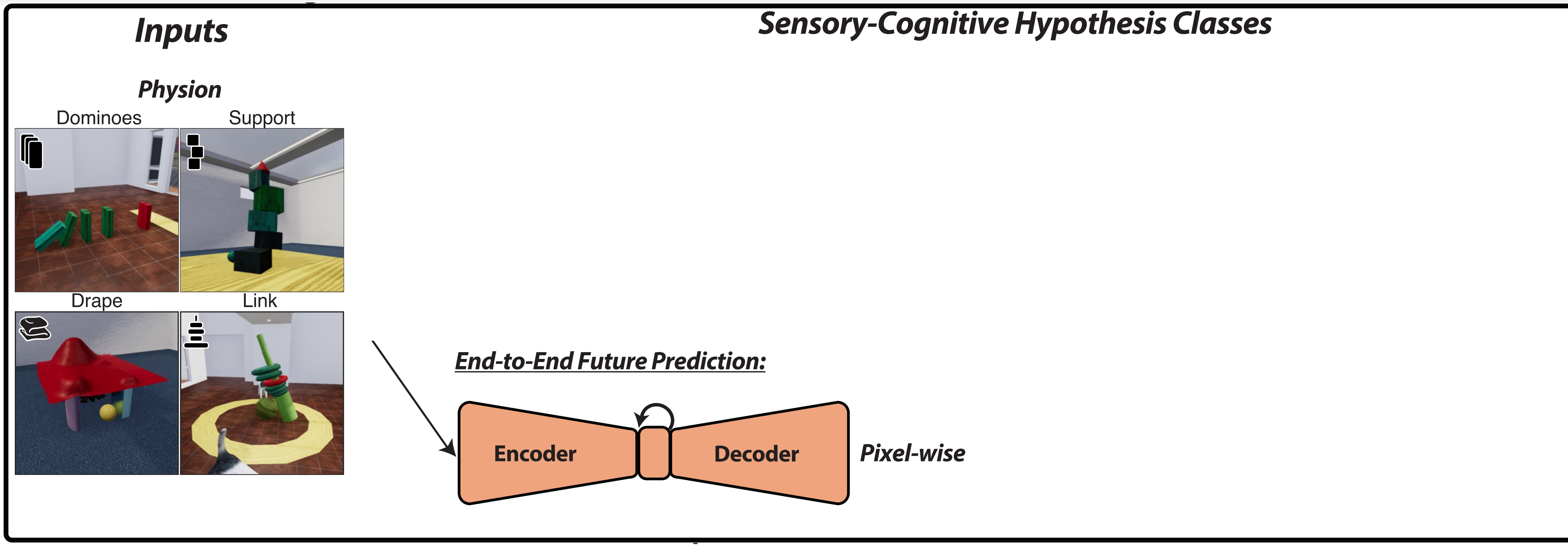
Inputs

Sensory-Cognitive Hypothesis Classes



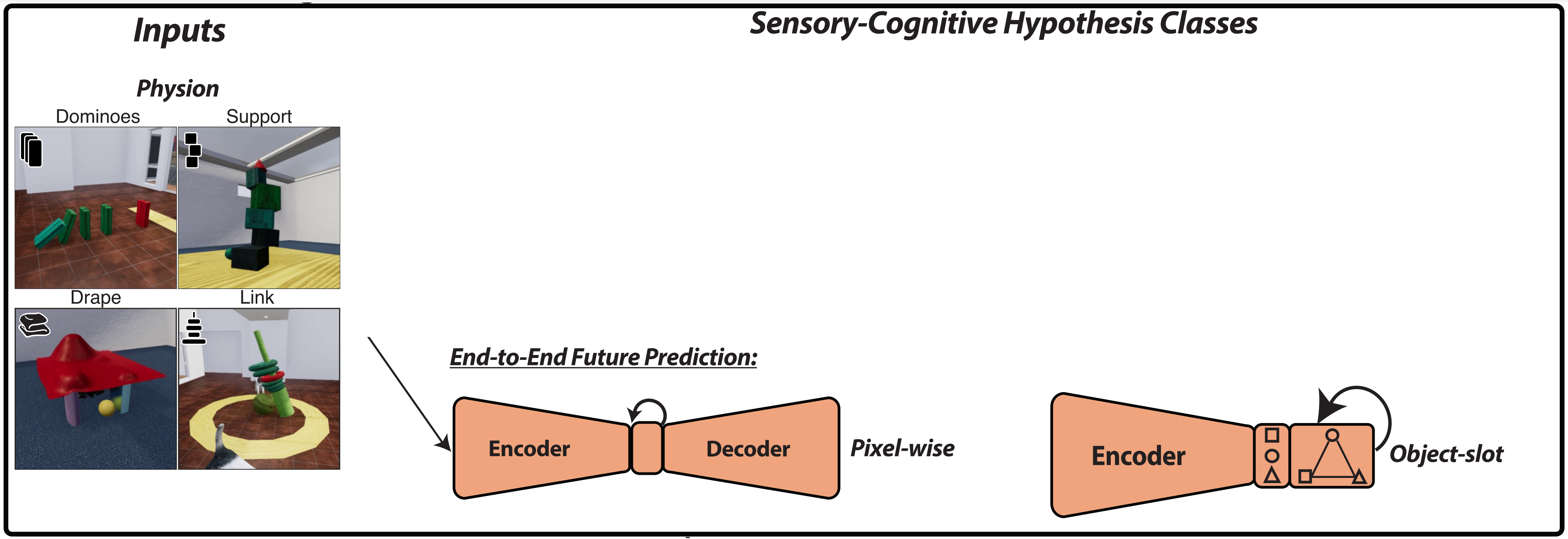
Neurobiological Puzzle: What are the functional constraints that enable us to predict the future state of our environment *across* diverse settings?

Task-Optimized Approach



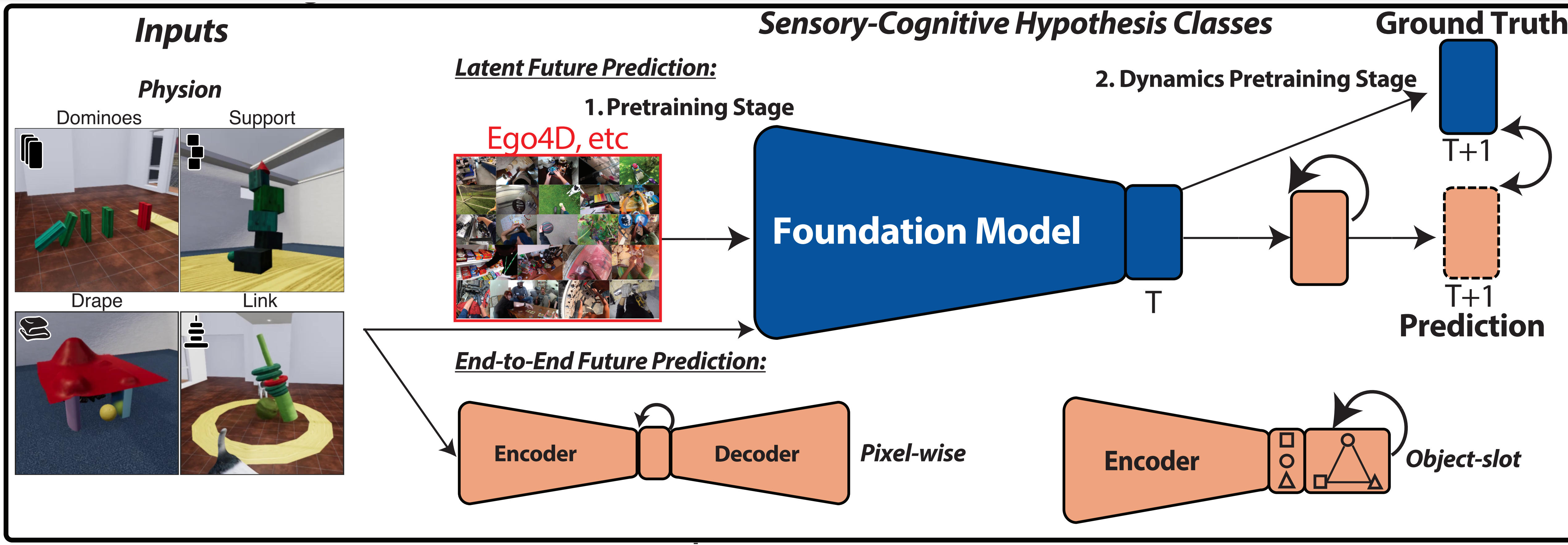
Neurobiological Puzzle: What are the functional constraints that enable us to predict the future state of our environment *across* diverse settings?

Task-Optimized Approach



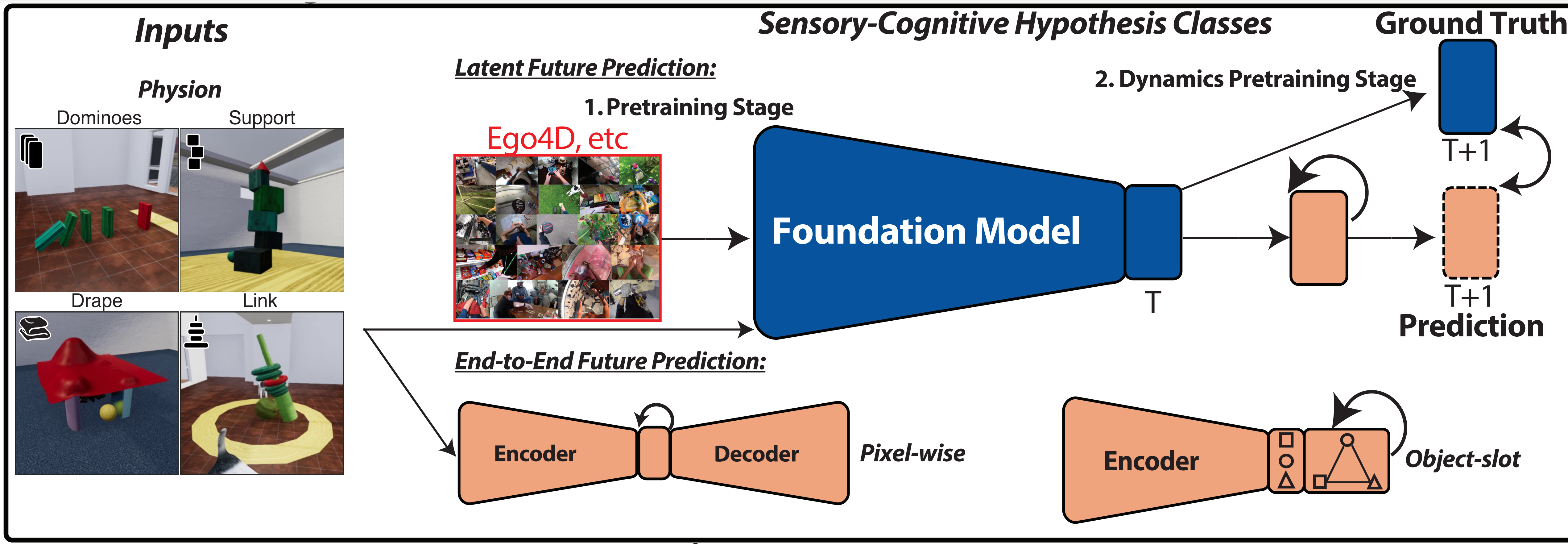
Neurobiological Puzzle: What are the functional constraints that enable us to predict the future state of our environment *across* diverse settings?

Task-Optimized Approach



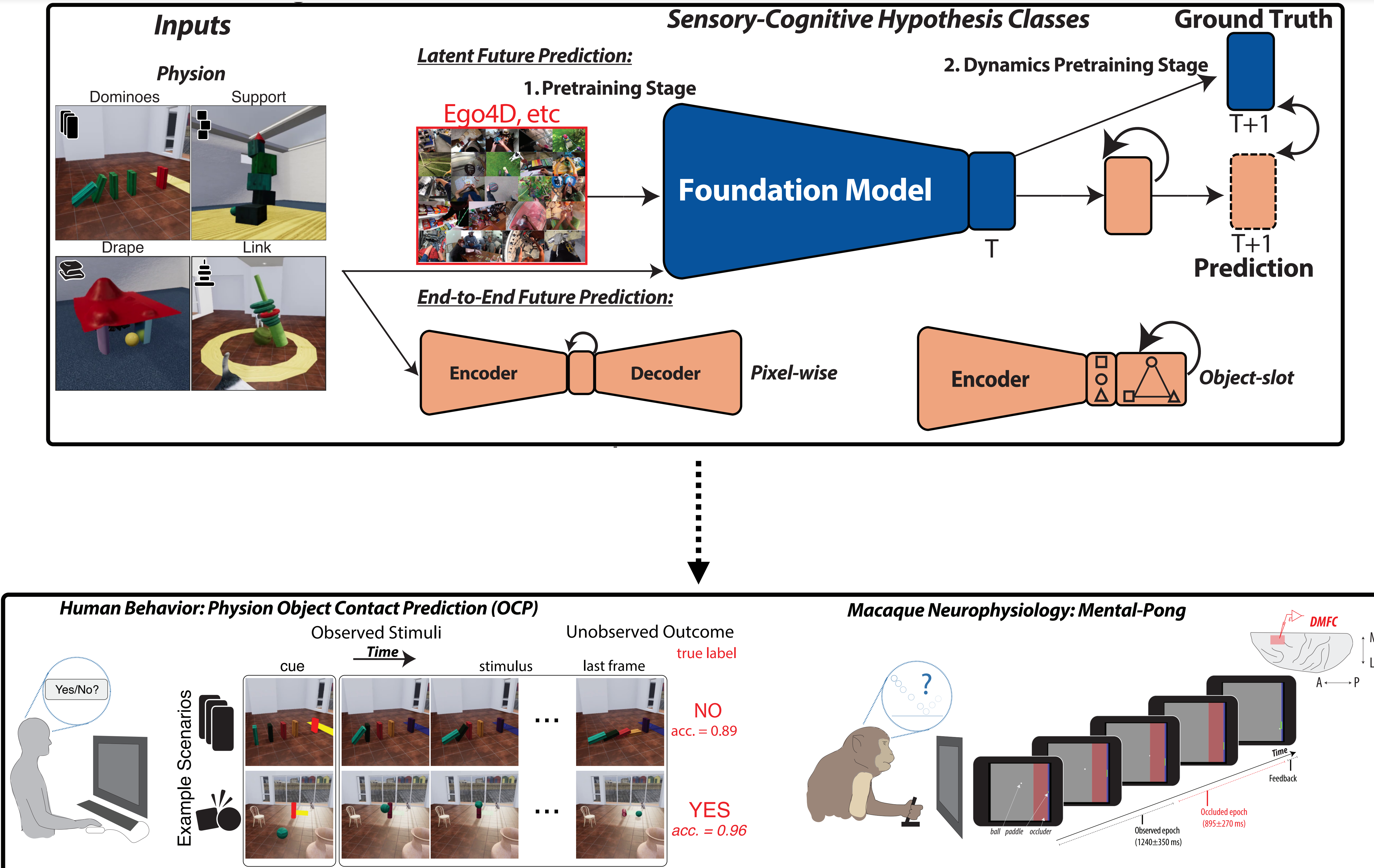
Neurobiological Puzzle: What are the functional constraints that enable us to predict the future state of our environment *across* diverse settings?

Task-Optimized Approach

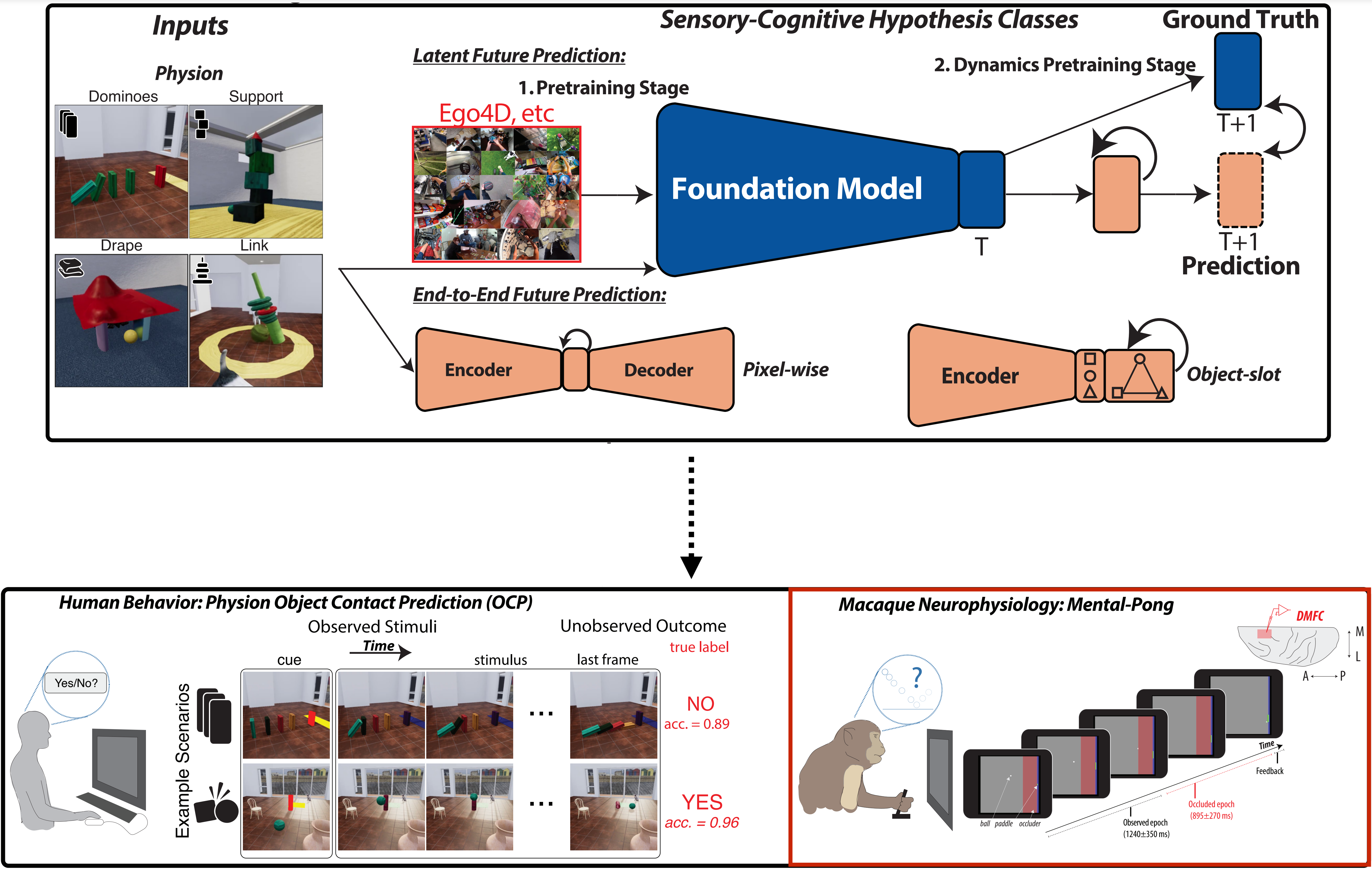


Neurobiological Puzzle: What are the functional constraints that enable us to predict the future state of our environment *across* diverse settings?

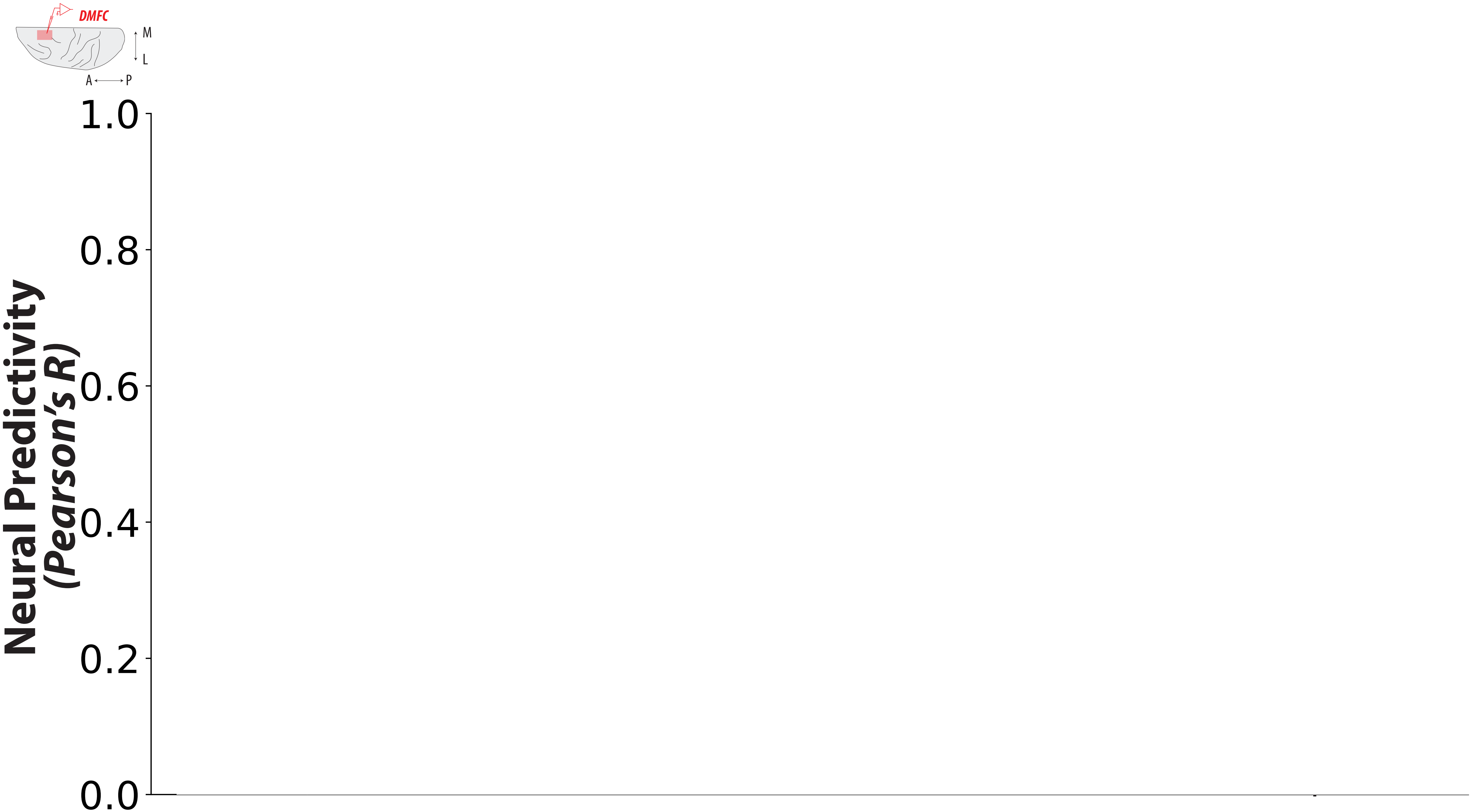
Task-Optimized Approach



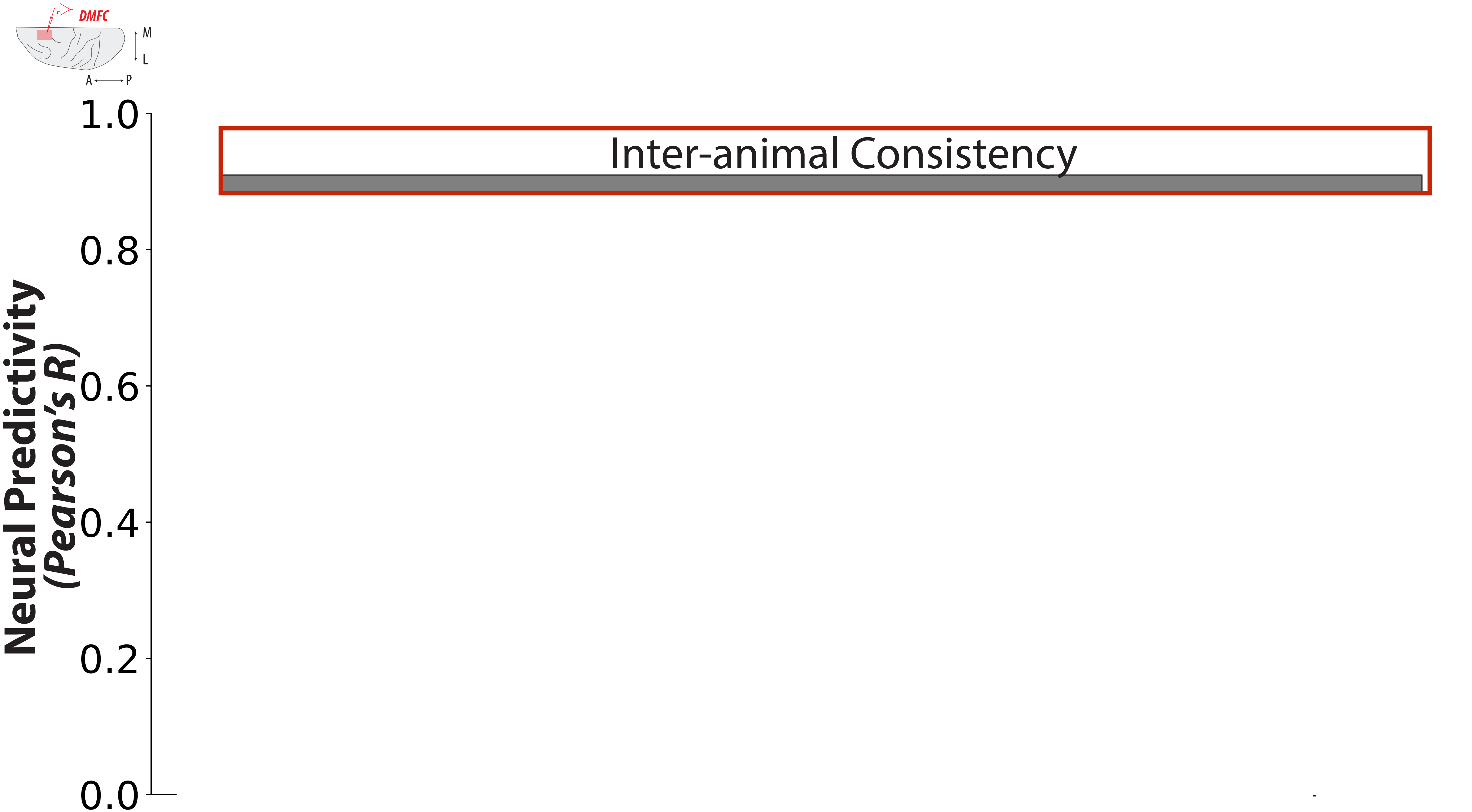
Macaque Neurophysiology: Mental Pong



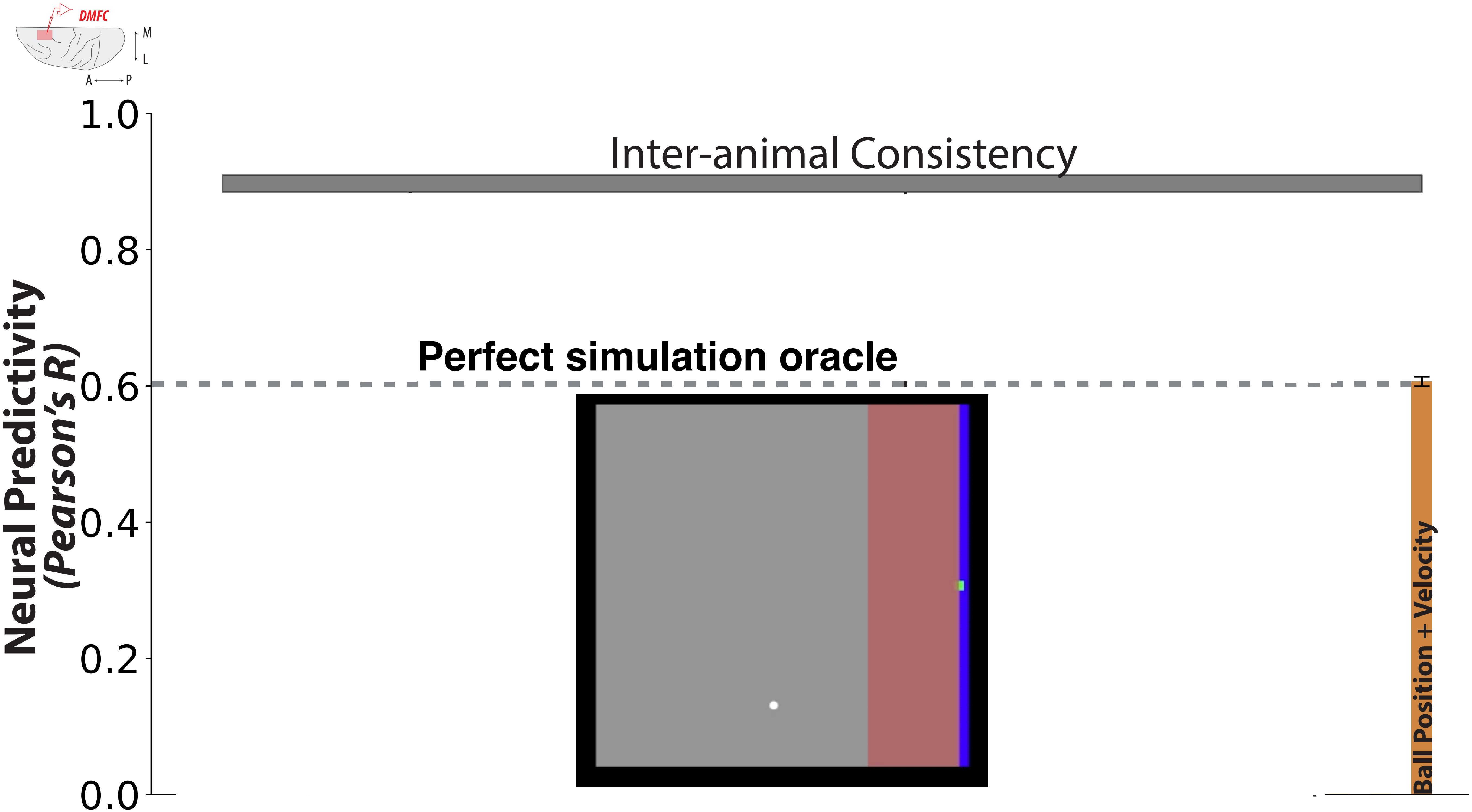
Macaque Neurophysiology: Mental Pong



Macaque Neurophysiology: Mental Pong



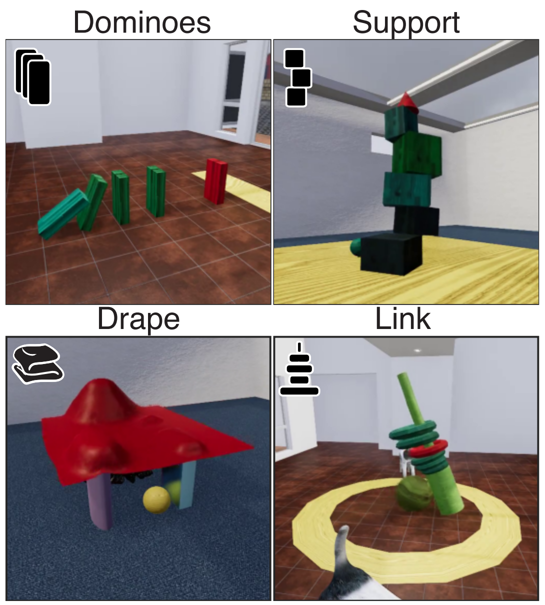
Perfect Simulation Oracle Predicts Neural Data Well



Functional Constraint Hypotheses

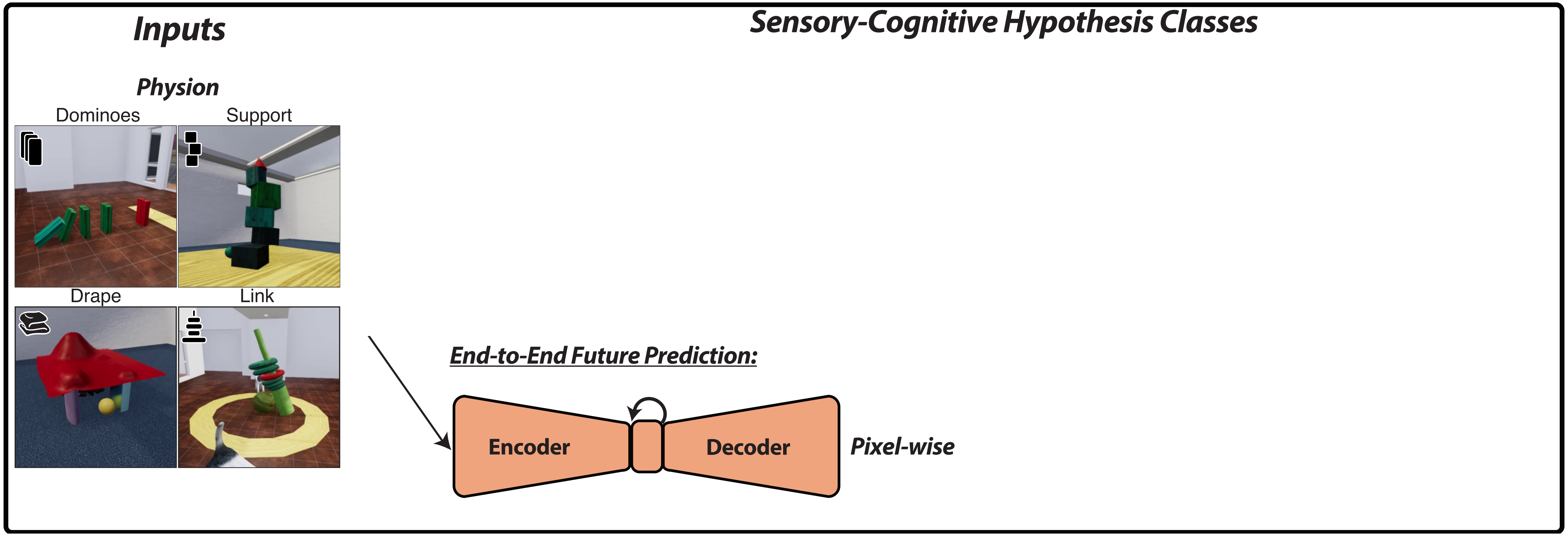
Inputs

Physion



Sensory-Cognitive Hypothesis Classes

Hypothesis Class 1: Pixel-wise Future Prediction

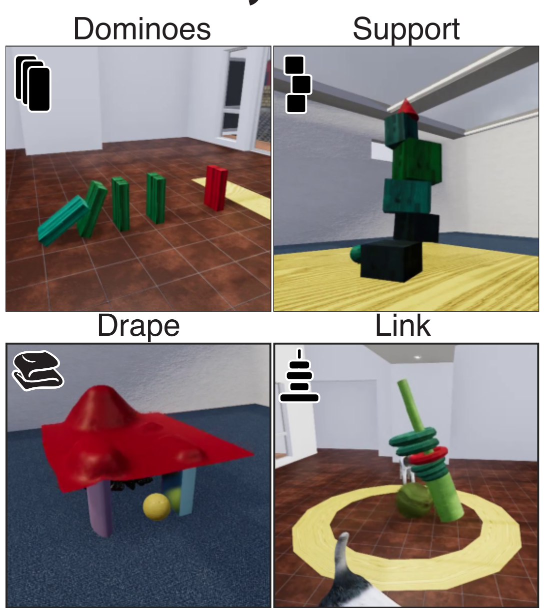


Hypothesis Class I: Pixel-wise Future Prediction

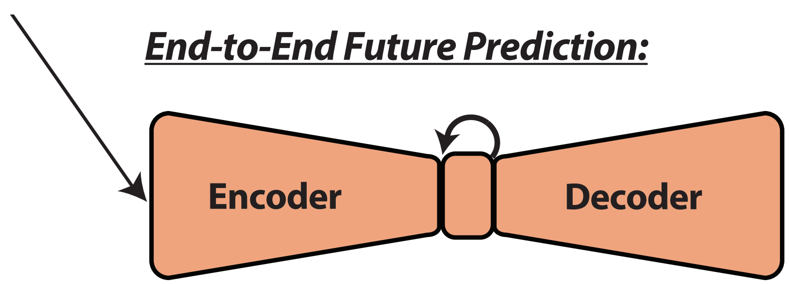
Inputs

Sensory-Cognitive Hypothesis Classes

Physion

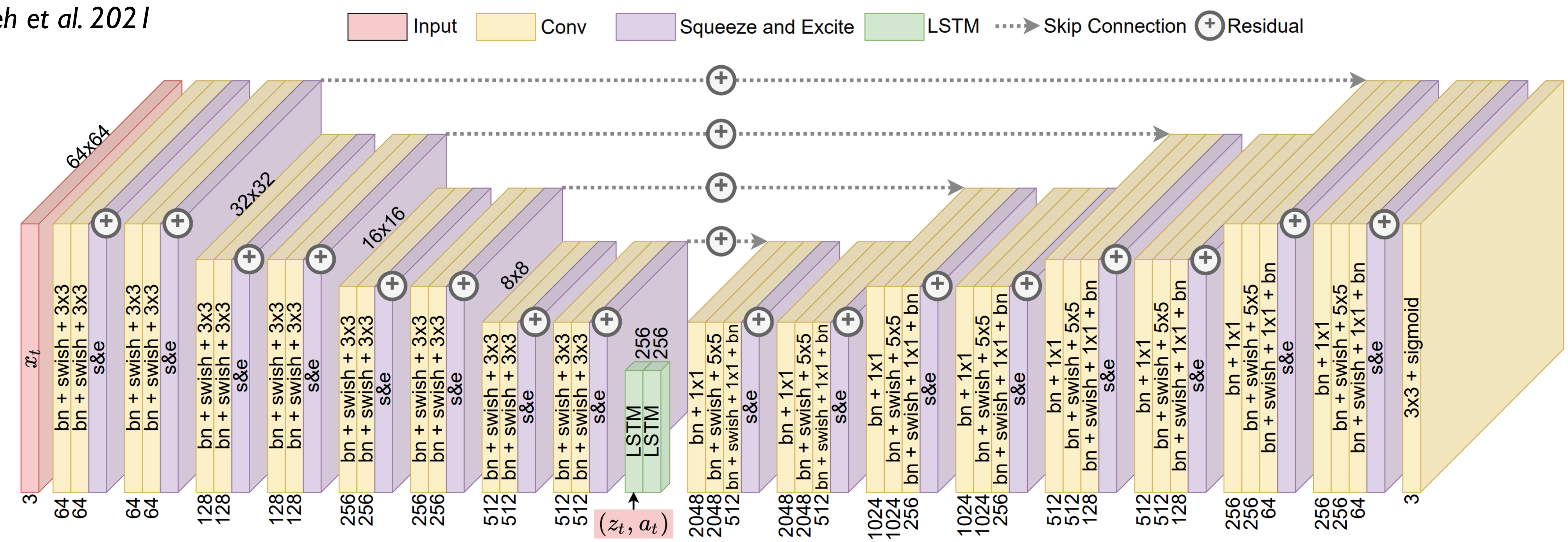


End-to-End Future Prediction:



Pixel-wise

Babaeizadeh et al. 2021

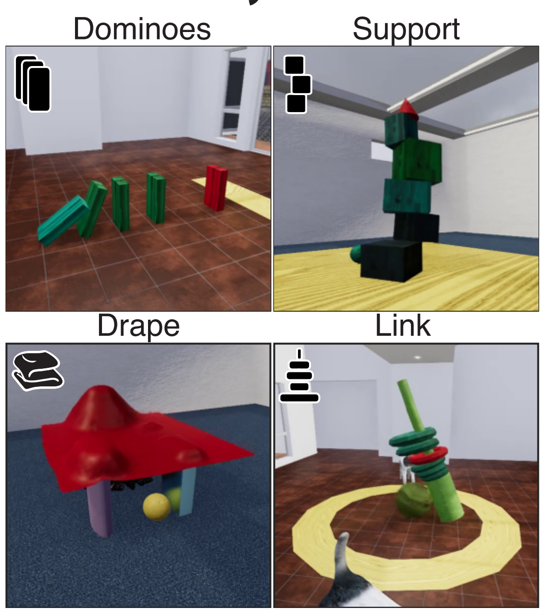


Hypothesis Class I: Pixel-wise Future Prediction

Inputs

Sensory-Cognitive Hypothesis Classes

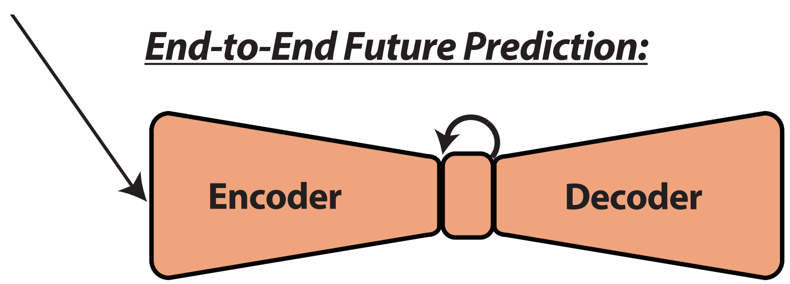
Physion



Predicts the future at the *pixel* resolution of the sensory input **(very detailed)**

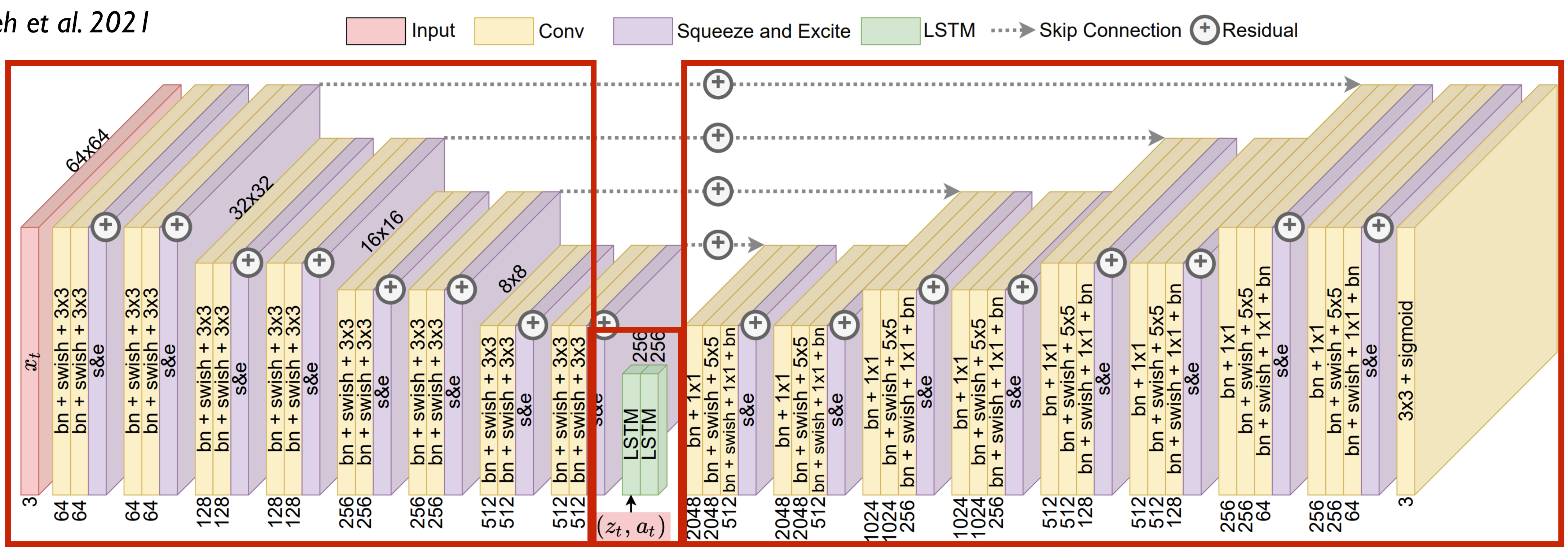
$$\mathcal{L} = \|\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1}\|_2^2 + \beta D_{KL}(\mathcal{N}(\mu, \sigma)\|\mathcal{N}(0, \mathbf{I}))$$

End-to-End Future Prediction:



Pixel-wise

Babaeizadeh et al. 2021

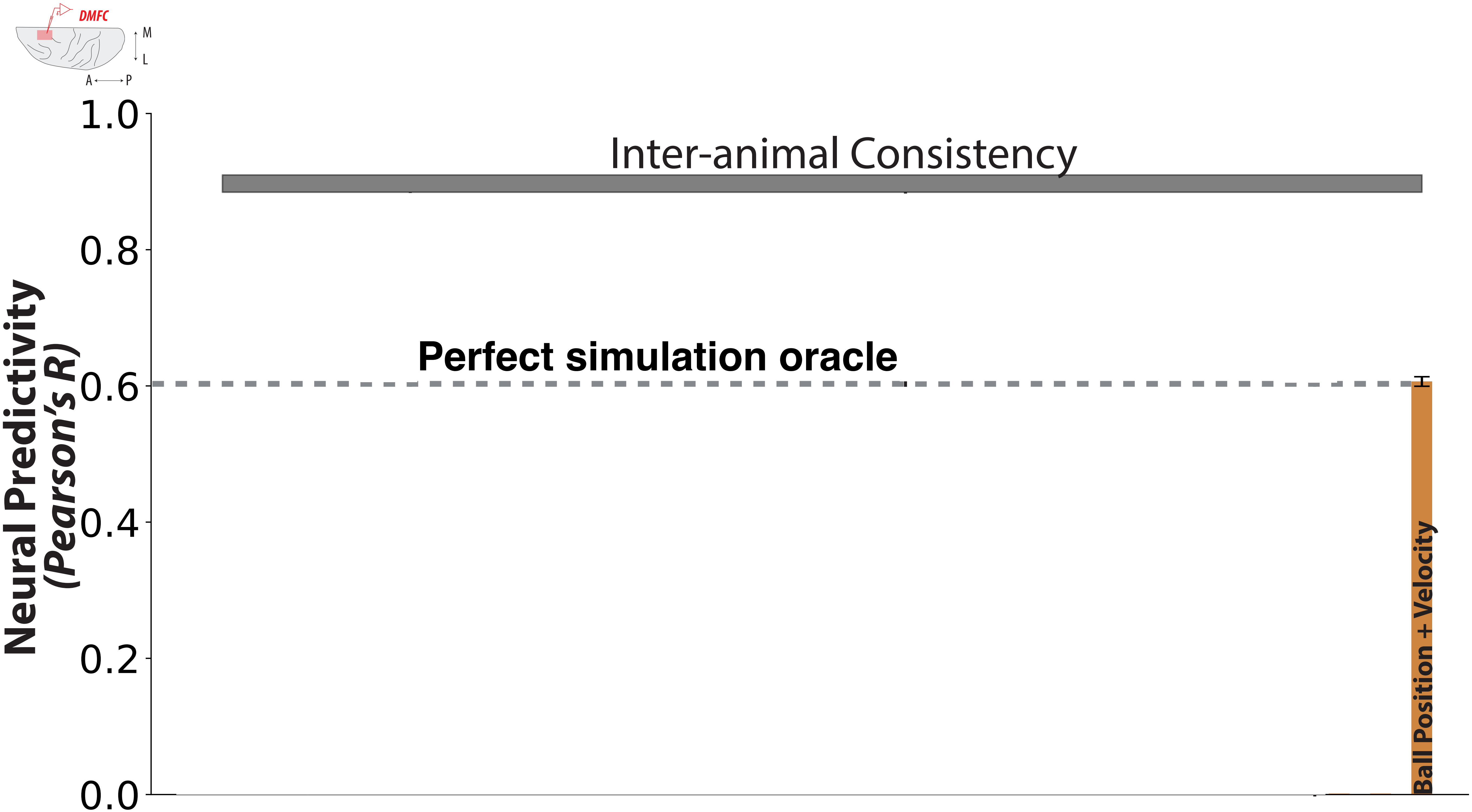


Visual Encoder ("Sensory")

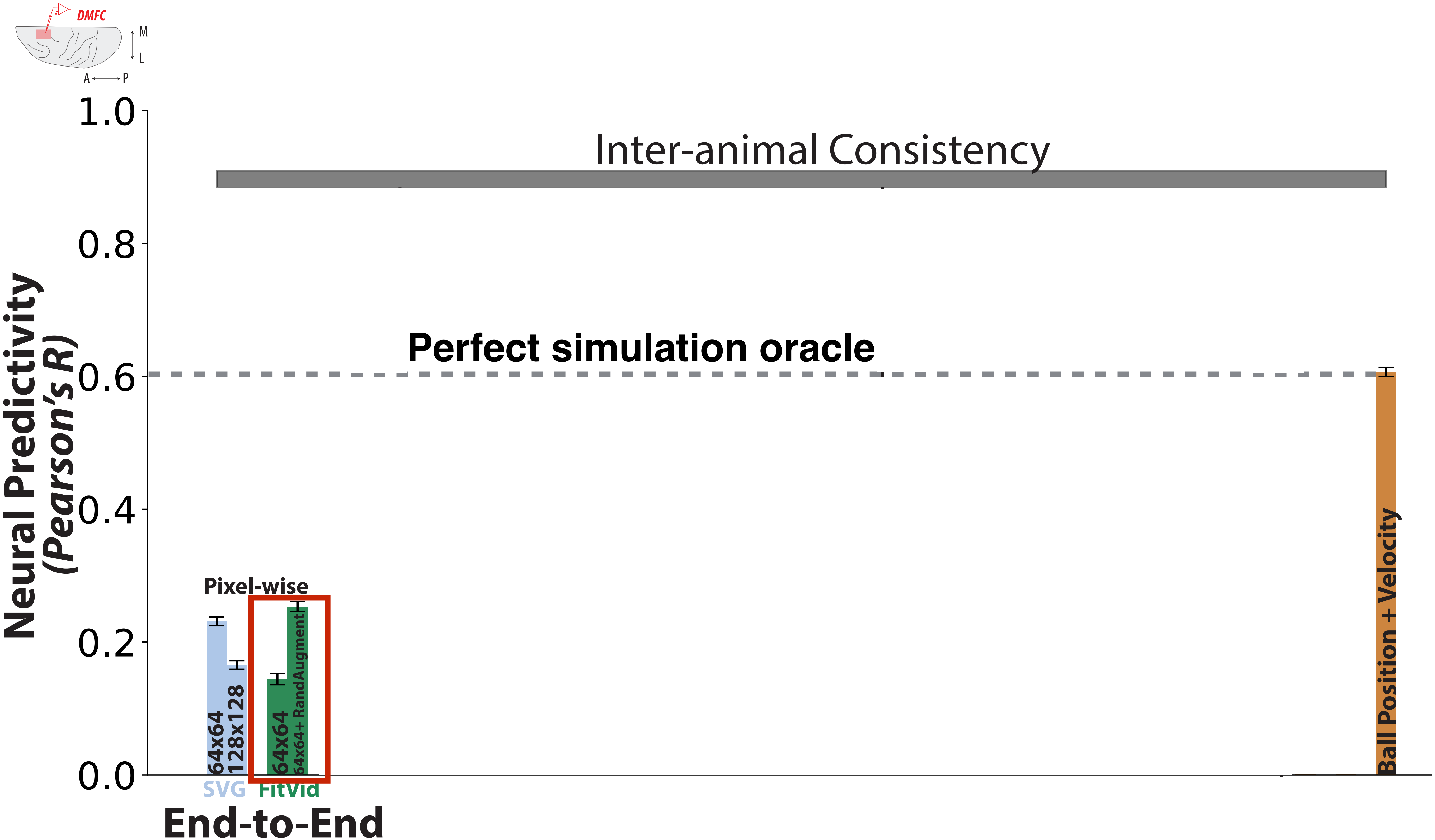
Dynamics Predictor ("Cognitive")

Frame Decoder ("Objective/Behavior")

Physical Simulation Oracles Predict Neural Data Well

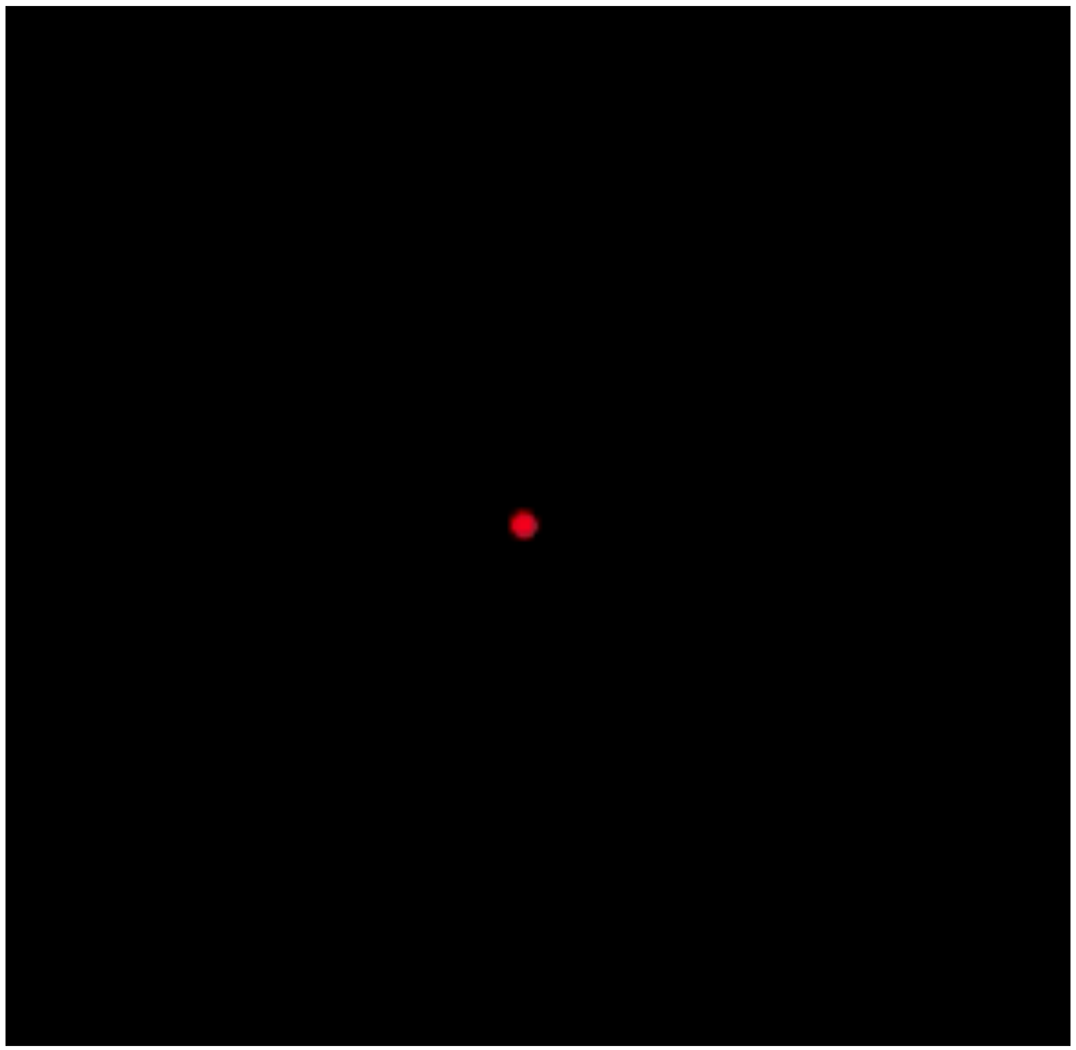


Pixel-wise Future Prediction Poorly Predicts Neurons

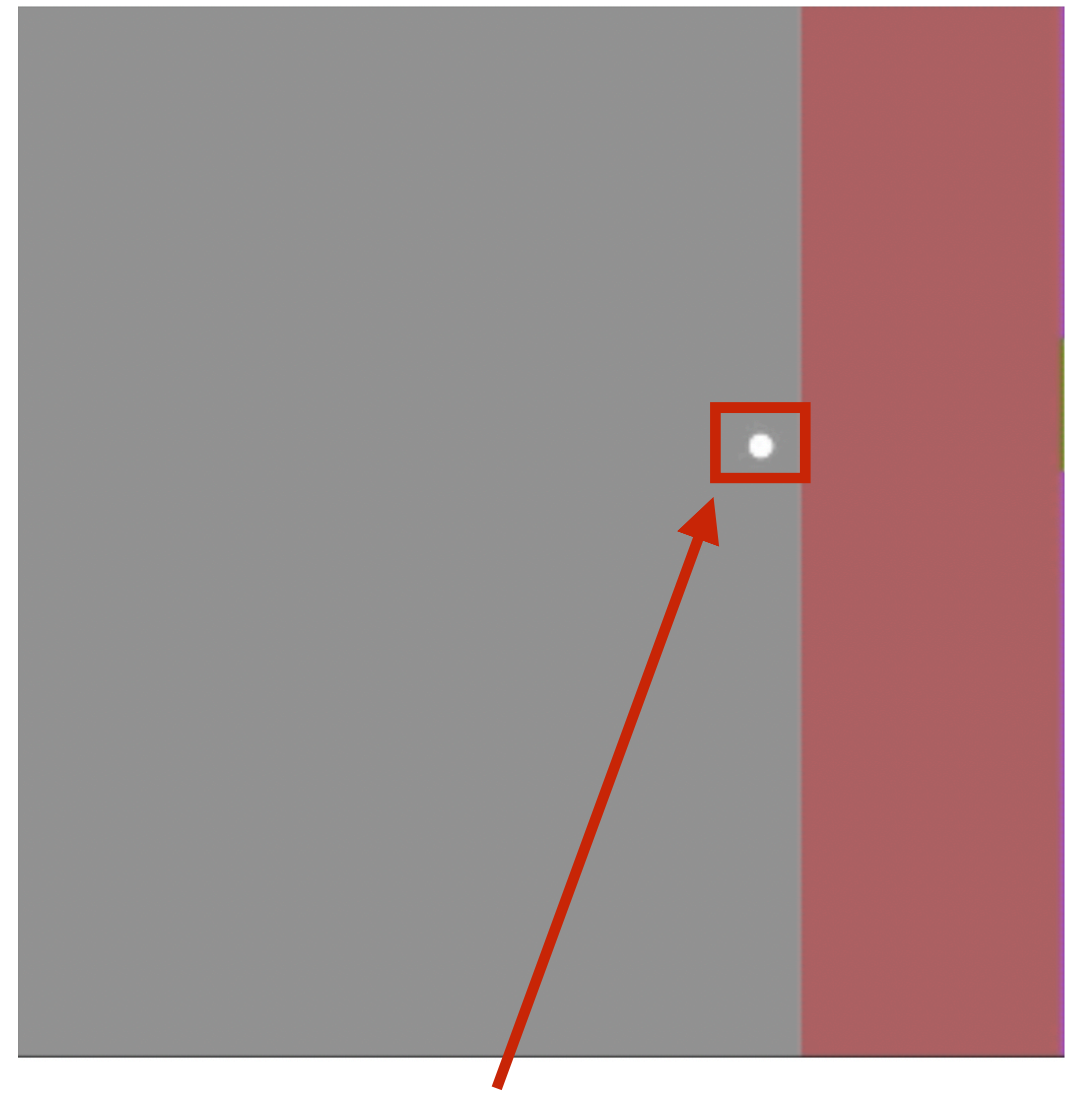


…and they struggle to generalize to Pong

Input Frames



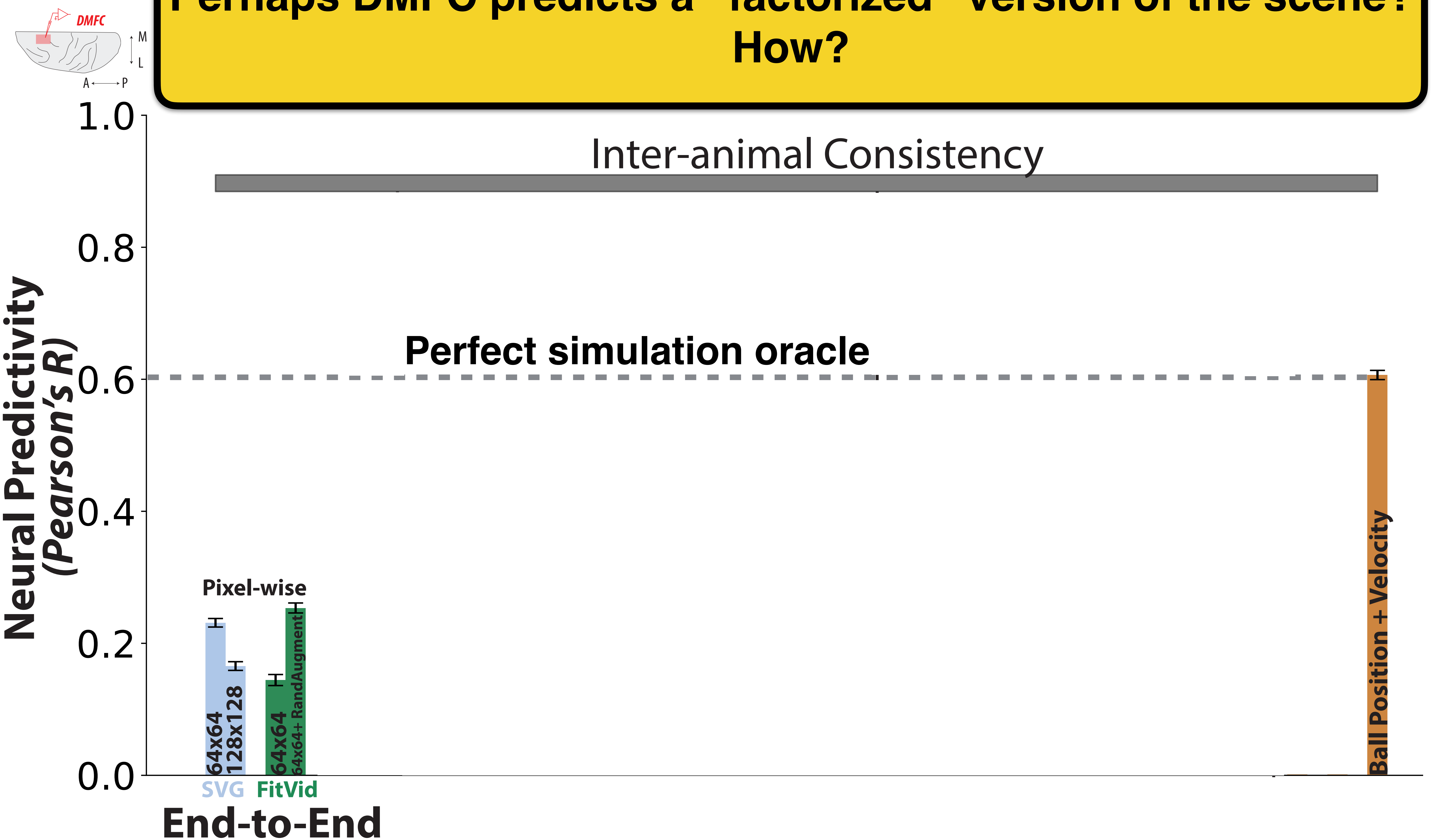
Predicted Frames



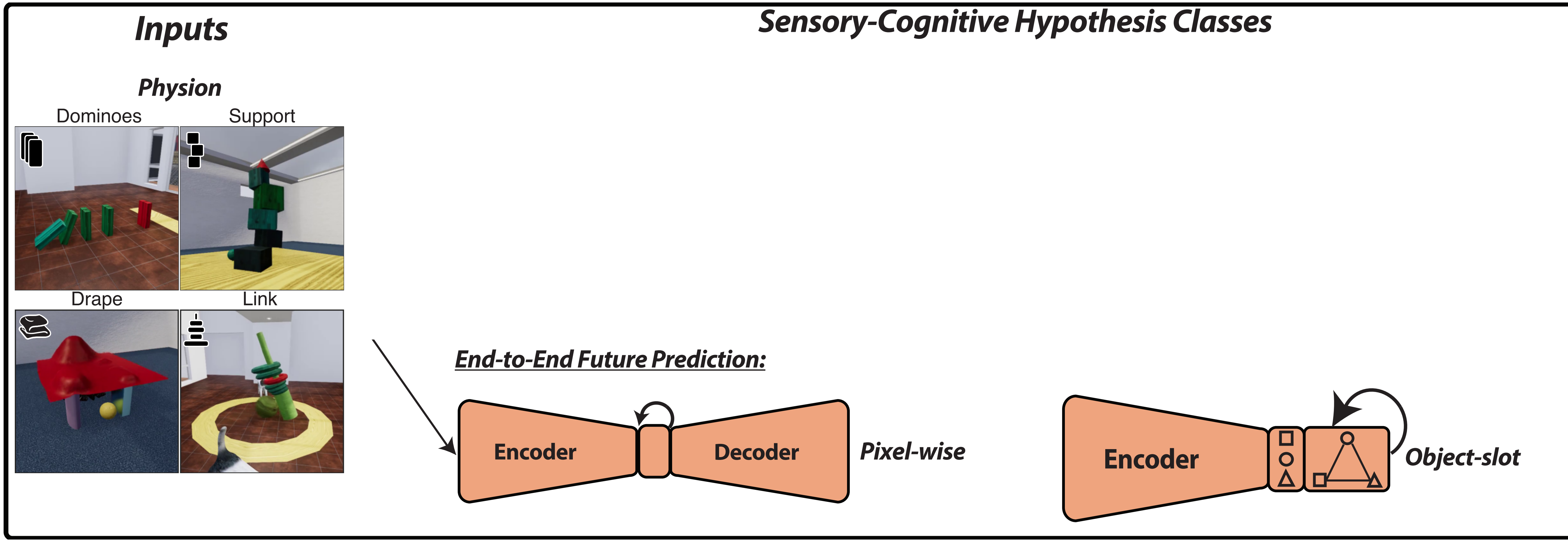
Ball stops at final input frame, in the model's "imagination"

Pixel-wise Future Prediction Poorly Predicts Neurons

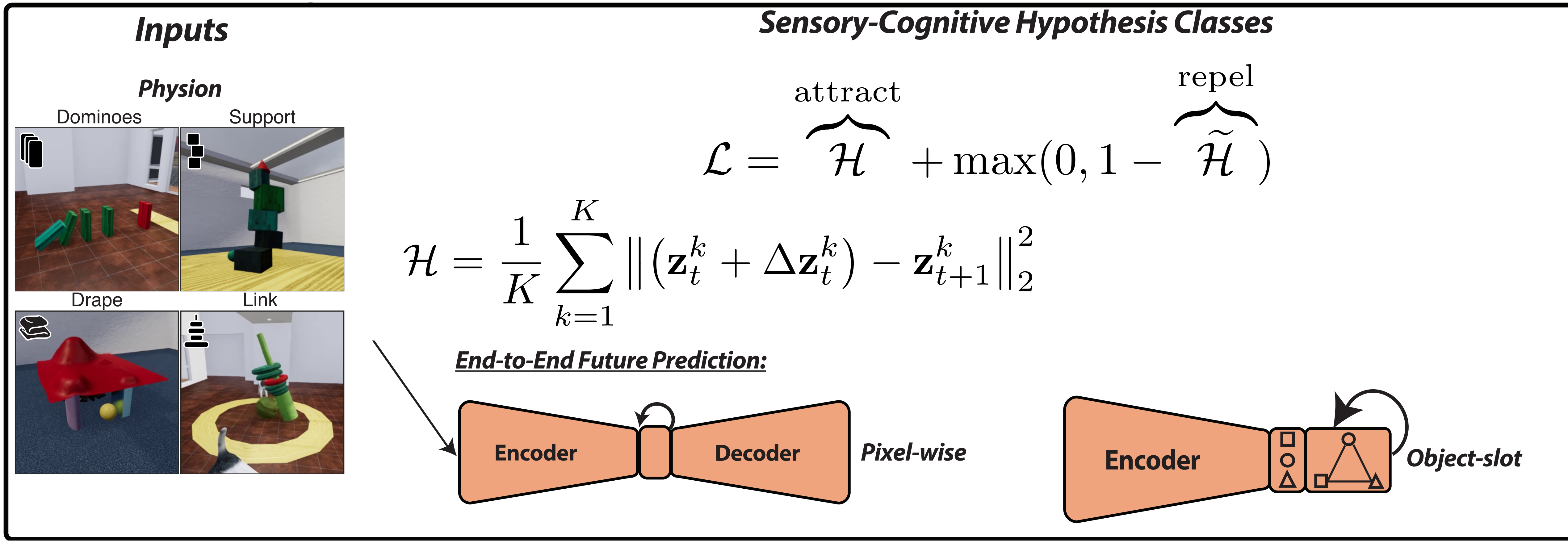
Perhaps DMFC predicts a "factorized" version of the scene? How?



Hypothesis Class 2: Object Slots

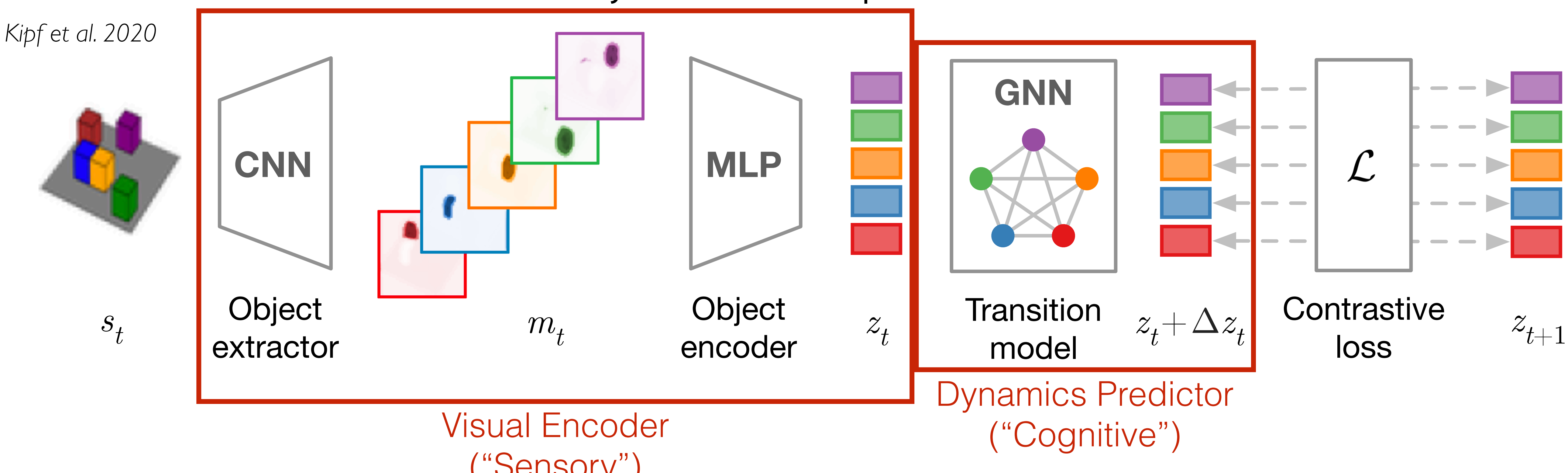


Hypothesis Class 2: Object Slots



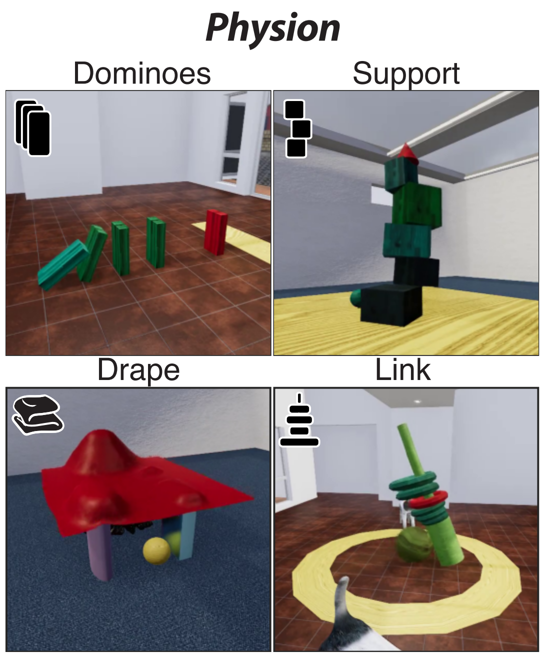
Predicts at the level of object slot representations and their relations

Kipf et al. 2020



Hypothesis Class 2: Object Slots

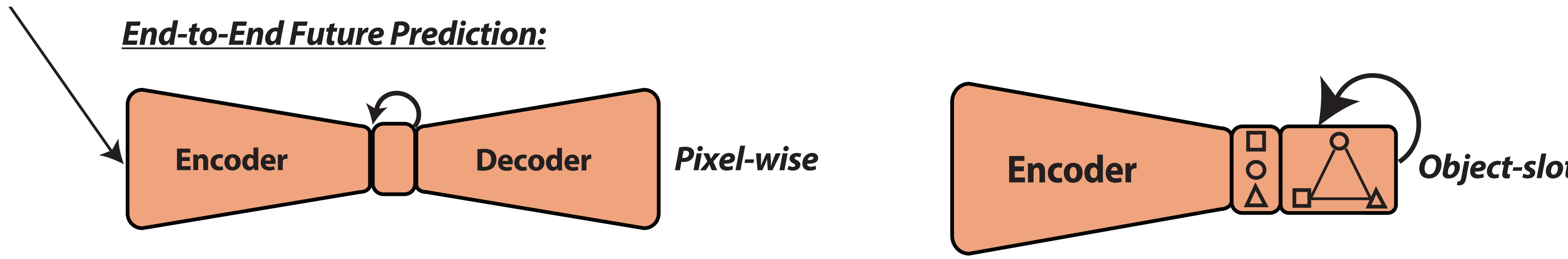
Inputs



Sensory-Cognitive Hypothesis Classes

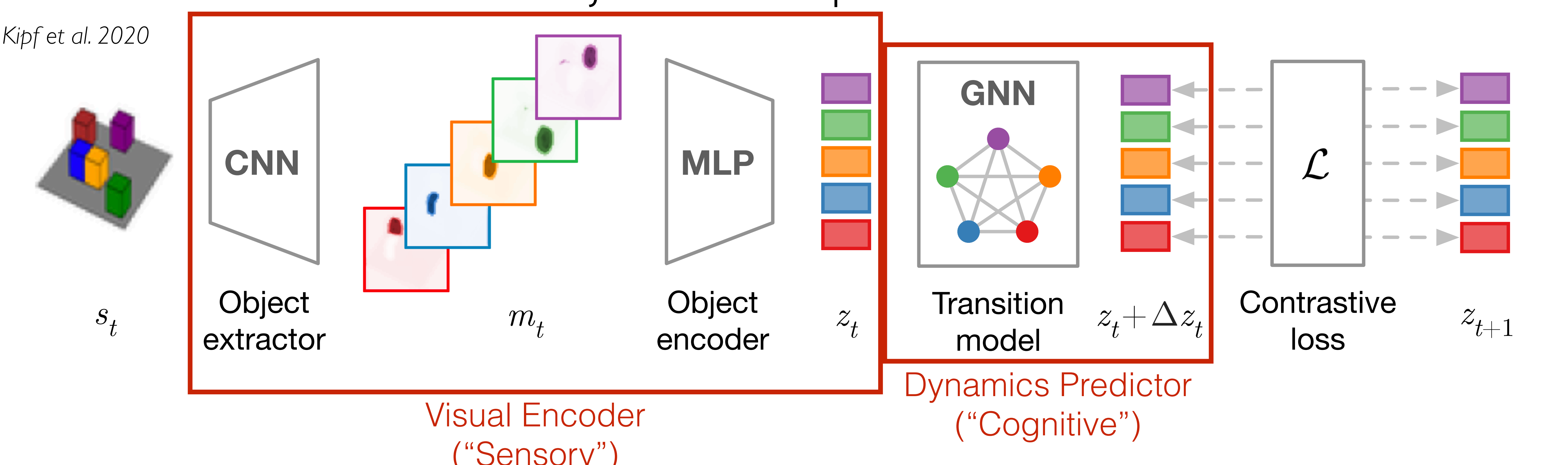
$$\mathcal{L} = \overbrace{\mathcal{H}}^{\text{attract}} + \max(0, 1 - \overbrace{\widetilde{\mathcal{H}}}^{\text{repel}})$$

$$\mathcal{H} = \frac{1}{K}\sum_{k=1}^{K} \left\| (\mathbf{z}_t^k + \Delta \mathbf{z}_t^k) - \mathbf{z}_{t+1}^k \right\|_2^2, \quad \widetilde{\mathcal{H}} = \frac{1}{K}\sum_{k=1}^{K} \left\| \widetilde{\mathbf{z}}_t^k - \mathbf{z}_{t+1}^k \right\|_2^2$$

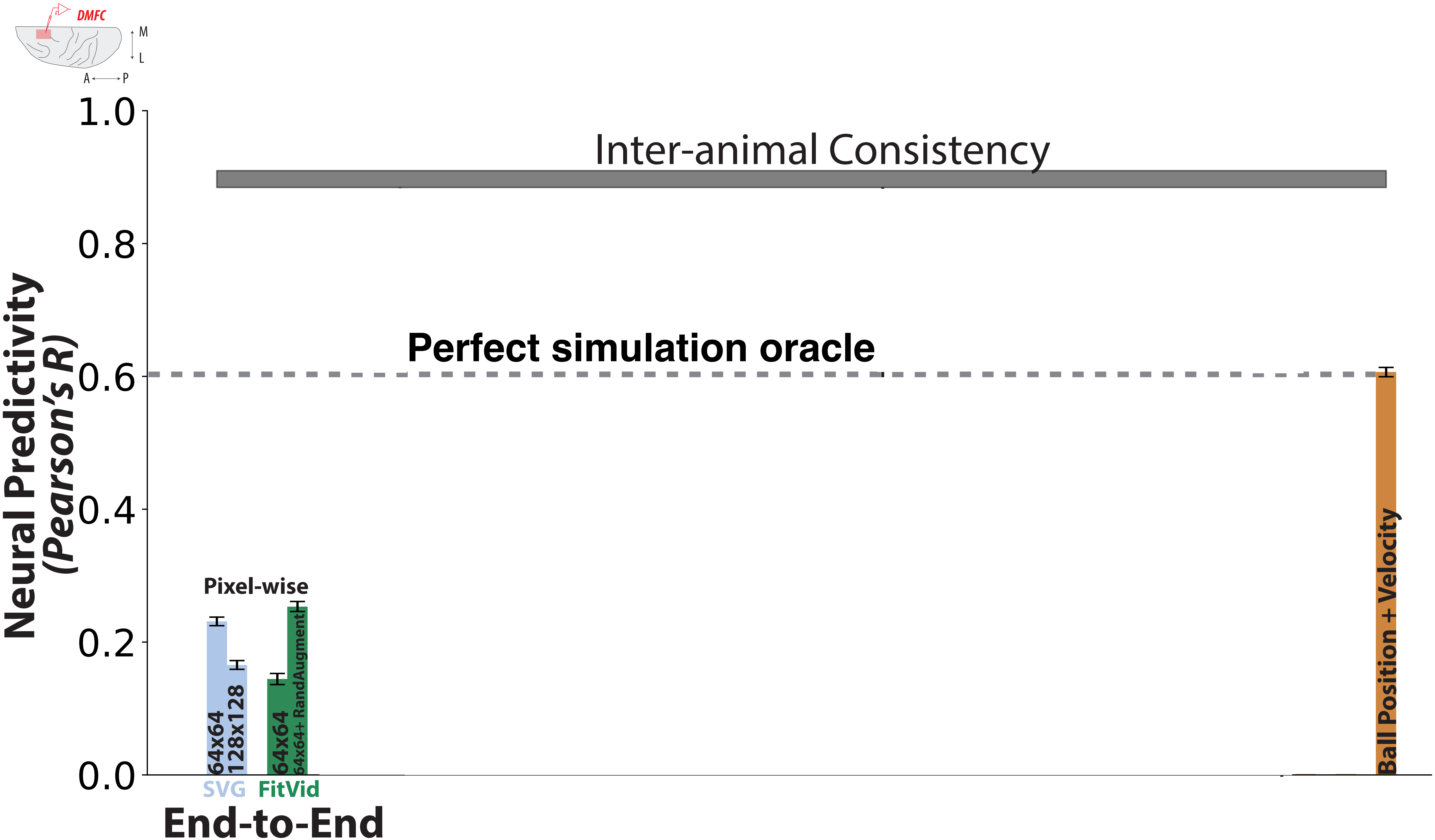


Predicts at the level of object slot representations and their relations

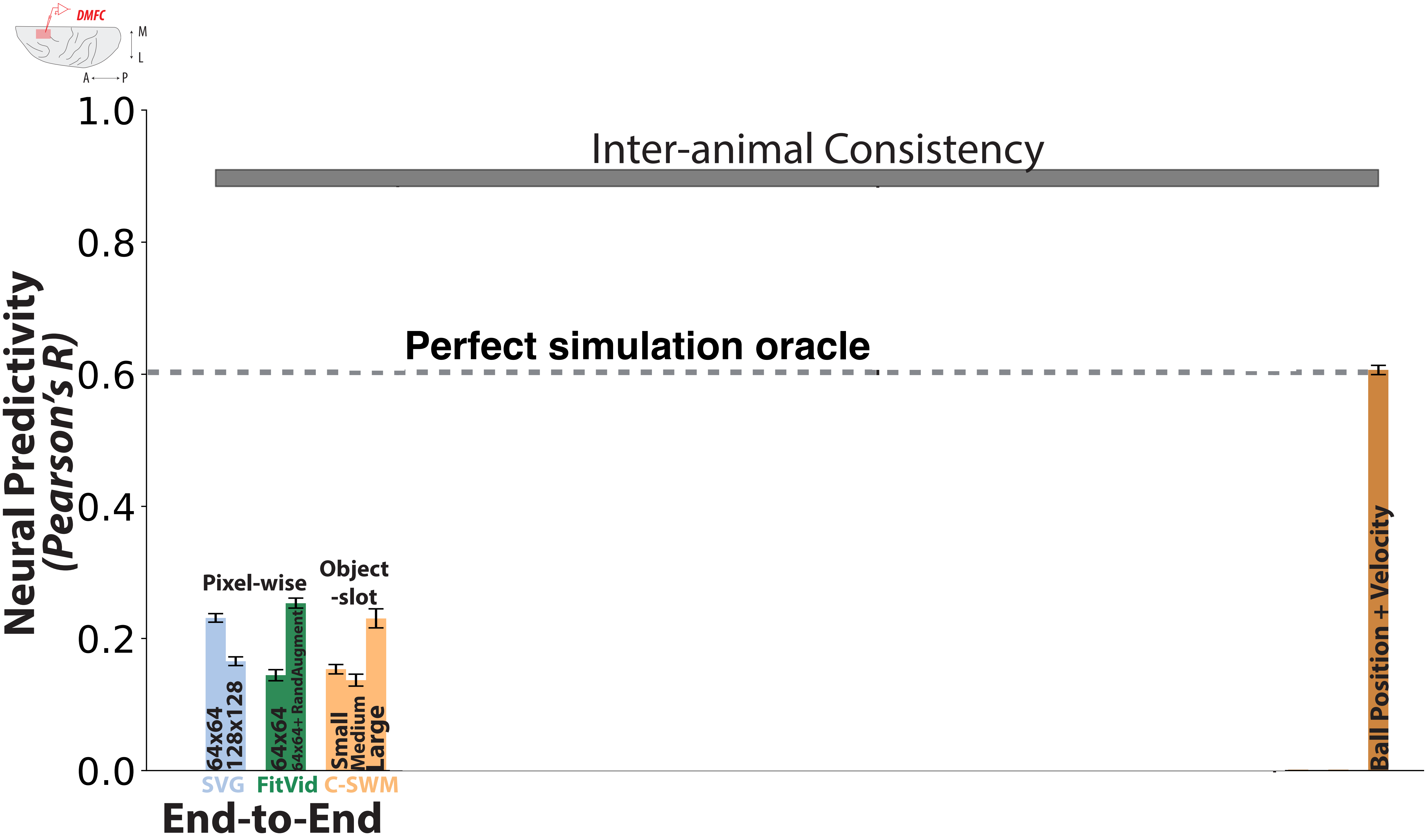
Kipf et al. 2020



Pixel-wise Future Prediction Poorly Predicts Neurons

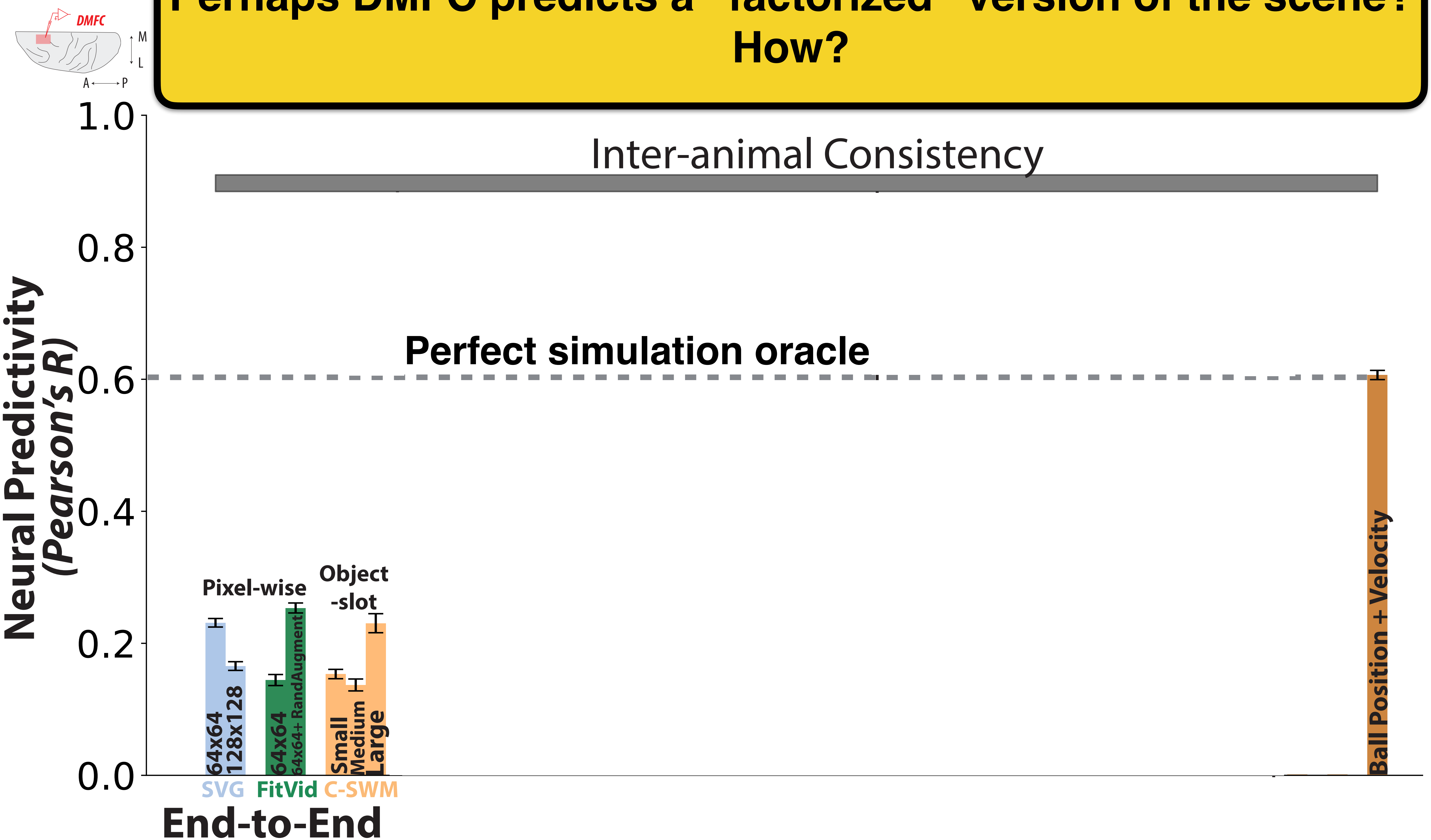


Object Slot Future Prediction Poorly Predicts Neurons



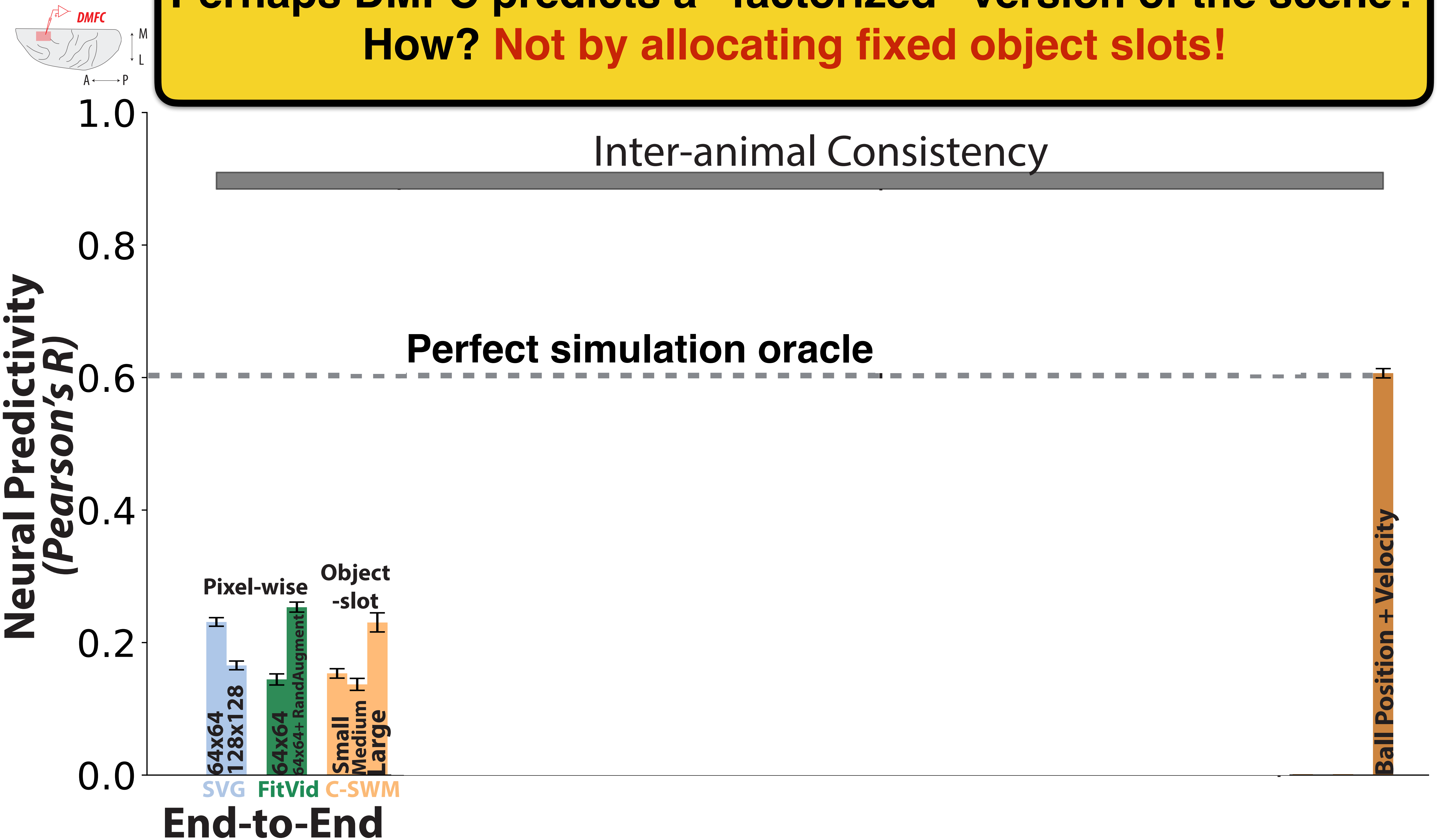
Object Slot Future Prediction Poorly Predicts Neurons

Perhaps DMFC predicts a "factorized" version of the scene? How?

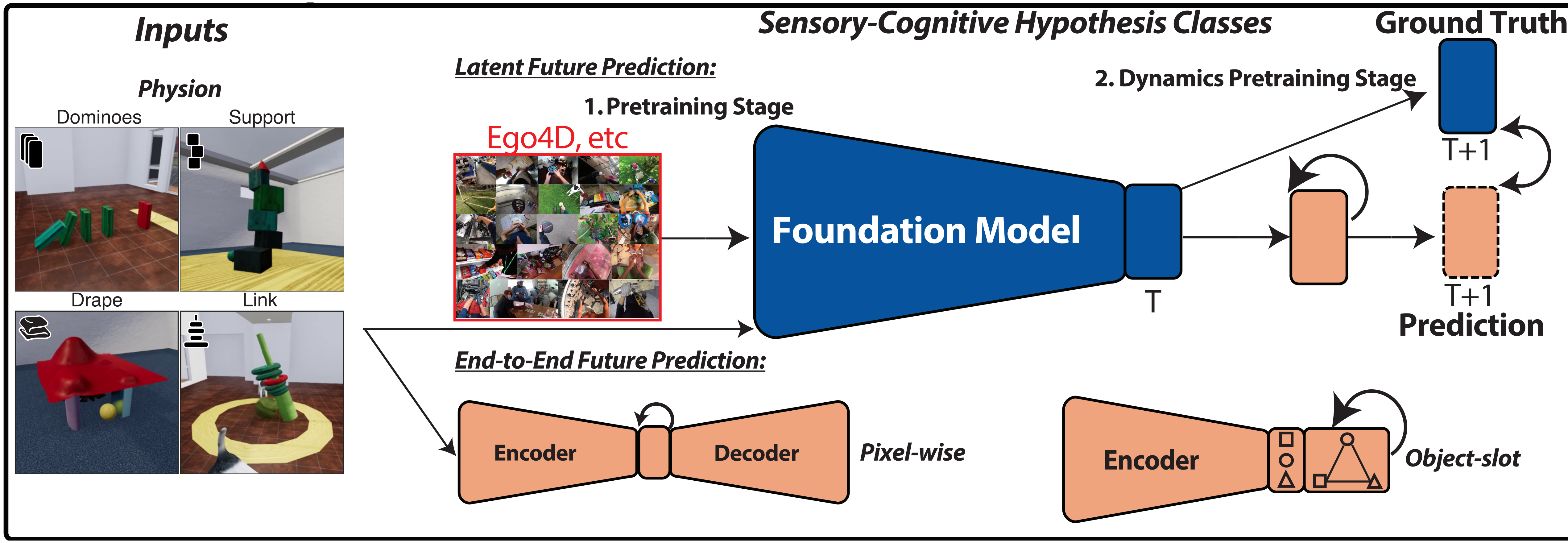


Object Slot Future Prediction Poorly Predicts Neurons

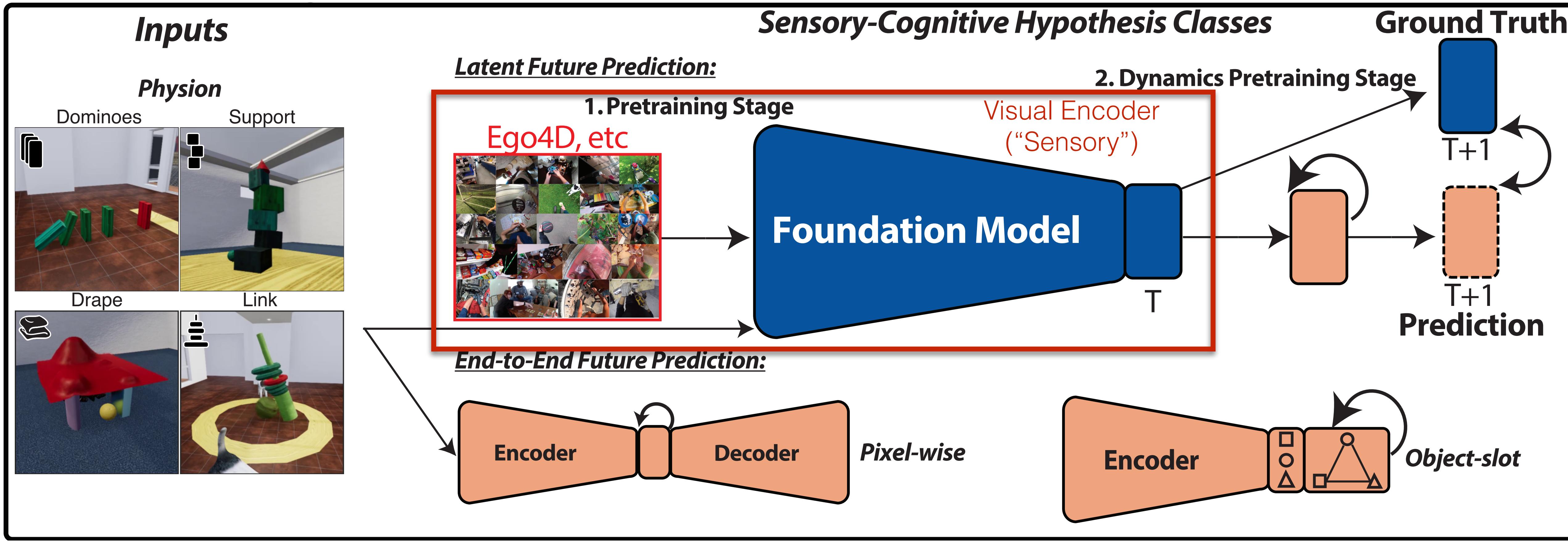
Perhaps DMFC predicts a "factorized" version of the scene?
How? Not by allocating fixed object slots!



Hypothesis Class 3: Latent Future Prediction

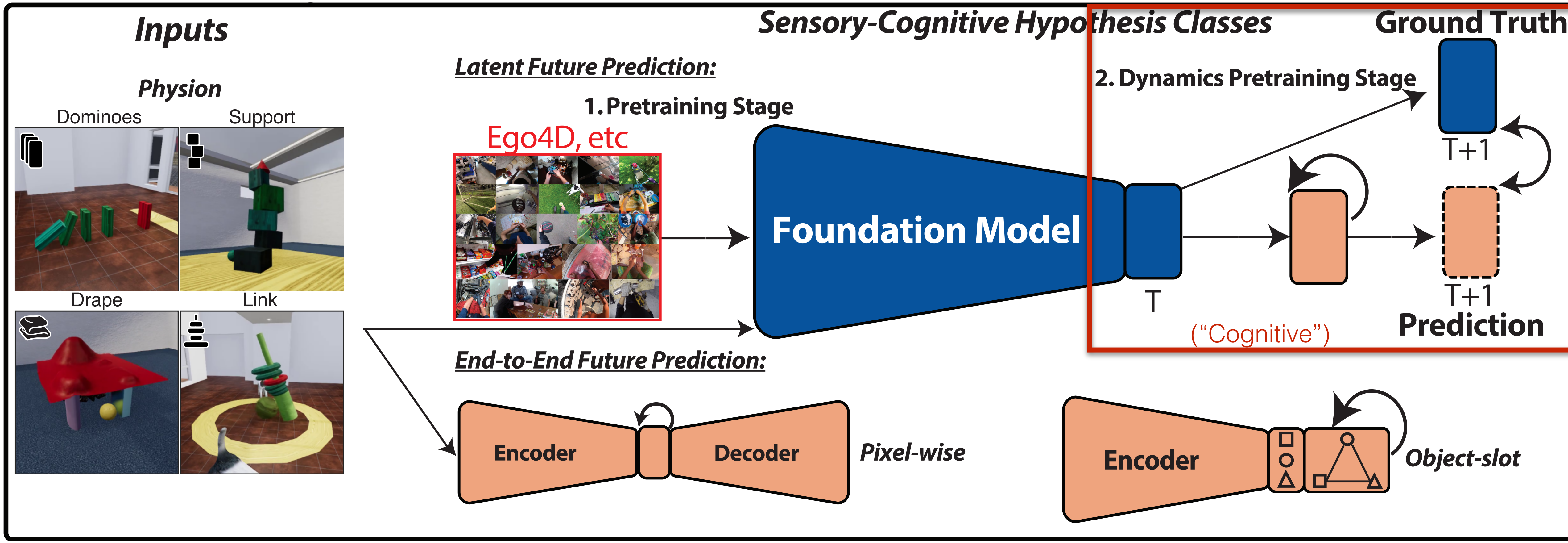


Hypothesis Class 3: Latent Future Prediction



Learn a partial, *implicit* representation of the physical world by performing a challenging vision task ("foundation model")

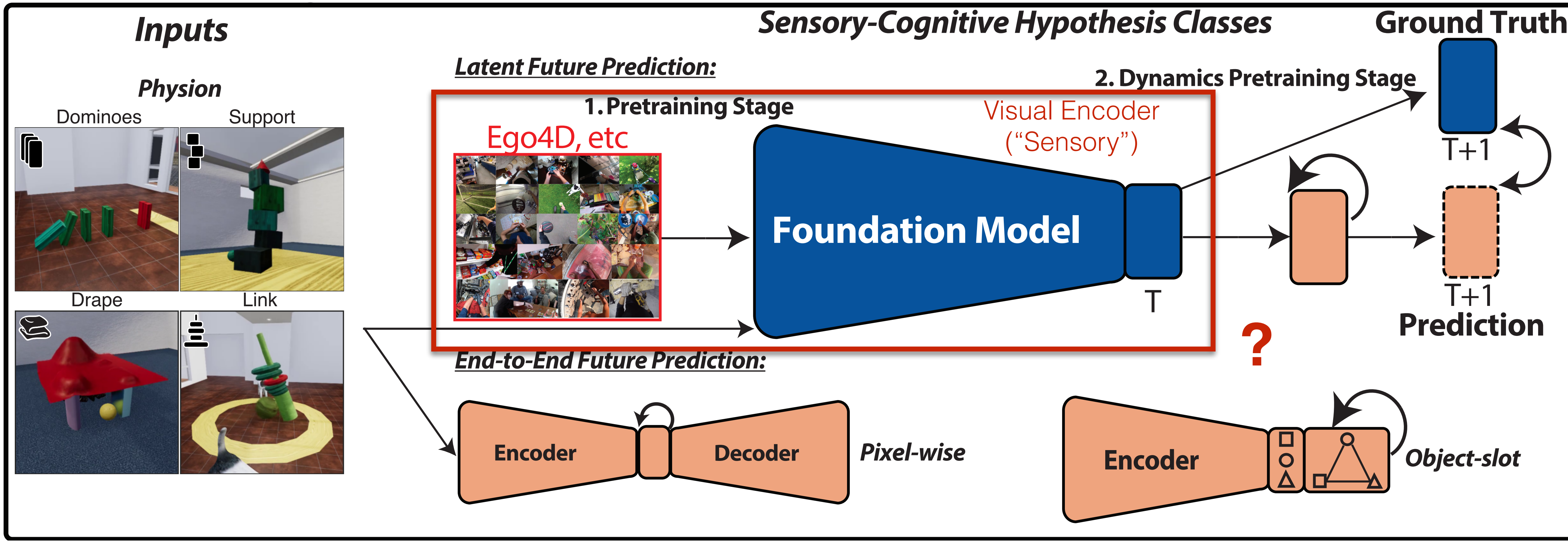
Hypothesis Class 3: Latent Future Prediction



Learn a partial, *implicit* representation of the physical world by performing a challenging vision task ("foundation model")

Leverage these dynamics to do explicit future prediction

Hypothesis Class 3: Foundation Models

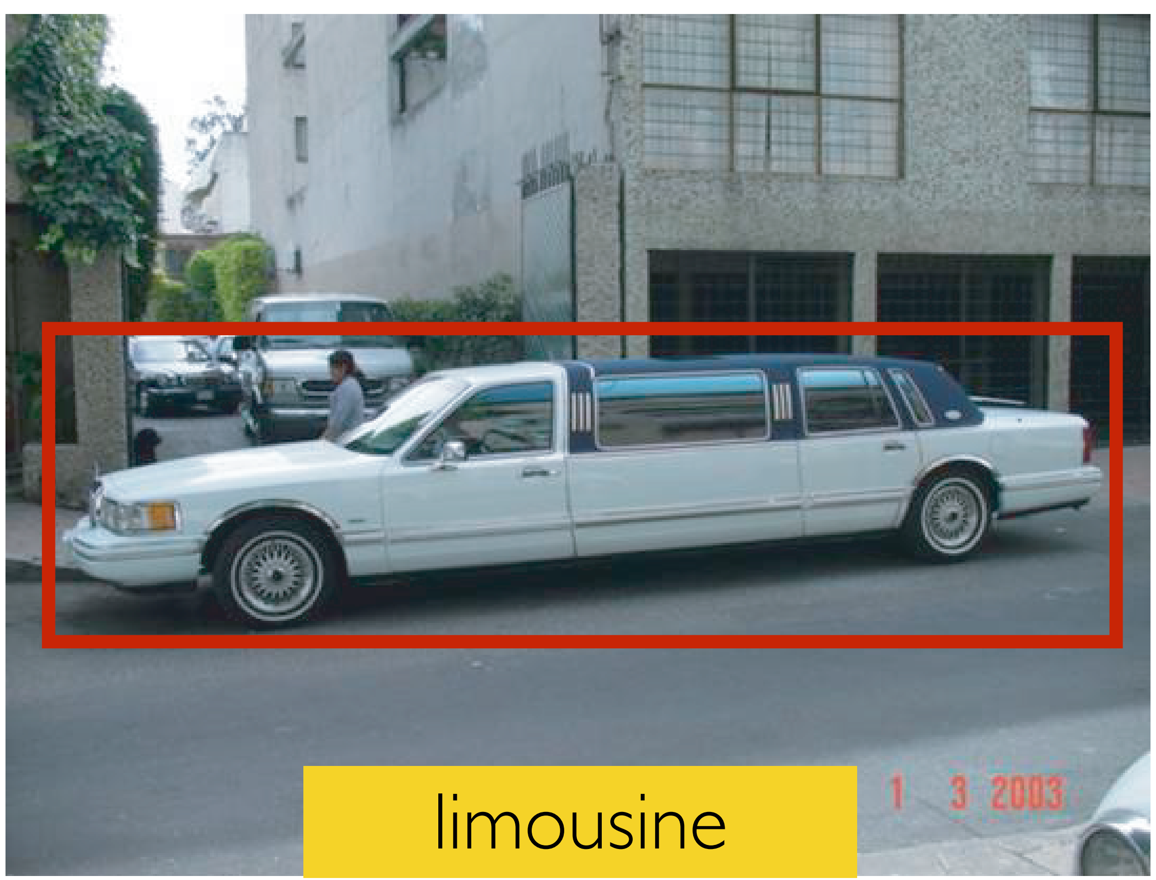


Learn a partial, *implicit* representation of the physical world by performing a challenging vision task ("foundation model")

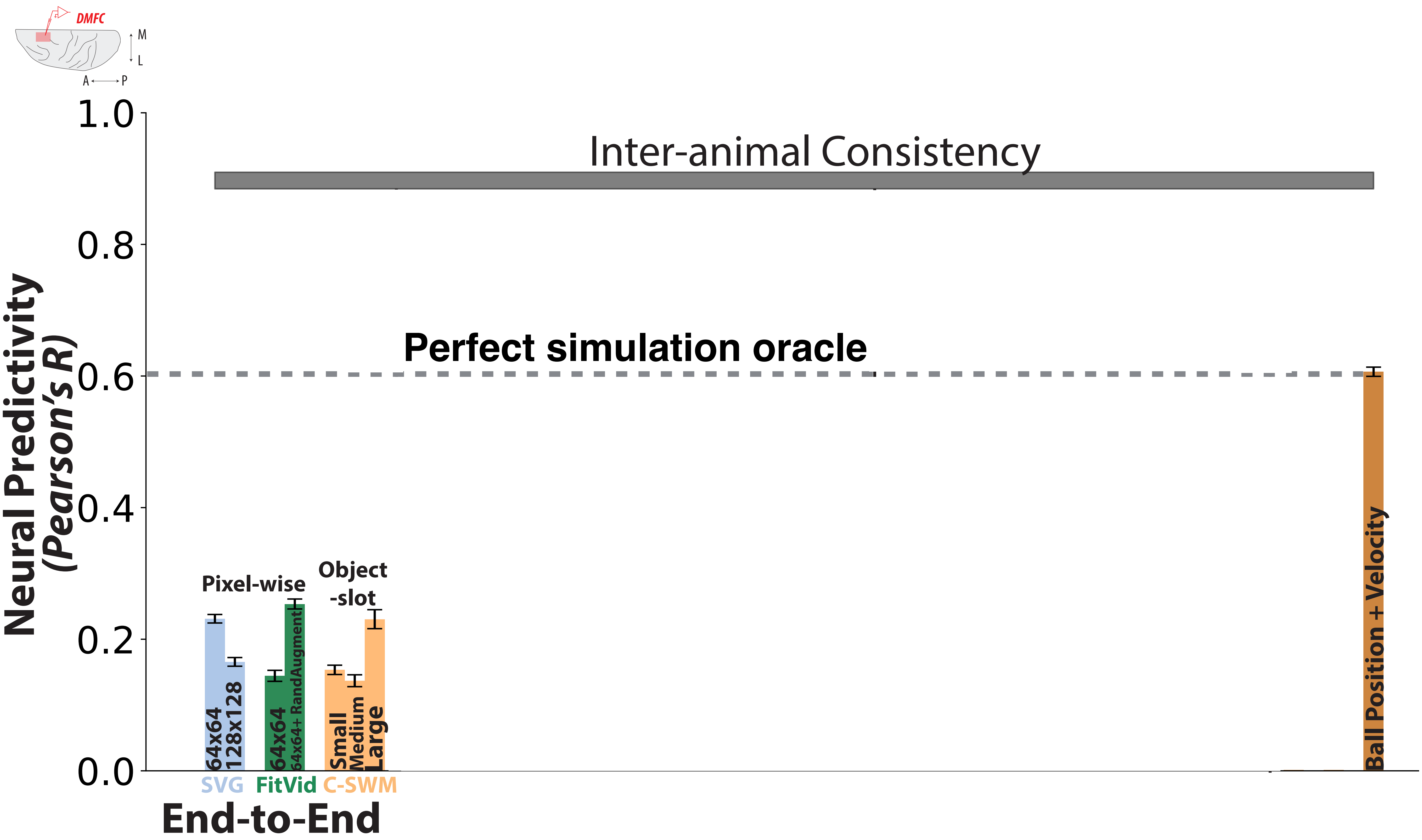
What vision task?

Leverage these dynamics to do explicit future prediction

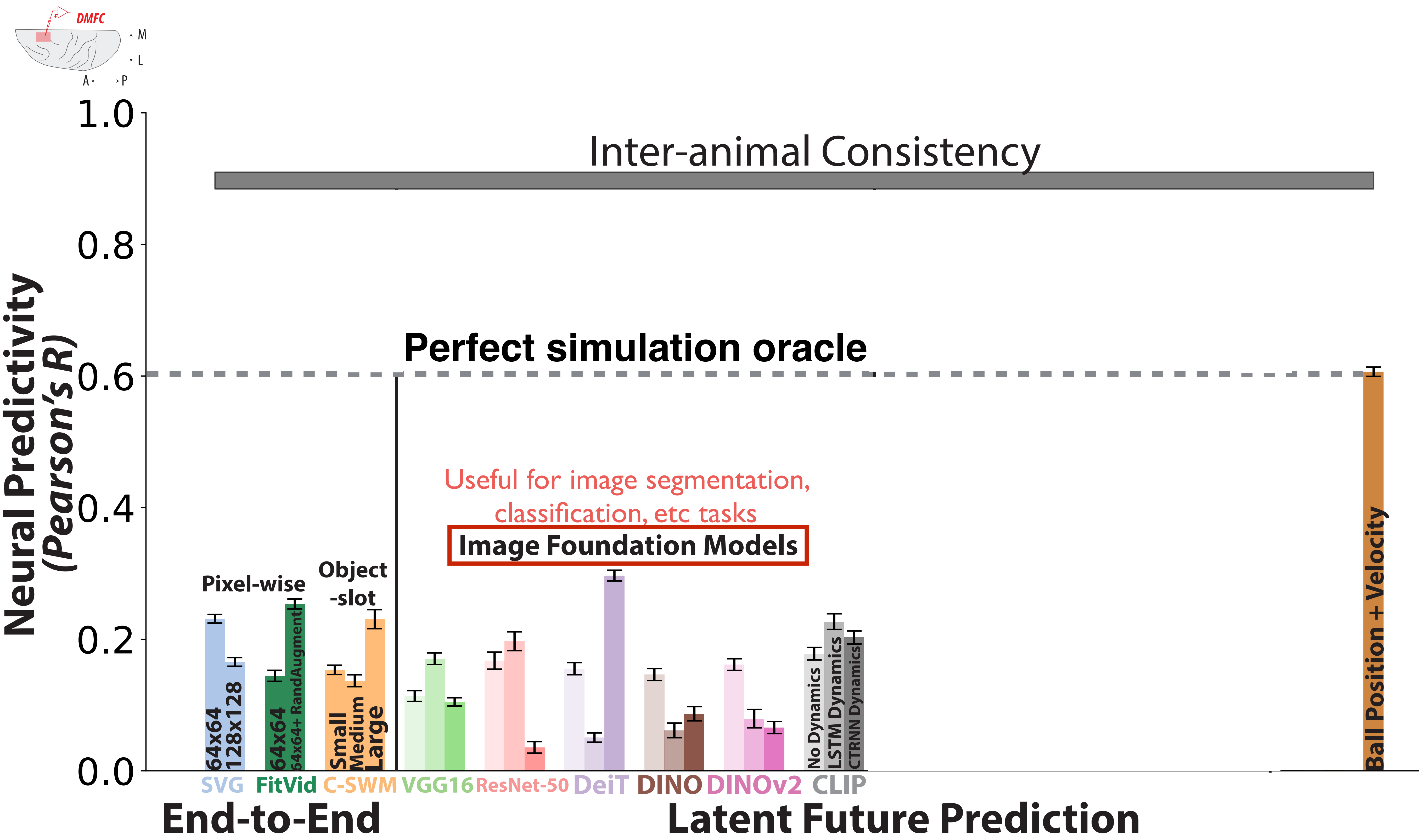
Hypothesis Class 3: Static Image Foundation Models



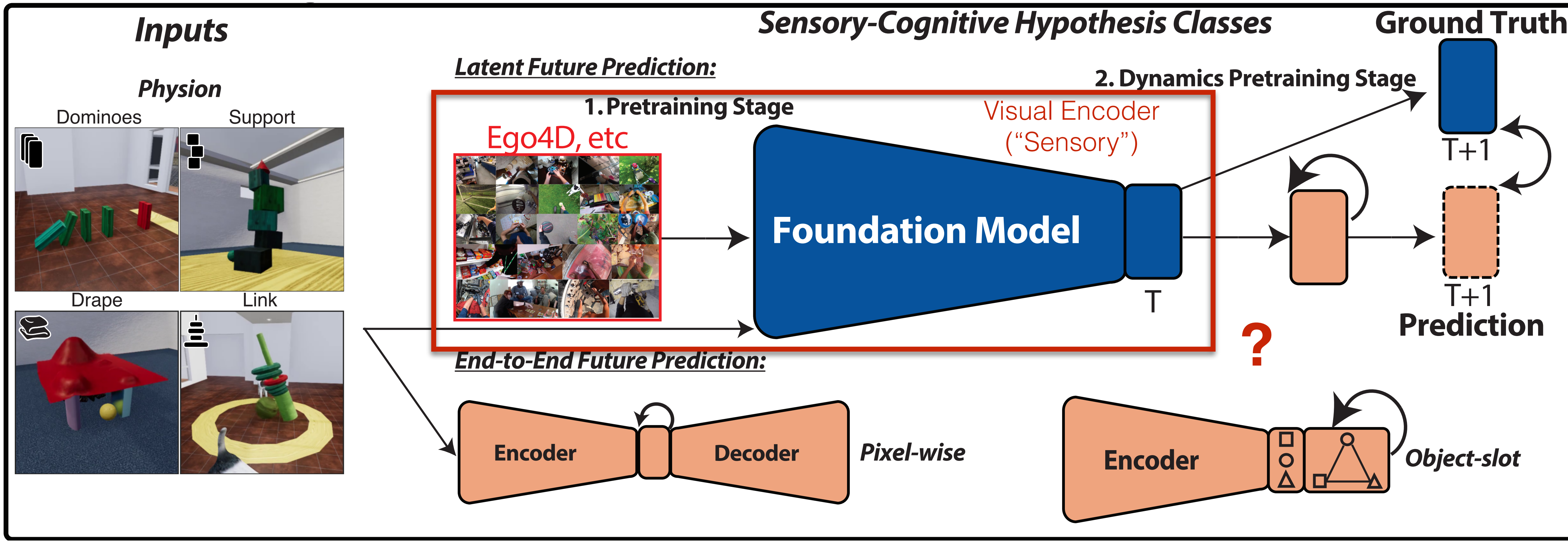
Object Slot Future Prediction Poorly Predicts Neurons



Static Image Foundation Future Prediction Poorly Predicts Neurons



Hypothesis Class 3: Foundation Models



Learn a partial, *implicit* representation of the physical world by performing a challenging vision task ("foundation model")

What vision task?

We do far more than engage with static images!

Leverage these dynamics to do explicit future prediction

Hypothesis Class 3: Video Foundation Models

Ego4D: everyday activity around the world

$$\mathcal{L}_{contrastive} = -\sum_{b\in B}\log \frac{\overbrace{e^{\mathcal{S}(\mathbf{z}_i^b,\mathbf{z}_j^b)}}^{\text{attract}}}{e^{\mathcal{S}(\mathbf{z}_i^b,\mathbf{z}_j^b)} + \overbrace{e^{\mathcal{S}(\mathbf{z}_i^b,\mathbf{z}_k^b)}}^{\text{repel}} + \overbrace{e^{\mathcal{S}(\mathbf{z}_i^b,\widetilde{\mathbf{z}}_i^b)}}^{\text{repel}}}$$

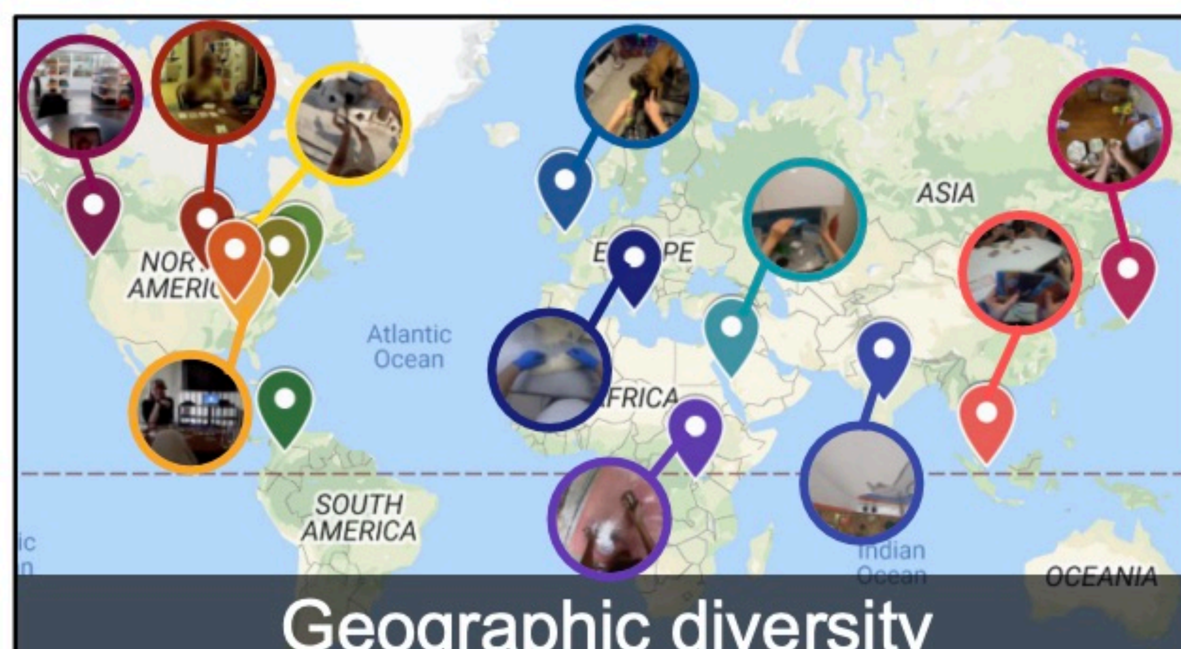
$$[I_i, I_{j>i}, I_{k>j}]^{1:B}$$

Ego4D: A massive-scale egocentric dataset

3,670 hours of in-the-wild daily life activity

931 participants from 74 worldwide locations

Multimodal: audio, 3D scans, IMU, stereo, multi-camera

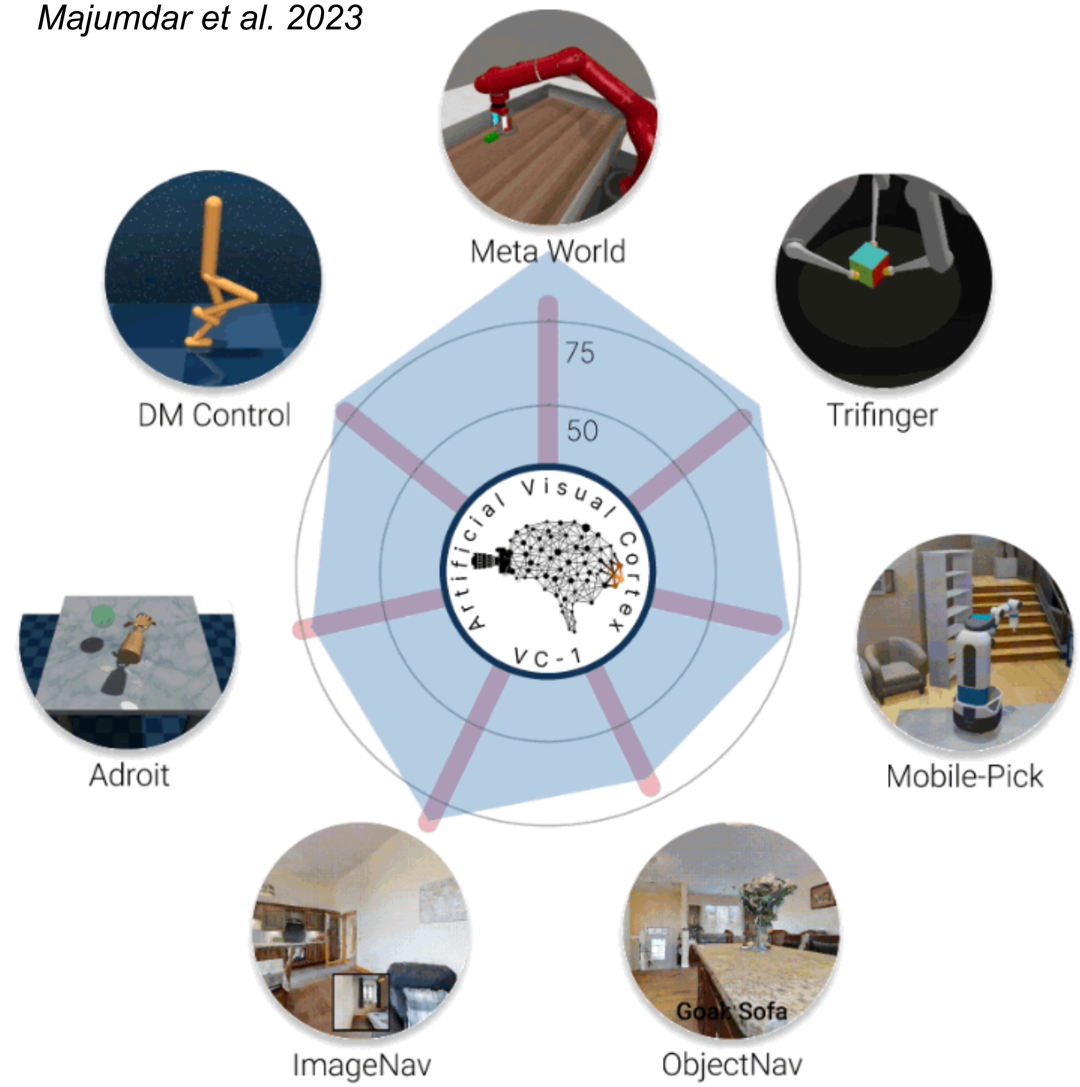


Grauman et al. 2022

Hypothesis Class 3: Video Foundation Models

Ego4D: everyday activity around the world

Majumdar et al. 2023

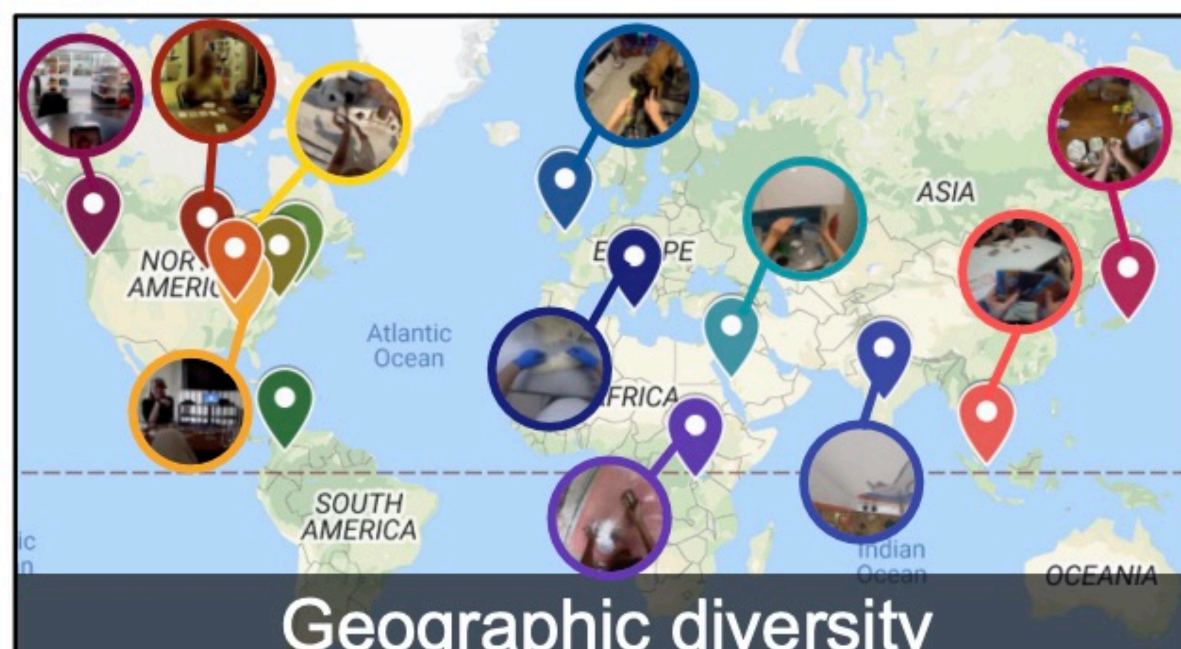


Ego4D: A massive-scale egocentric dataset

3,670 hours of in-the-wild daily life activity

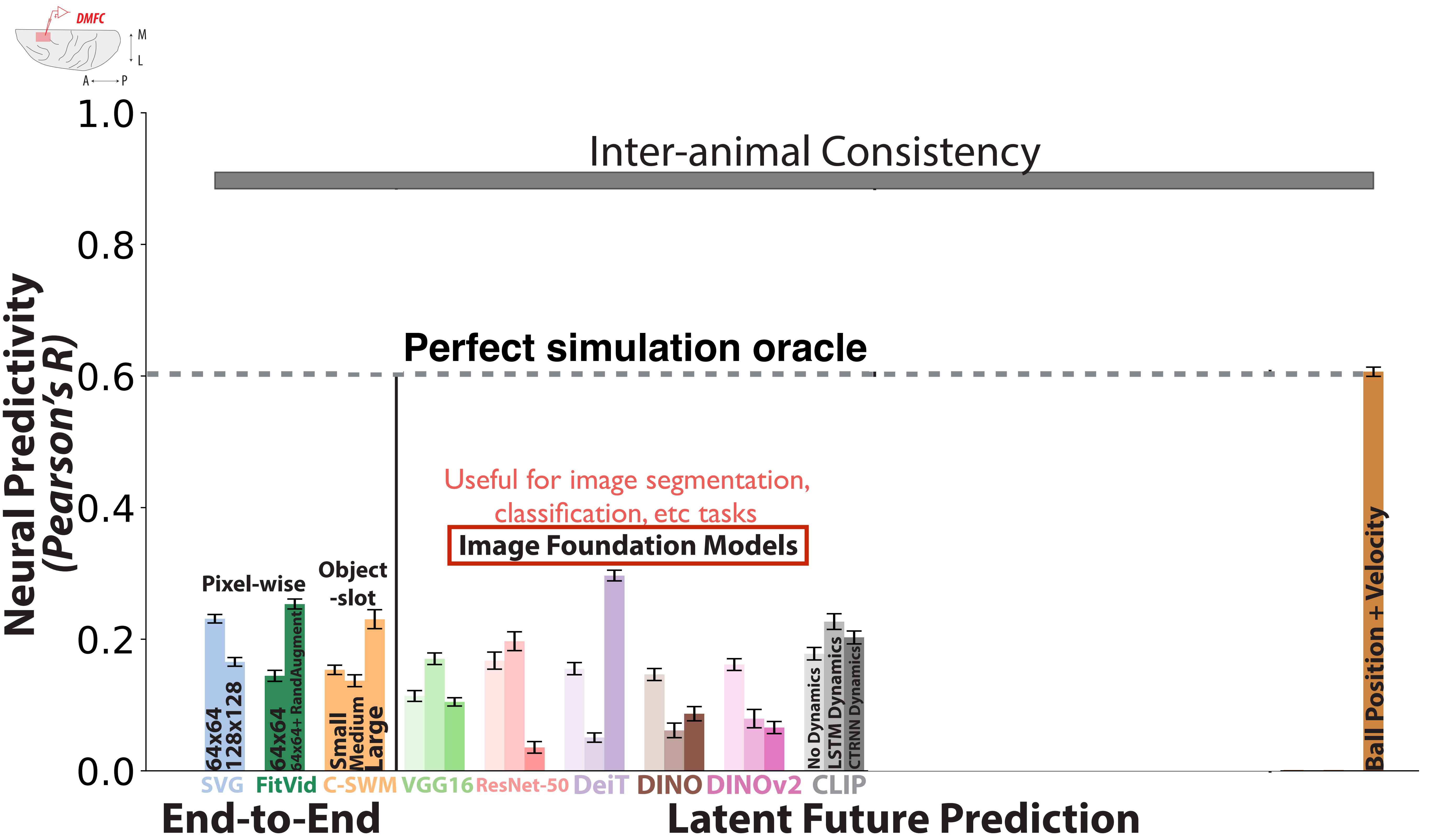
931 participants from 74 worldwide locations

Multimodal: audio, 3D scans, IMU, stereo, multi-camera

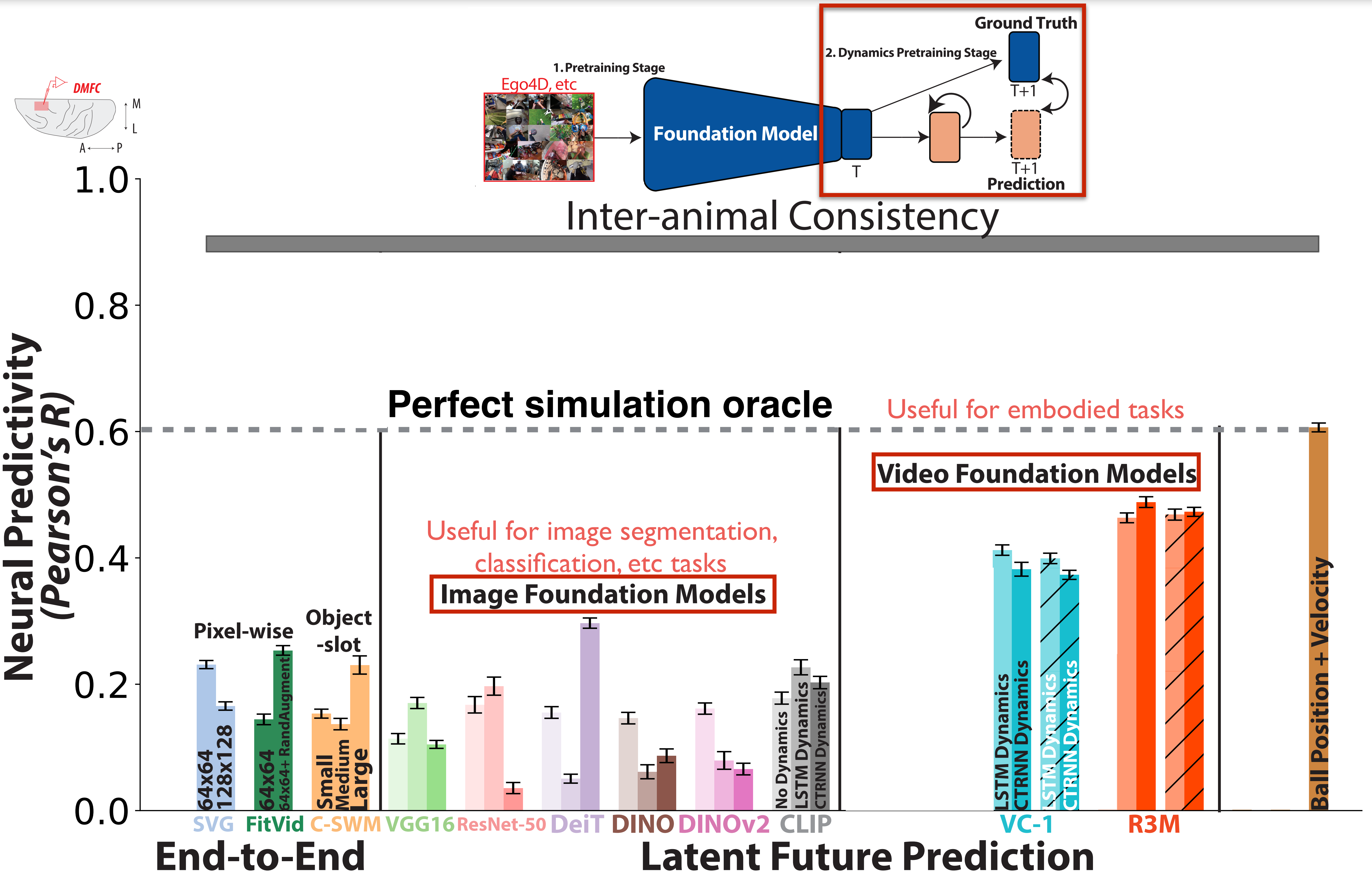


Grauman et al. 2022

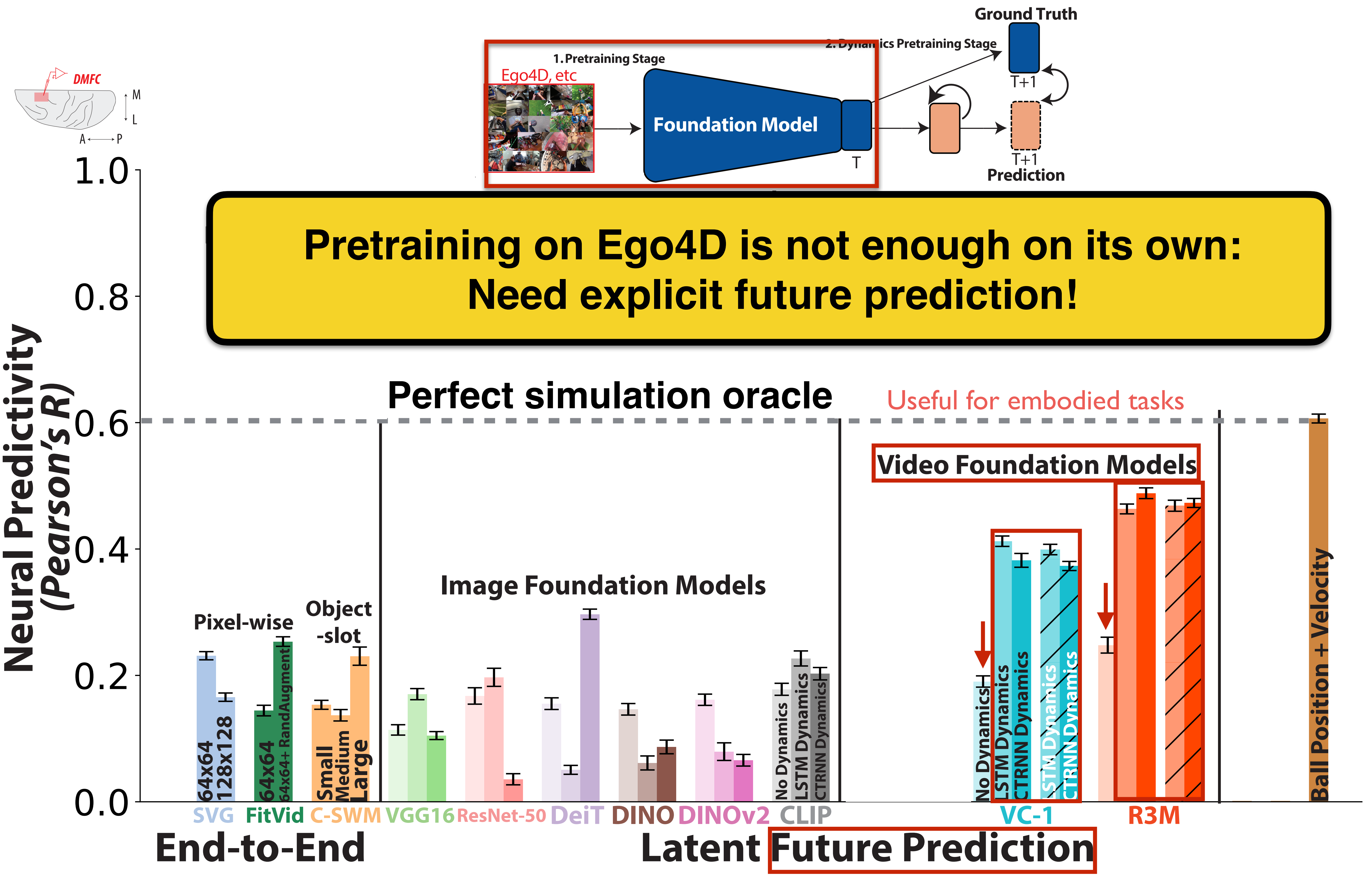
Static Image Foundation Future Prediction Poorly Predicts Neurons



Video Foundation Future Prediction Best Predict Neurons

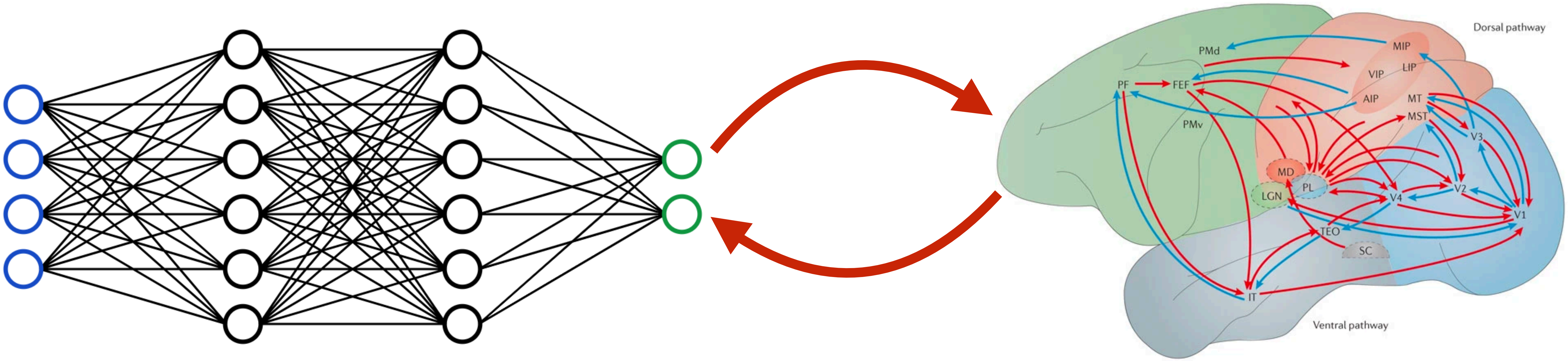


Video Foundation Future Prediction Best Predict Neurons



Task-Optimized Modeling

Design ML Algorithms Optimized to Perform Organism's Behavior under Organism's Constraints



Yields:

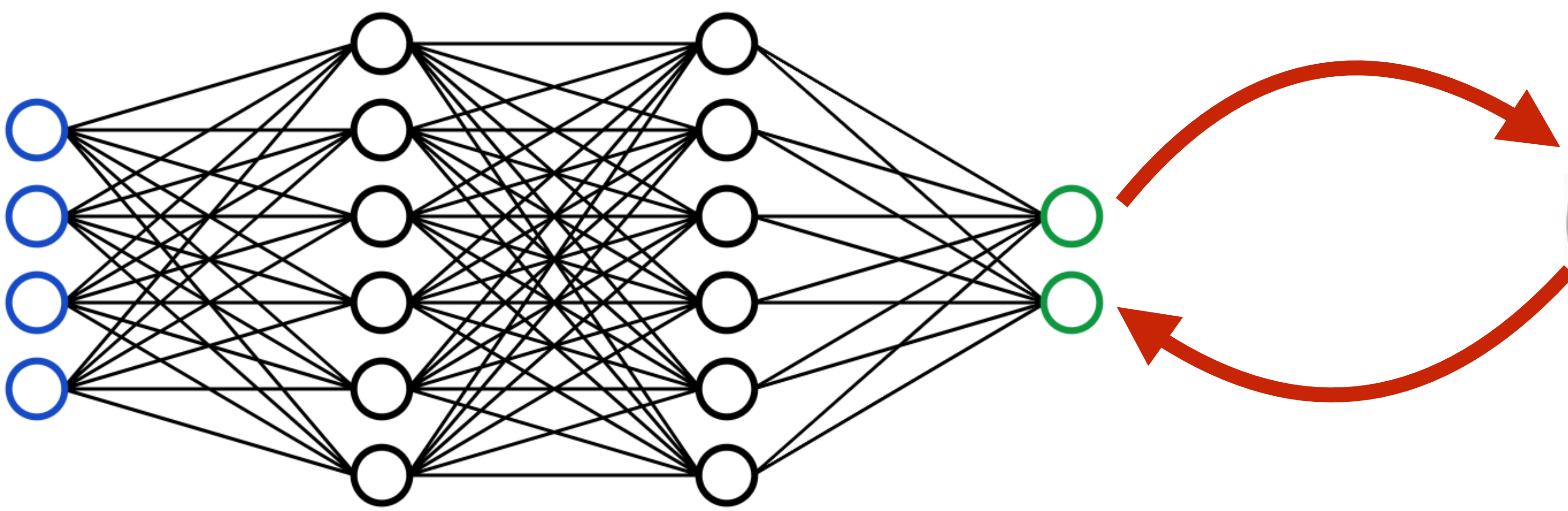
Quantitatively Accurate & Practically Useful Brain Models

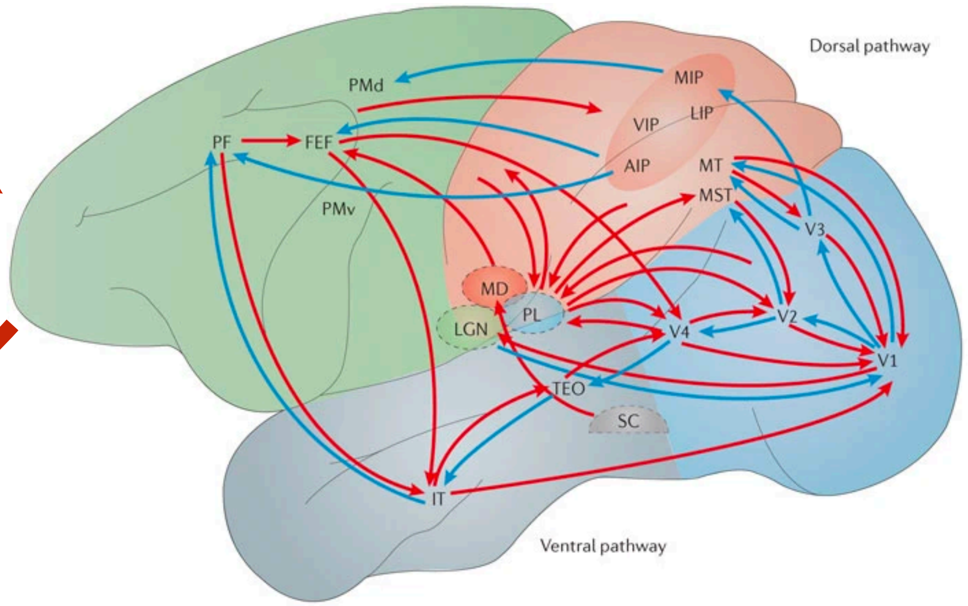
AND

Principles of *Why* Neural Responses Are As They Are

Task-Optimized Modeling

Design ML Algorithms Optimized to Perform Organism's Behavior under Organism's Constraints





Yields:

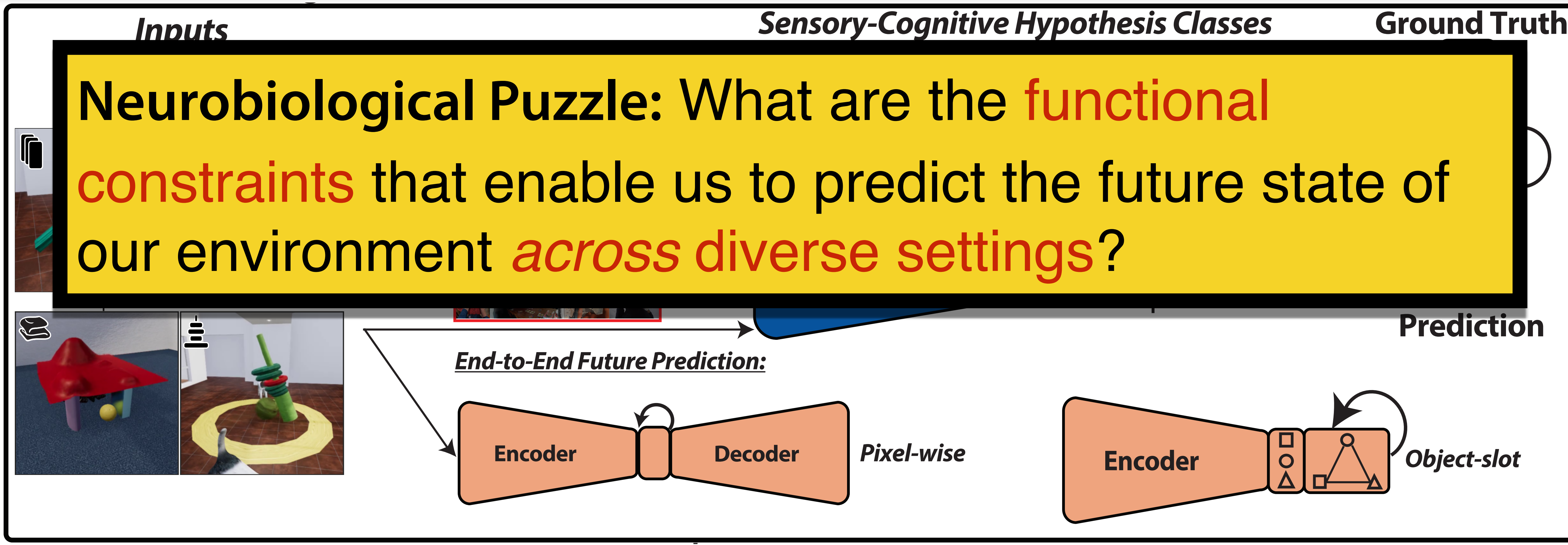
Quantitatively Accurate & Practically Useful Brain Models

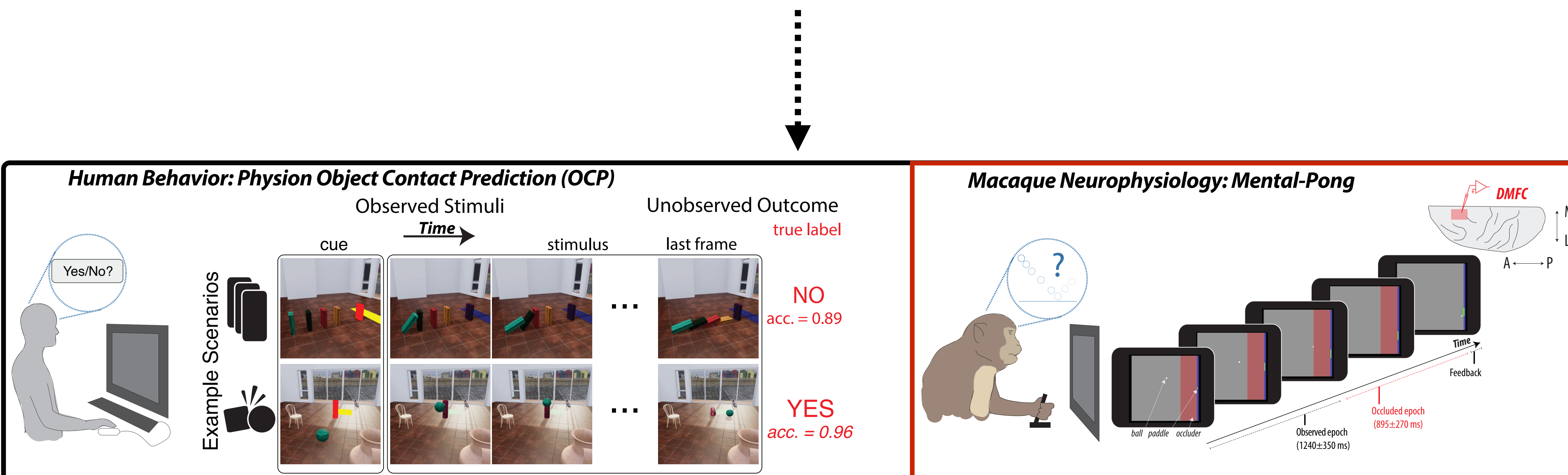
AND

Principles of *Why* Neural Responses Are As They Are

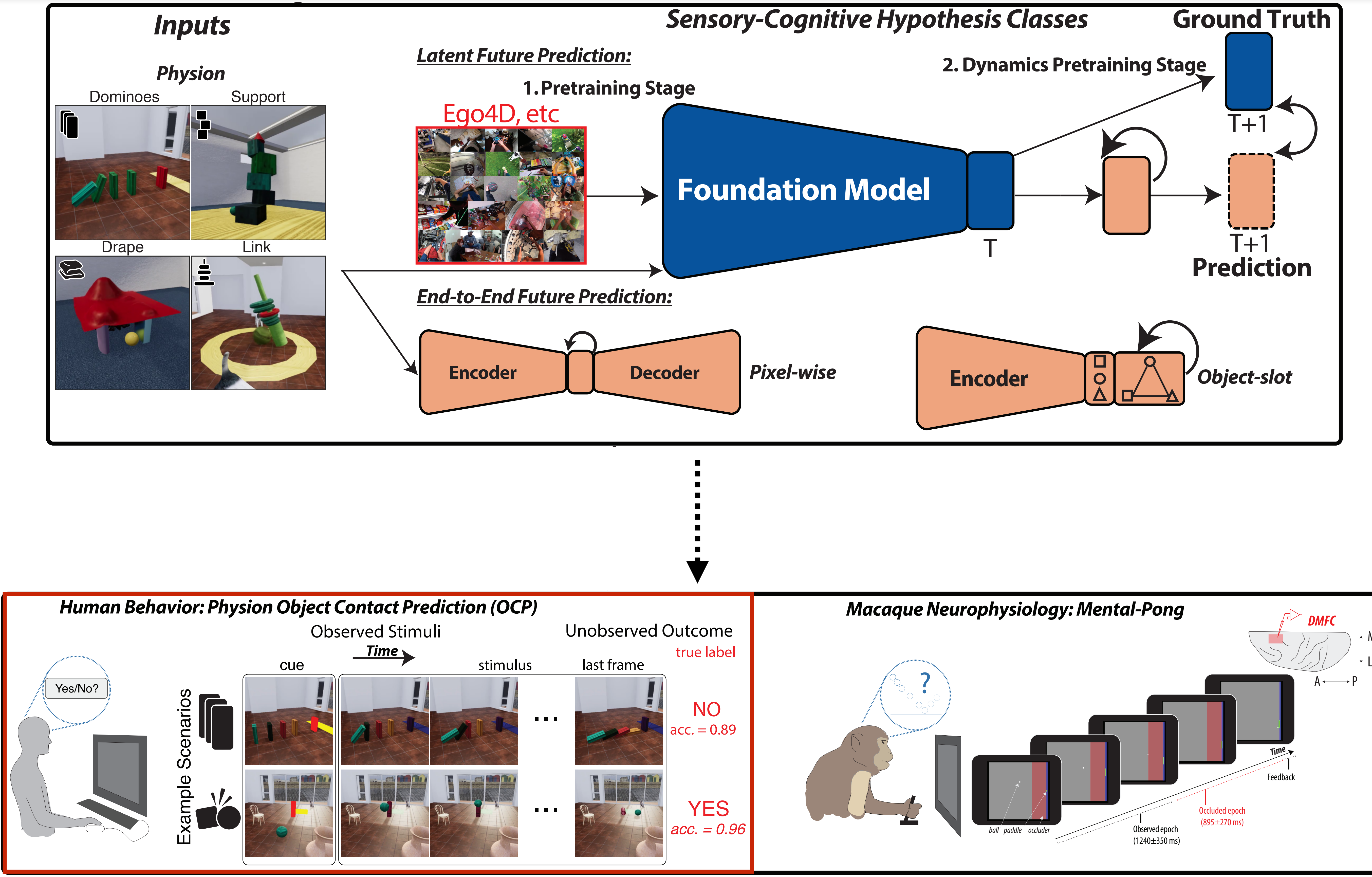
Macaque Neurophysiology: Mental Pong

Neurobiological Puzzle: What are the functional constraints that enable us to predict the future state of our environment *across* diverse settings?





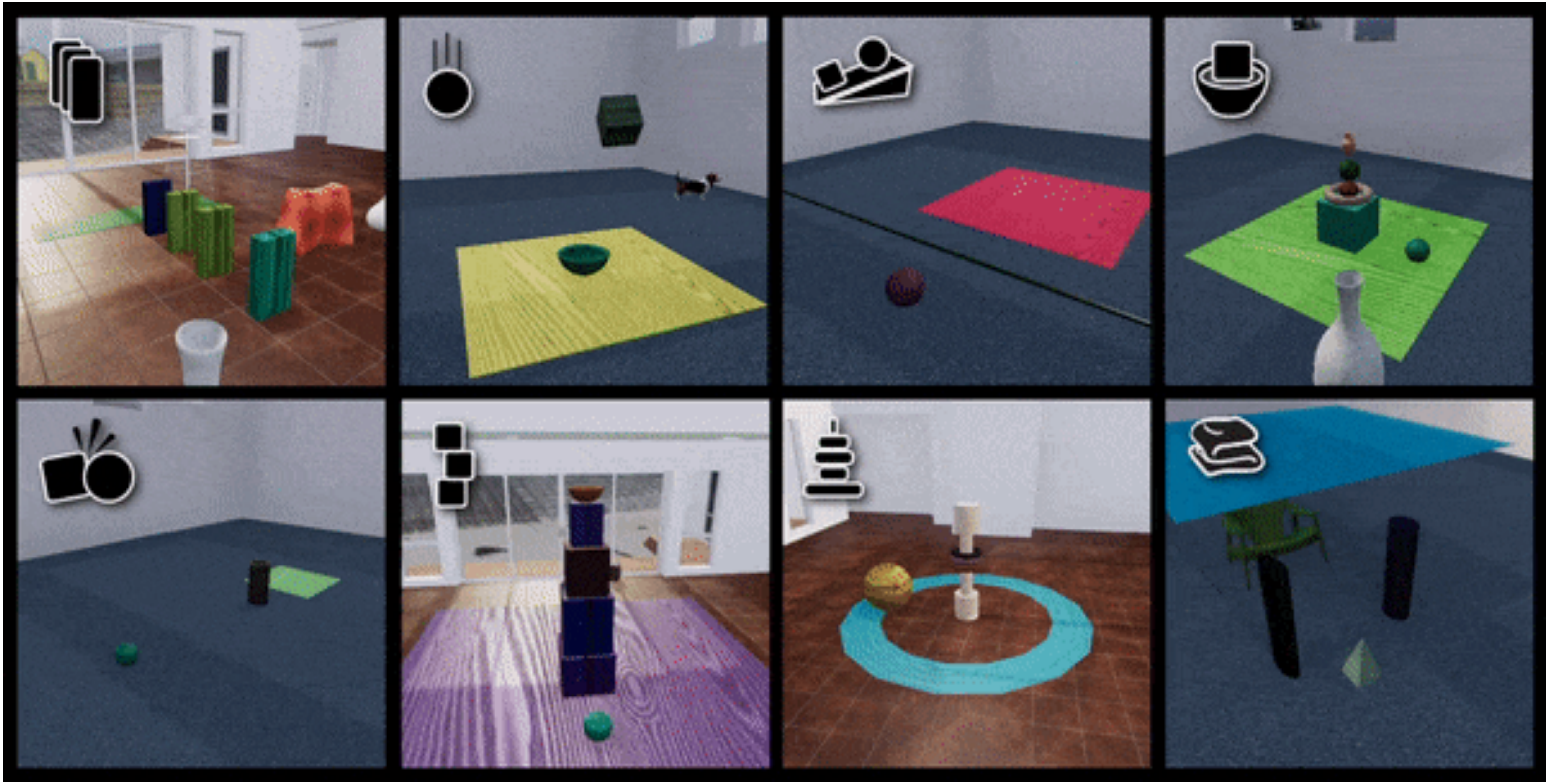
Human Behavior: Object Contact Prediction



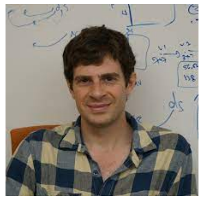
Object Contact Prediction Environment

Physion/ThreeD World (TDW)

Bear et al. 2021



Focus on everyday physical understanding



Daniel Bear

Joshua Tenenbaum

Daniel Yamins

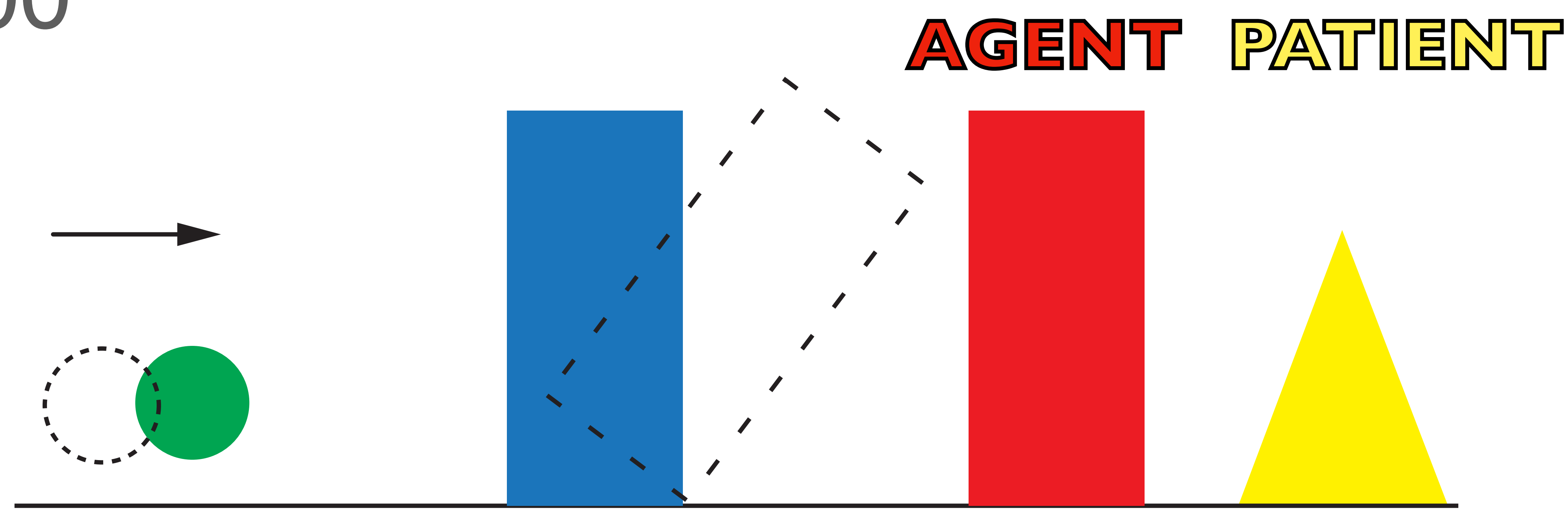
Judith Fan

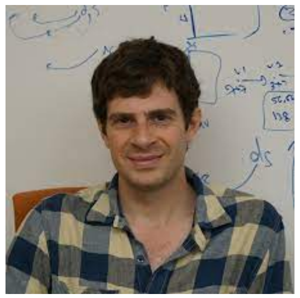
Human Behavior: Object Contact Prediction

Bear et al. 2021

"Will the agent object contact the patient object?"

n=100




Daniel Bear

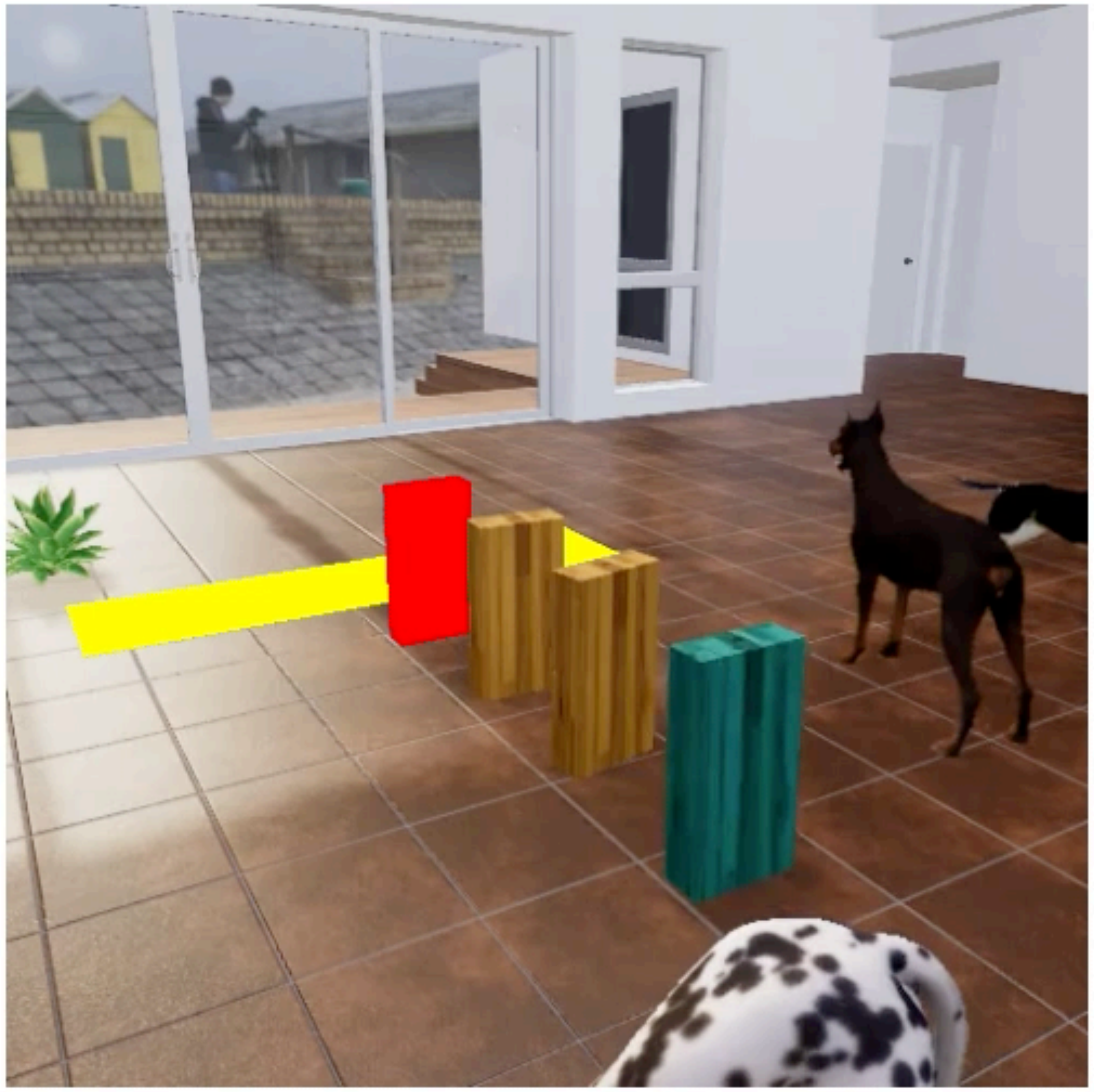

Joshua Tenenbaum


Daniel Yamins


Judith Fan

Completion Progress

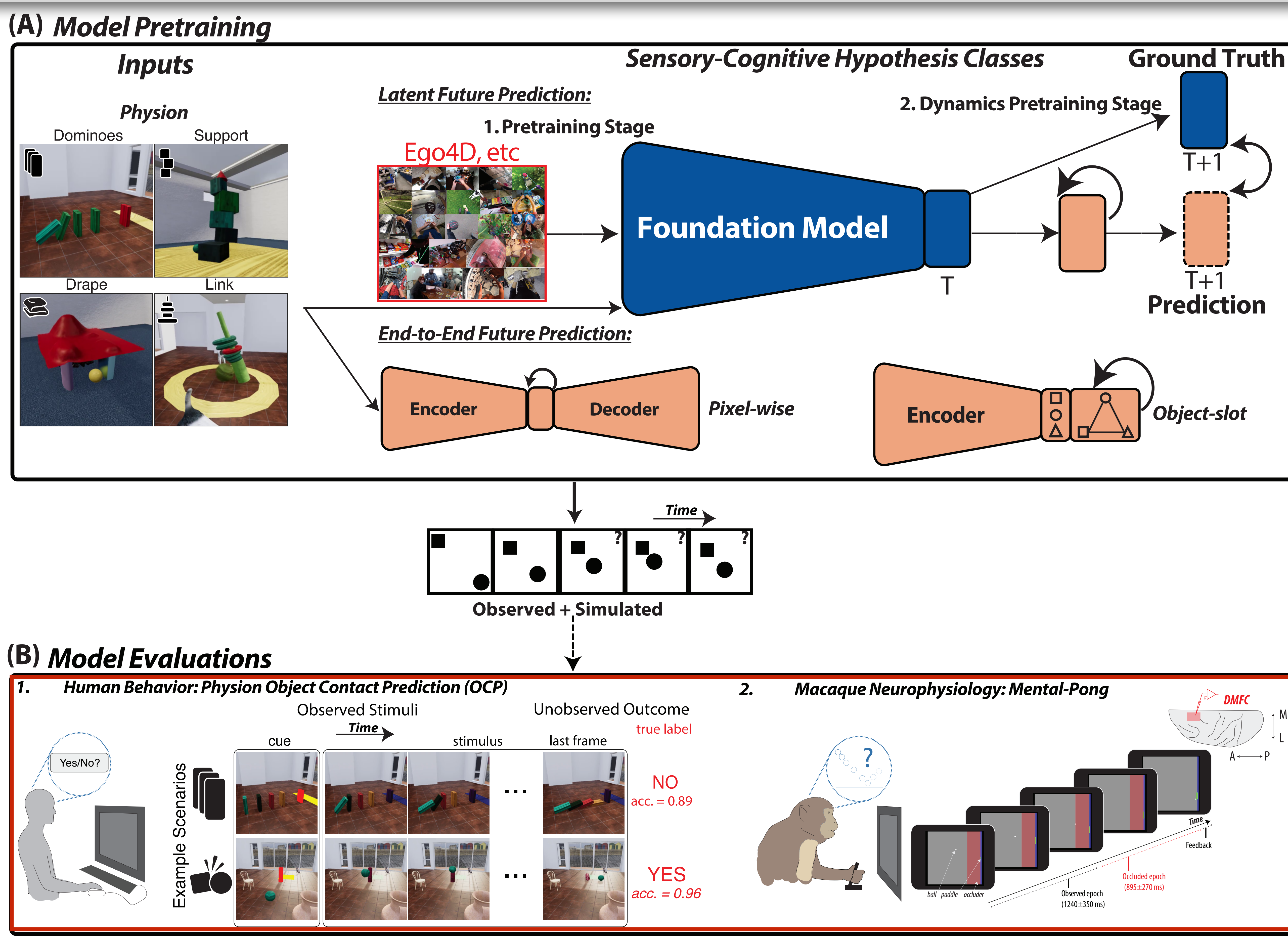
Bear et al. 2021



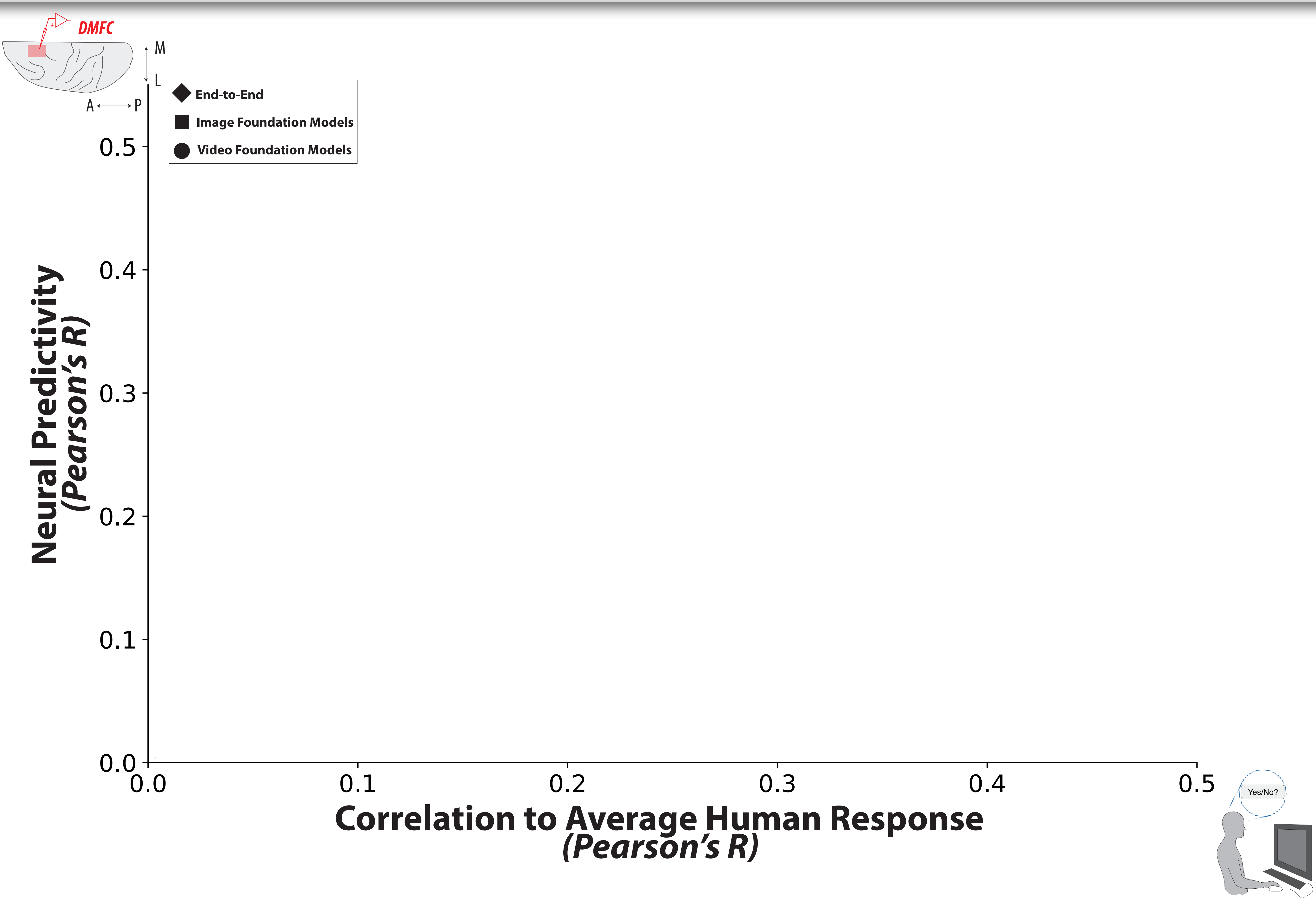
YES NO

Is the red object going to hit the yellow area?

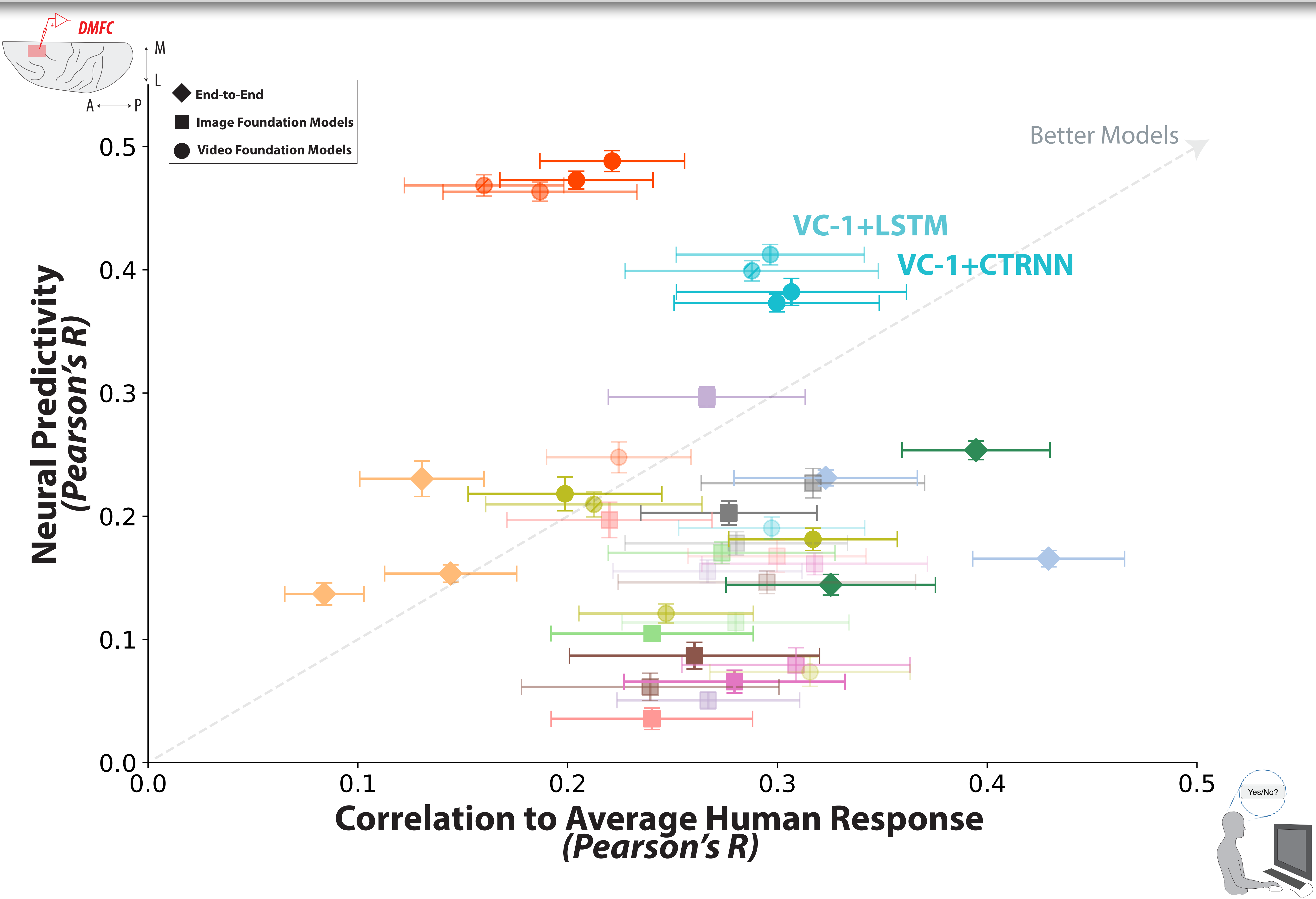
Model Evaluations: What About Both Metrics?



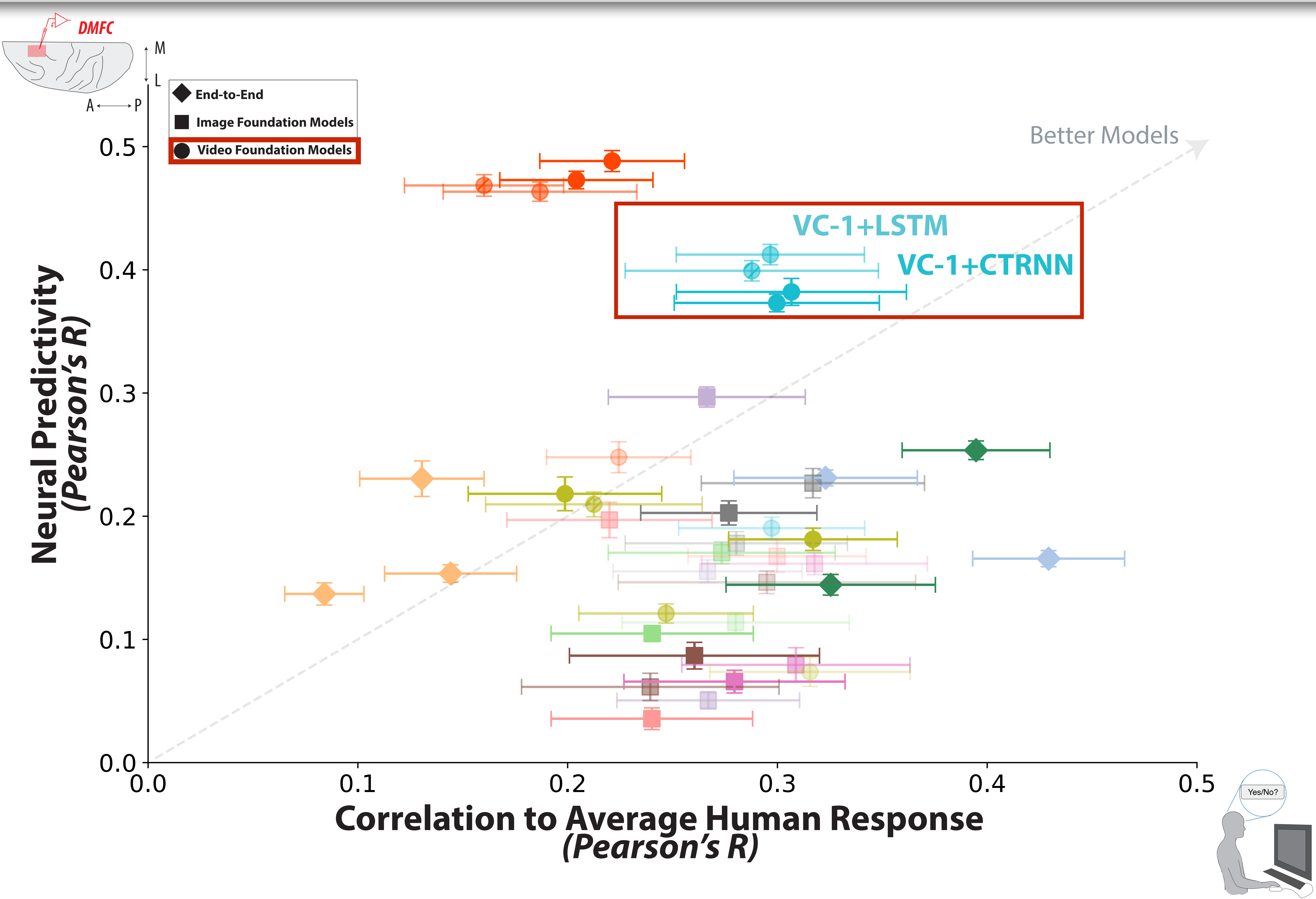
Comparing to *Both* Human Behavioral and Neural Response Patterns



Dynamically-Equipped Video Foundation Models Can Match Both

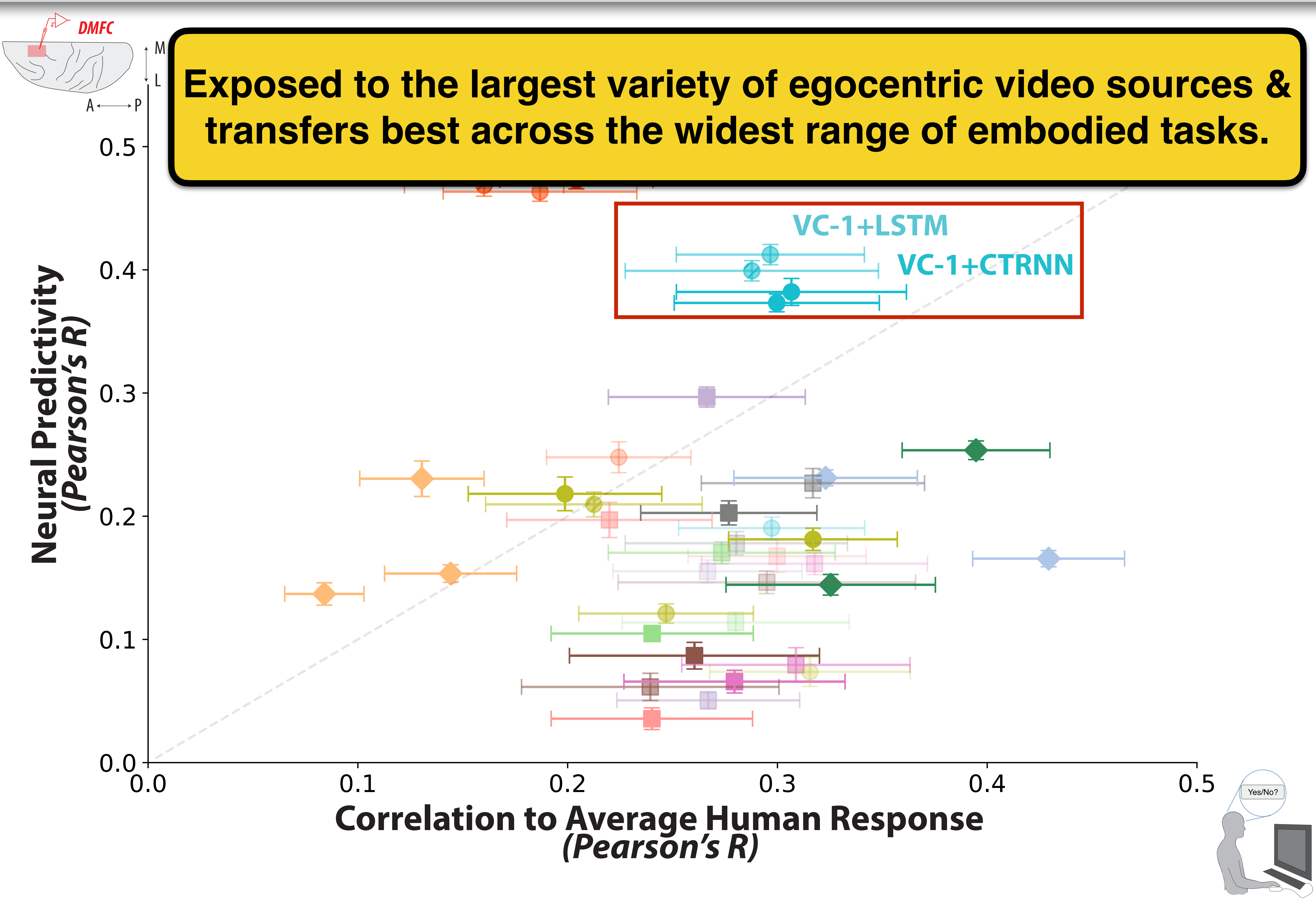


Dynamically-Equipped Video Foundation Models Can Match Both

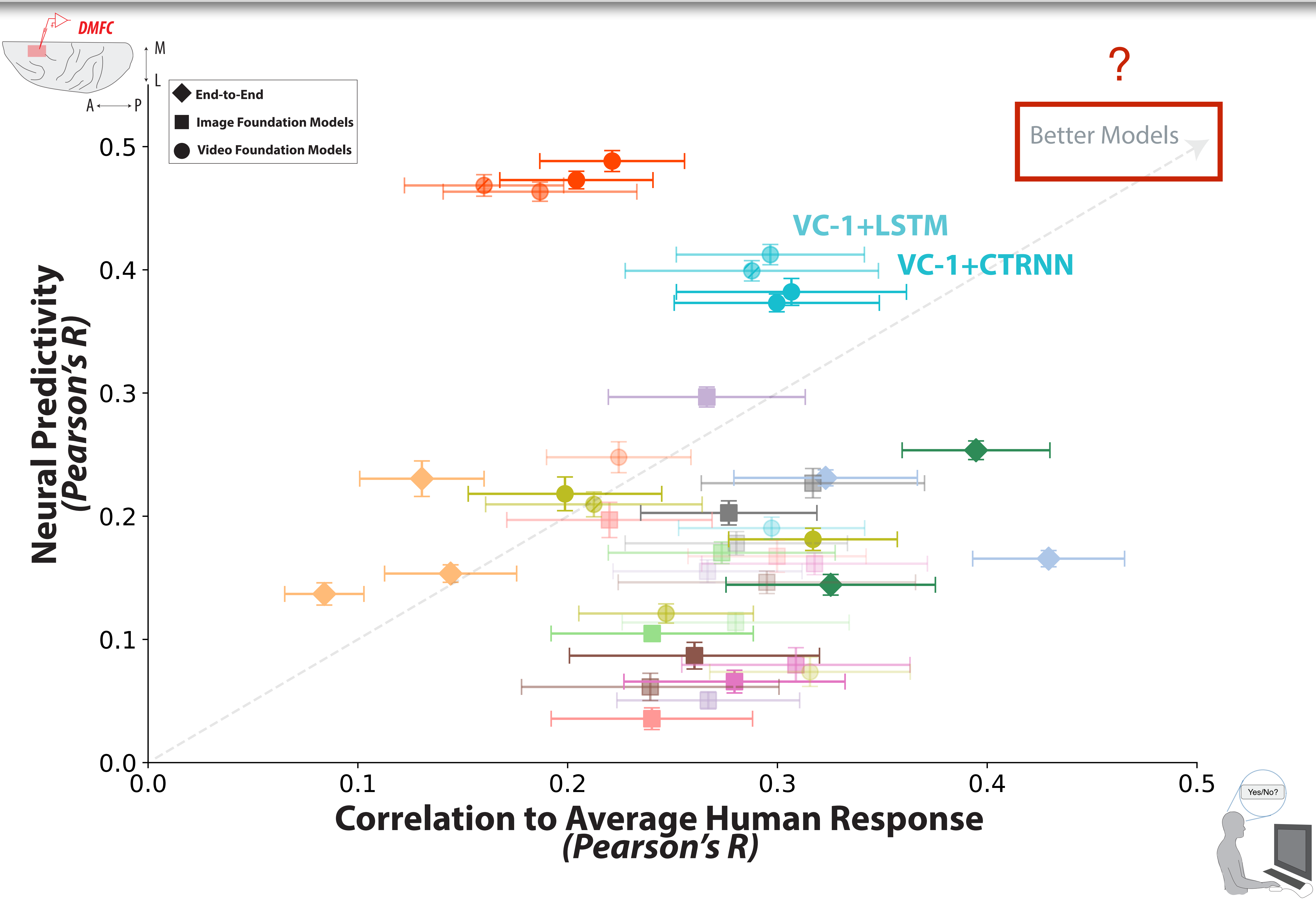


Dynamically-Equipped Video Foundation Models Can Match Both

Exposed to the largest variety of egocentric video sources & transfers best across the widest range of embodied tasks.



Towards More Robust Future Inference



Takeaways

L *= learning rule*

"Natural selection + plasticity"

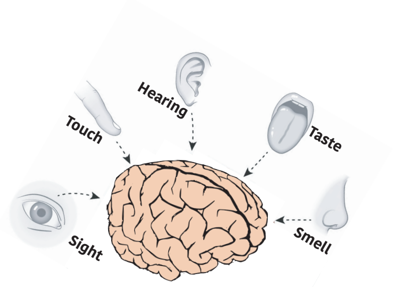
Backpropagation

T *= task loss*

"Ecological niche/ behavior"

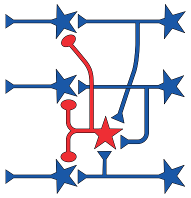
Neurobiological Puzzle:

What are the functional constraints that enable us to predict the future state of our environment *across* diverse settings?



"Environment"

D *= data stream*



"Circuit"

A *= architecture class*

Takeaways

L = *learning rule*

"Natural selection + plasticity"

Backpropagation

T = *task loss*

"Ecological niche/ behavior"

latent future prediction

<u>Neurobiological Puzzle</u>:

What are the functional constraints that enable us to predict the future state of our environment *across* diverse settings?

egocentric videos

"Environment"

D = *data stream*

video foundation encoder + recurrent neural network

"Circuit"

A = *architecture class*

Takeaways

Neurobiological Puzzle:

What are the functional constraints that enable us to predict the future state of our environment *across* diverse settings?

Findings:

The brain's mental simulations crucially involve explicit future prediction of a visual scene description.

Takeaways

Neurobiological Puzzle:

What are the functional constraints that enable us to predict the future state of our environment *across* diverse settings?

Findings:

The brain's mental simulations crucially involve explicit future prediction of a visual scene description.

The scene description used by the brain is *not* fine-grained at the level of pixels, but must be "factorized" by the brain somehow.

Takeaways

Neurobiological Puzzle:

What are the functional constraints that enable us to predict the future state of our environment *across* diverse settings?

Findings:

The brain's mental simulations crucially involve explicit future prediction of a visual scene description.

The scene description used by the brain is *not* fine-grained at the level of pixels, but must be "factorized" by the brain somehow.

This factorization is strongly constrained. It does *not* appear to represent fixed object slots, but rather a critical component is for it to enable a wide range of embodied abilities.

Outline

- Role of Recurrent Processing During Object Recognition
- Visually-Grounded Mental Simulation
- Vision and Navigation in Rodents

- Future Directions

Mouse Visual Cortex as a Domain-General, Limited Resource System

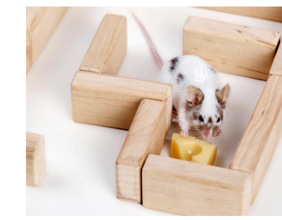
L = *learning rule*

"Natural selection + plasticity"

Backpropagation

T = *task loss*

"Ecological niche/ behavior"

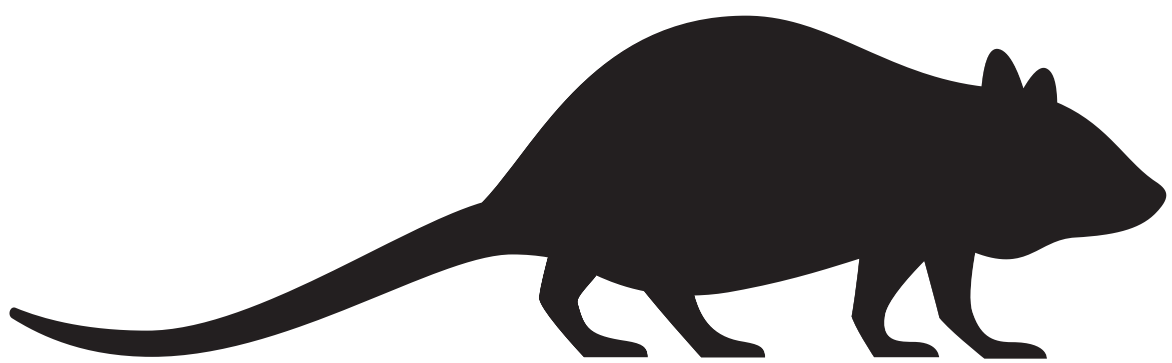


A. Nayebi*, N.C.L. Kong*, C. Zhuang, J.L. Gardner, A.M. Norcia, D.L.K. Yamins

Mouse visual cortex as a limited resource system that self-learns an ecologically-general representation

PLOS Computational Biology 2023

Nathan C.L. Kong*

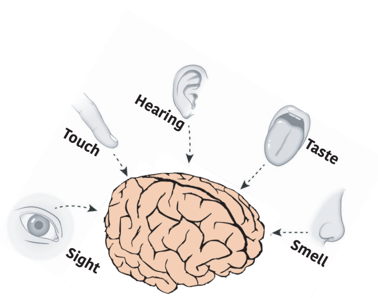


Chengxu Zhuang

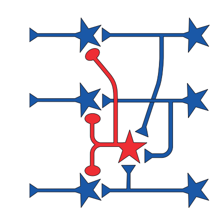
Justin L. Gardner

Anthony M. Norcia

Daniel Yamins



"Environment"



"Circuit"

D = *data stream*

A = *architecture class*

Heterogeneity in Rodent Medial Entorhinal Cortex

L *= learning rule*

"Natural selection + plasticity"

Backpropagation

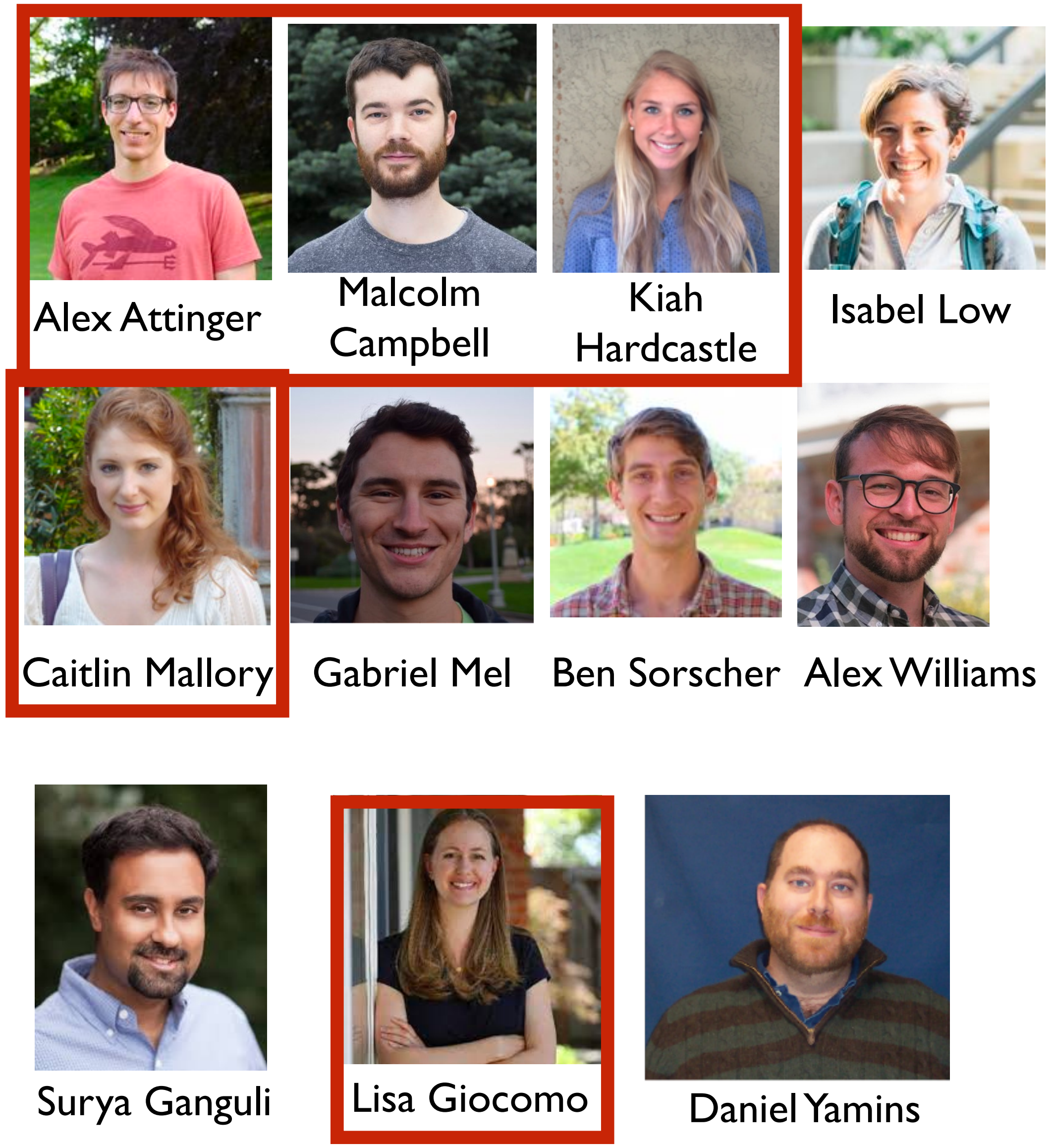
T *= task loss*

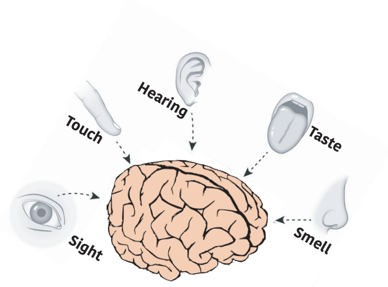
"Ecological niche/ behavior"

A. Nayebi, *et al.*

Explaining Heterogeneity in Medial Entorhinal Cortex with Task-Driven Neural Networks

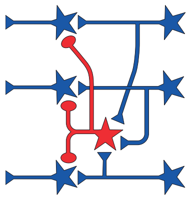
NeurIPS 2021 (spotlight)





"Environment"

D *= data stream*



"Circuit"

A *= architecture class*

CNNs as Models of Primate Object Recognition

L = *learning rule*

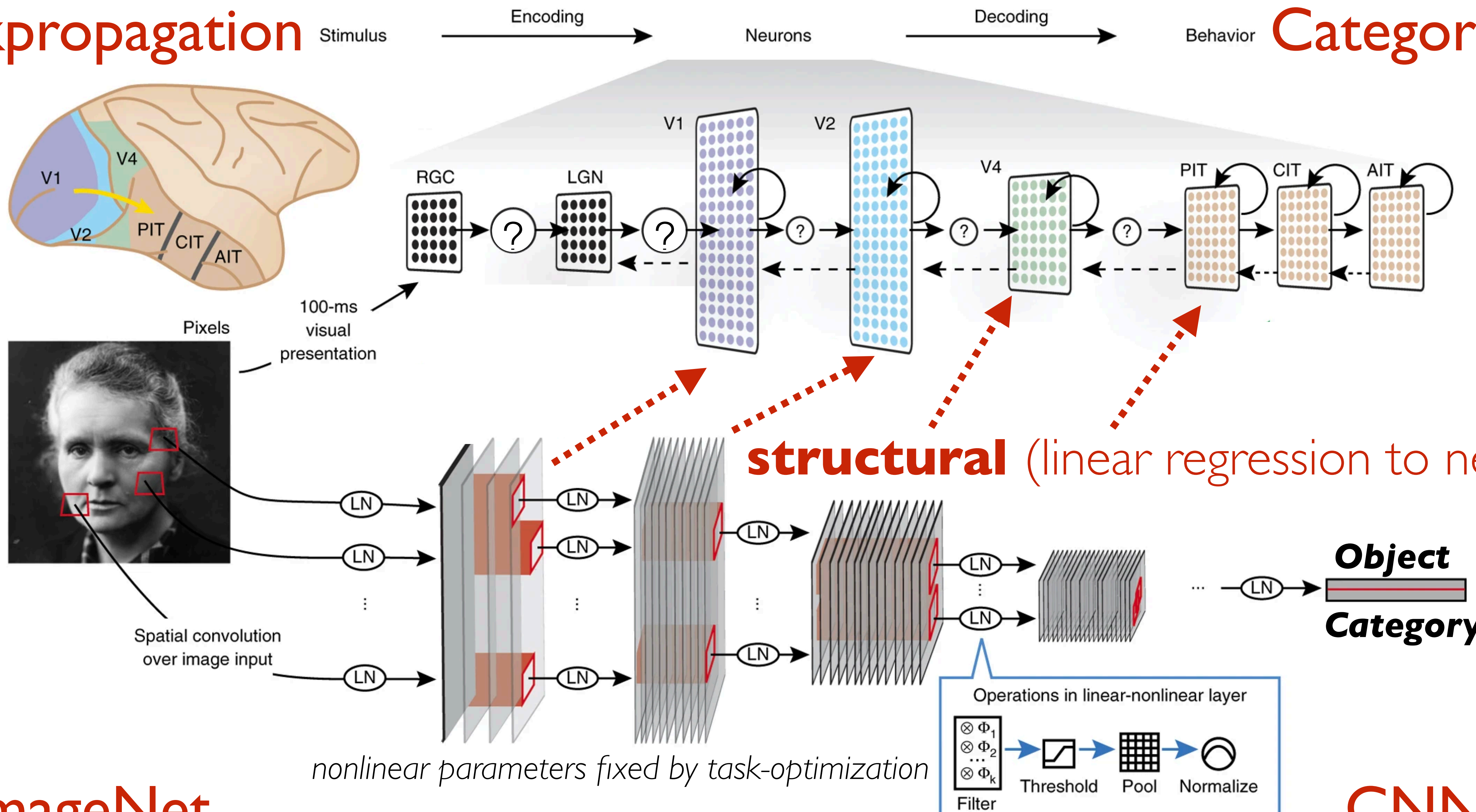
"Natural selection + plasticity"

Backpropagation

T = *task loss*

"Ecological niche/ behavior"

Categorization



structural (linear regression to neurons)

nonlinear parameters fixed by task-optimization

functional (performs behavior)

ImageNet

"Environment"

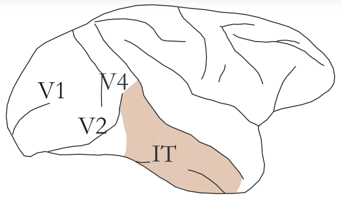
D = *data stream*

CNNs

"Circuit"

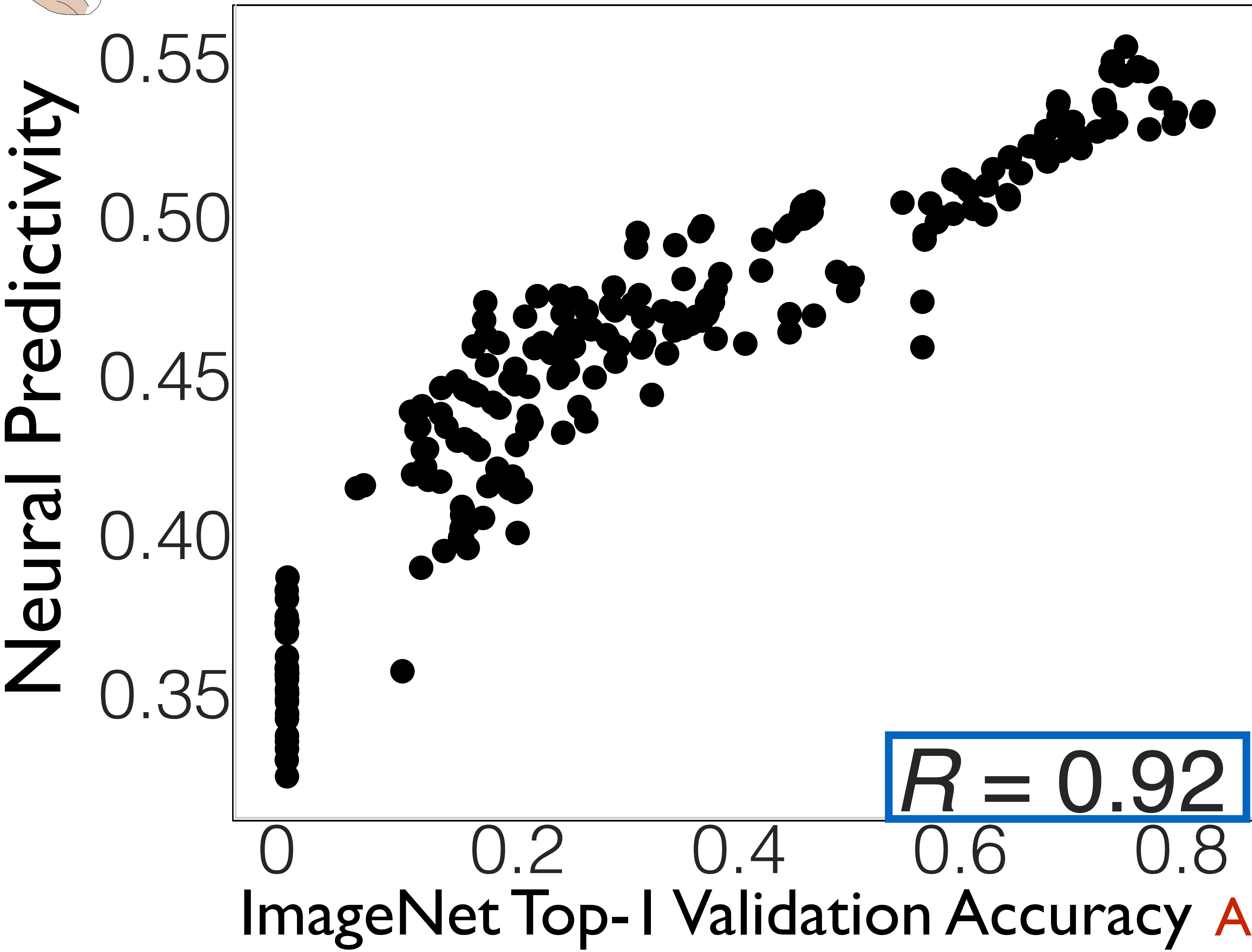
A = *architecture class*

Task Performance Correlated with Neural Predictivity



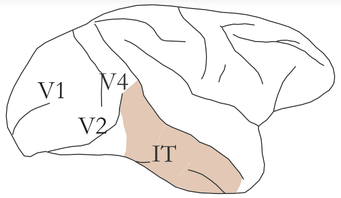
A Neuroscience Goal

Schrimpf, Kubilius* et al. 2018*

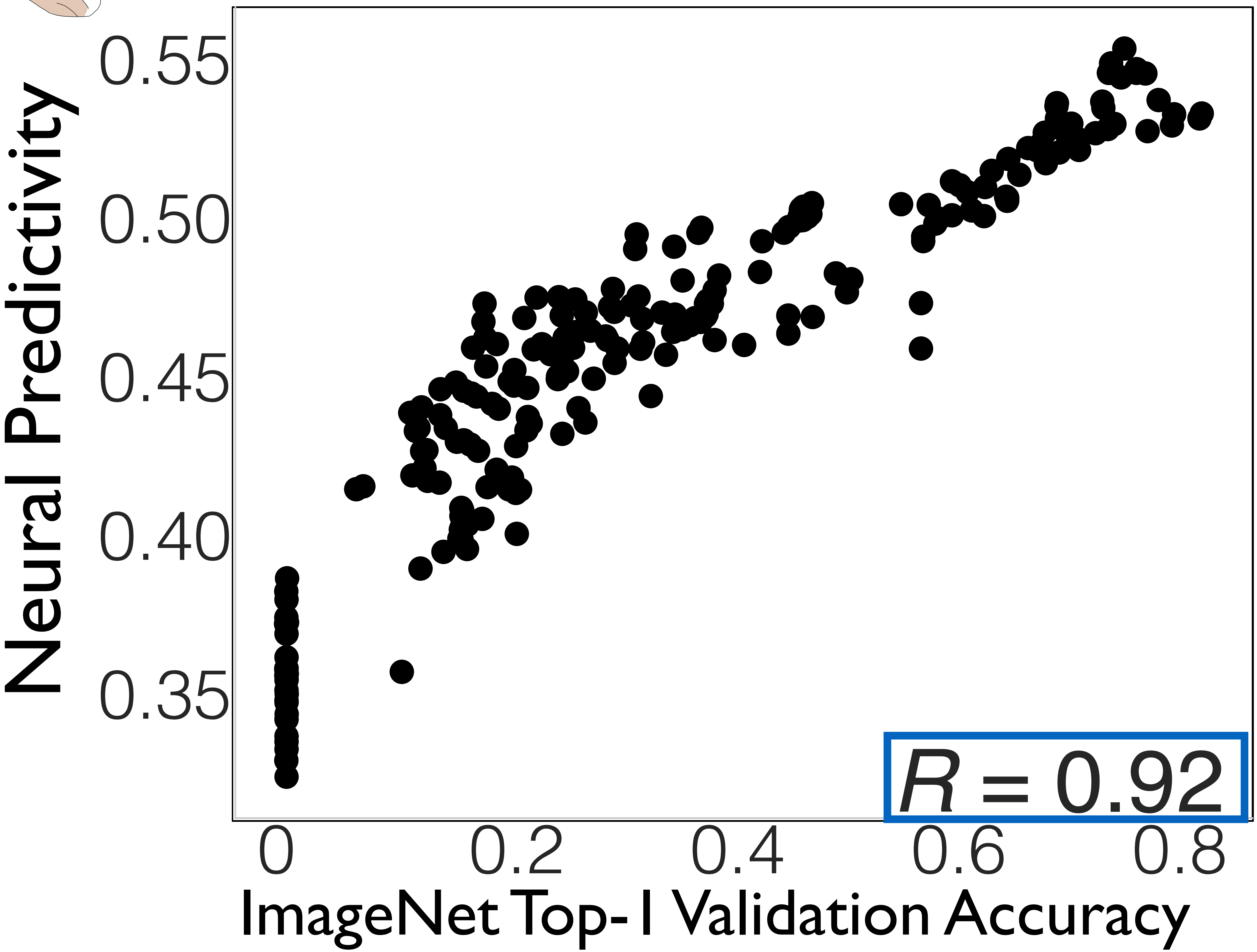


An AI Goal

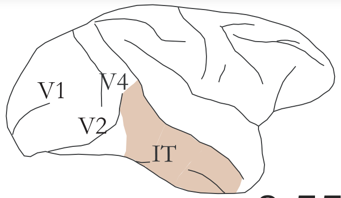
Task Performance Correlated with Neural Predictivity



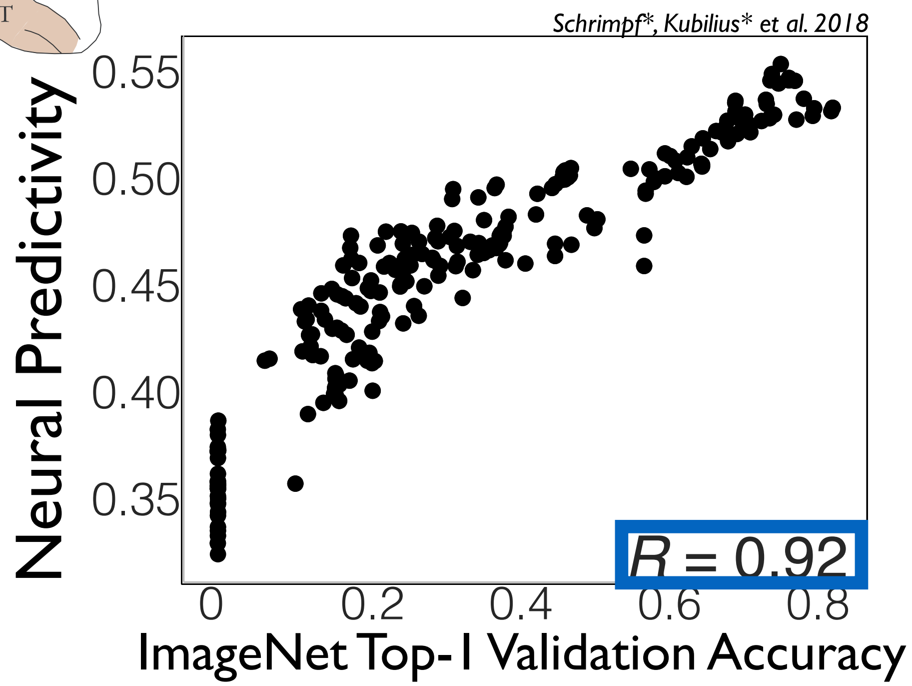
Schrimpf*, Kubilius* et al. 2018

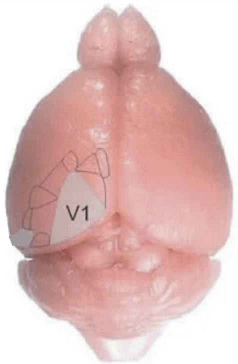


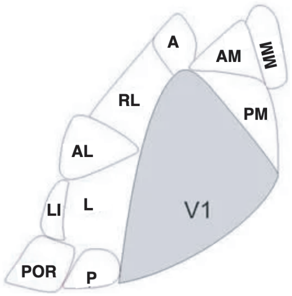
Task Performance Correlated with Neural Predictivity



Primates

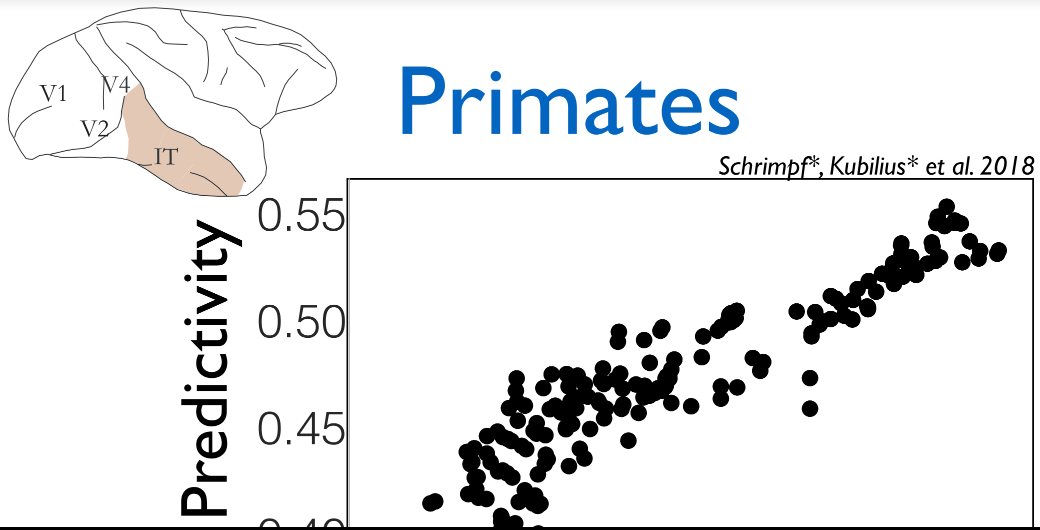


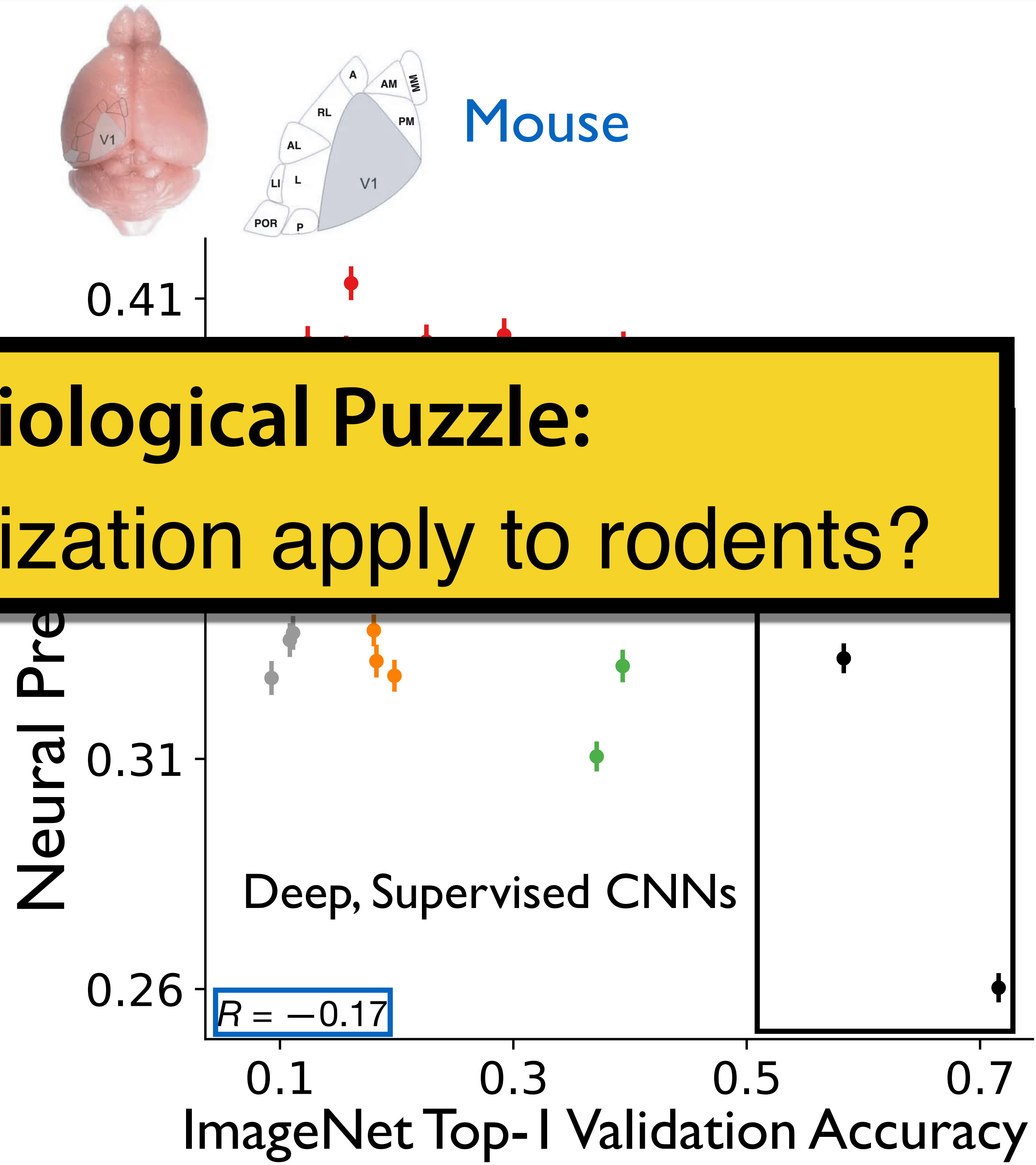




Mouse?

Object Categorization Ability **NOT** Correlated with Neural Predictivity

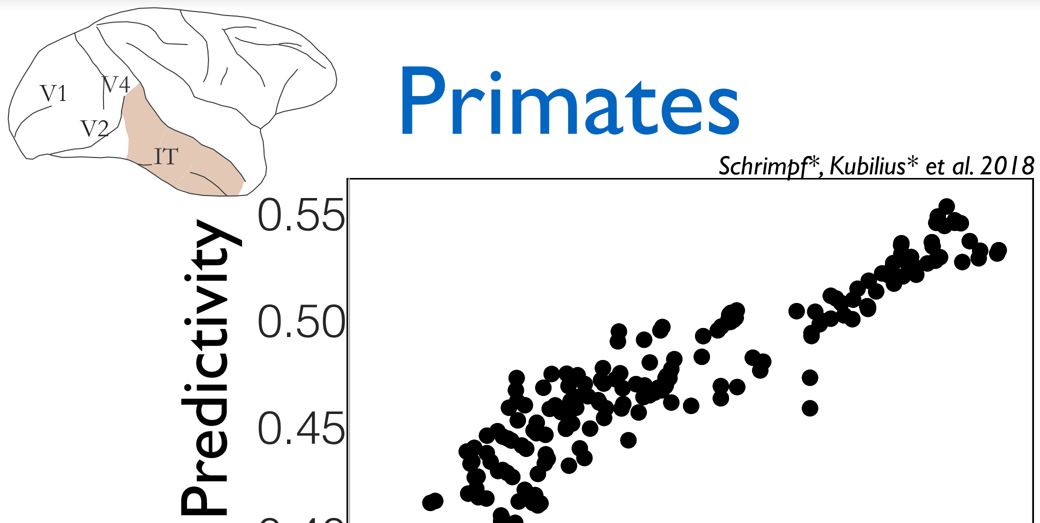


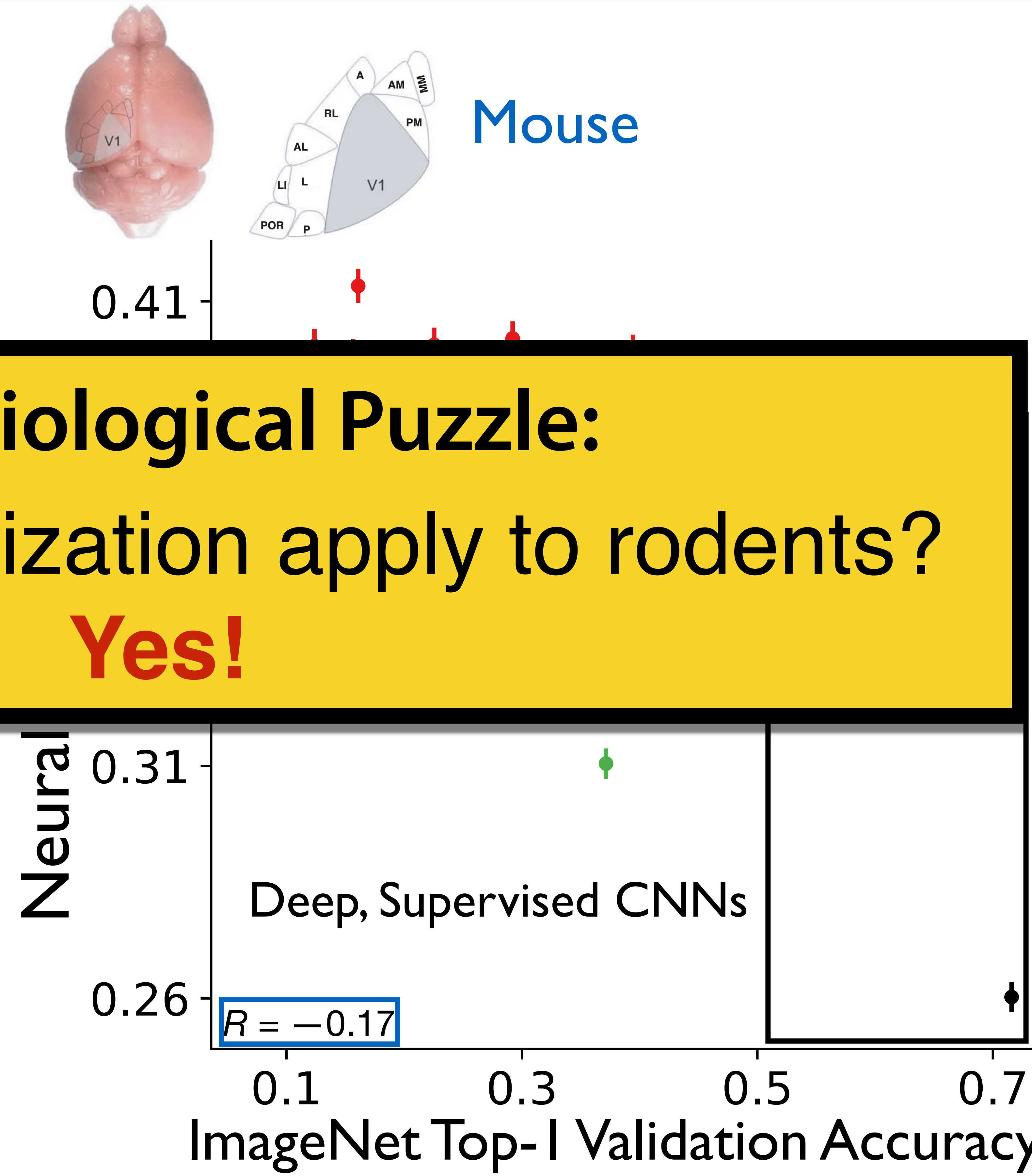


Neurobiological Puzzle:

Does task-optimization apply to rodents?

Object Categorization Ability **NOT** Correlated with Neural Predictivity





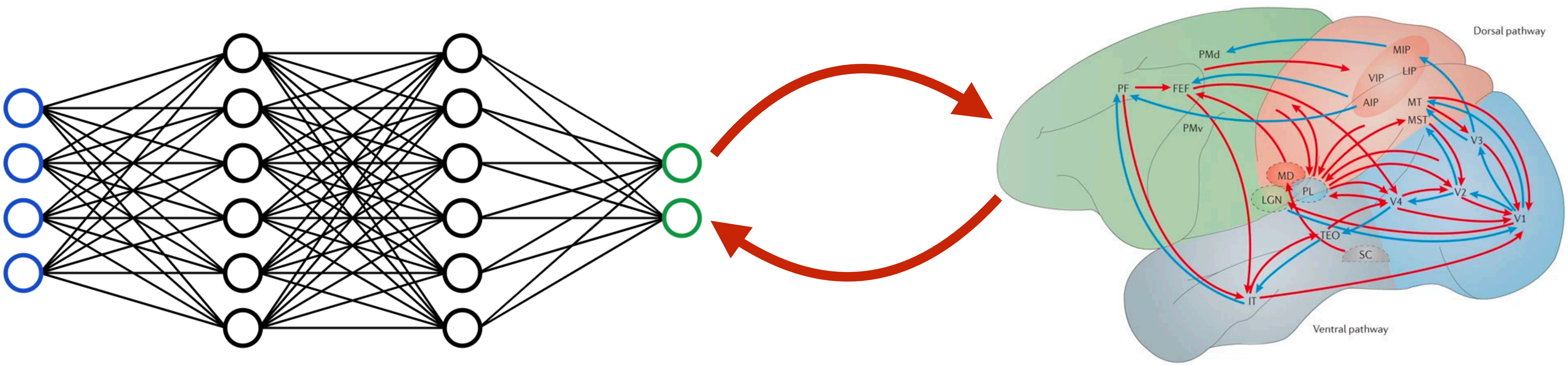
Neurobiological Puzzle:

Does task-optimization apply to rodents?

Yes!

Task-Optimized Modeling

Design ML Algorithms Optimized to Perform Organism's Behavior under Organism's Constraints



Yields:

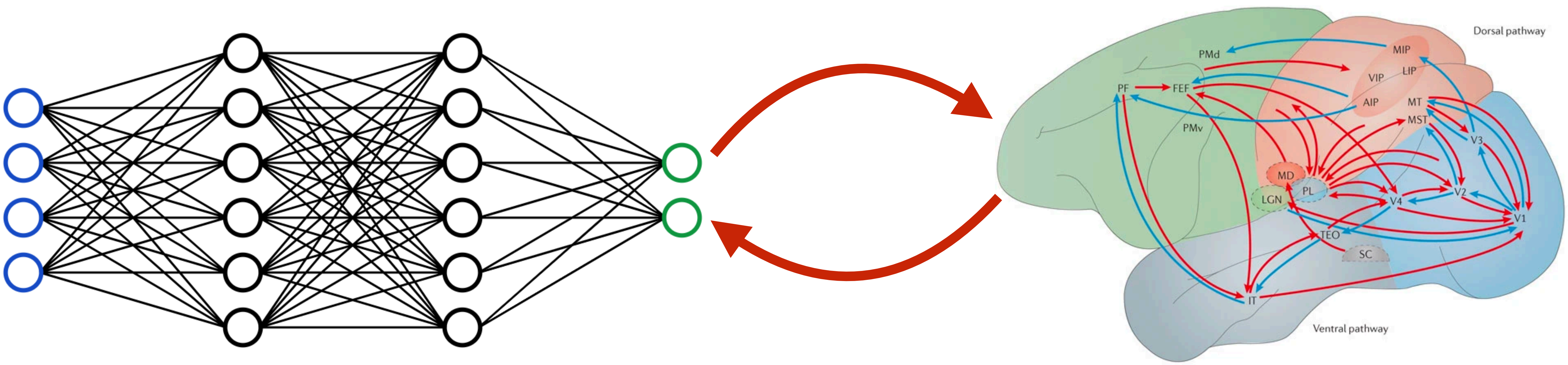
Quantitatively Accurate & Practically Useful Brain Models

AND

Principles of *Why* Neural Responses Are As They Are

Task-Optimized Modeling

Design ML Algorithms Optimized to Perform Organism's Behavior under Organism's Constraints



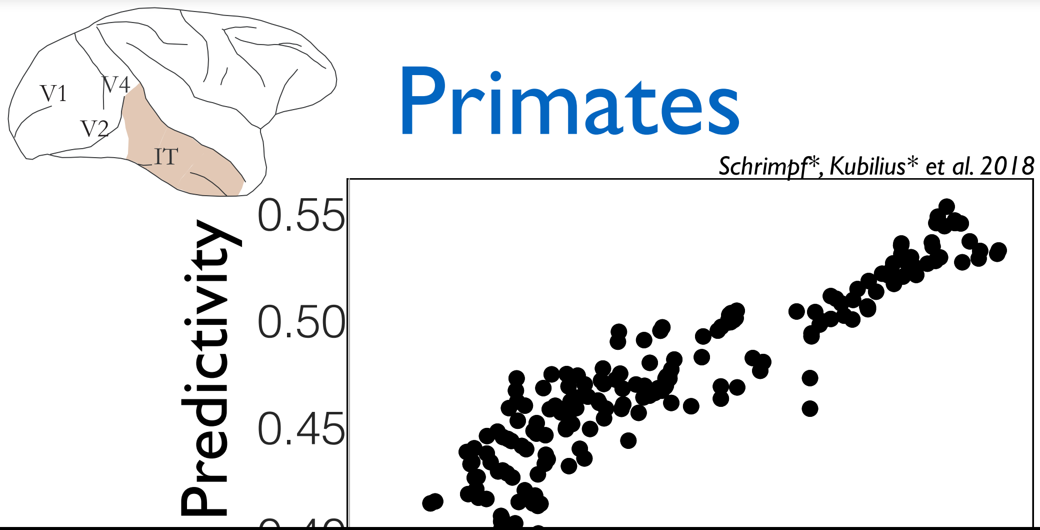
Yields:

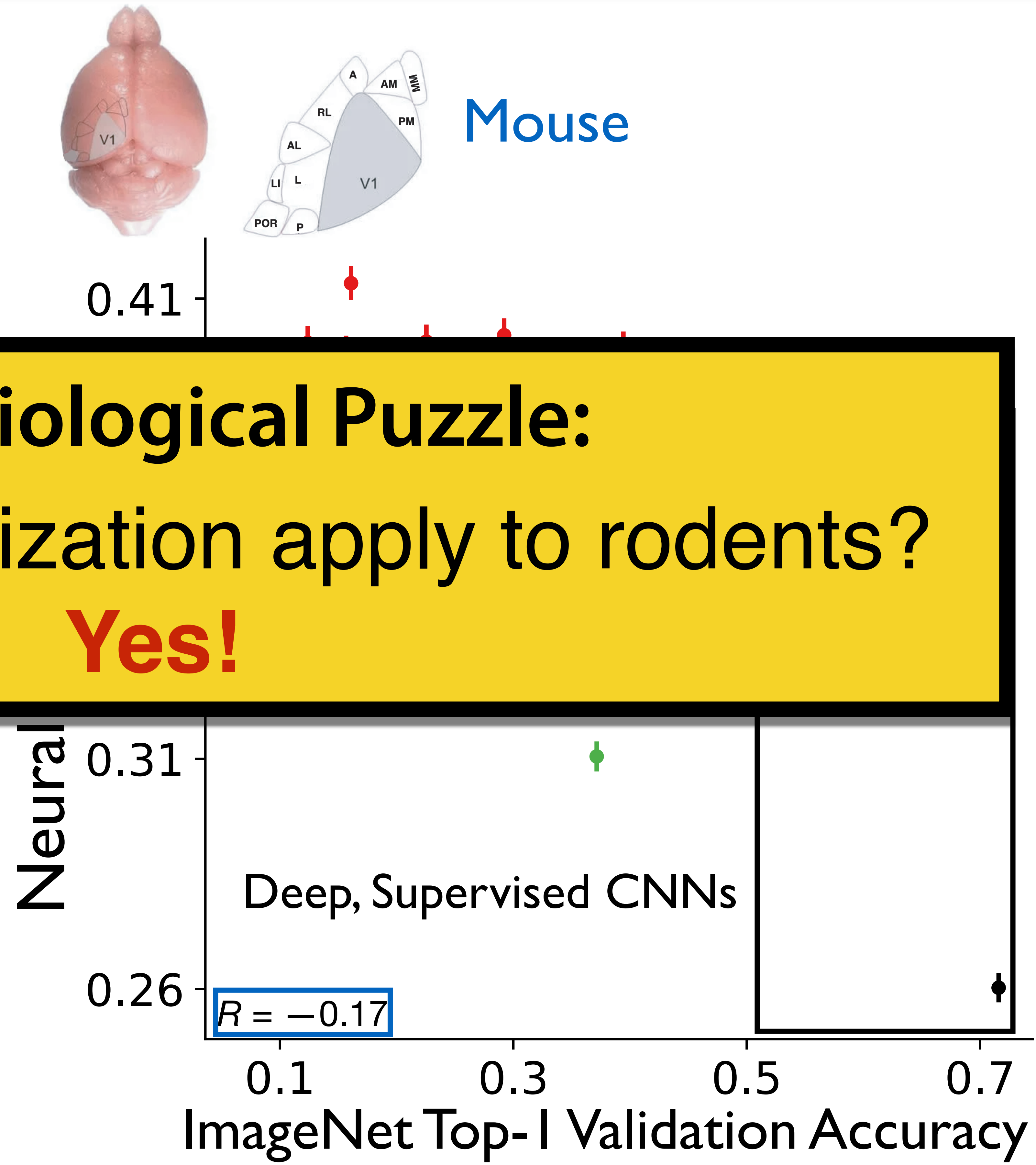
Quantitatively Accurate & Practically Useful Brain Models

AND

Principles of *Why* Neural Responses Are As They Are

Object Categorization Ability **NOT** Correlated with Neural Predictivity



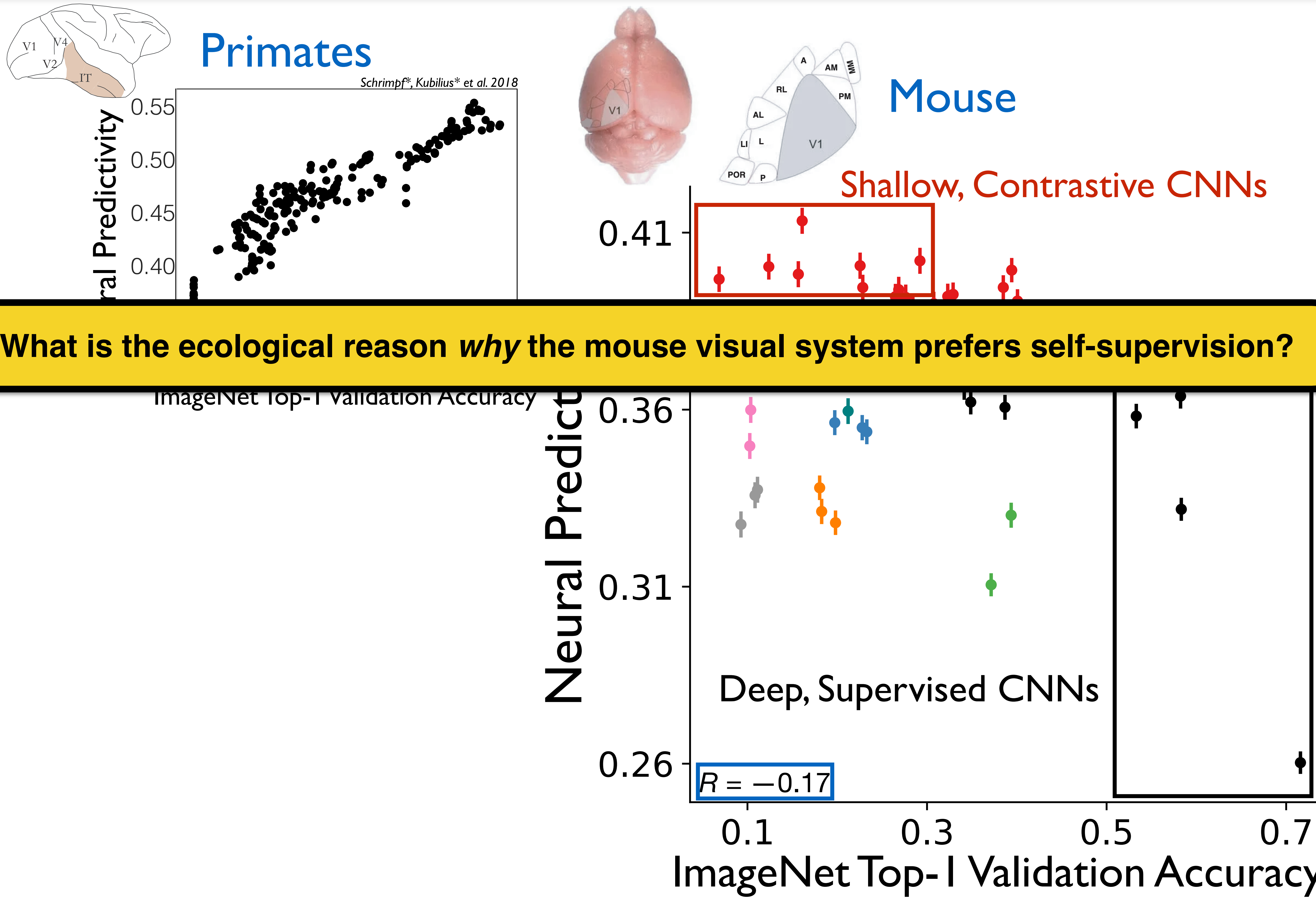


Neurobiological Puzzle:

Does task-optimization apply to rodents?

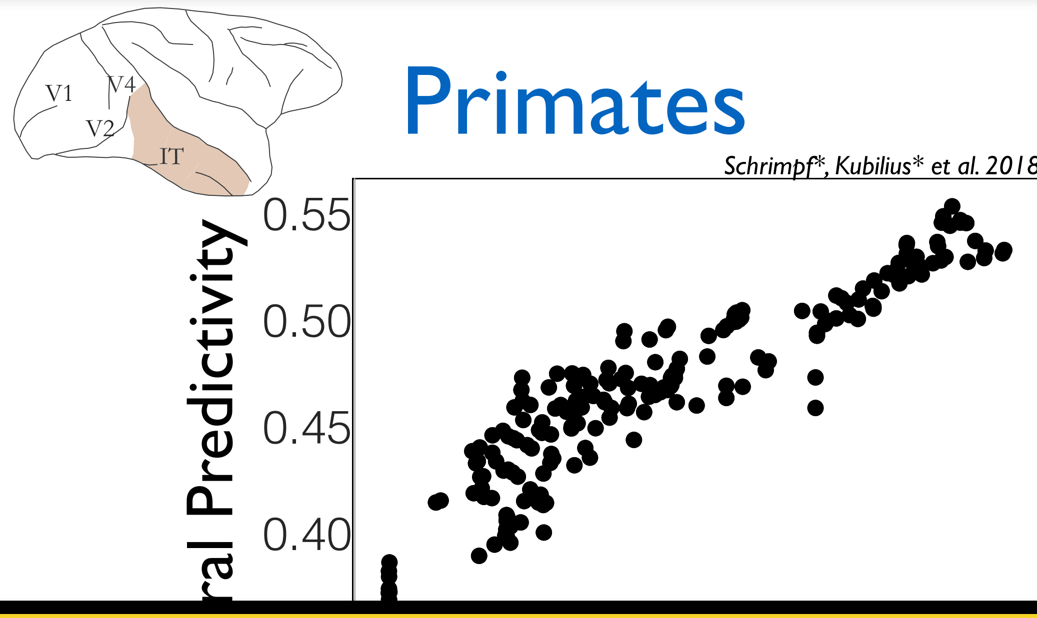
Yes!

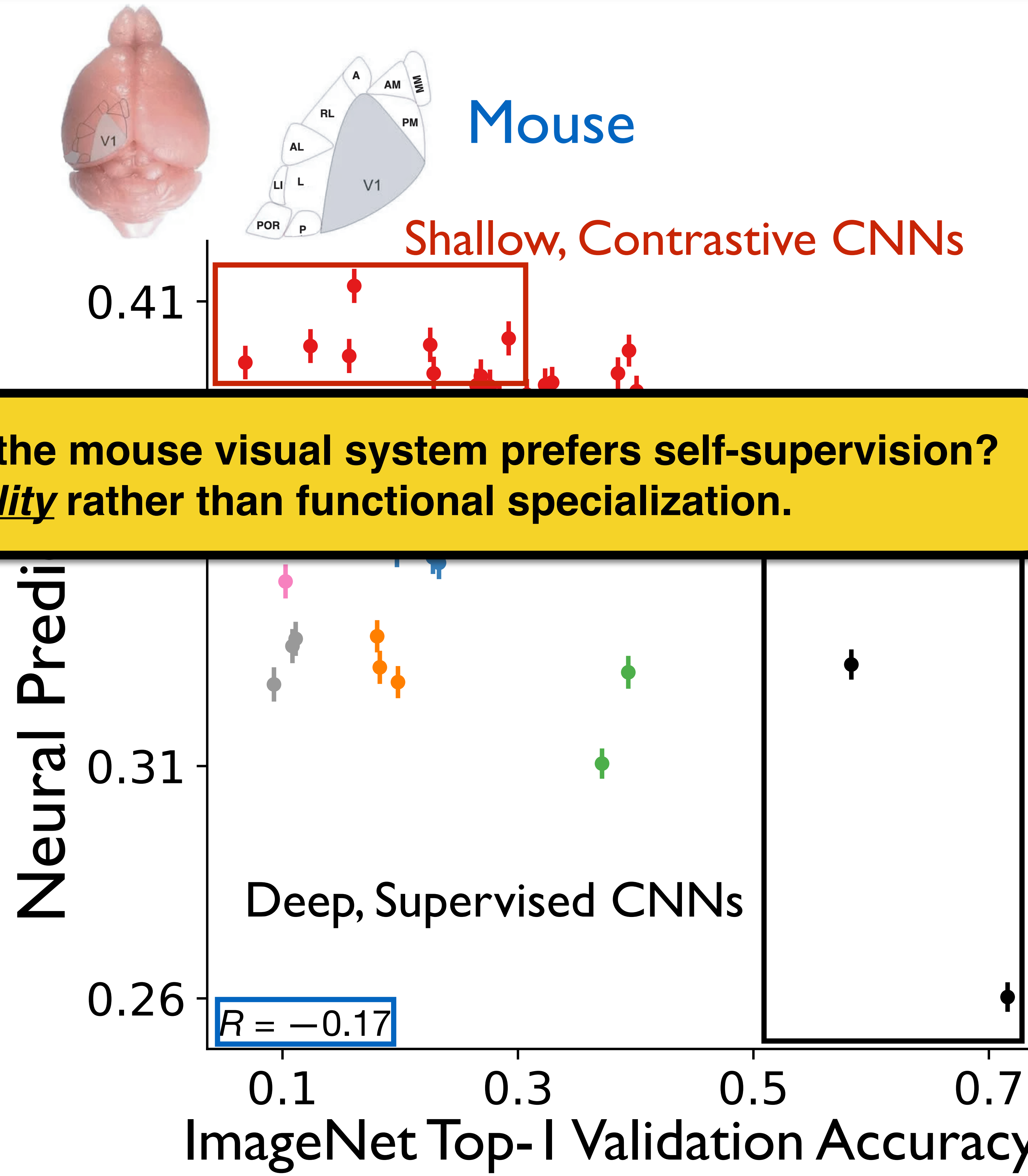
Object Categorization Ability **NOT** Correlated with Neural Predictivity



What is the ecological reason *why* the mouse visual system prefers self-supervision?

Object Categorization Ability **NOT** Correlated with Neural Predictivity





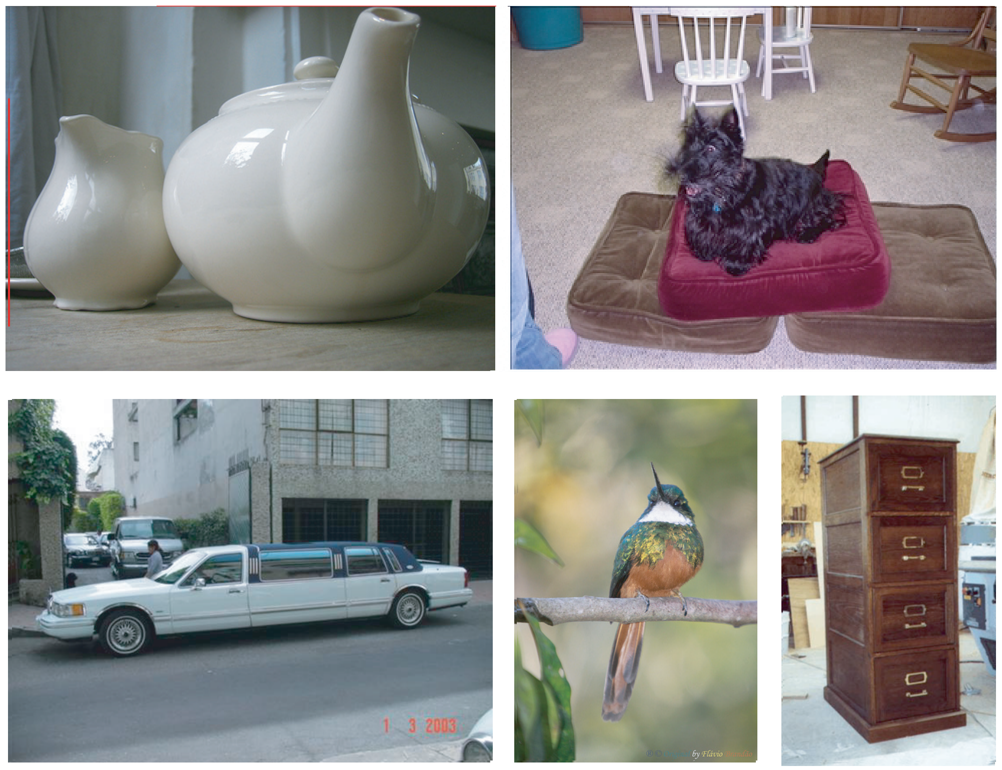
What is the ecological reason *why* the mouse visual system prefers self-supervision?
Hypothesis: *task-generality* rather than functional specialization.

Assessing Task-Generality

Assessing Task-Generality

Train

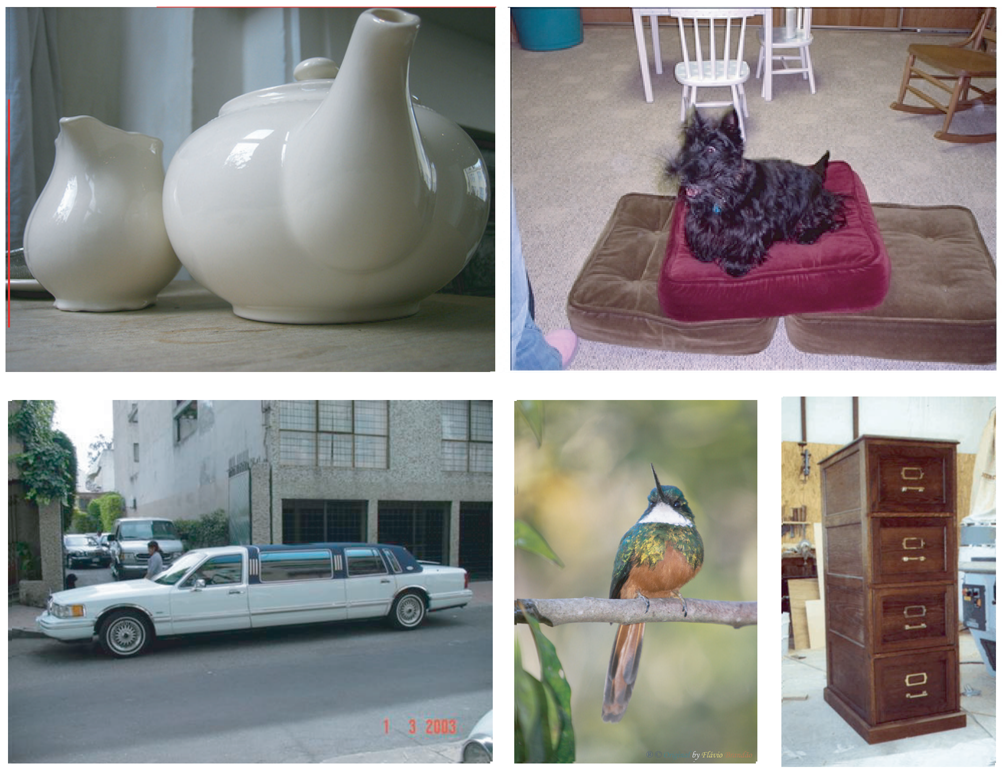
ImageNet



Assessing Task-Generality

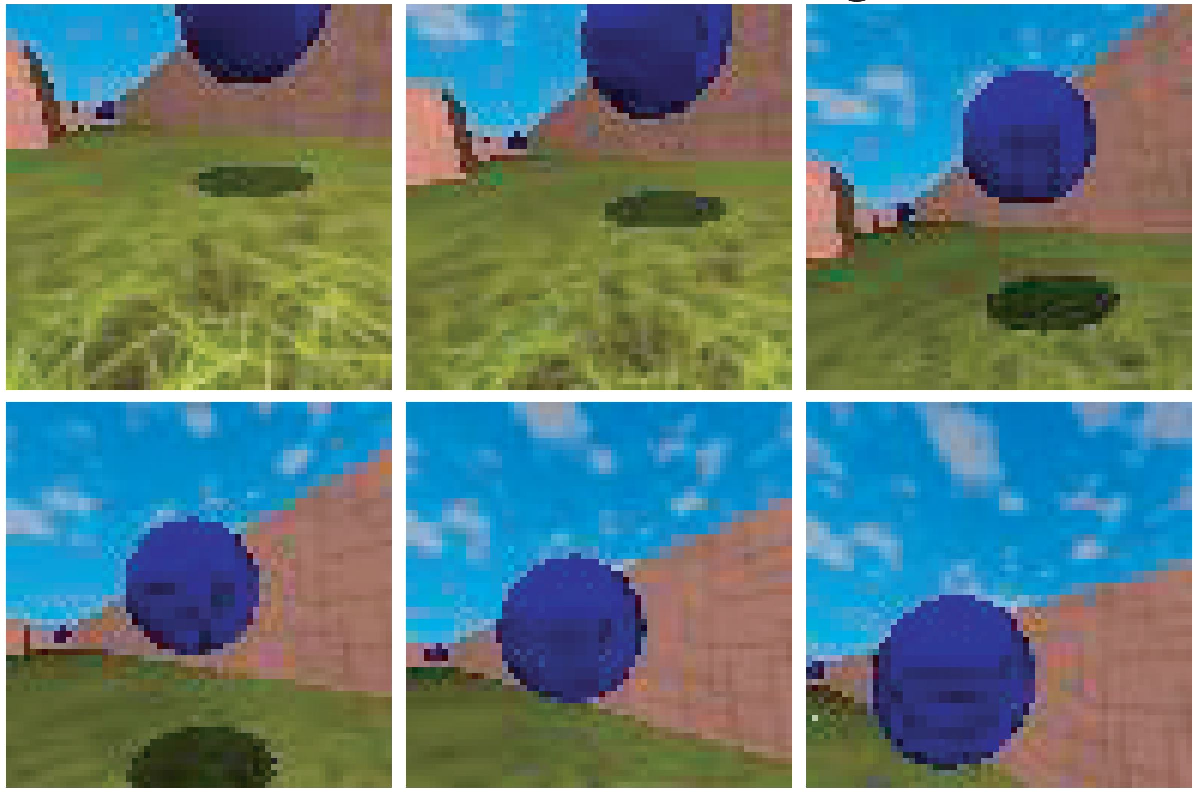
Train

ImageNet

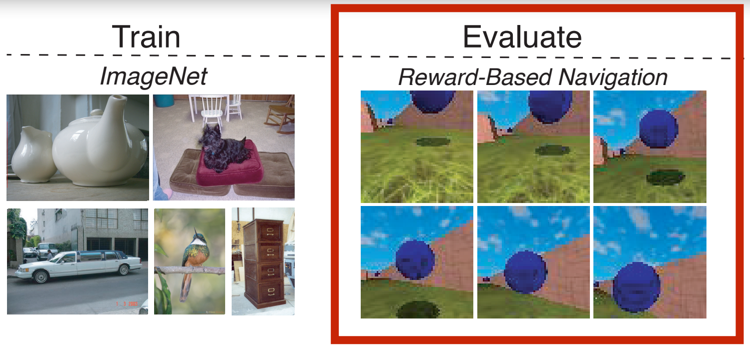


Evaluate

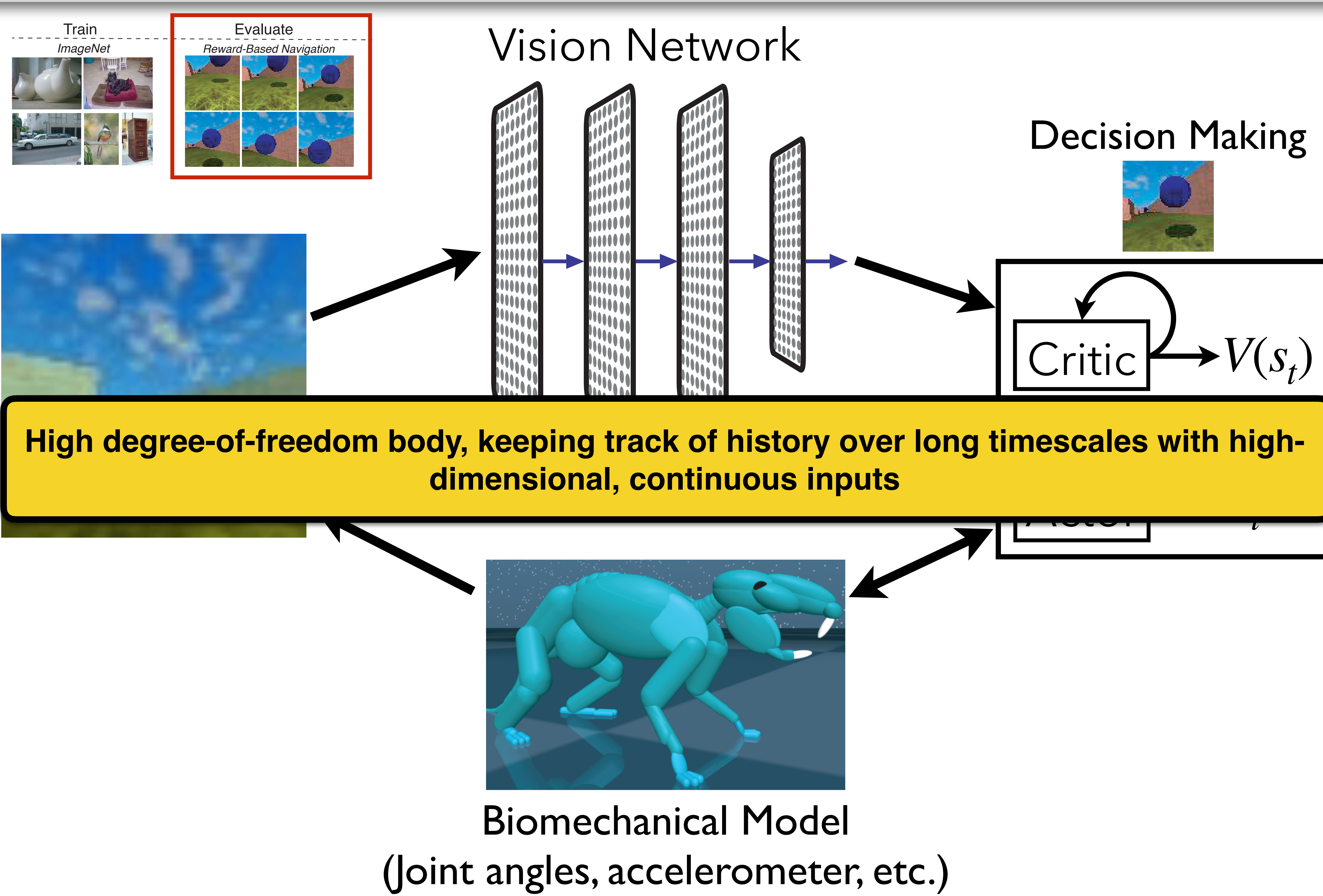
Reward-Based Navigation



Assessing Task-Generality

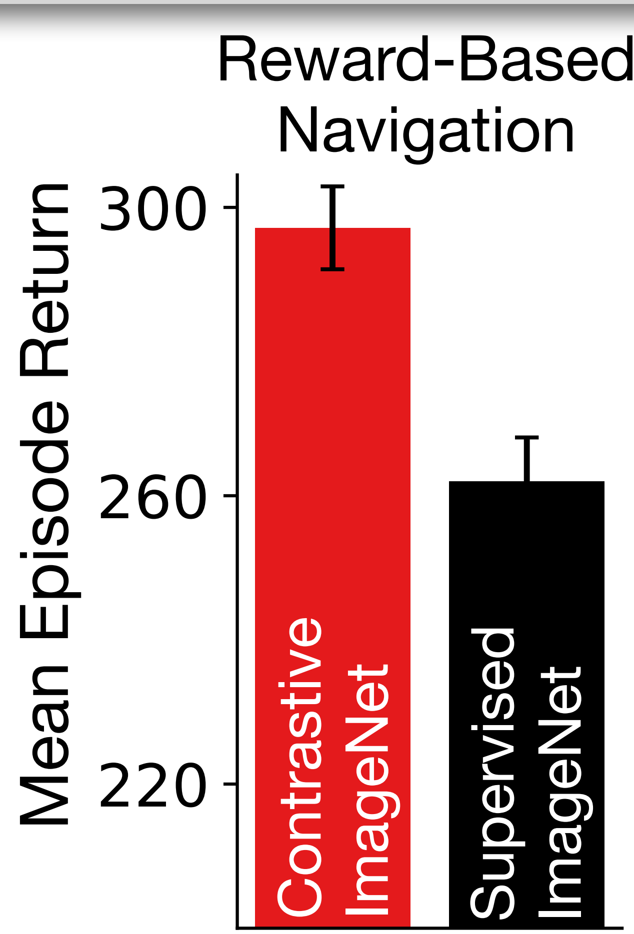


Embodied Virtual Rodent Navigation



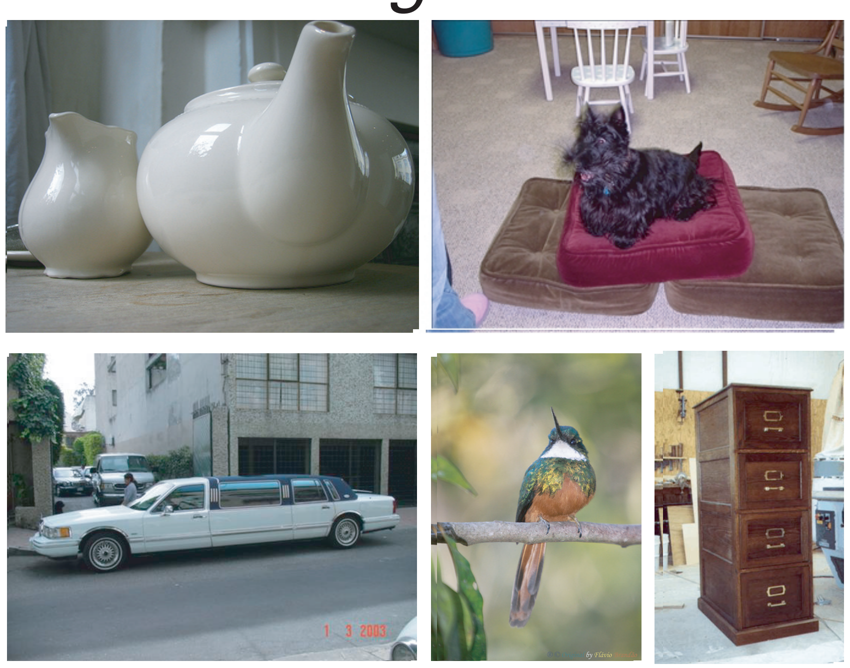
Contrastive Models Yield Better Transfer Performance

Contrastive Models Yield Better Transfer Performance

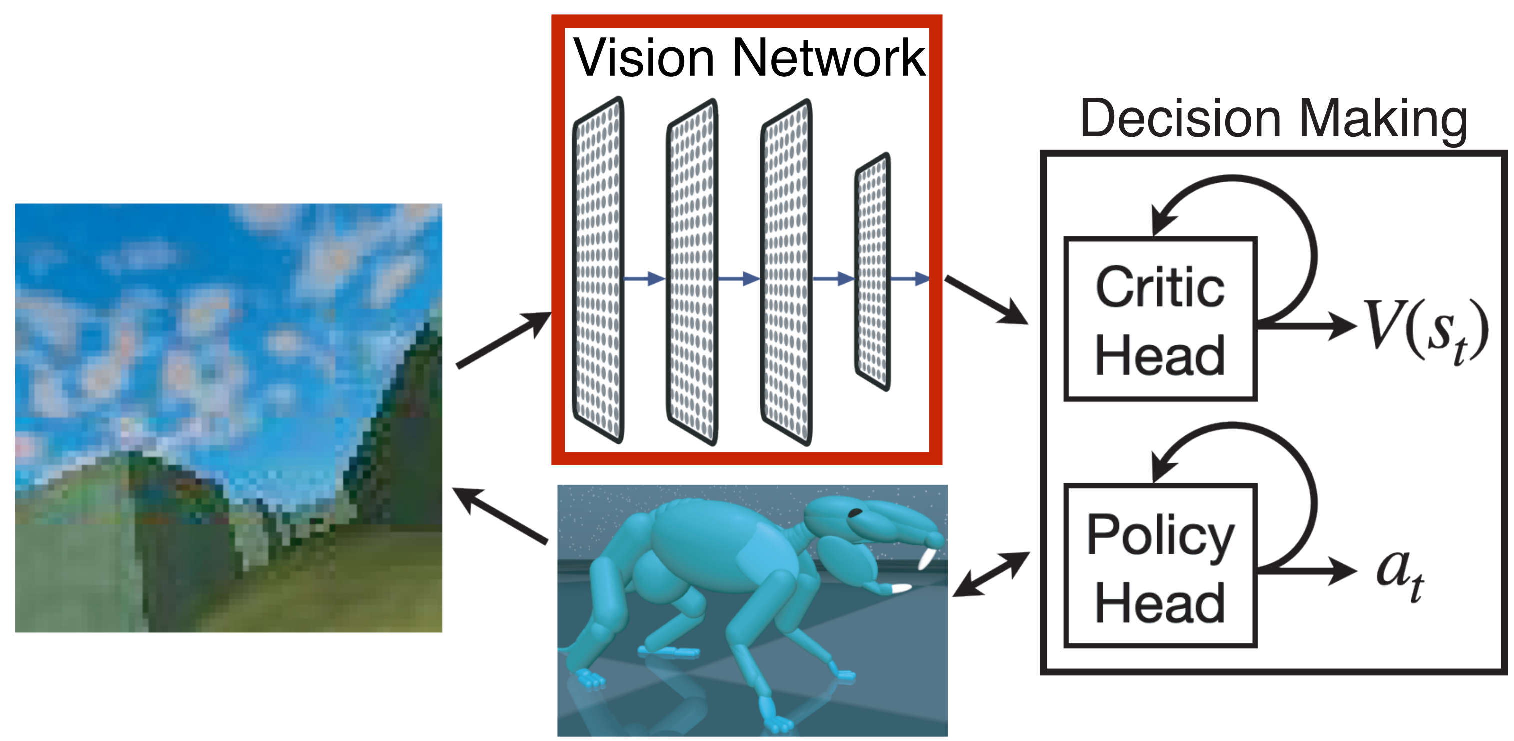


Train

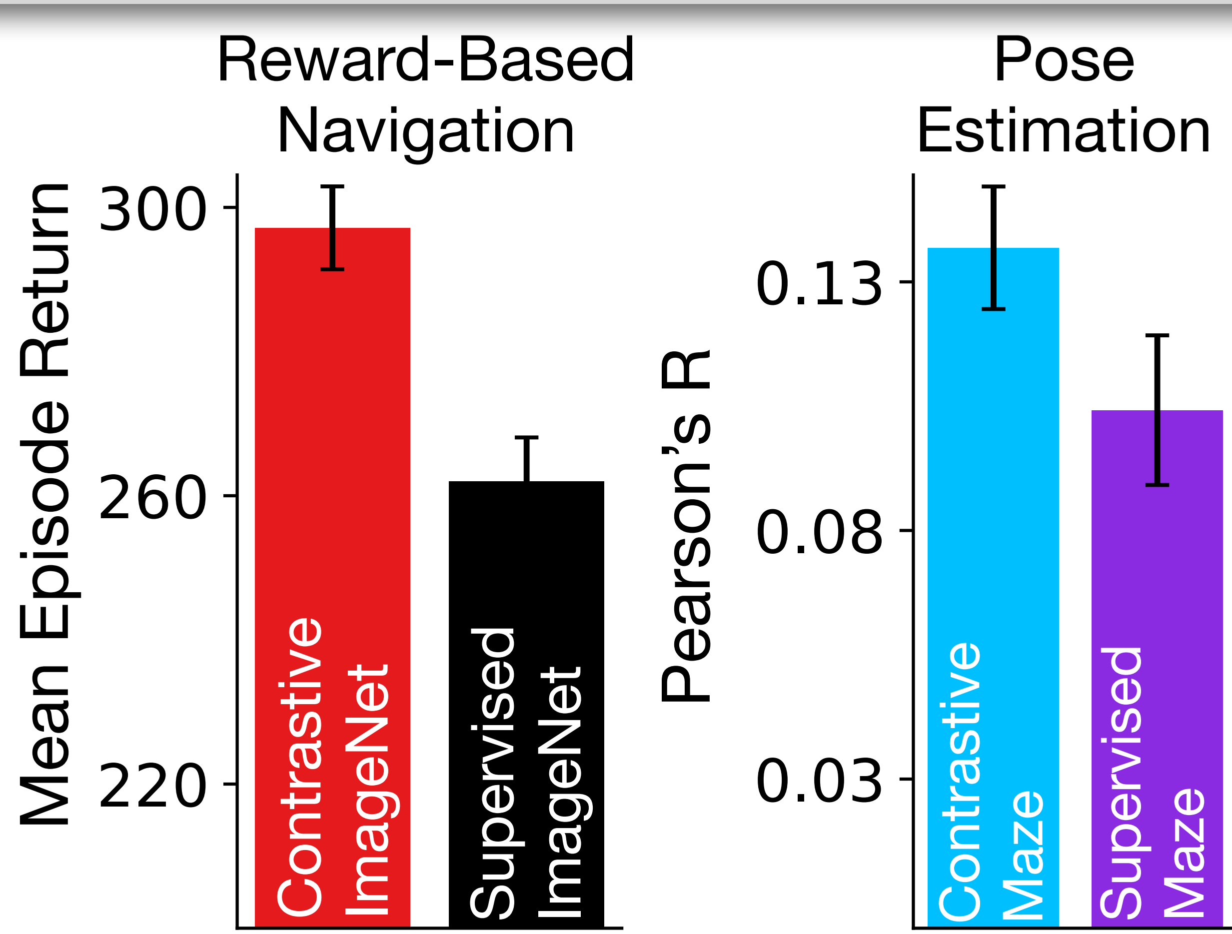
ImageNet



Evaluate

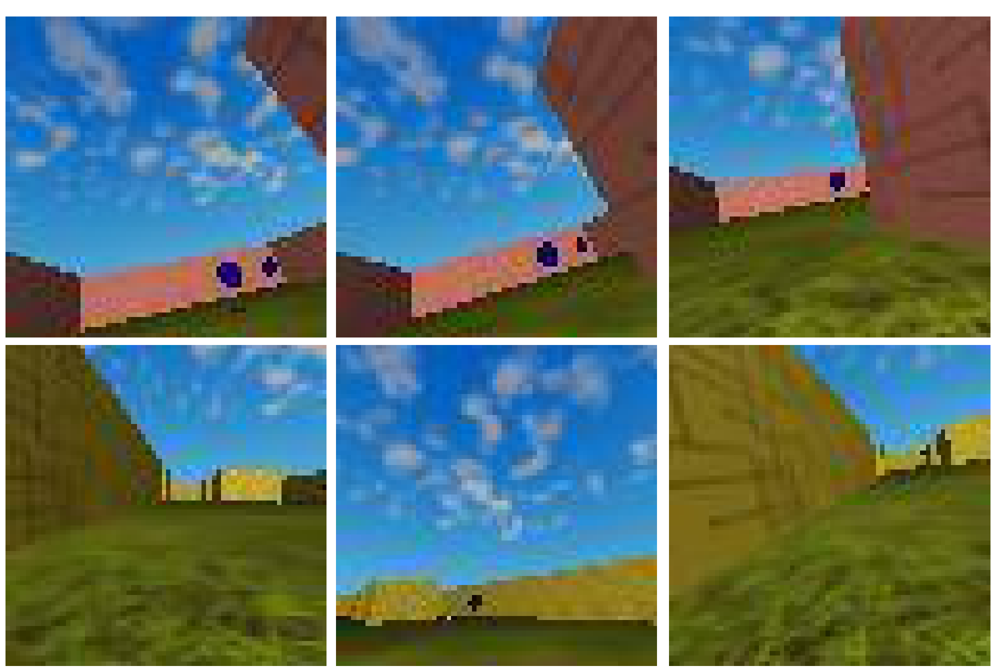


Contrastive Models Yield Better Transfer Performance



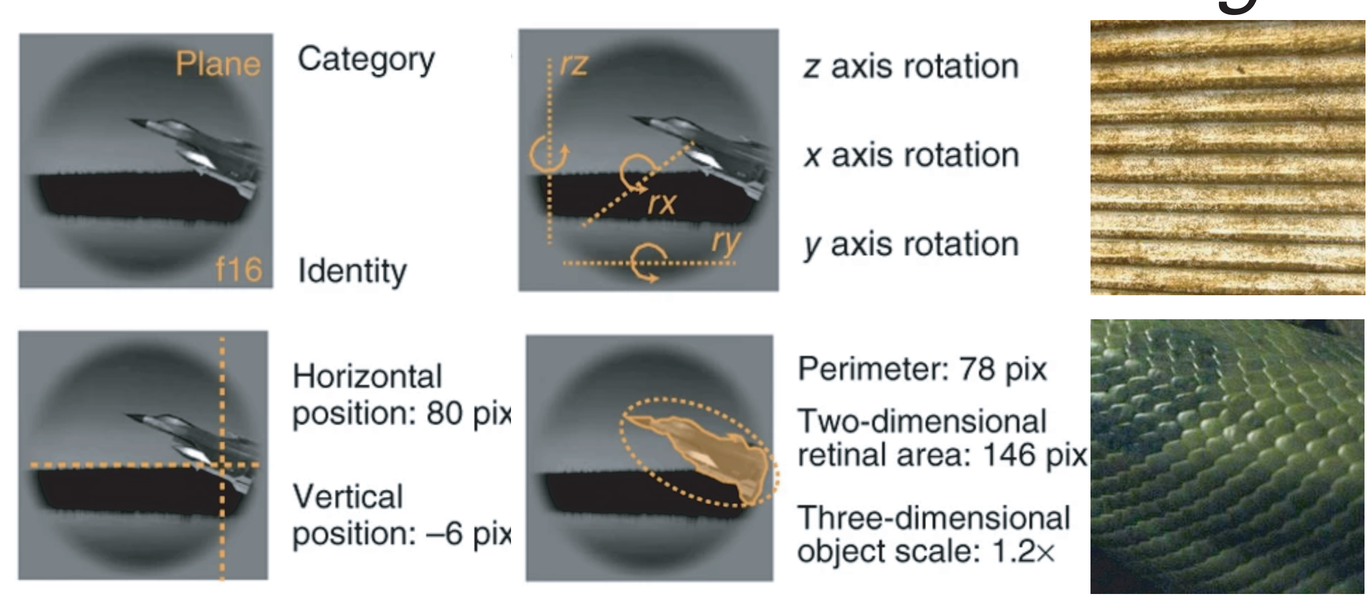
Train

Maze Environment



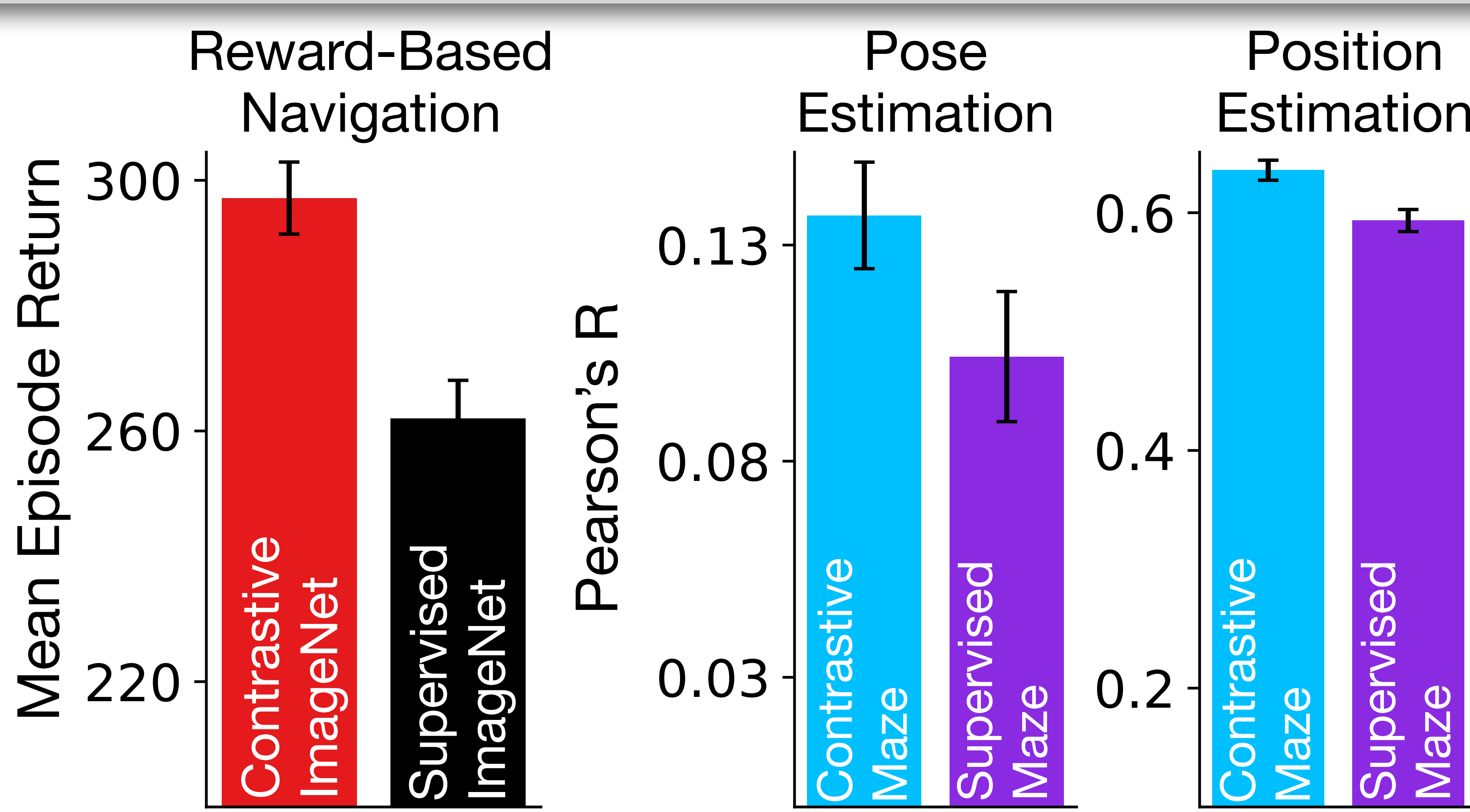
Evaluate

Visual Scene Understanding



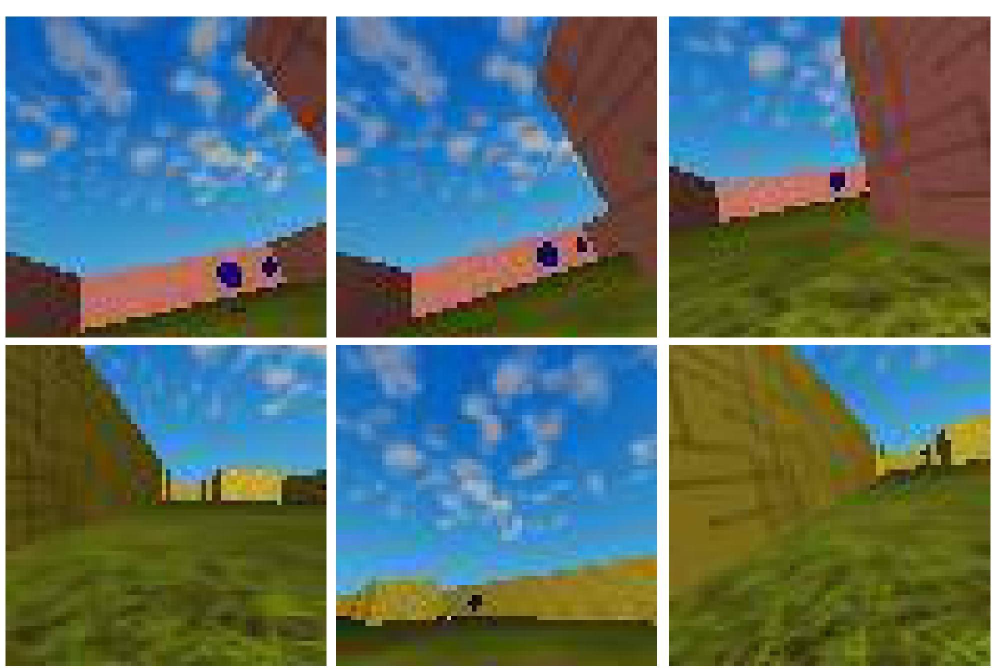
Object properties *Texture*

Contrastive Models Yield Better Transfer Performance



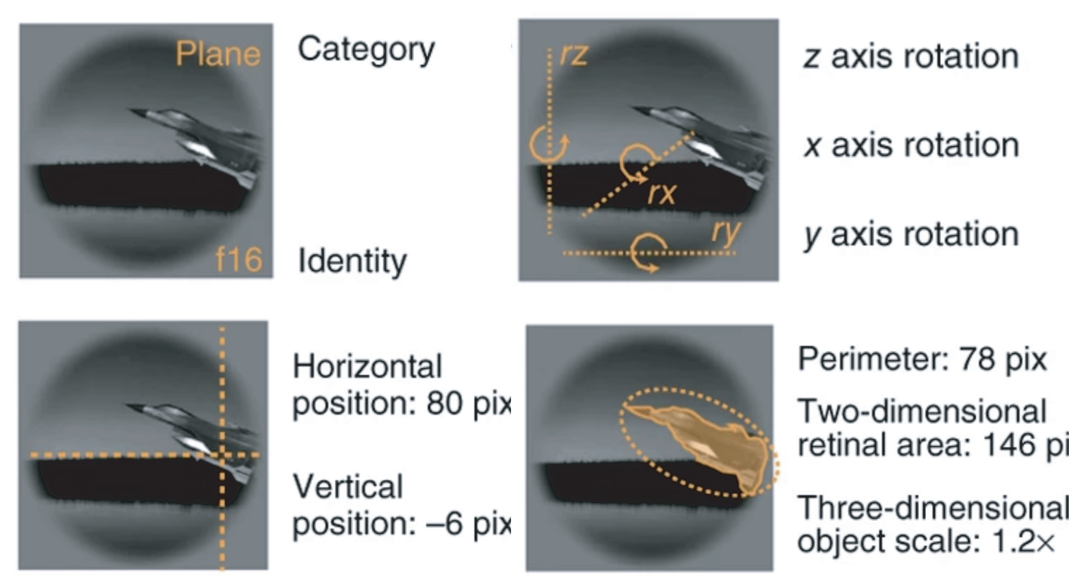
Train

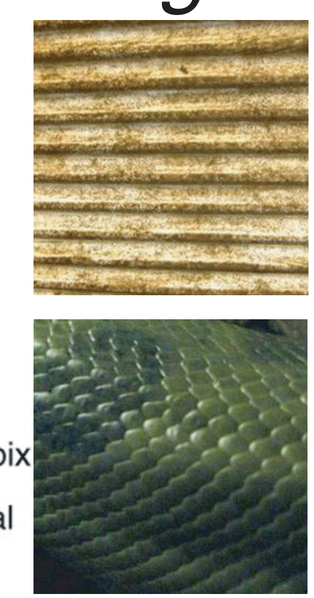
Maze Environment



Evaluate

Visual Scene Understanding

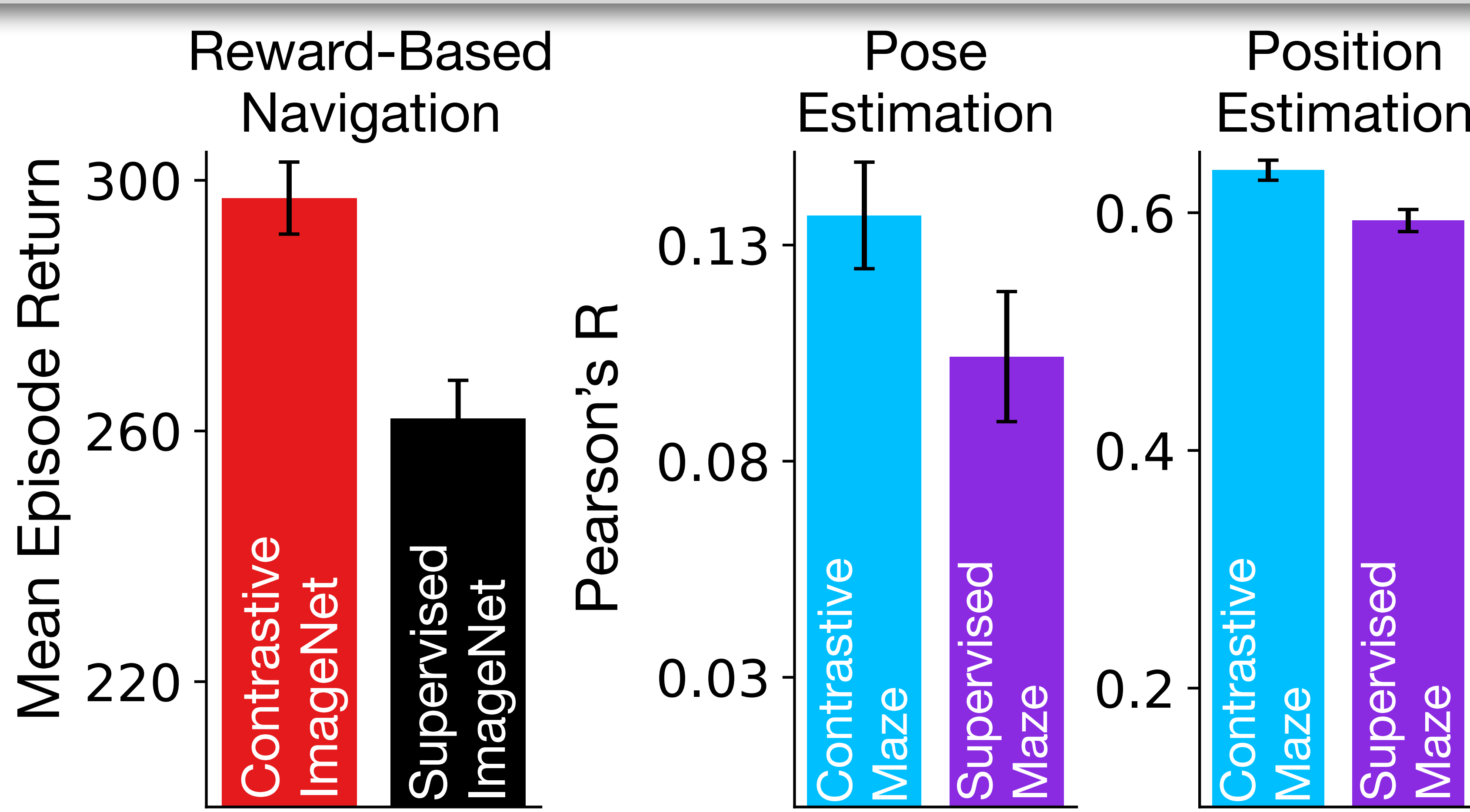


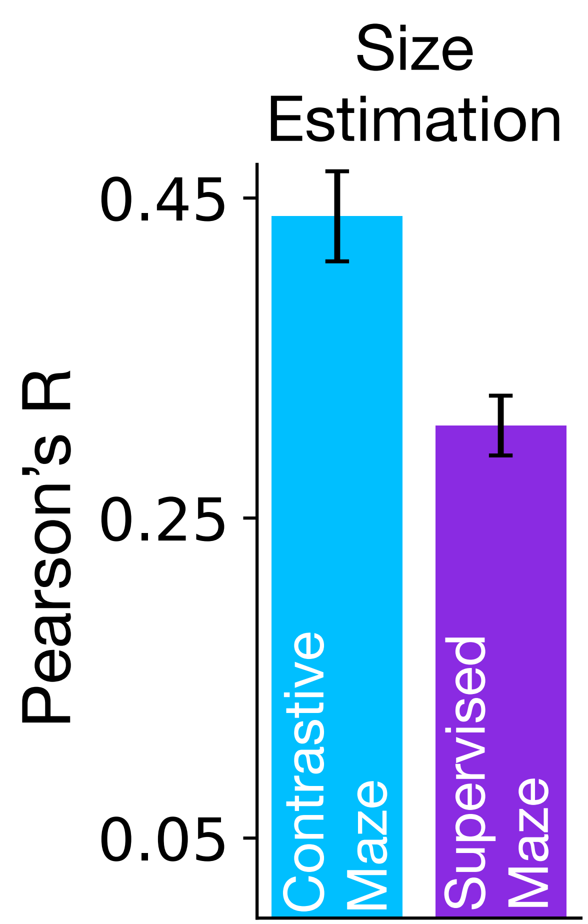


Object properties

Texture

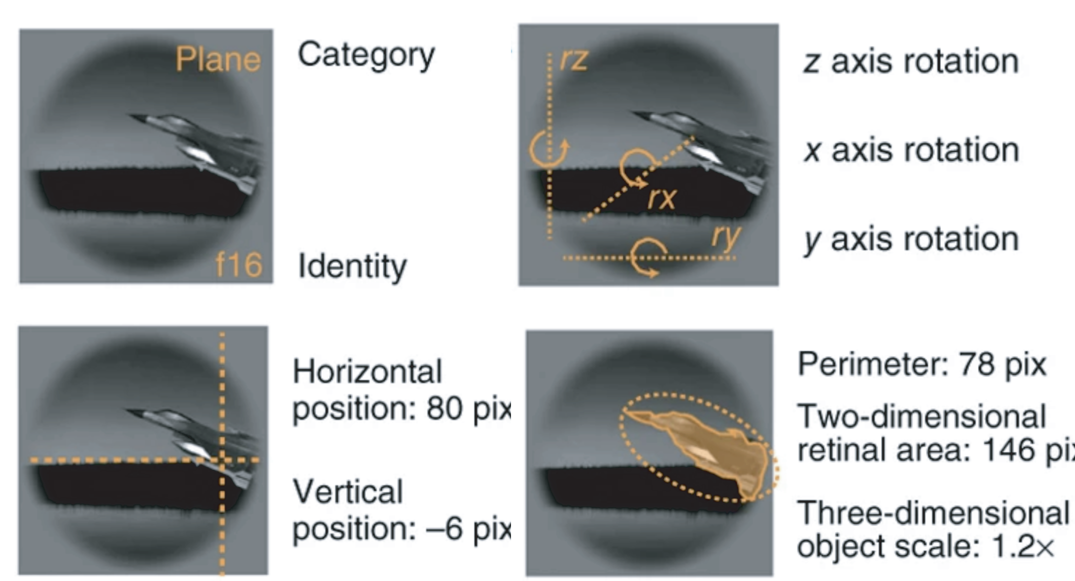
Contrastive Models Yield Better Transfer Performance



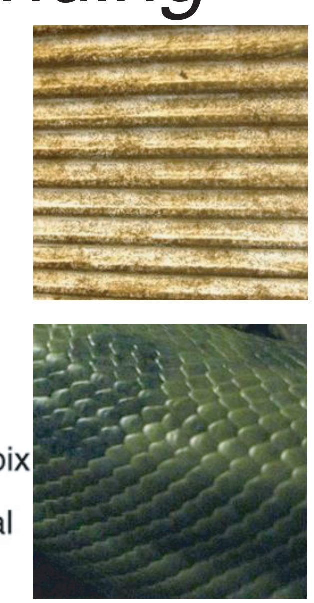


Evaluate

Visual Scene Understanding

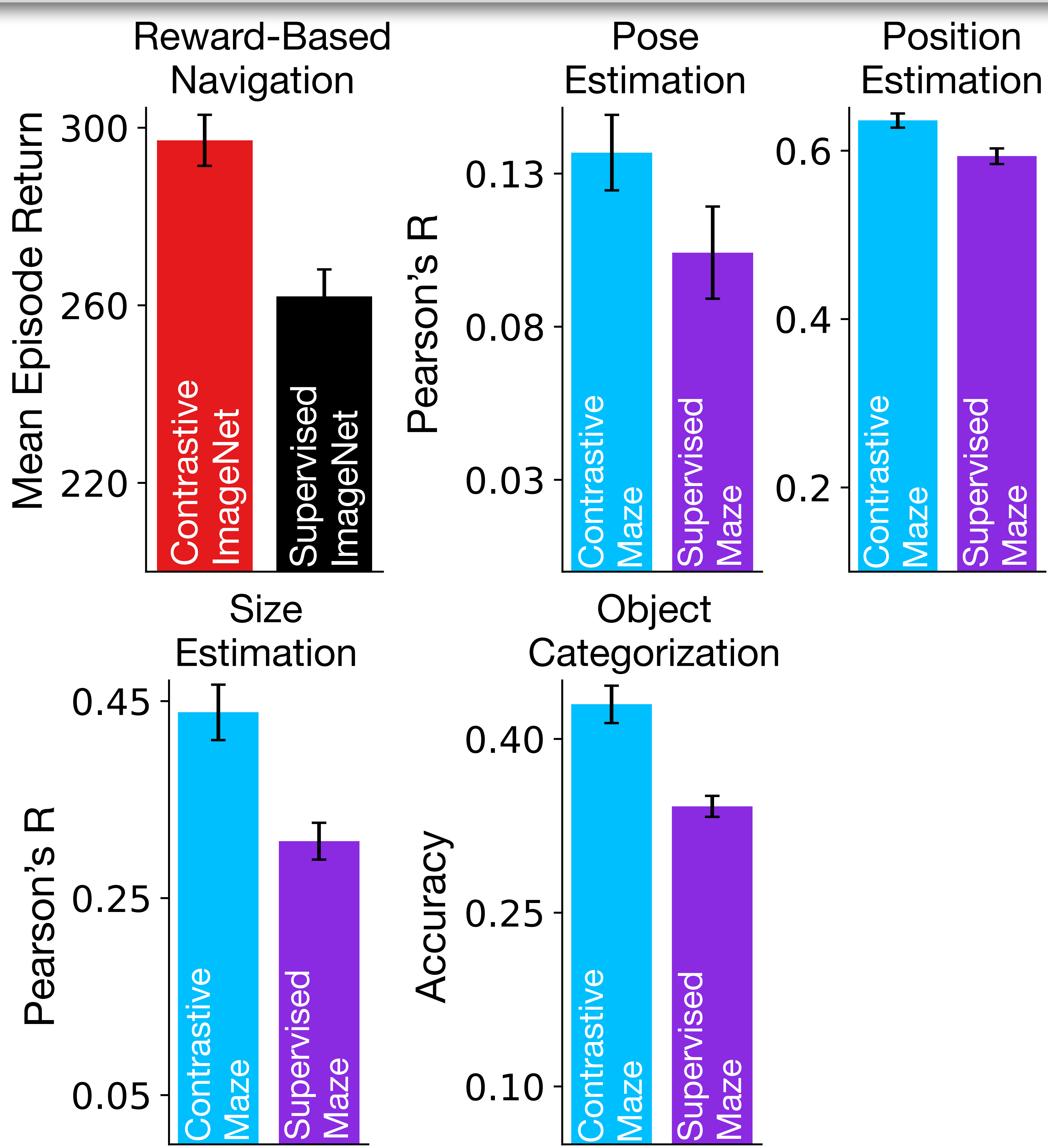


Object properties

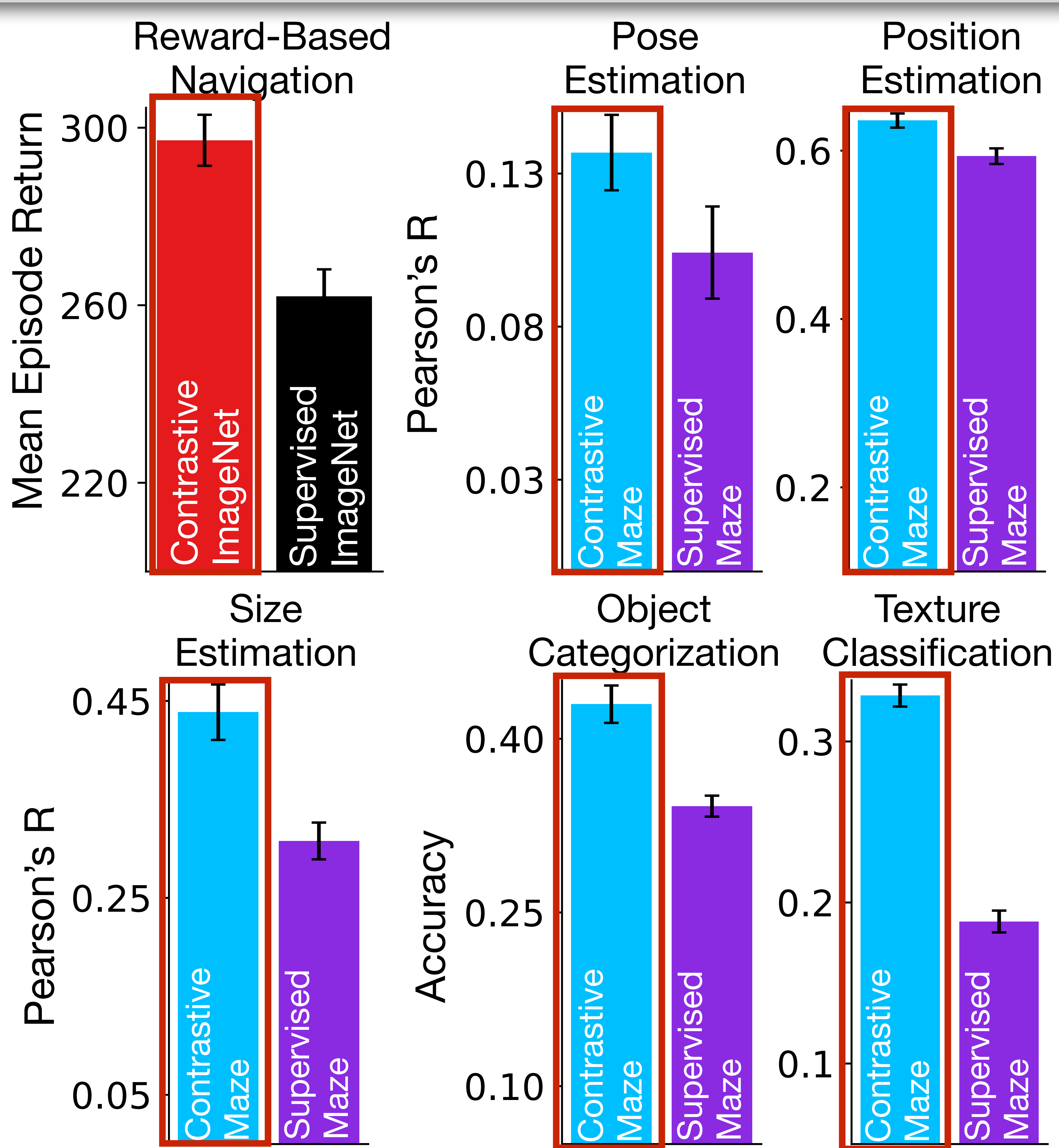


Texture

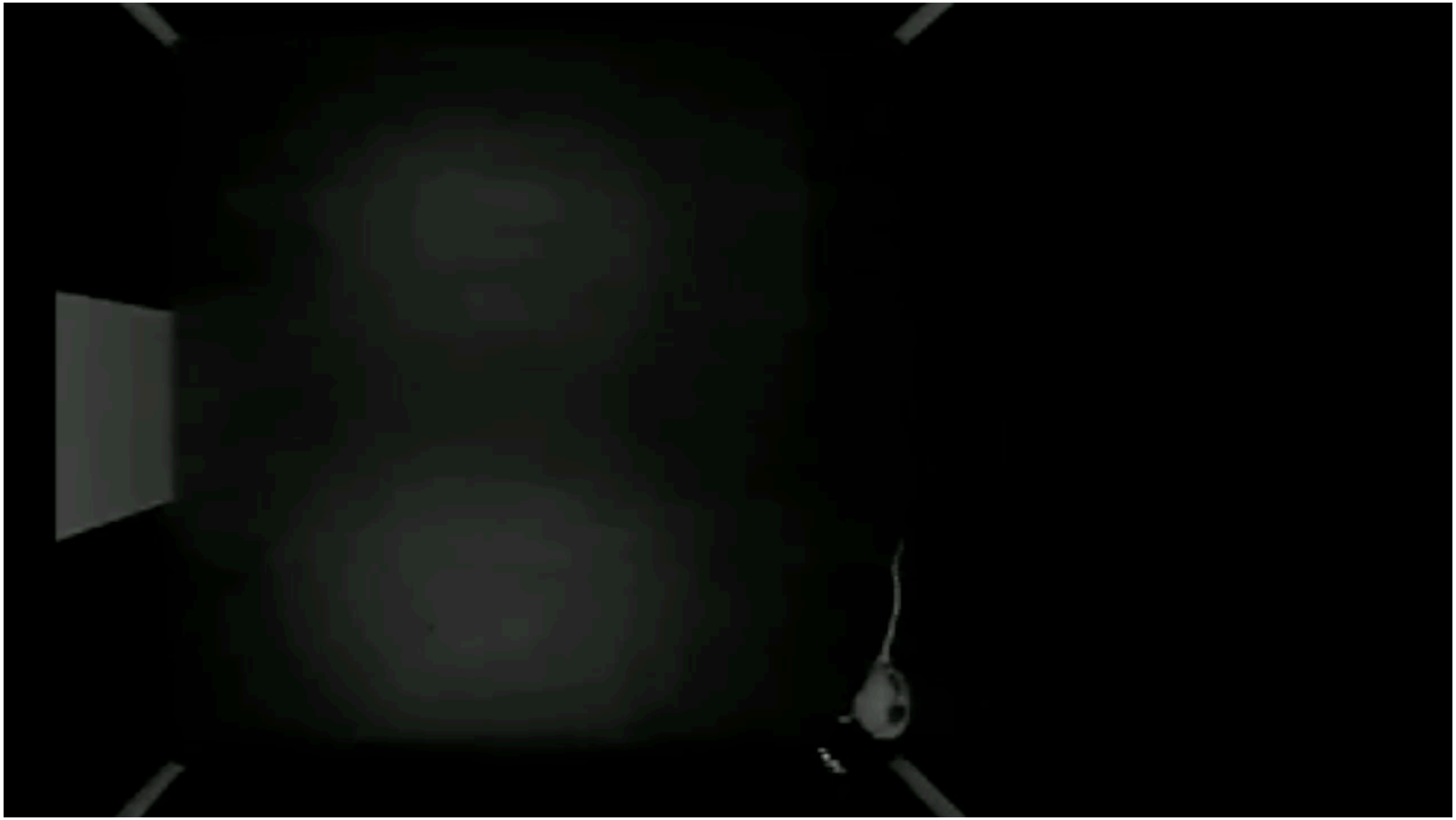
Contrastive Models Yield Better Transfer Performance



Contrastive Models Yield Better Transfer Performance

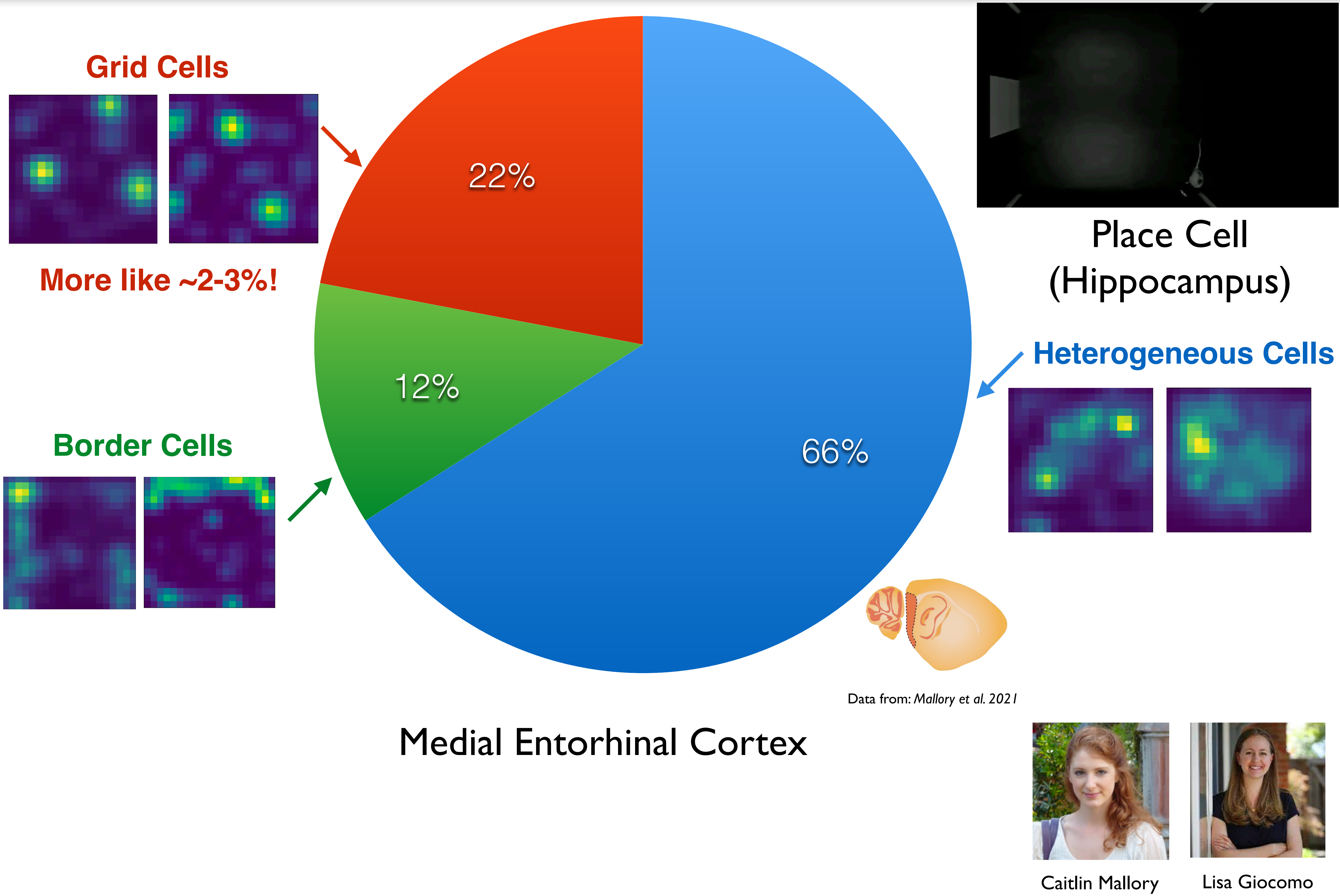


Hippocampal-Entorhinal Spatial Map

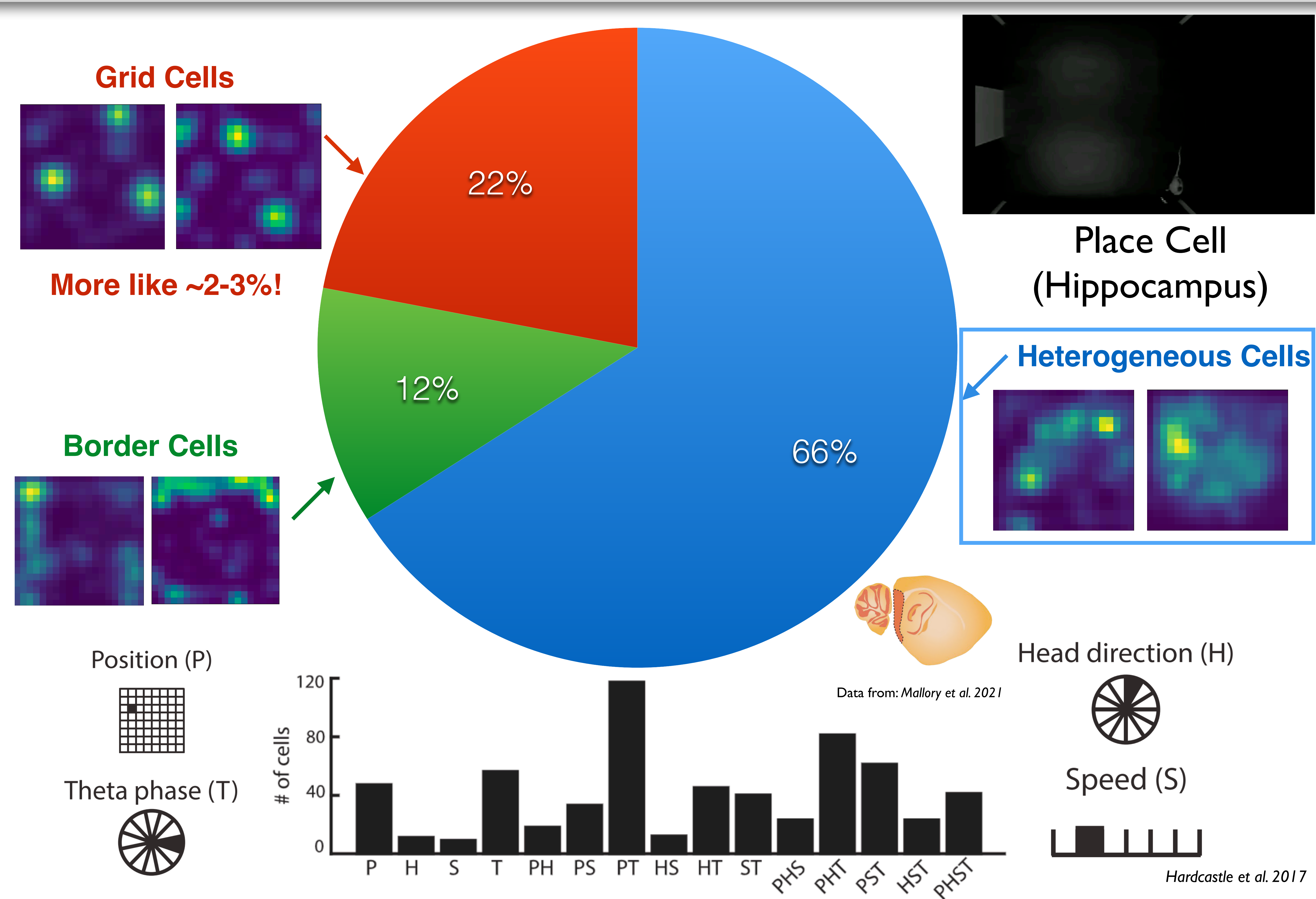


Place Cell
(Hippocampus)

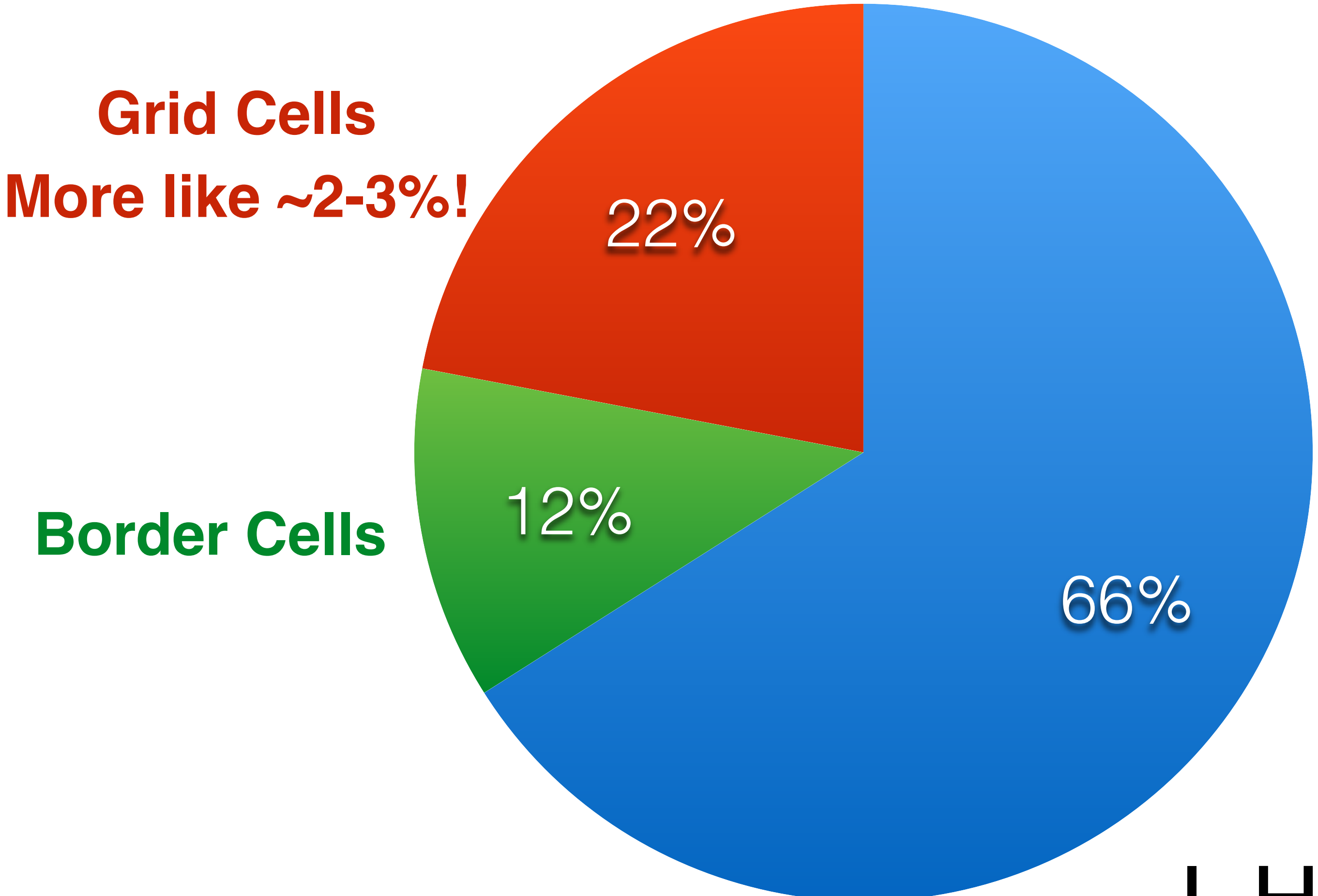
Hippocampal-Entorhinal Spatial Map



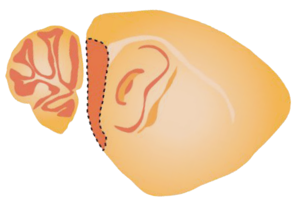
Accounting for Heterogeneous Code?



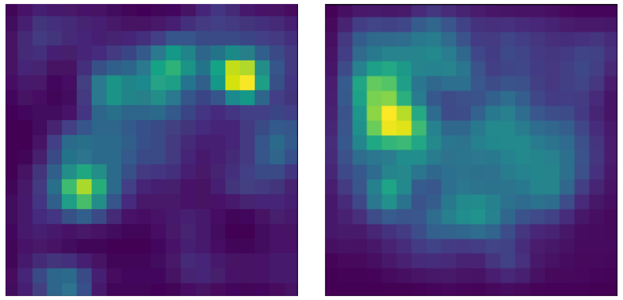
Accounting for Heterogeneous Code?



Data from: *Mallory et al. 2021*



Heterogeneous Cells

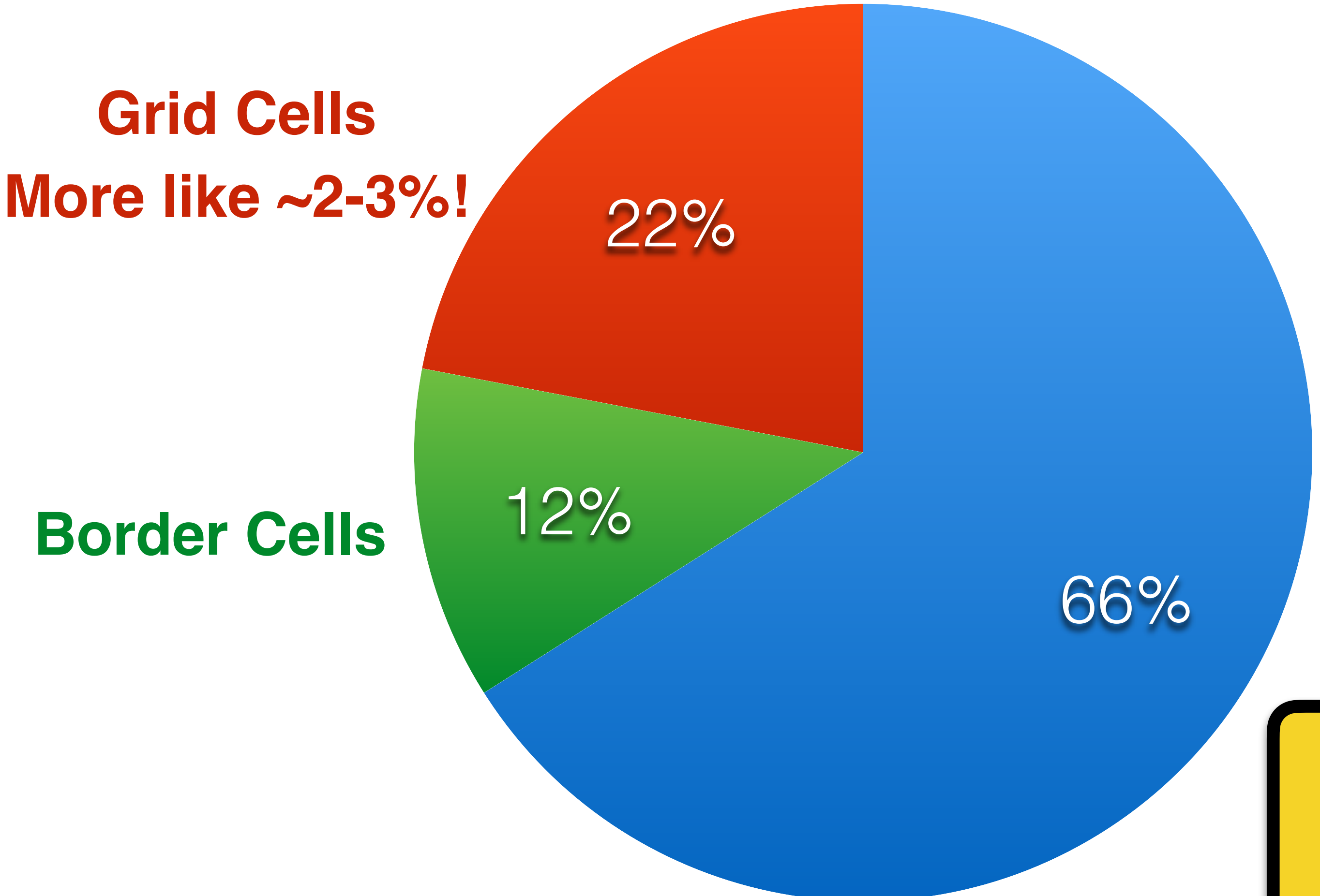


<u>Neurobiological Puzzle(s):</u>

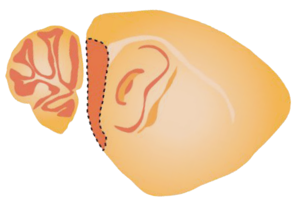
1. How might we characterize what these heterogeneous cells do?

2. What functional role do these cells serve in the circuit, if any?

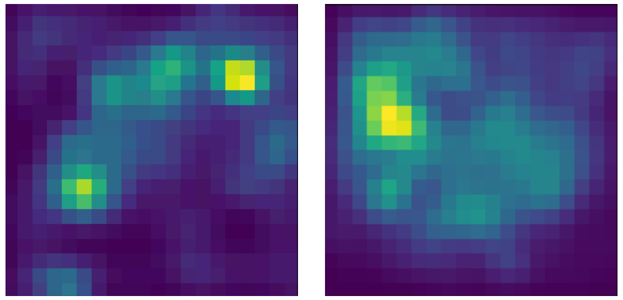
A Task-Optimized Account of Heterogeneity



Data from: *Mallory et al. 2021*

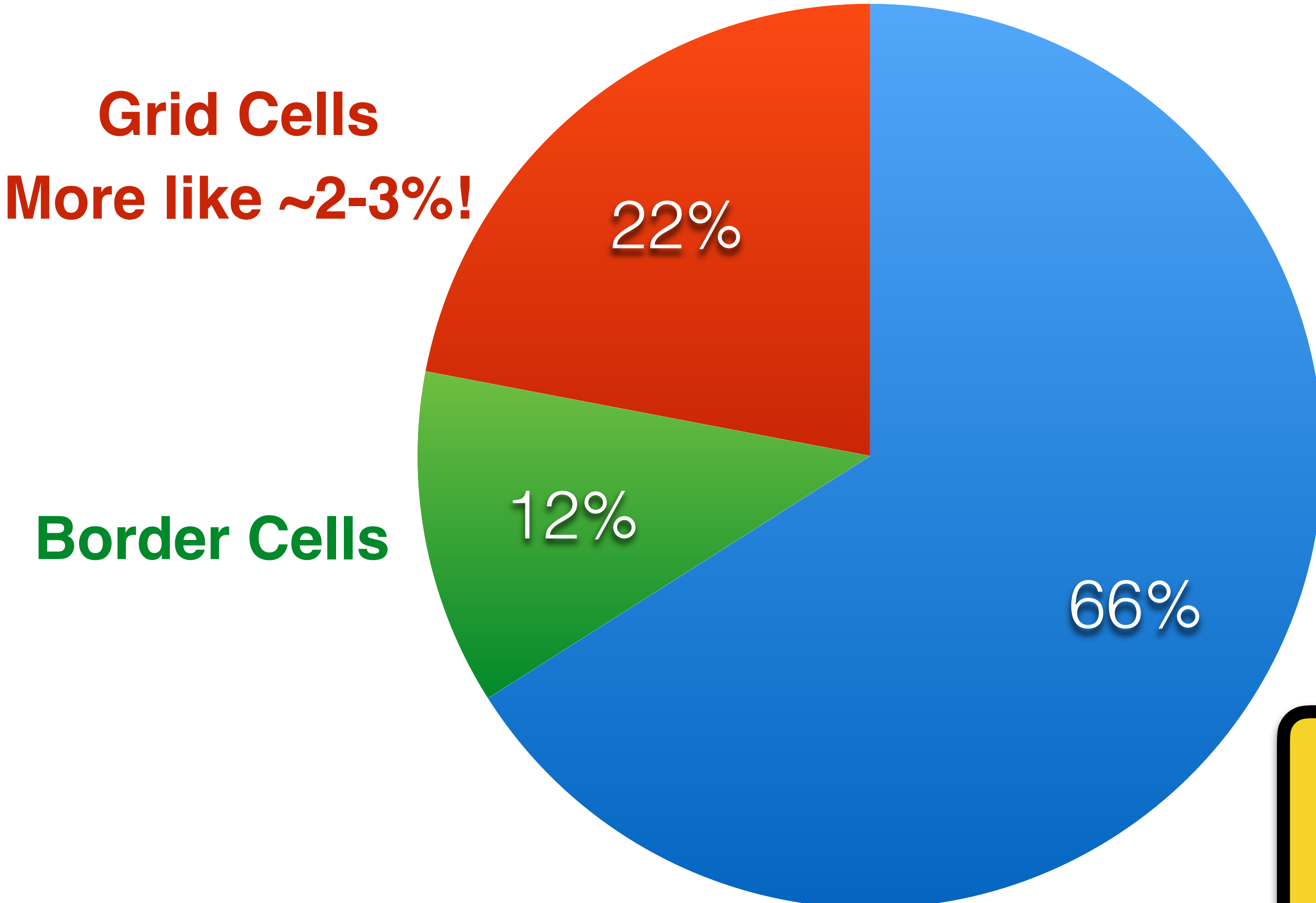


Heterogeneous Cells

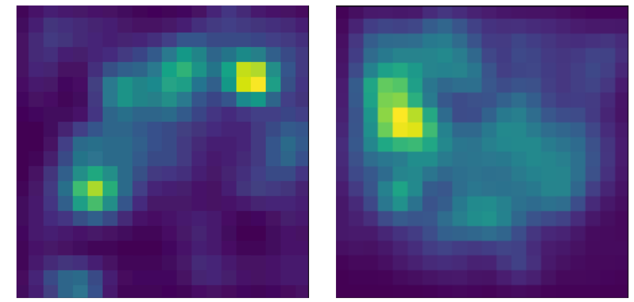


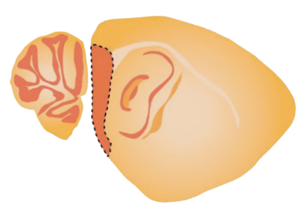
Heterogeneous cell types emerge in networks optimized for path integration!

A Task-Optimized Account of Heterogeneity



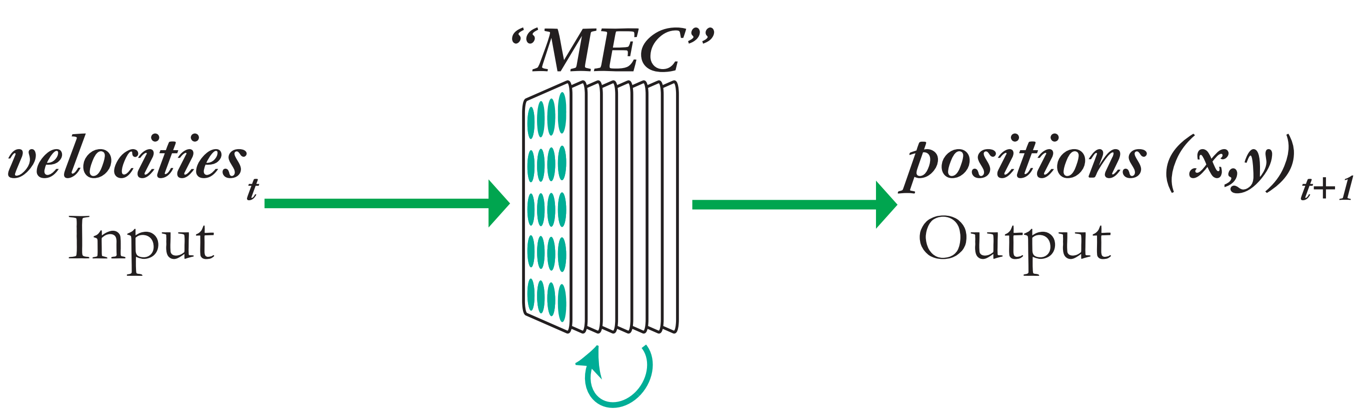
Heterogeneous Cells



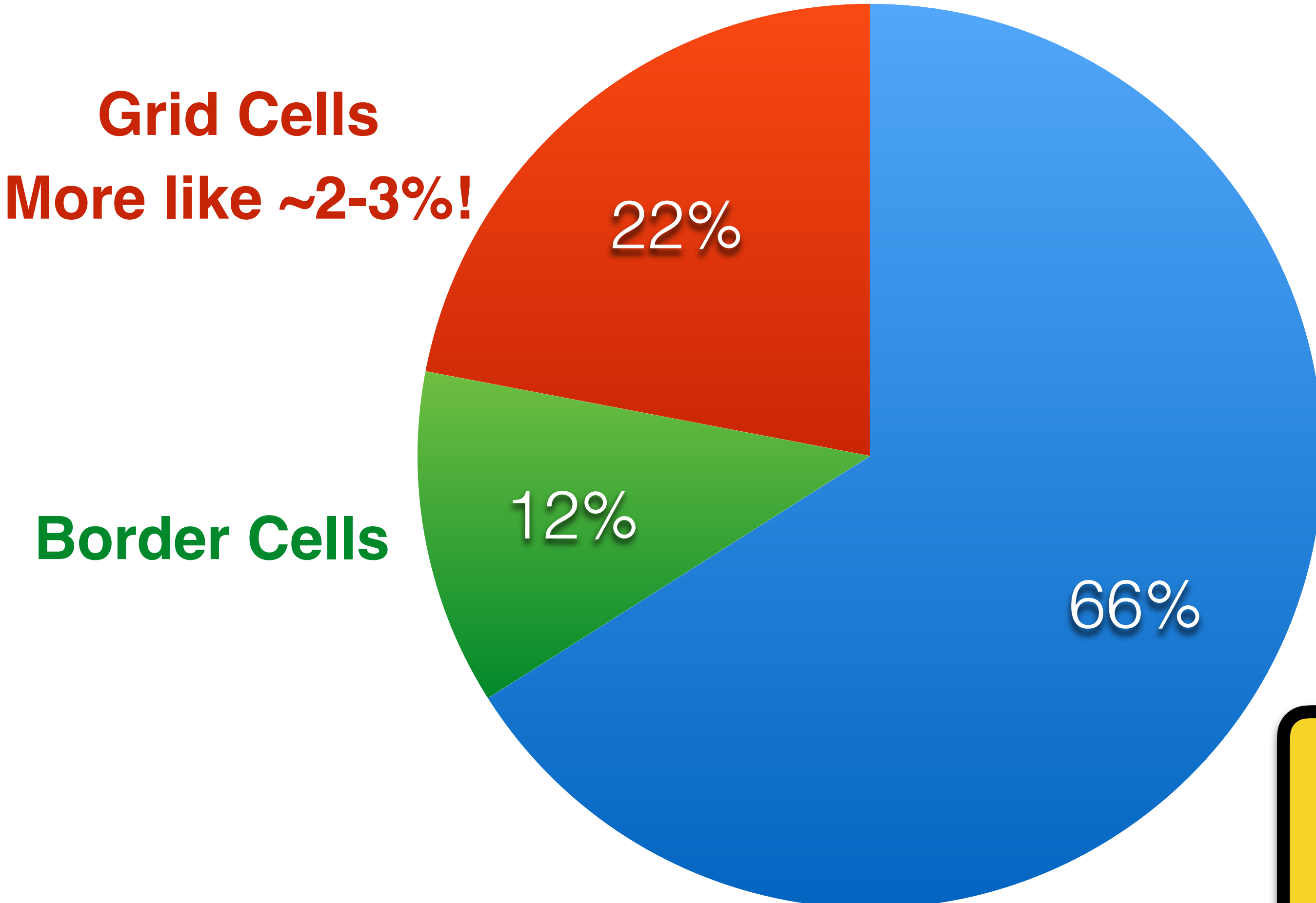


Data from: *Mallory et al.* 2021

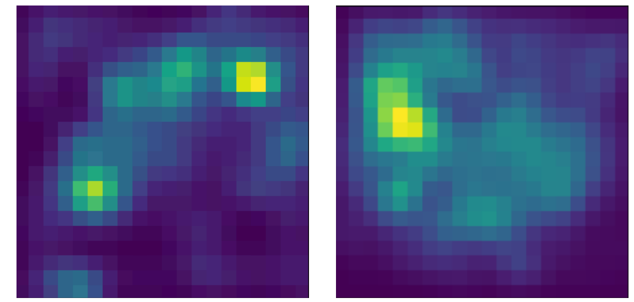
Heterogeneous cell types emerge in networks optimized for path integration!



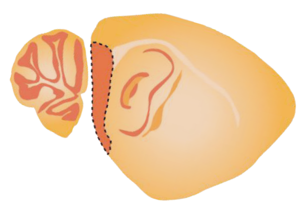
A Task-Optimized Account of Heterogeneity



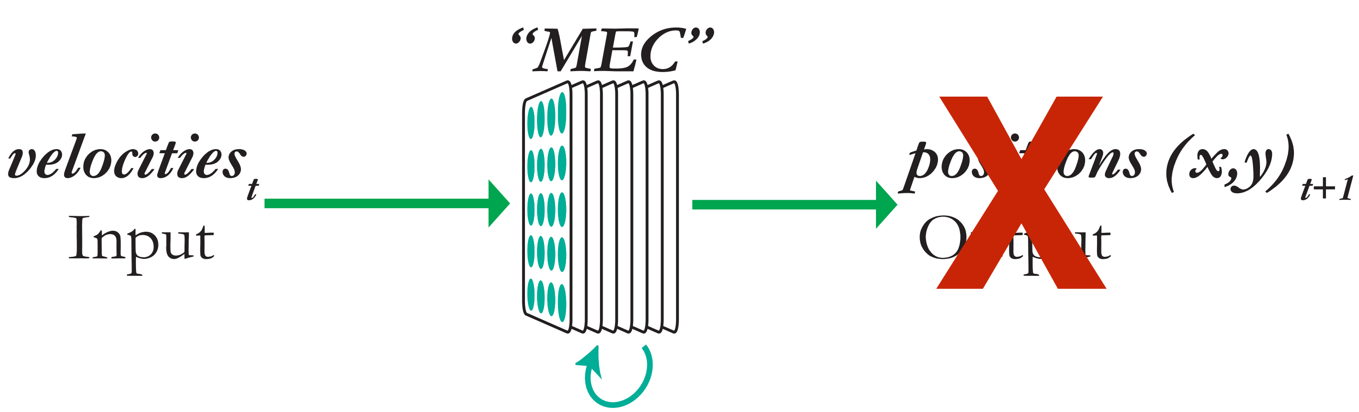
Heterogeneous Cells

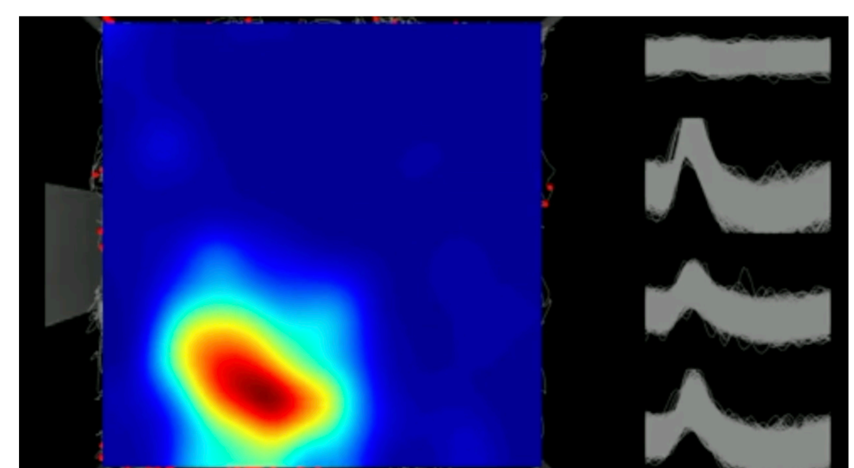


Data from: *Mallory et al. 2021*



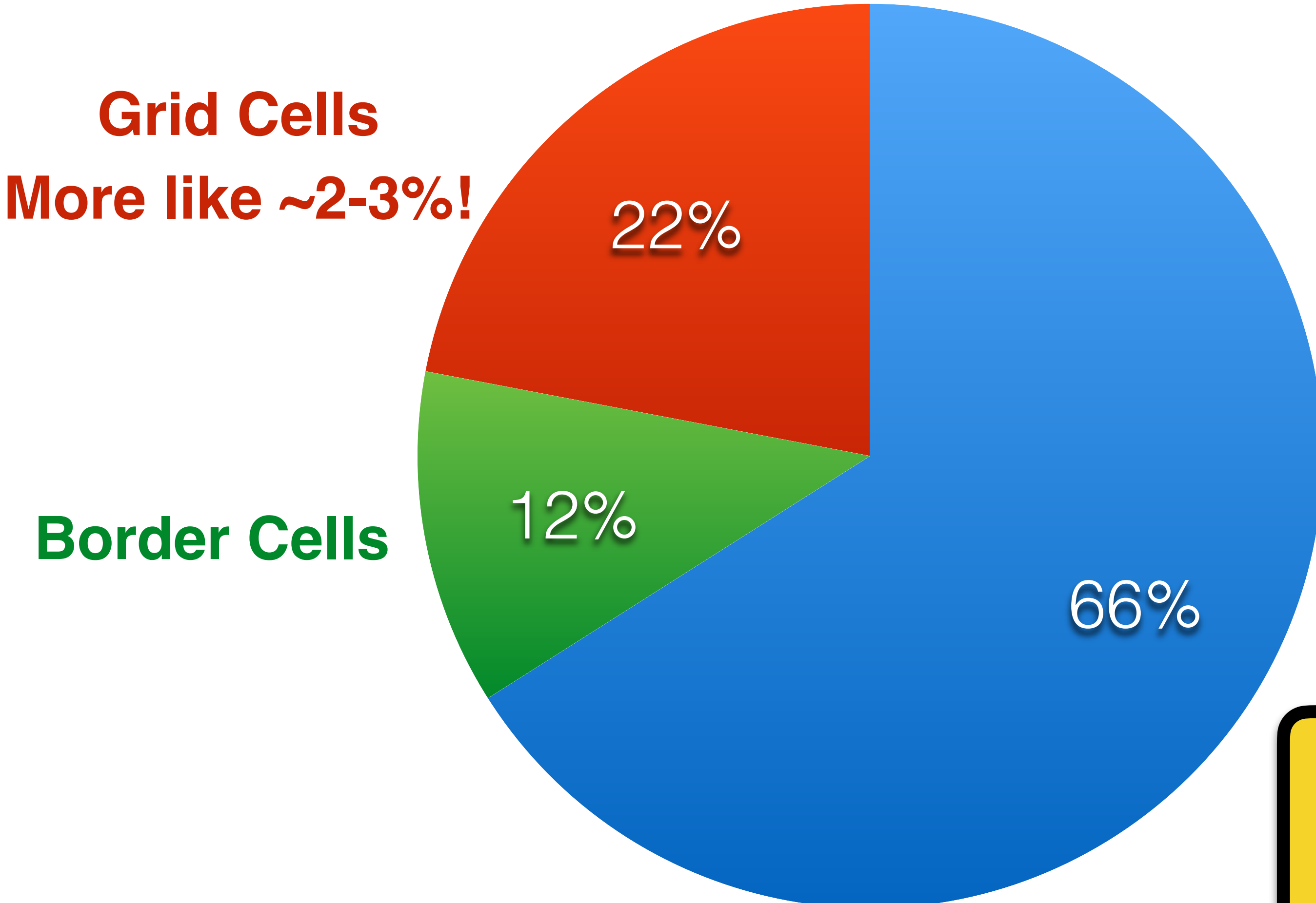
Heterogeneous cell types emerge in networks optimized for path integration!



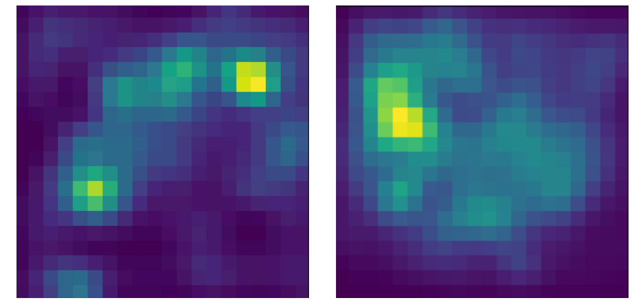


Place Cell (Hippocampus)

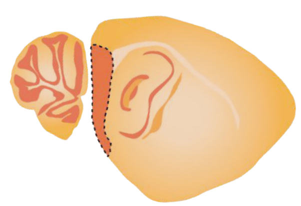
A Task-Optimized Account of Heterogeneity



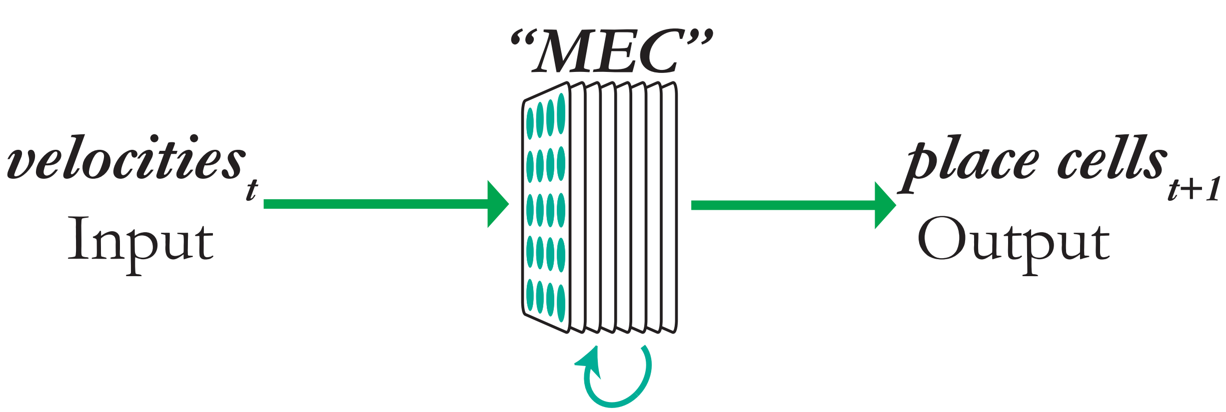
Heterogeneous Cells

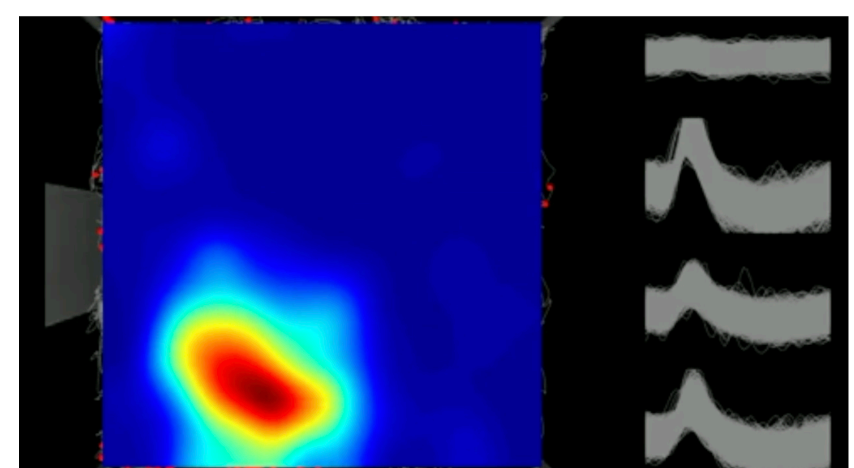


Data from: *Mallory et al. 2021*



Heterogeneous cell types emerge in networks optimized for path integration!

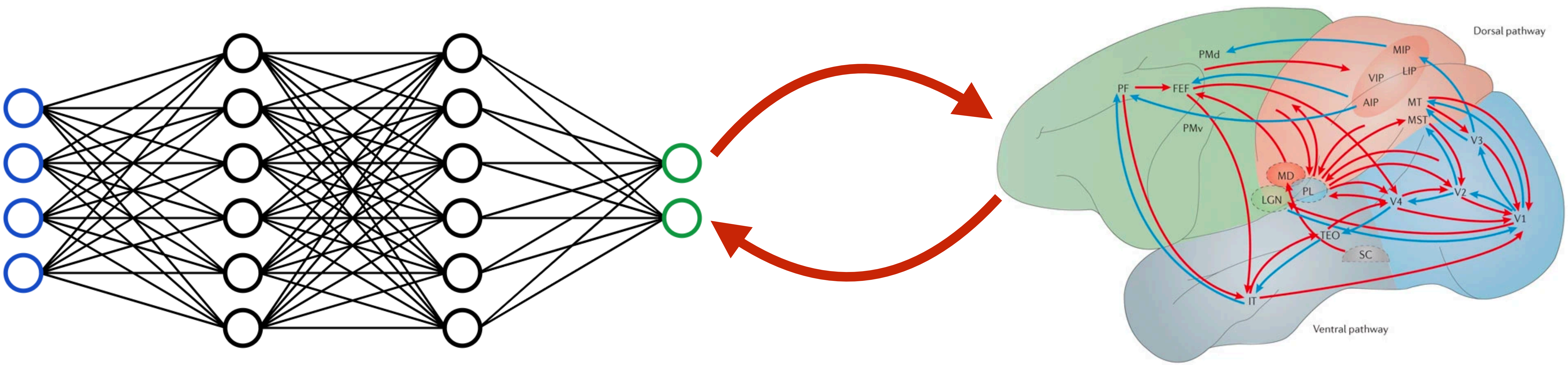




Place Cell (Hippocampus)

Task-Optimized Modeling

Design ML Algorithms Optimized to Perform Organism's Behavior under Organism's Constraints



Yields:

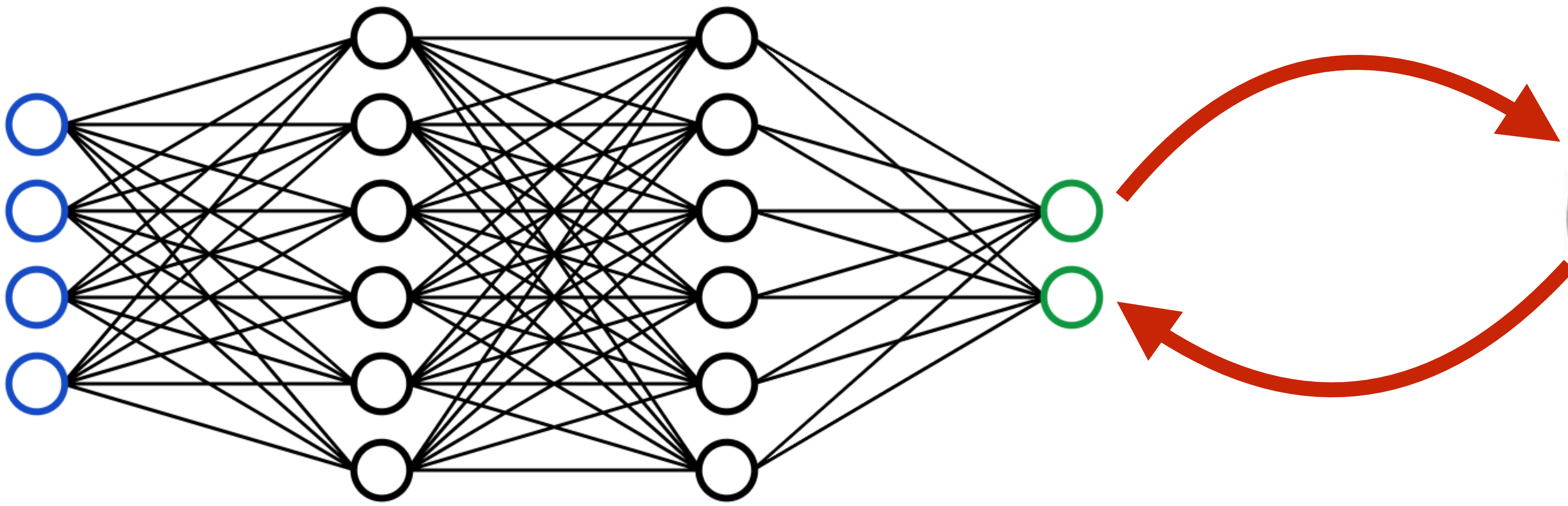
Quantitatively Accurate & Practically Useful Brain Models

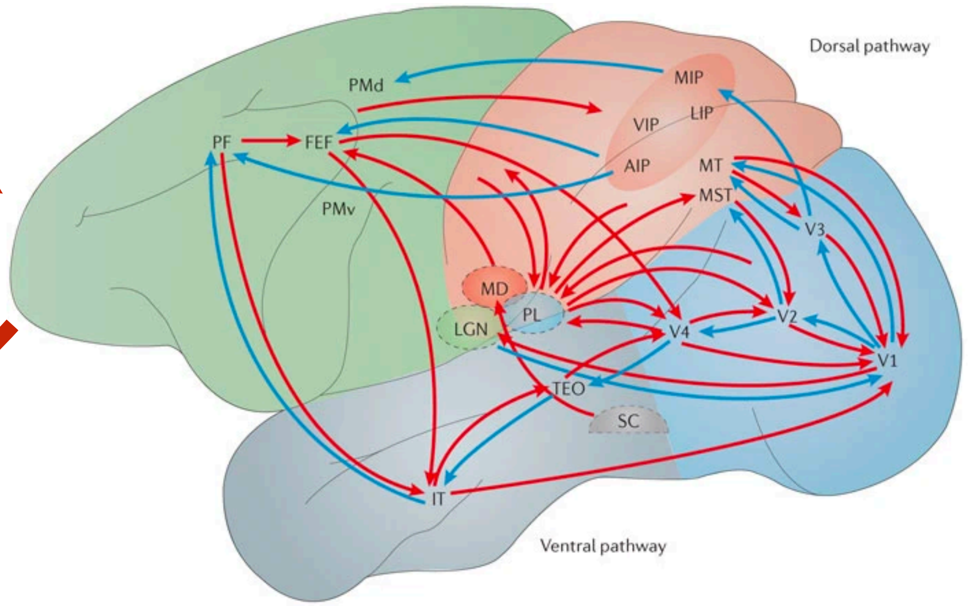
AND

Principles of *Why* Neural Responses Are As They Are

Task-Optimized Modeling

Design ML Algorithms Optimized to Perform Organism's Behavior under Organism's Constraints





Yields:

Quantitatively Accurate & Practically Useful Brain Models

AND

Principles of *Why* Neural Responses Are As They Are

Takeaways

L = *learning rule*

"Natural selection + plasticity"

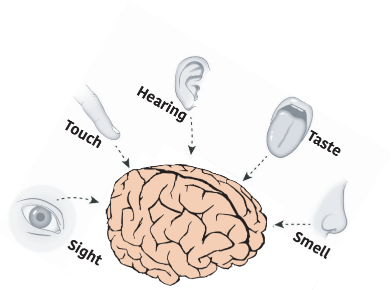
Backpropagation

T = *task loss*

"Ecological niche/ behavior"

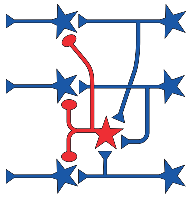
<u>Neurobiological Puzzle</u>:

Does task-optimization apply to rodents?



"Environment"

D = *data stream*



"Circuit"

A = *architecture class*

Distilling Constraints: Putting it all together

L *= learning rule*

"Natural selection + plasticity"

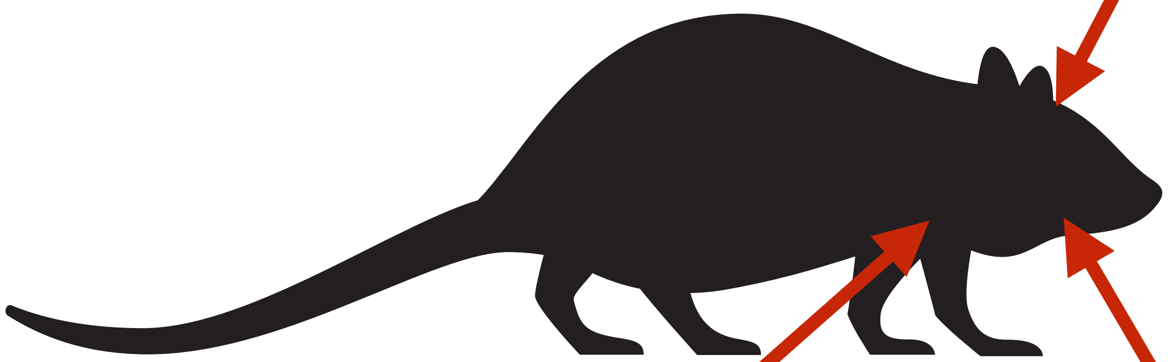
Backpropagation

T *= task loss*

"Ecological niche/ behavior"

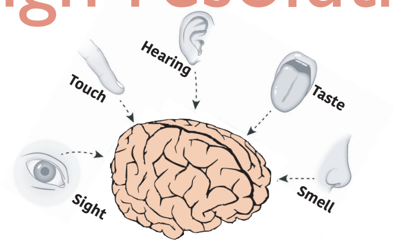
contrastive self-supervised

~~categorization supervised~~



low resolution

~~high resolution~~

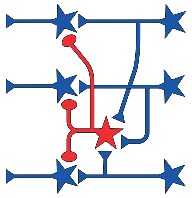


"Environment"

D *= data stream*

shallow

~~deep~~



"Circuit"

A *= architecture class*

Takeaways

Neurobiological Puzzle:

Does task-optimization apply to rodents?

Findings:

Yes!

Takeaways

Neurobiological Puzzle:

Does task-optimization apply to rodents?

Findings:

Yes!

1. Mouse visual cortex: Makes best use of the mouse's limited resources to create a general-purpose visual system.

Takeaways

Neurobiological Puzzle:

Does task-optimization apply to rodents?

Findings:

Yes!

1. Mouse visual cortex: Makes best use of the mouse's limited resources to create a general-purpose visual system.

2. Rodent medial entorhinal cortex: Grid cells are not uniquely relevant to navigation. Both heterogeneous and grid cells arise jointly through task-optimization.

Broad Takeaways

- Role of Recurrent Processing During Object Recognition

Enables the primate ventral stream to attain high object recognition ability through temporal rather than spatial complexity, specifically conserving on the number of neurons.

- Visually-Grounded Mental Simulation

Mental simulation crucially relies on explicit future prediction of a "factorized description" of visual scenes, where this "factorized description" is strongly constrained and must enable a wide range of dynamic sensorimotor abilities.

- Vision and Navigation in Rodents

Both mouse visual cortex and rodent medial entorhinal cortex are best explained by a process of biological performance optimization on a suitable task objective.

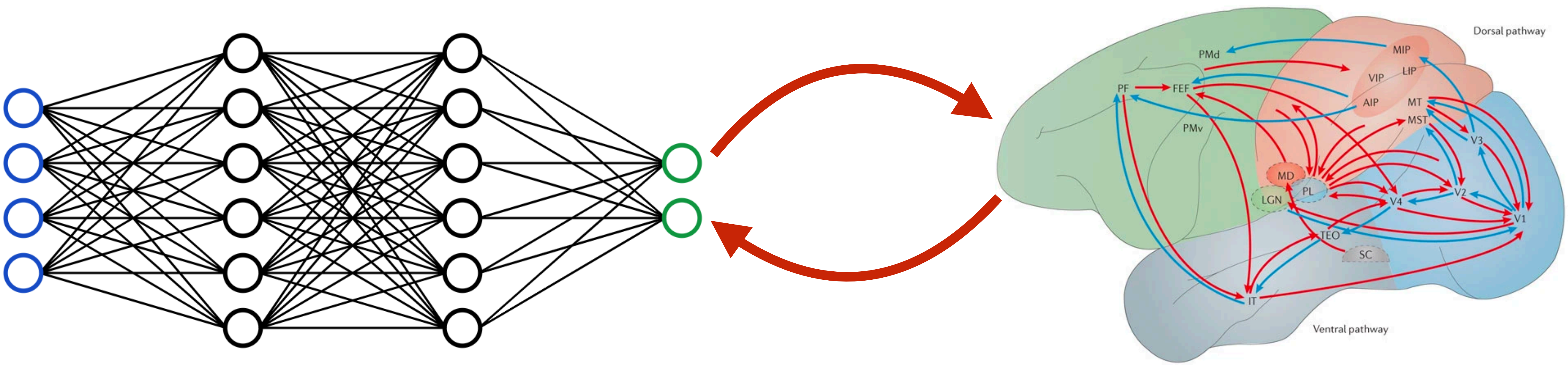
Outline

- Role of Recurrent Processing During Object Recognition
- Visually-Grounded Mental Simulation
- Vision and Navigation in Rodents

- Future Directions

Task-Optimized Modeling

Design ML Algorithms Optimized to Perform Organism's Behavior under Organism's Constraints



Yields:

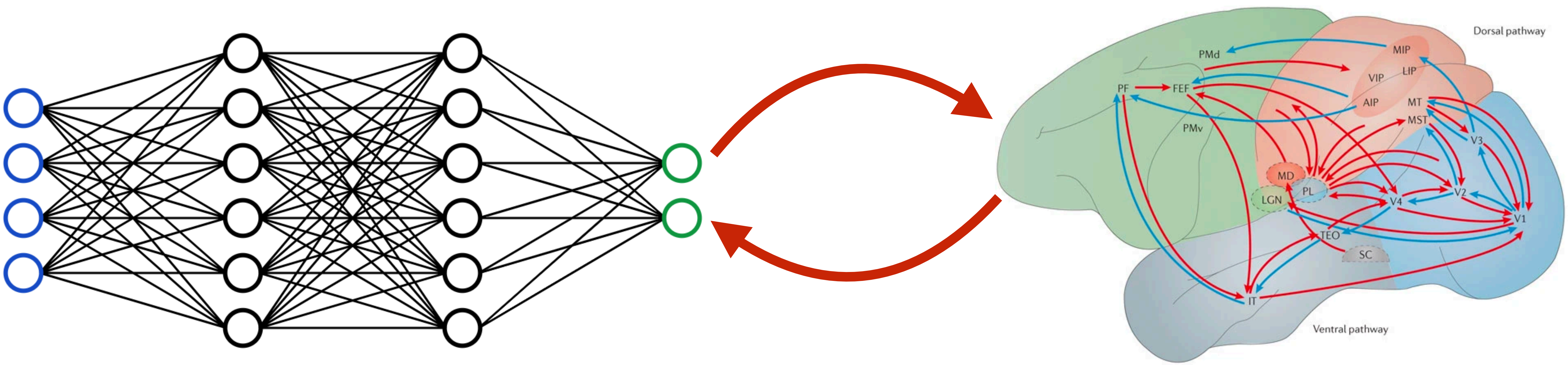
Quantitatively Accurate & Practically Useful Brain Models

AND

Principles of *Why* Neural Responses Are As They Are

Task-Optimized Modeling

Design ML Algorithms Optimized to Perform Organism's Behavior under Organism's Constraints



Yields:

Quantitatively Accurate & Practically Useful Brain Models

AND

Principles of *Why* Neural Responses Are As They Are

Task-Optimized Models of **Individual Areas**

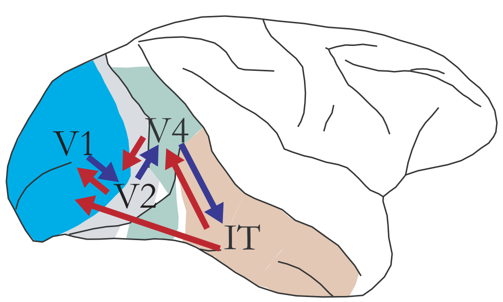
L = *learning rule*

"Natural selection + plasticity"

T = *task loss*

"Ecological niche/ behavior"

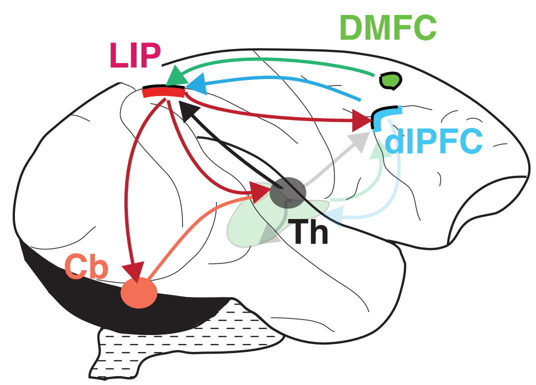
primate vision



Nayebi*, Bear*, Kubilius* et al NeurIPS (2018)

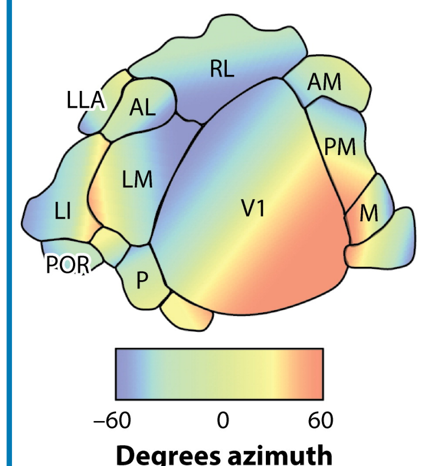
Nayebi et al Neural Computation (2022)

primate & human mental simulation



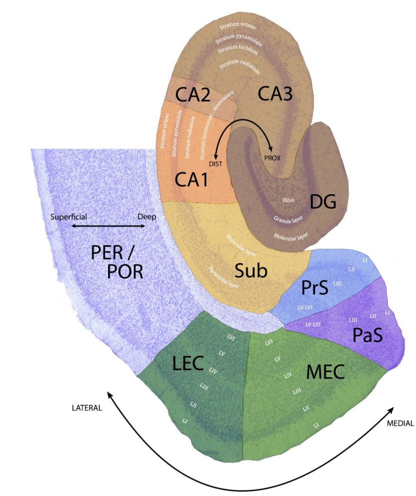
Nayebi et al NeurIPS (2023)

mouse vision



Nayebi*, Kong* et al PLOS Comp. Biol. (2023)

rodent MEC/HPC



Nayebi et al NeurIPS (2021)

1. Recurrent Processing During Object Recognition

2. Visually-Grounded Mental Simulation

3. Vision and Navigation in Rodents

"Environment"

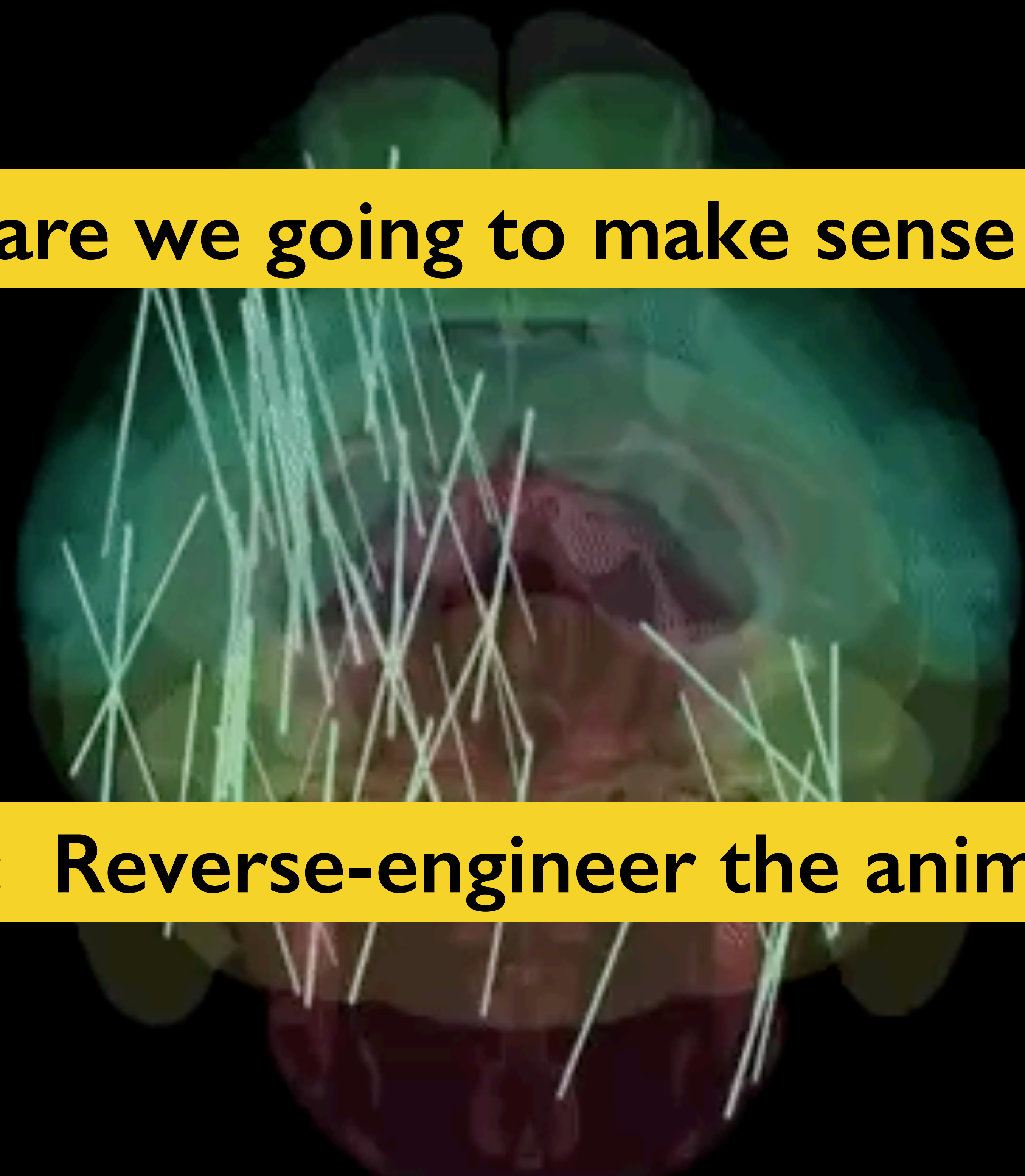
D = *data stream*

"Circuit"

A = *architecture class*

The Brave New World of Large-Scale Neuroscience

Whole brain. . .



Q: How are we going to make sense of all this data?

A: Reverse-engineer the animal brain.

International Brain Laboratory 2022

. . . awake, behaving animals (rodents)

Next Steps: Build Artificial Organisms

Next Steps: Build Artificial **Rodents**

Integrated, Task-Optimized Model of the Rodent

Why?

The *de facto* organism of choice in neuroscience.

Next Steps: Build Artificial **Rodents**

Integrated, Task-Optimized Model of the Rodent

Why?

The *de facto* organism of choice in neuroscience.

Rodents perform interesting *embodied* behaviors:

- Navigation & planning
- Flexible motor control
- Autonomous (and trained) decision making

Next Steps: Build Artificial **Rodents**

Integrated, Task-Optimized Model of the Rodent

Why?

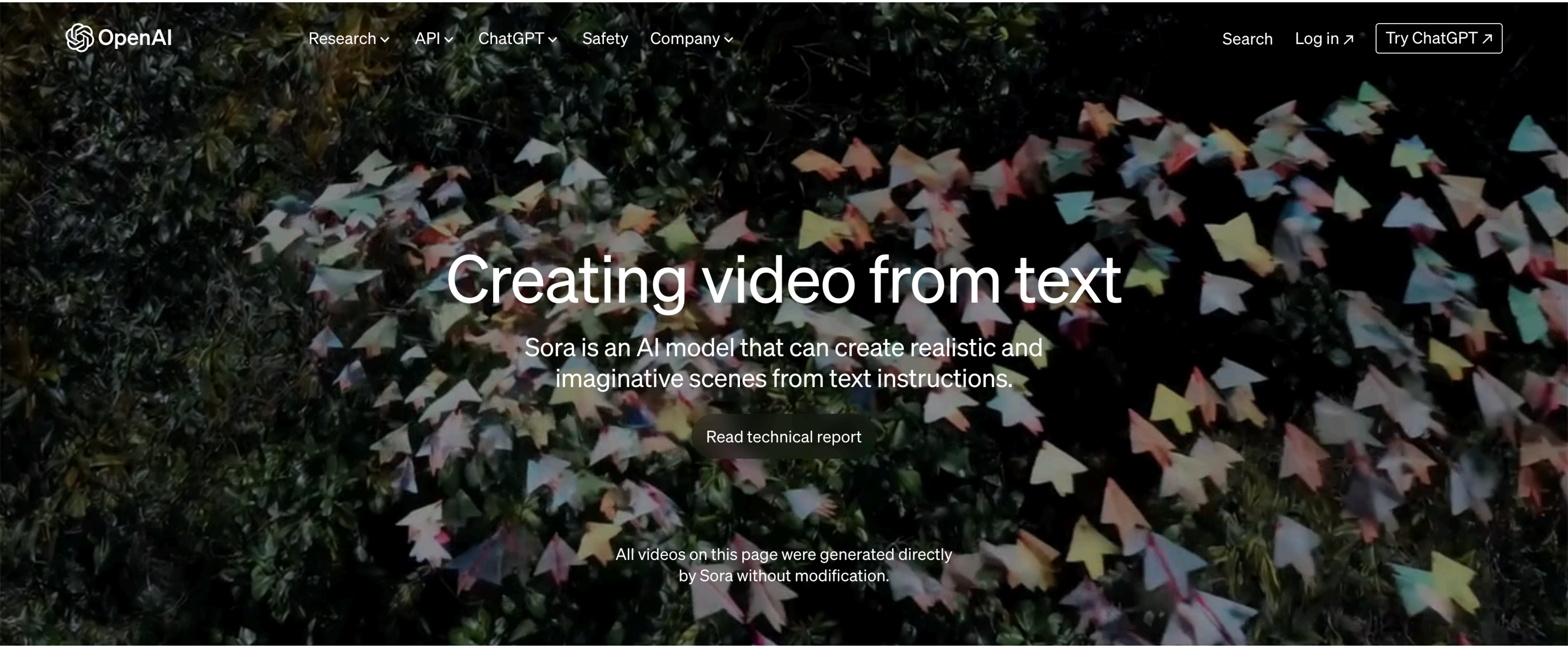
The *de facto* organism of choice in neuroscience.

Rodents perform interesting *embodied* behaviors:

- Navigation & planning
- Flexible motor control
- Autonomous (and trained) decision making

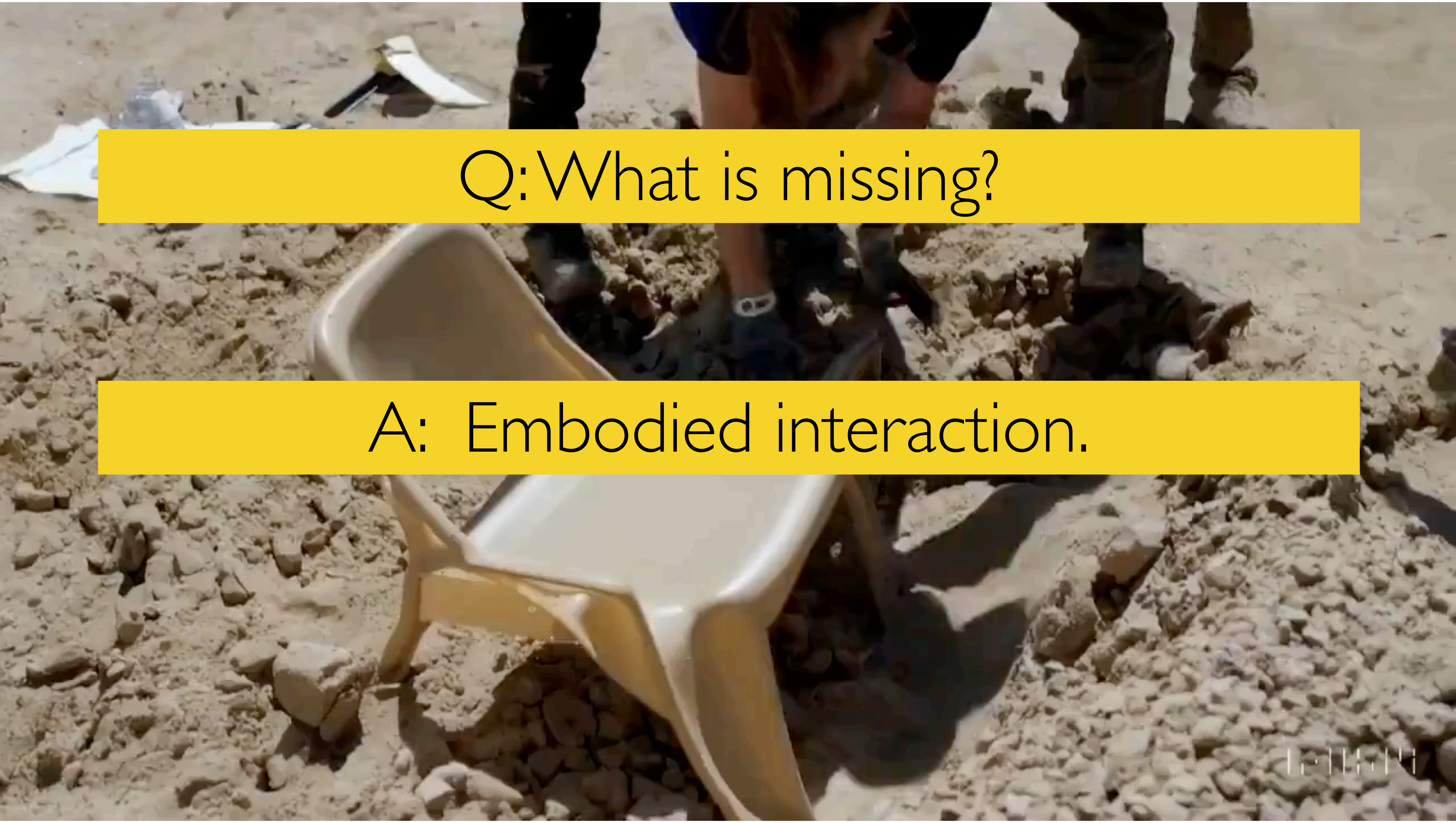
Current AI Struggles to Understand the Physical World

OpenAI Sora,
February 2024



Current AI Struggles to Understand the Physical World

OpenAI Sora,
February 2024



Q: What is missing?

A: Embodied interaction.

Next Steps: Build Artificial **Rodents**

Integrated, Task-Optimized Model of the Rodent

Why?

The *de facto* organism of choice in neuroscience.

Rodents perform interesting *embodied* behaviors:

- Navigation & planning
- Flexible motor control
- Autonomous (and trained) decision making

Main Directions

Integrated, Task-Optimized Model of the Rodent

Direction 1:
Building the Embodied Agent

Direction 2:
Applying the Embodied Agent to Neuroscience Questions

Main Directions

Integrated, Task-Optimized Model of the Rodent

Direction 1:
Building the Embodied Agent

Direction 2:
Applying the Embodied Agent to Neuroscience Questions

Next Steps: Building the Embodied Agent

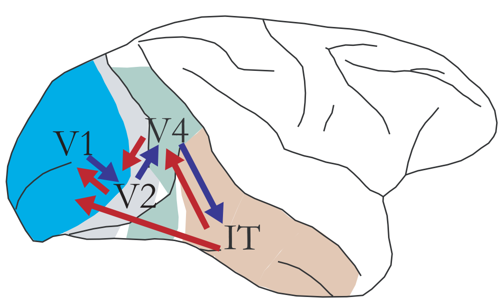
L = *learning rule*

"Natural selection + plasticity"

T = *task loss*

"Ecological niche/ behavior"

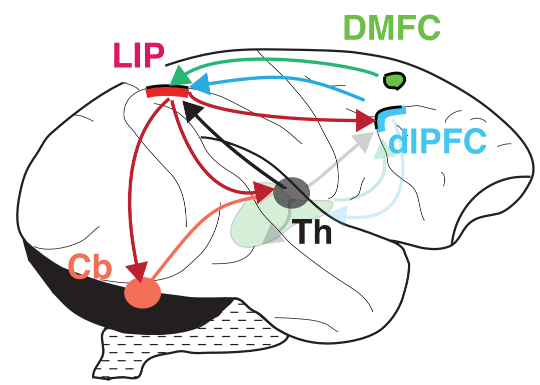
primate vision



Nayebi*, Bear*, Kubilius* et al NeurIPS (2018)

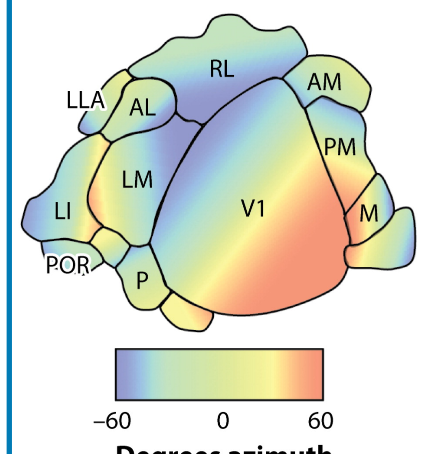
Nayebi et al Neural Computation (2022)

primate & human mental simulation



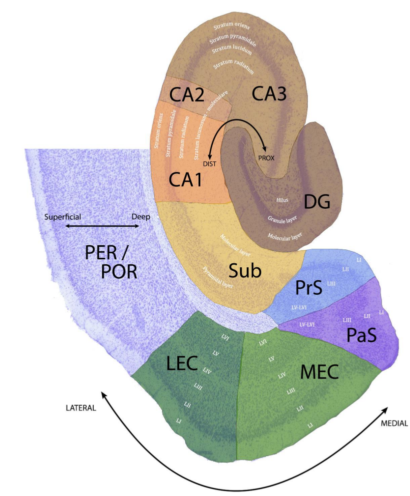
Nayebi et al NeurIPS (2023)

mouse vision



Nayebi*, Kong* et al PLOS Comp. Biol. (2023)

rodent MEC/HPC



Nayebi et al NeurIPS (2021)

1. Recurrent Processing During Object Recognition

2. Visually-Grounded Mental Simulation

3. Vision and Navigation in Rodents

"Environment"

D = *data stream*

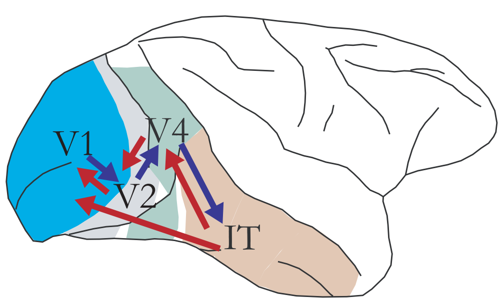
"Circuit"

A = *architecture class*

Next Steps: Building the Embodied Agent

How does the brain build and use world models?

primate vision

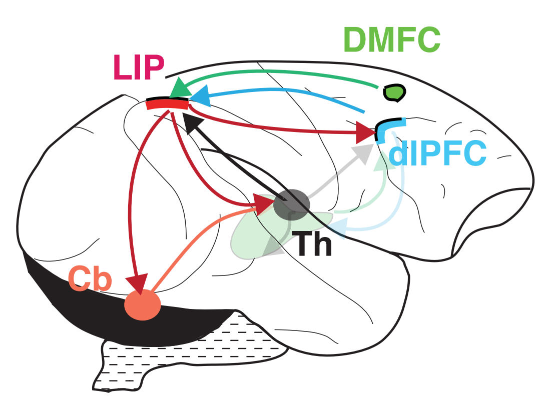


Nayebi*, Bear*, Kubilius* et al NeurIPS (2018)

Nayebi et al Neural Computation (2022)

1. Recurrent Processing During Object Recognition

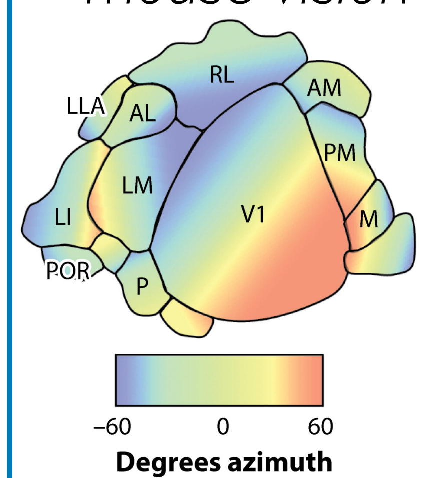
primate & human mental simulation



Nayebi et al NeurIPS (2023)

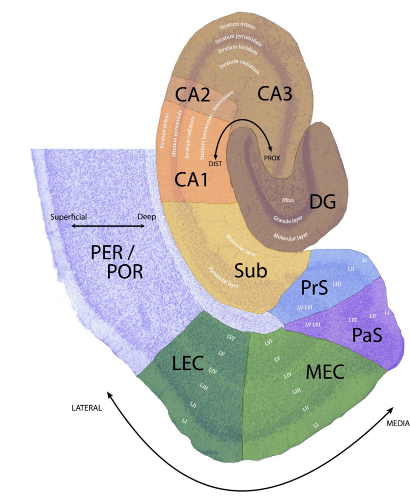
2. Visually-Grounded Mental Simulation

mouse vision



Nayebi*, Kong* et al PLOS Comp. Biol. (2023)

rodent MEC/HPC



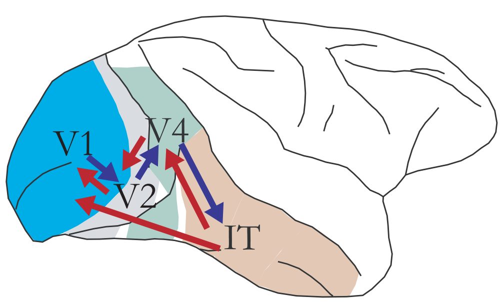
Nayebi et al NeurIPS (2021)

3. Vision and Navigation in Rodents

Next Steps: Building the Embodied Agent

How does the brain build and use world models?

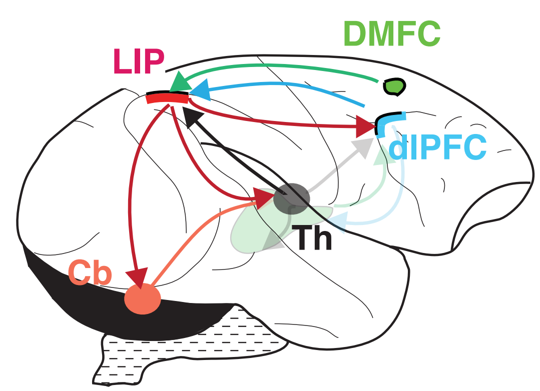
primate vision



Nayebi*, Bear*, Kubilius* et al NeurIPS (2018)

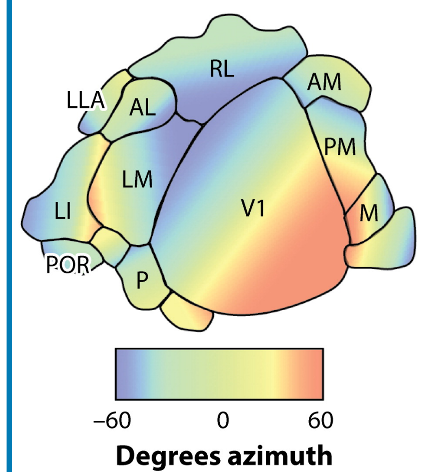
Nayebi et al Neural Computation (2022)

primate & human mental simulation



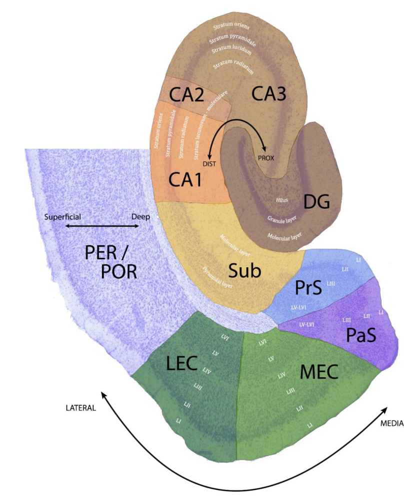
Nayebi et al NeurIPS (2023)

mouse vision



Nayebi*, Kong* et al PLOS Comp. Biol. (2023)

rodent MEC/HPC



Nayebi et al NeurIPS (2021)

1. Recurrent Processing During Object Recognition

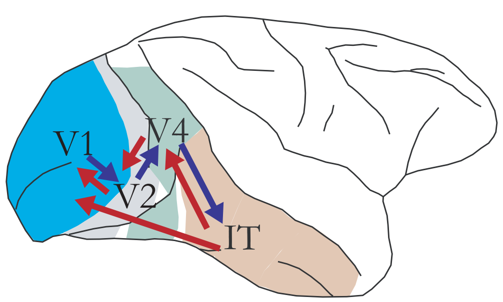
2. Visually-Grounded Mental Simulation

3. Vision and Navigation in Rodents

Next Steps: Building the Embodied Agent

How does the brain *represent*, *predict*, *plan*, and enable *action*?

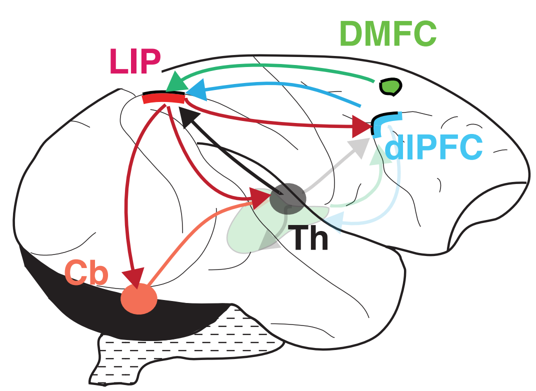
primate vision



Nayebi*, Bear*, Kubilius* et al NeurIPS (2018)

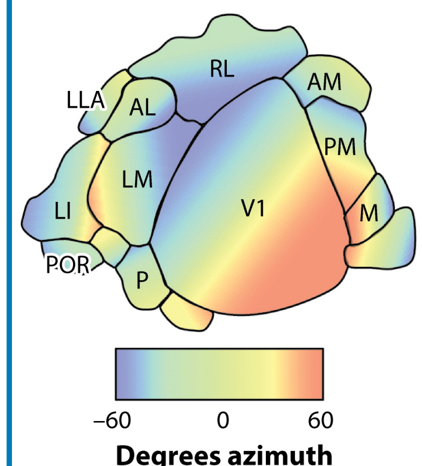
Nayebi et al Neural Computation (2022)

primate & human mental simulation



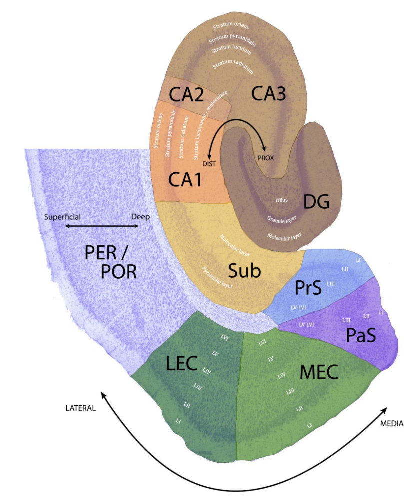
Nayebi et al NeurIPS (2023)

mouse vision



Nayebi*, Kong* et al PLOS Comp. Biol. (2023)

rodent MEC/HPC



Nayebi et al NeurIPS (2021)

1. Recurrent Processing During Object Recognition

2. Visually-Grounded Mental Simulation

3. Vision and Navigation in Rodents

Next Steps: Building the Embodied Agent

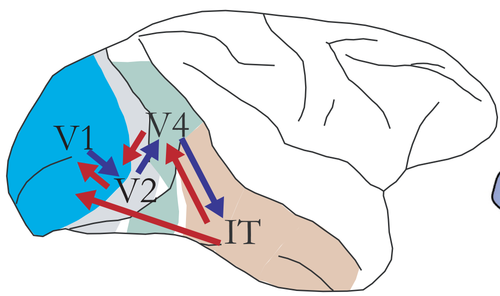
How does the brain *represent*, *predict*, *plan*, and enable *action*?

Sensory (Input) Stream

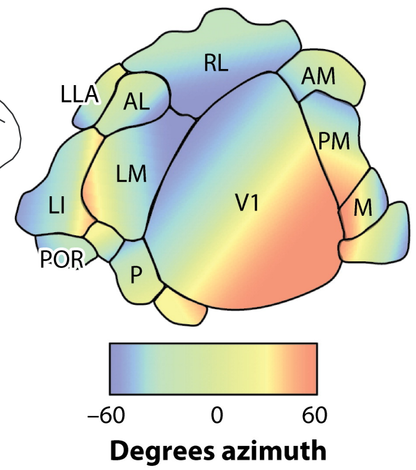
Environment

Perceptual Module

primate vision



mouse vision

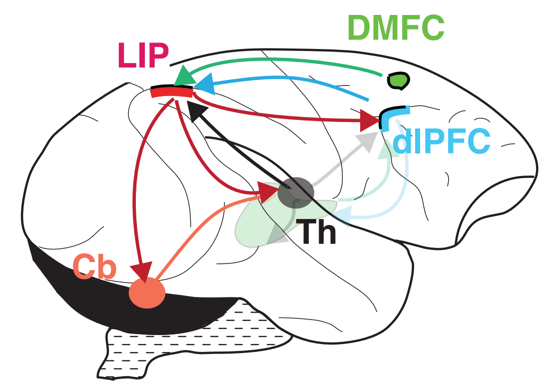


Nayebi*, Bear*, Kubilius* et al NeurIPS (2018)

Nayebi et al Neural Computation (2022)

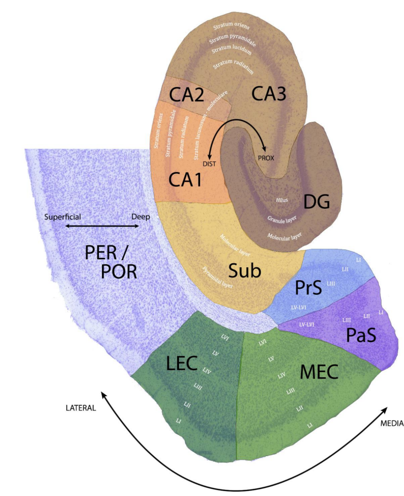
Nayebi*, Kong* et al PLOS Comp. Biol. (2023)

primate & human mental simulation



Nayebi et al NeurIPS (2023)

rodent MEC/HPC



Nayebi et al NeurIPS (2021)

2. Visually-Grounded Mental Simulation

3. Vision and Navigation in Rodents

Next Steps: Building the Embodied Agent

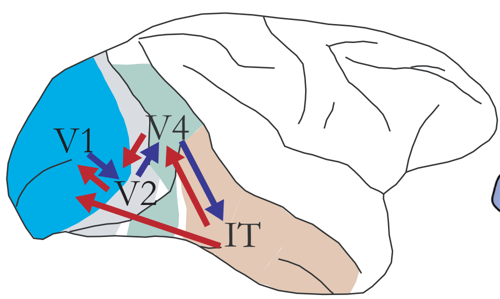
How does the brain *represent*, *predict*, *plan*, and enable *action*?

Sensory (Input) Stream

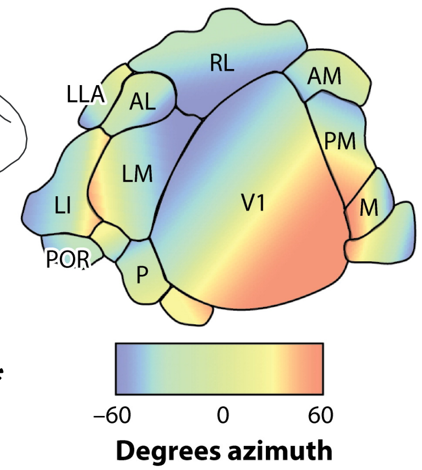
Environment

Perceptual Module

primate vision



mouse vision



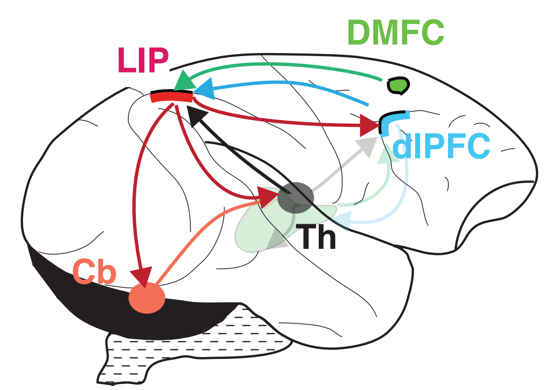
Nayebi*, Bear*, Kubilius* et al NeurIPS (2018)

Nayebi et al Neural Computation (2022)

Nayebi*, Kong* et al PLOS Comp. Biol. (2023)

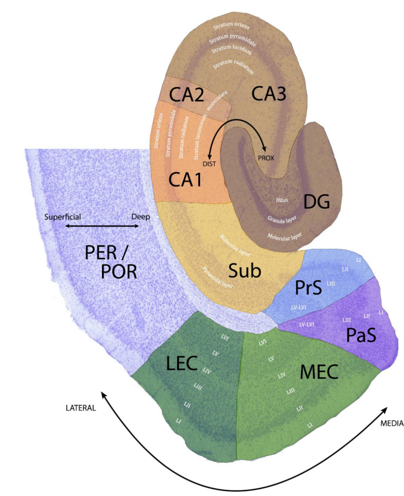
Future Inference Module

primate & human mental simulation



Nayebi et al NeurIPS (2023)

rodent MEC/HPC



Nayebi et al NeurIPS (2021)

3. Vision and Navigation in Rodents

Next Steps: Building the Embodied Agent

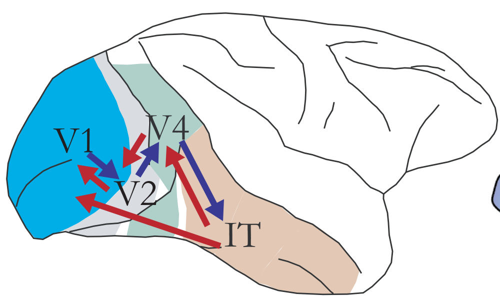
How does the brain *represent*, *predict*, *plan*, and enable *action*?

Sensory (Input) Stream

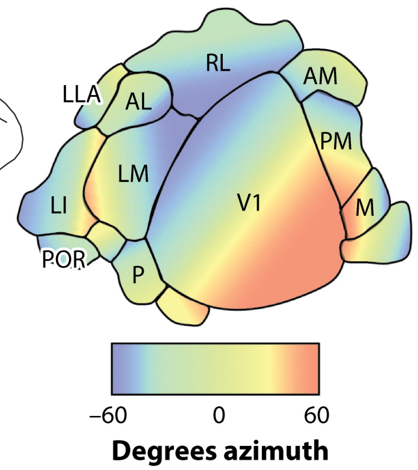
Environment

Perceptual Module

primate vision



mouse vision



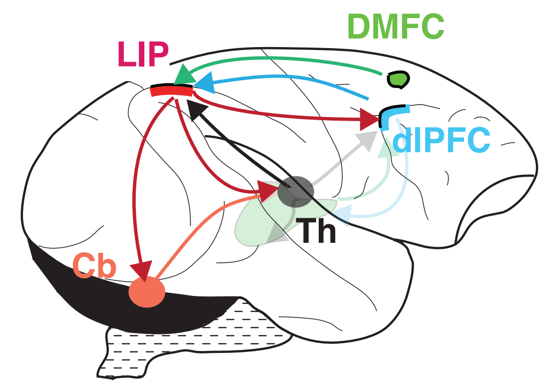
Nayebi*, Bear*, Kubilius* et al NeurIPS (2018)

Nayebi et al Neural Computation (2022)

Nayebi*, Kong* et al PLOS Comp. Biol. (2023)

Future Inference Module

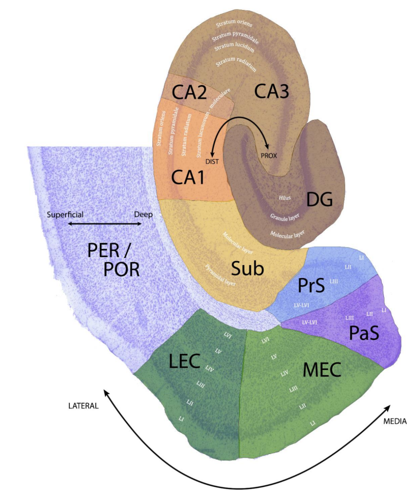
primate & human mental simulation



Nayebi et al NeurIPS (2023)

Planning Module

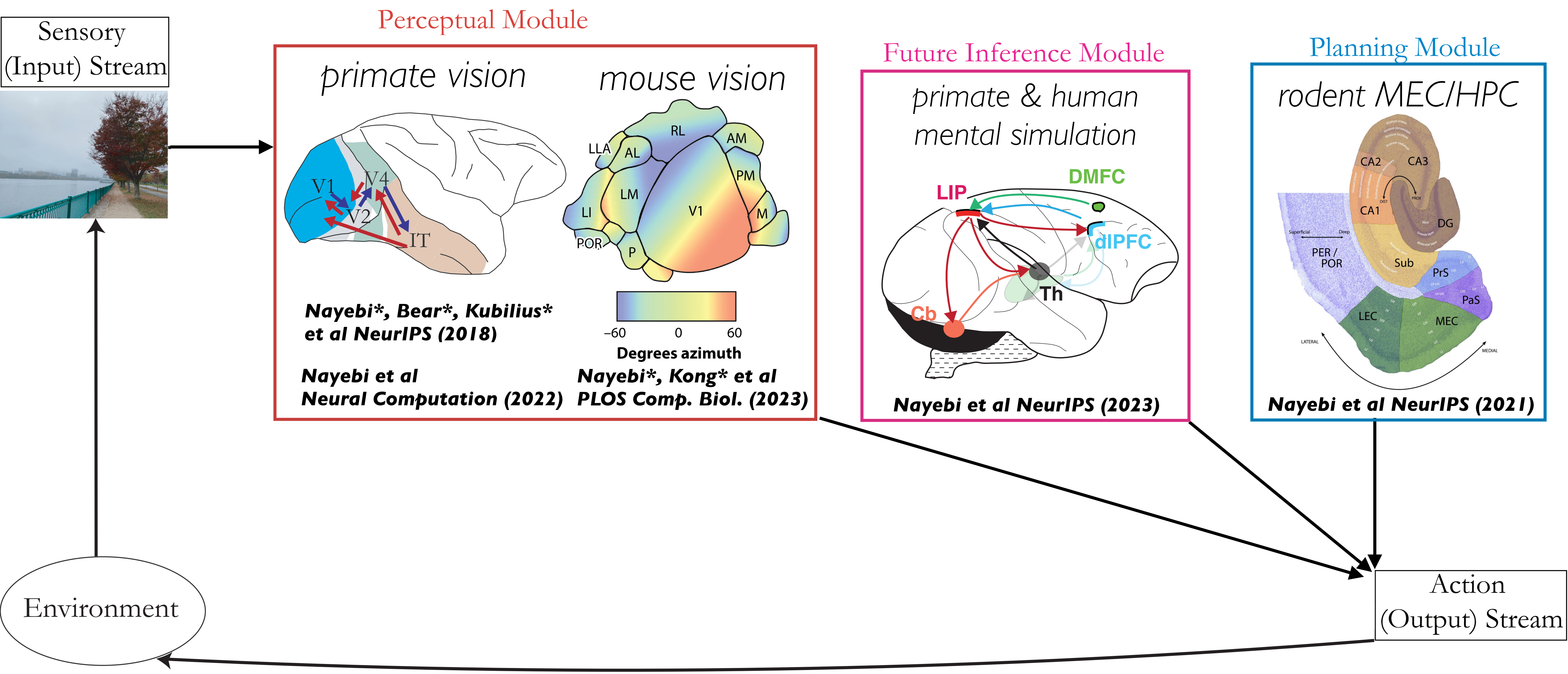
rodent MEC/HPC



Nayebi et al NeurIPS (2021)

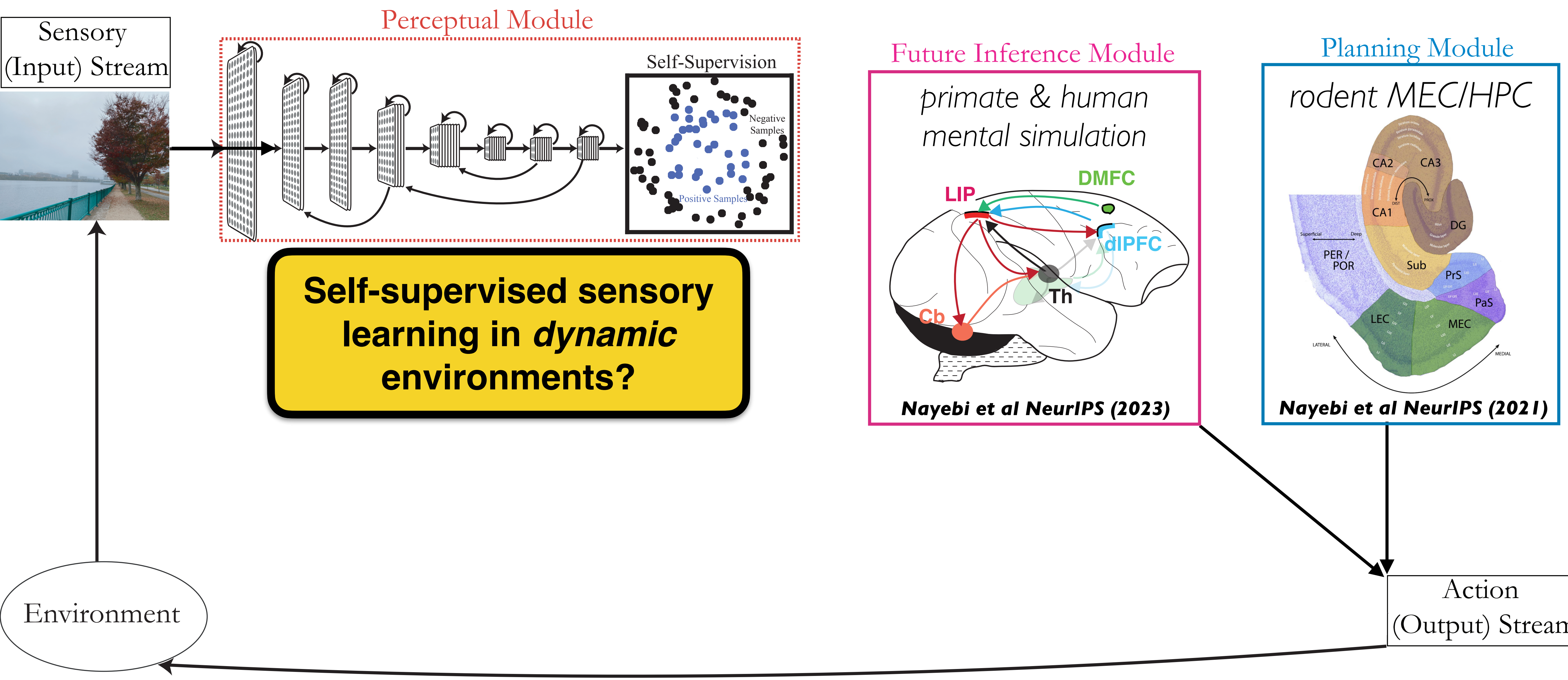
Next Steps: Building the Embodied Agent

How does the brain *represent*, *predict*, *plan*, and enable *action*?



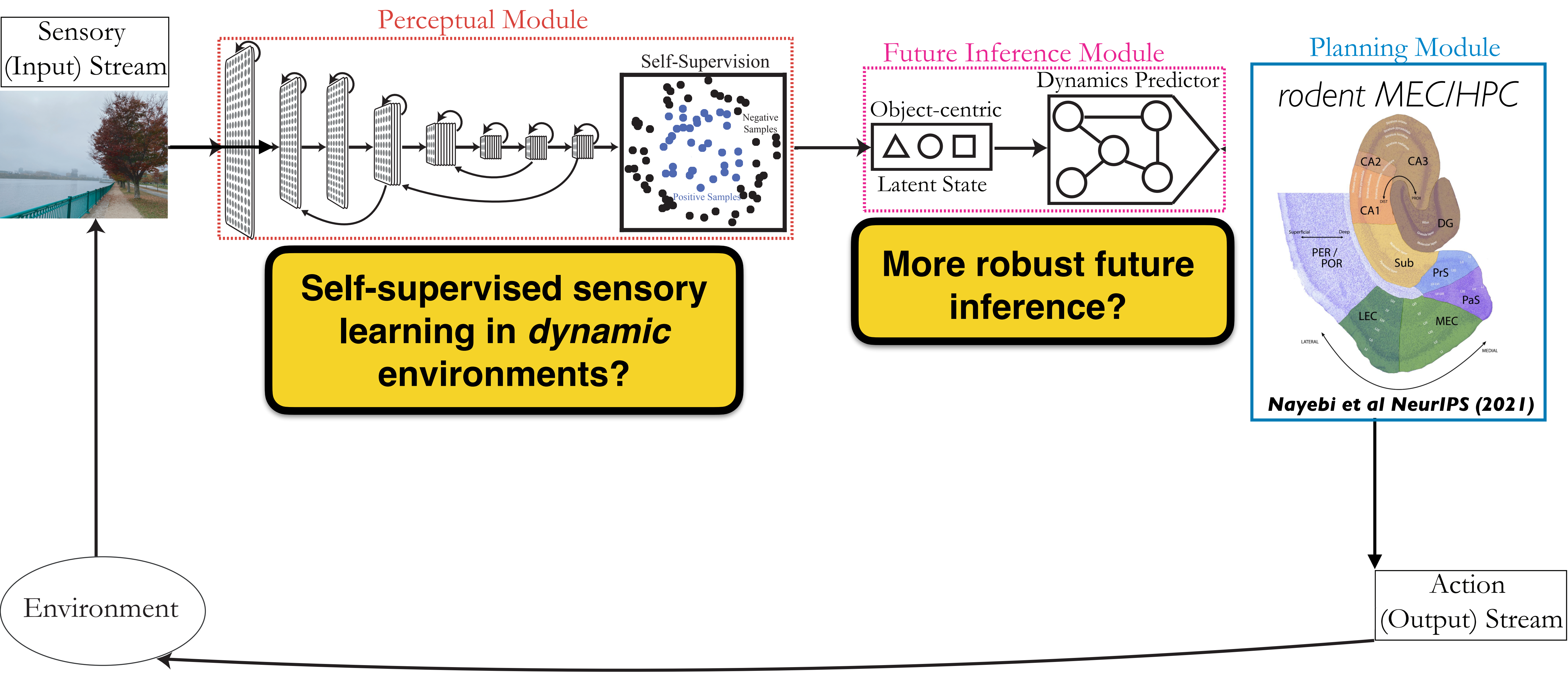
Next Steps: Building the Embodied Agent

How does the brain *represent*, *predict*, *plan*, and enable *action*?



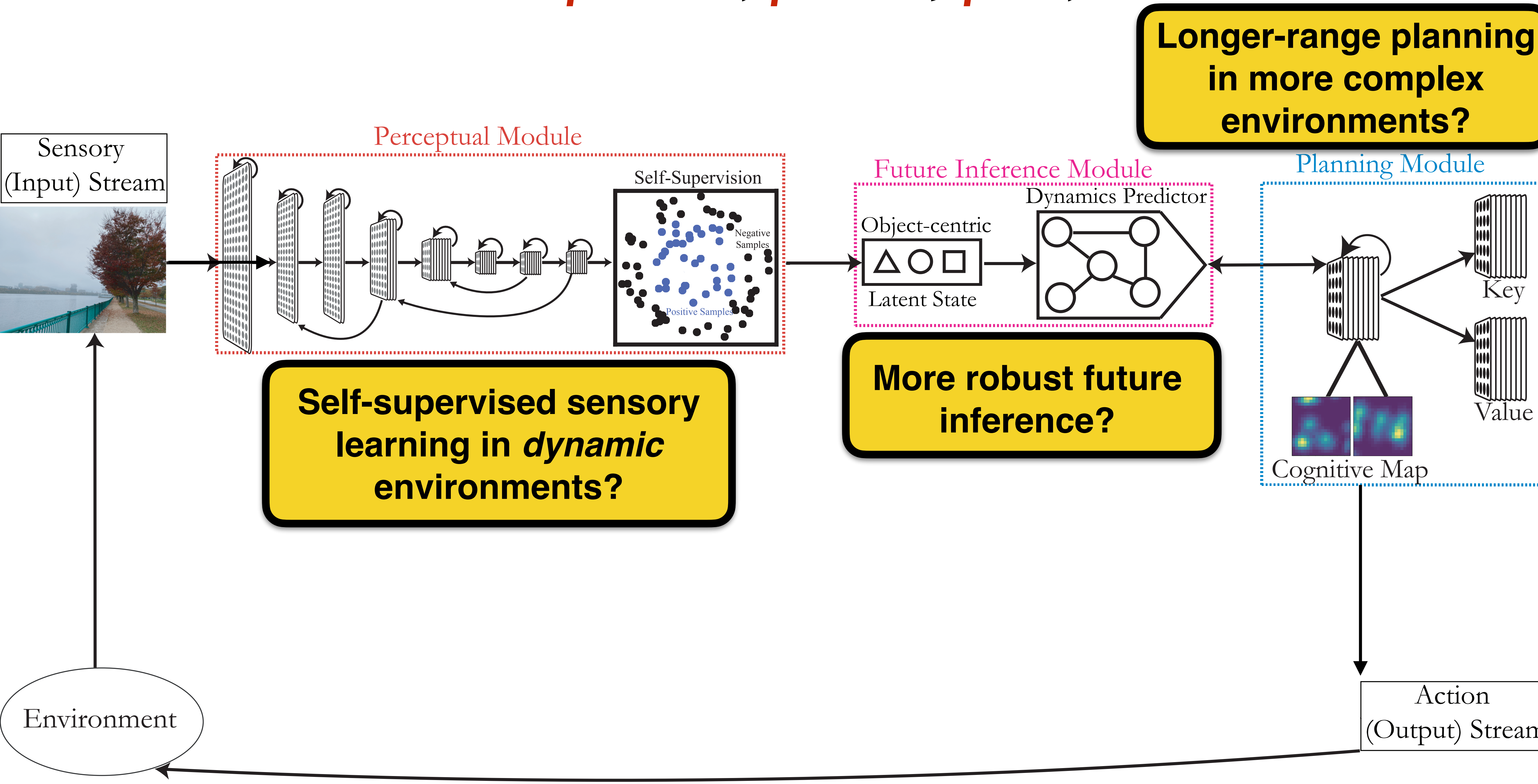
Next Steps: Building the Embodied Agent

How does the brain *represent*, *predict*, *plan*, and enable *action*?



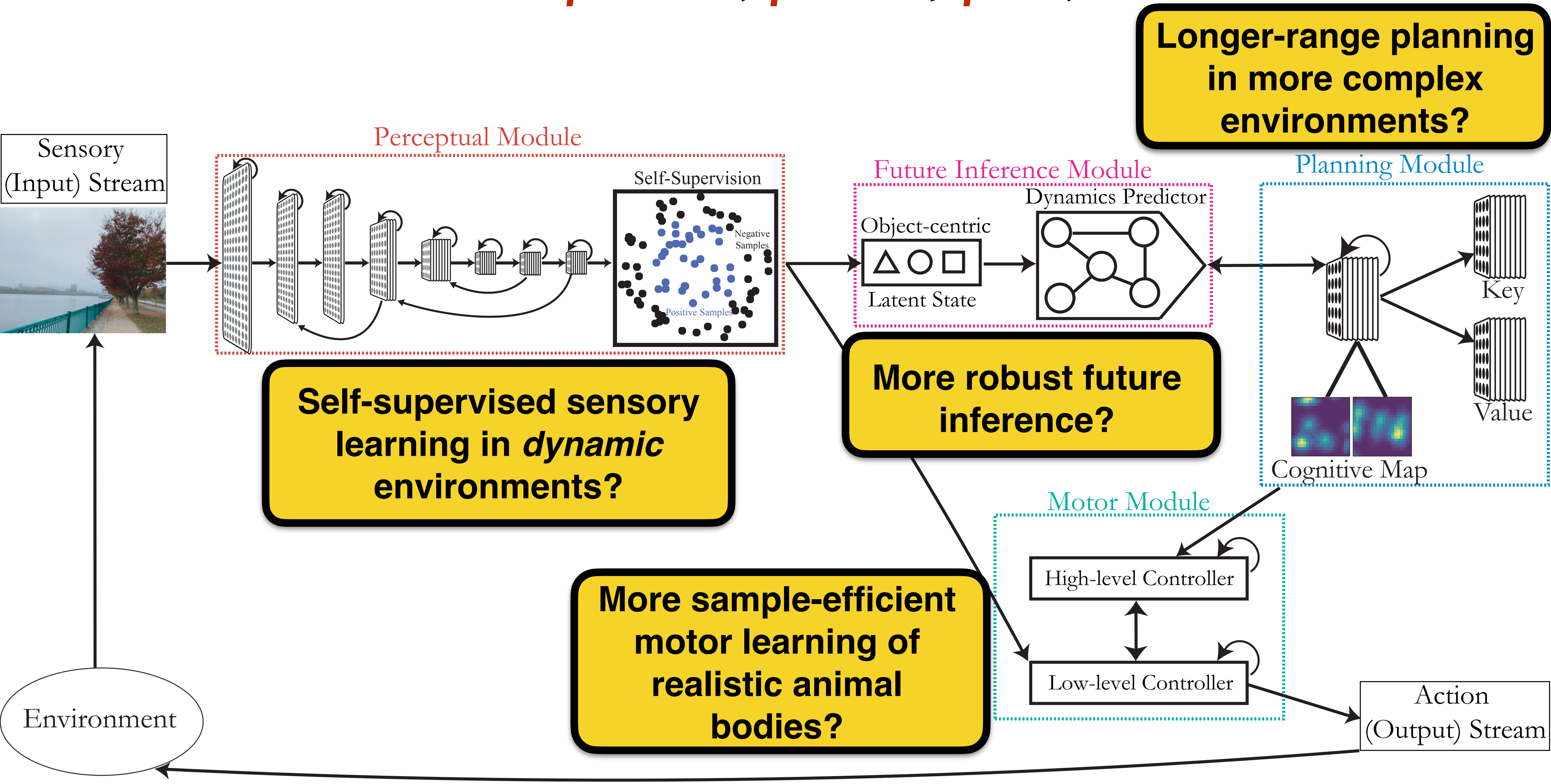
Next Steps: Building the Embodied Agent

How does the brain *represent*, *predict*, *plan*, and enable *action*?



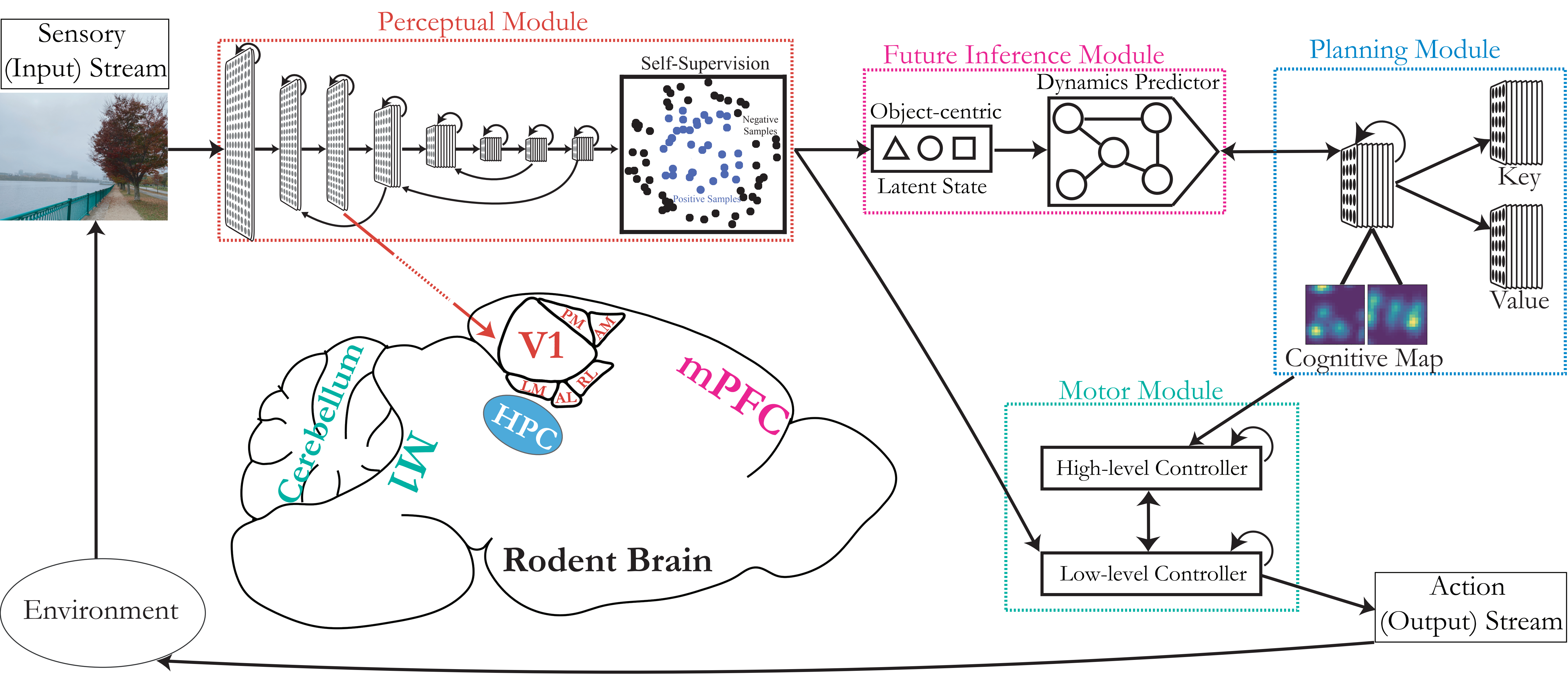
Next Steps: Building the Embodied Agent

How does the brain *represent*, *predict*, *plan*, and enable *action*?



Next Steps: Building the Embodied Agent

How does the brain *represent*, *predict*, *plan*, and enable *action*?



Main Directions

Integrated, Task-Optimized Model of the Rodent

Direction 1:
Building the Embodied Agent

Direction 2:
Applying the Embodied Agent to Neuroscience Questions

Main Directions

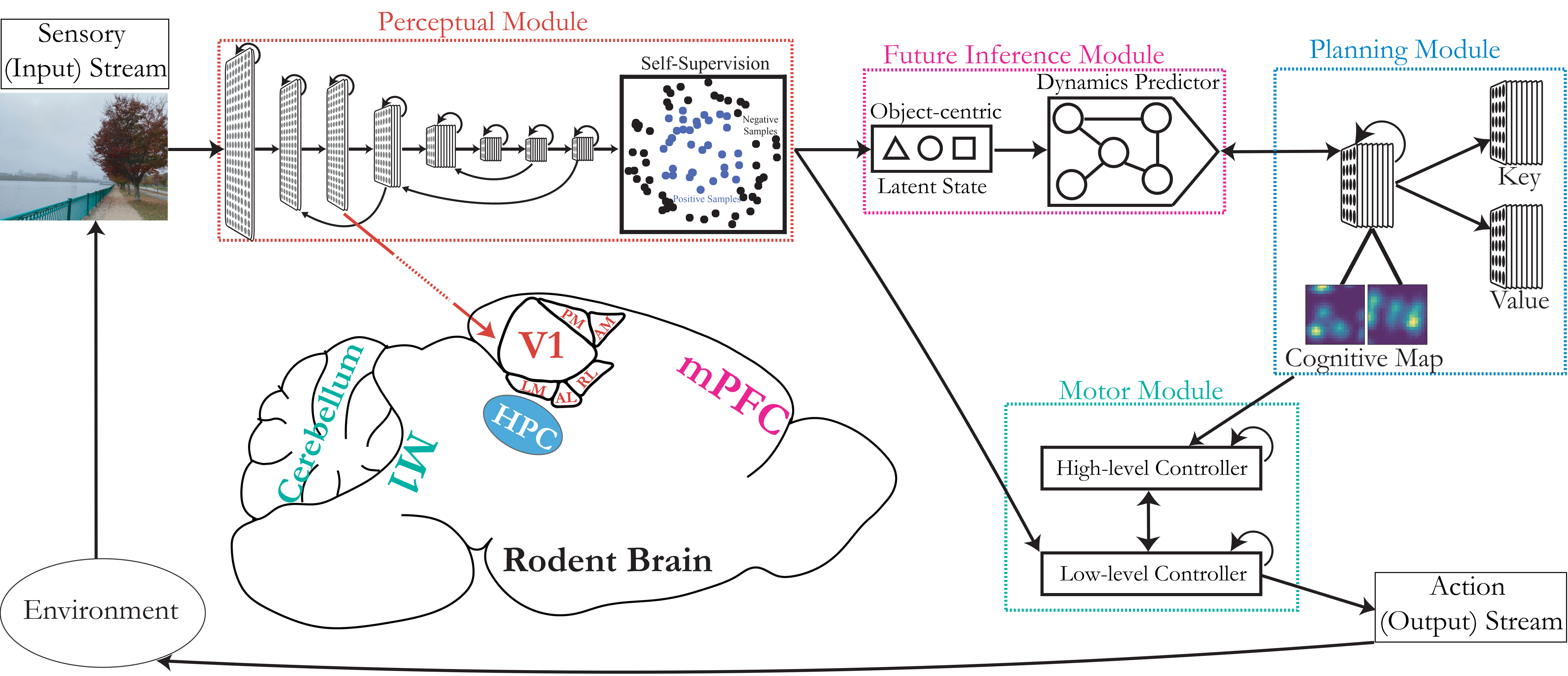
Integrated, Task-Optimized Model of the Rodent

Direction 1:
Building the Embodied Agent

Direction 2:
Applying the Embodied Agent to Neuroscience Questions

Next Steps: Applying the Embodied Agent

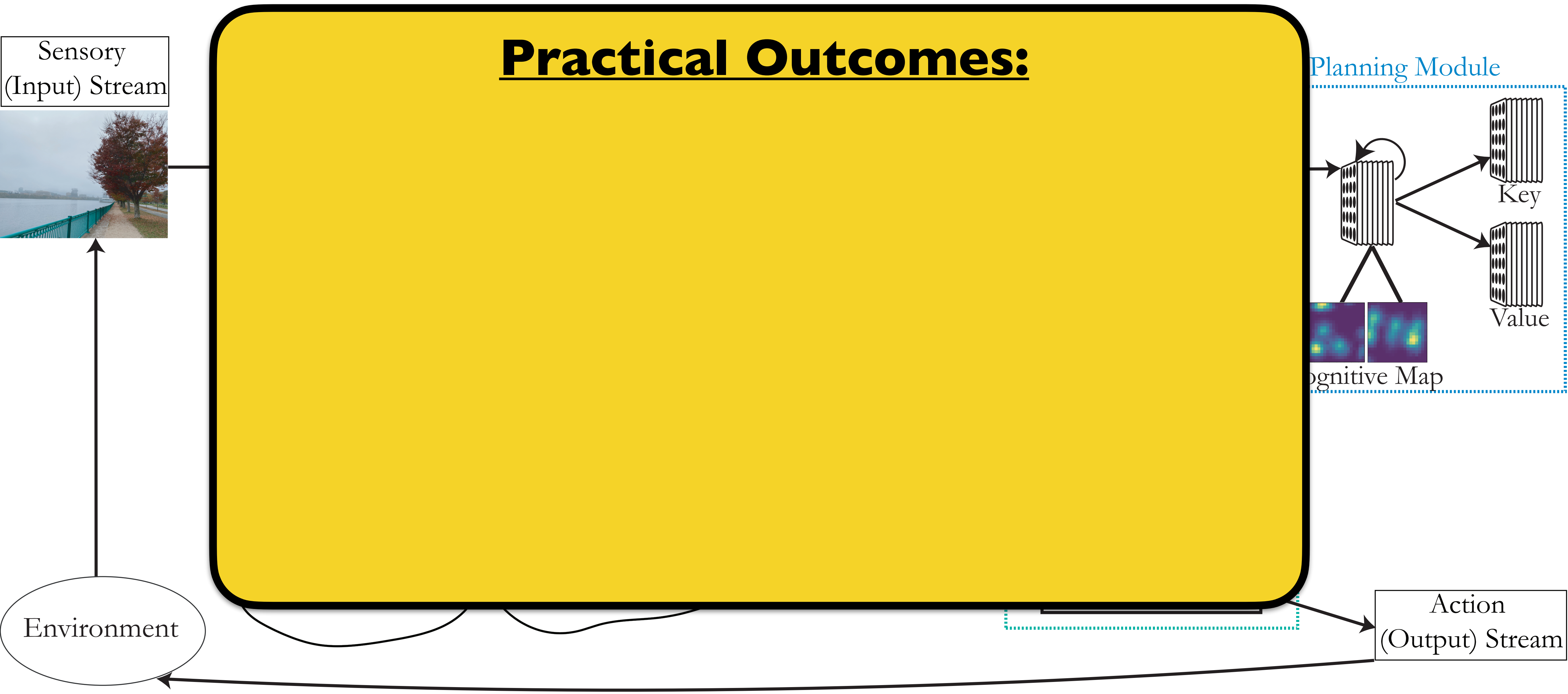
How does the brain *represent*, *predict*, *plan*, and enable *action*?



Next Steps: Applying the Embodied Agent

How does the brain *represent*, *predict*, *plan*, and enable *action*?

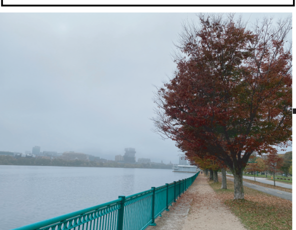
Practical Outcomes:



Next Steps: Applying the Embodied Agent

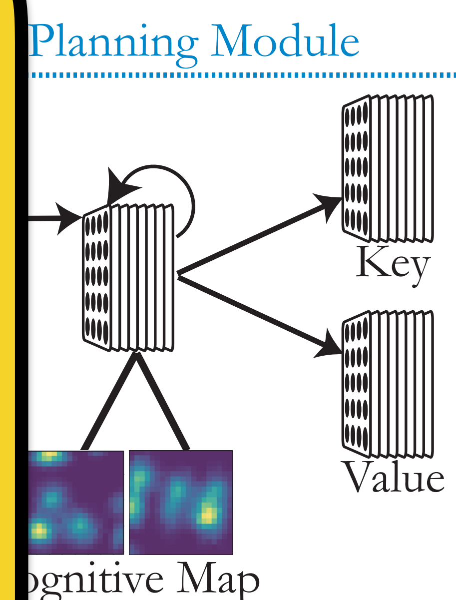
How does the brain *represent*, *predict*, *plan*, and enable *action*?

Sensory (Input) Stream



<u>Practical Outcomes:</u>

1. Open-source benchmarks

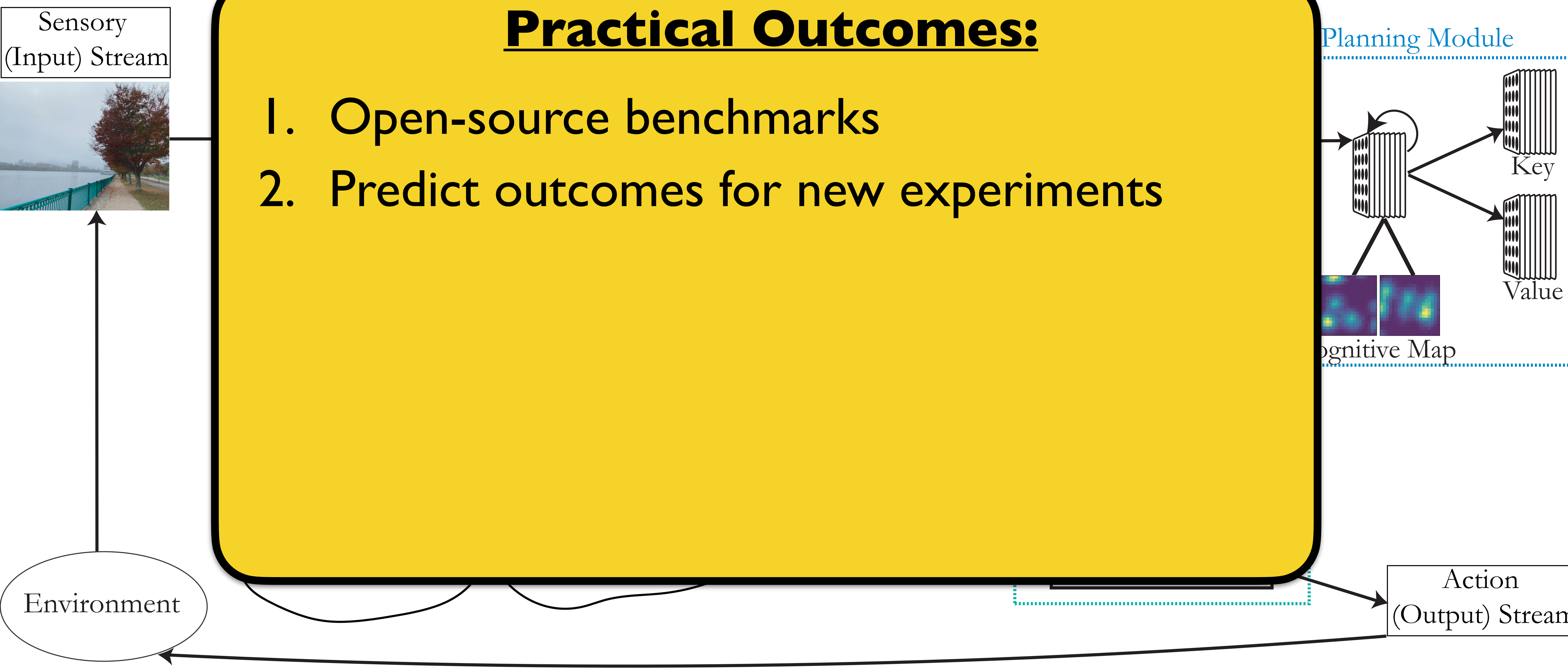


Environment

Action (Output) Stream

Next Steps: Applying the Embodied Agent

How does the brain *represent*, *predict*, *plan*, and enable *action*?

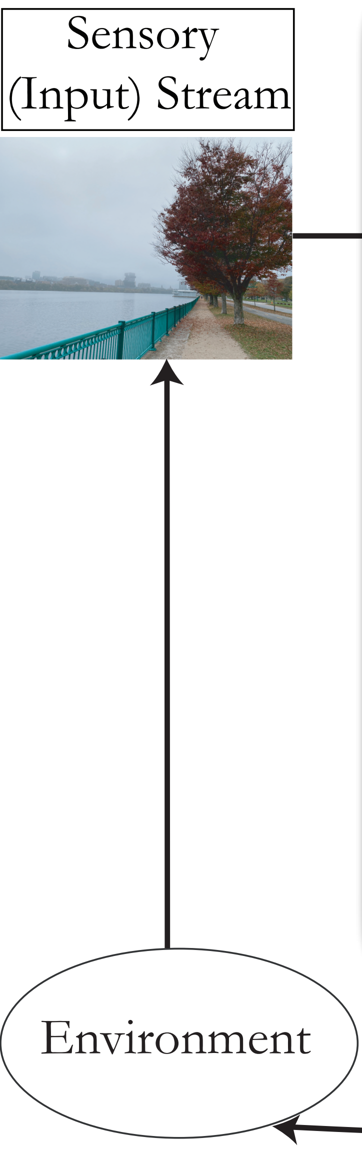


<u>Practical Outcomes:</u>

1. Open-source benchmarks
2. Predict outcomes for new experiments

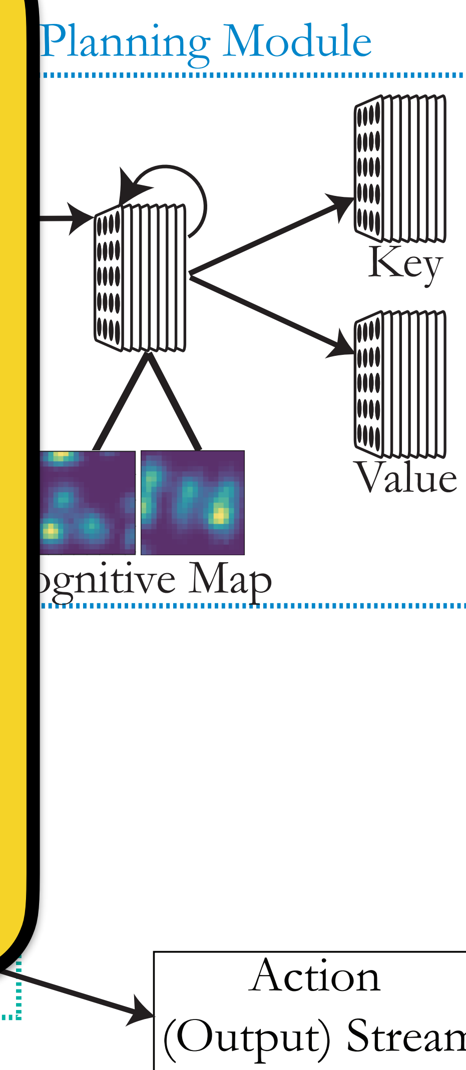
Next Steps: Applying the Embodied Agent

How does the brain *represent*, *predict*, *plan*, and enable *action*?



Practical Outcomes:

1. Open-source benchmarks
2. Predict outcomes for new experiments
3. Systems-level understanding of an organism



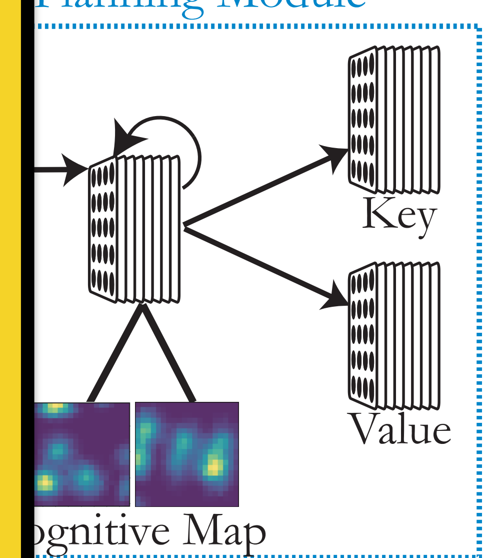
Next Steps: Applying the Embodied Agent

How does the brain *represent*, *predict*, *plan*, and enable *action*?

Sensory
(Input) Stream

Practical Outcomes:

1. Open-source benchmarks
2. Predict outcomes for new experiments
3. Systems-level understanding of an organism



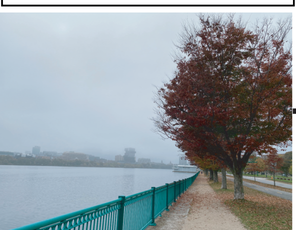
Environment

Action
(Output) Stream

Next Steps: Applying the Embodied Agent

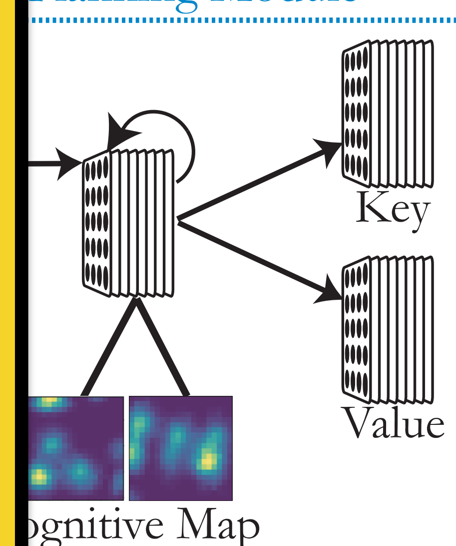
How does the brain *represent*, *predict*, *plan*, and enable *action*?

Sensory (Input) Stream



Practical Outcomes:

1. Open-source benchmarks
2. Predict outcomes for new experiments
3. Systems-level understanding of an organism
   - Hard-wired intrinsic goals



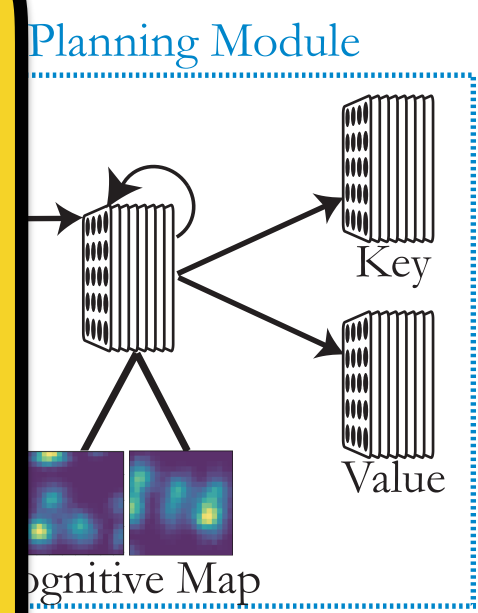
Environment

Action (Output) Stream

Next Steps: Applying the Embodied Agent

How does the brain *represent*, *predict*, *plan*, and enable *action*?

Sensory (Input) Stream

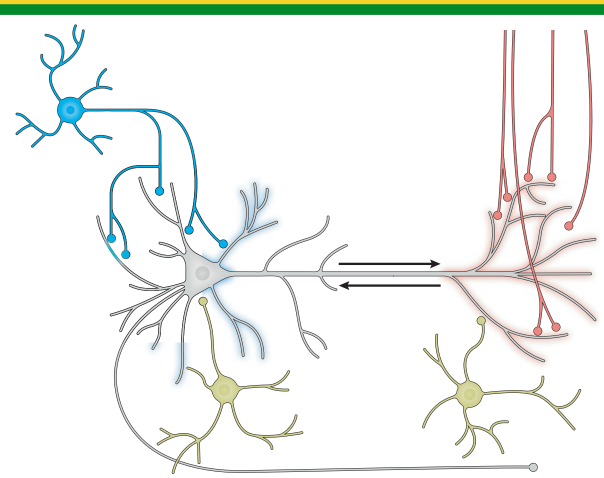


Practical Outcomes:

1. Open-source benchmarks
2. Predict outcomes for new experiments
3. Systems-level understanding of an organism
   - Hard-wired intrinsic goals
   - Identifying biological learning rules

Environment

Action (Output) Stream



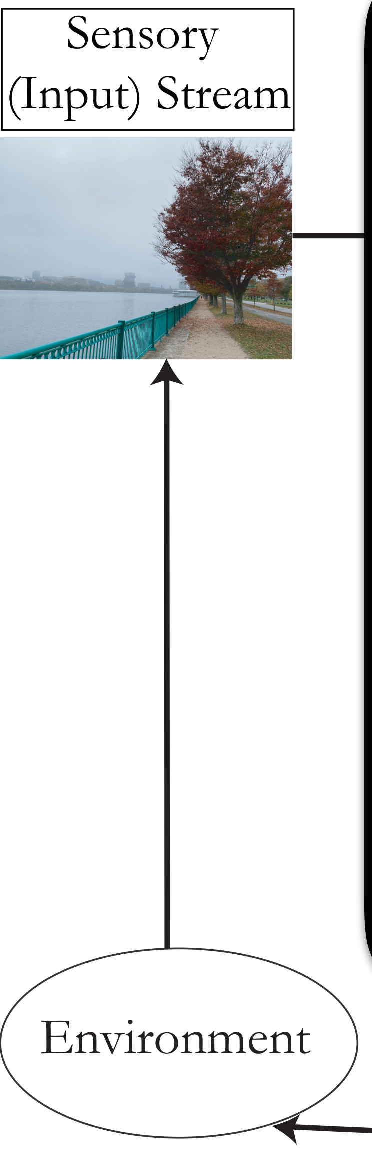
Kunin*, Nayebi*, Sagastuy-Brena* et al ICML (2020a)

Nayebi et al NeurIPS (2020b) Spotlight

Melander*, Nayebi* et al Cell Reports (2021)

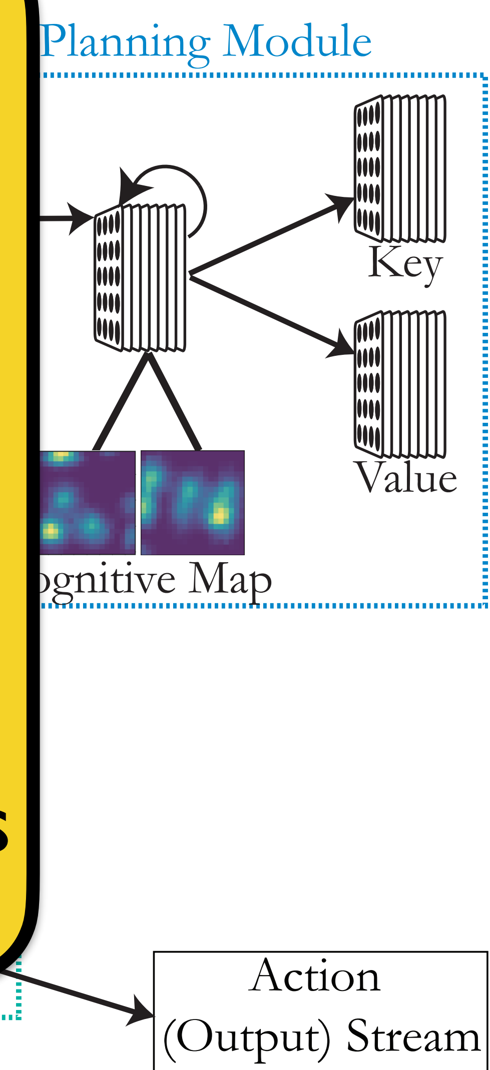
Next Steps: Applying the Embodied Agent

How does the brain *represent*, *predict*, *plan*, and enable *action*?



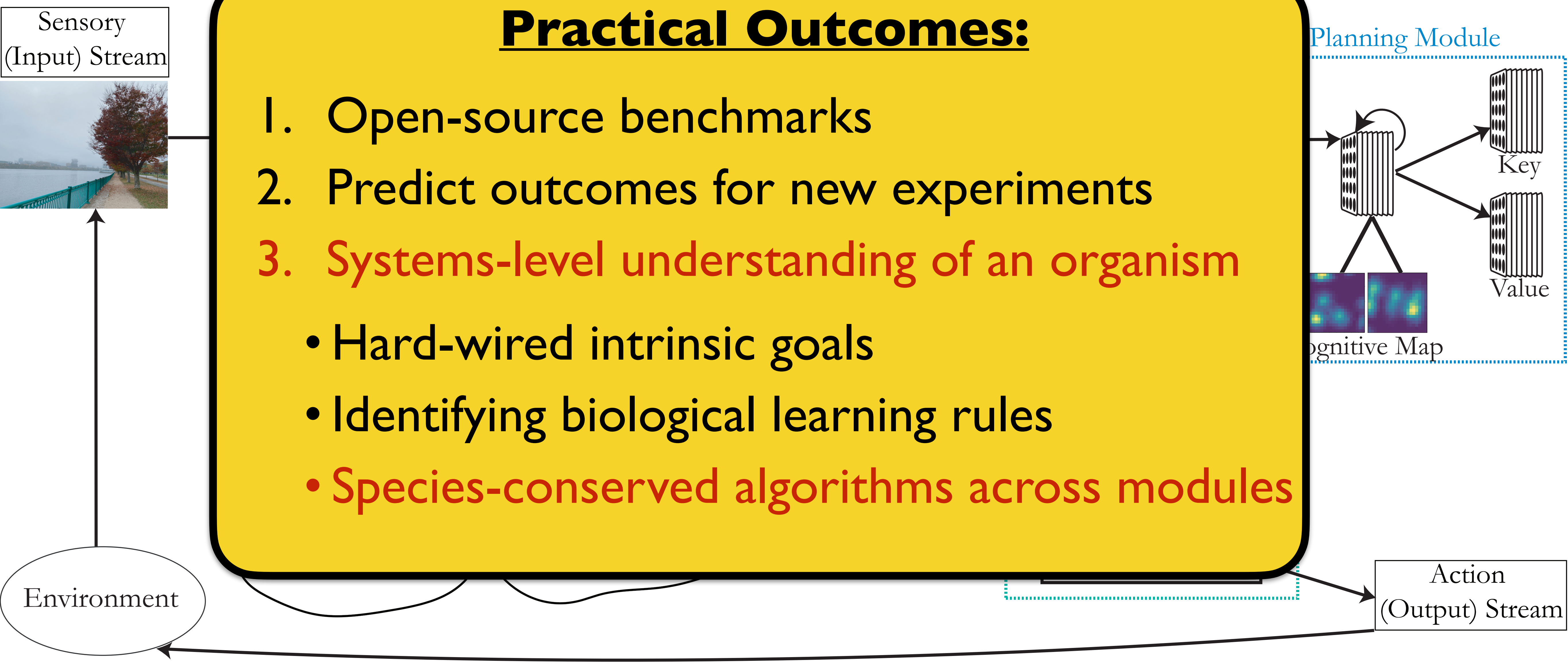
Practical Outcomes:

1. Open-source benchmarks
2. Predict outcomes for new experiments
3. Systems-level understanding of an organism
   - Hard-wired intrinsic goals
   - Identifying biological learning rules
   - Species-conserved algorithms across modules



Next Steps: Applying the Embodied Agent

How does the brain *represent*, *predict*, *plan*, and enable *action*?

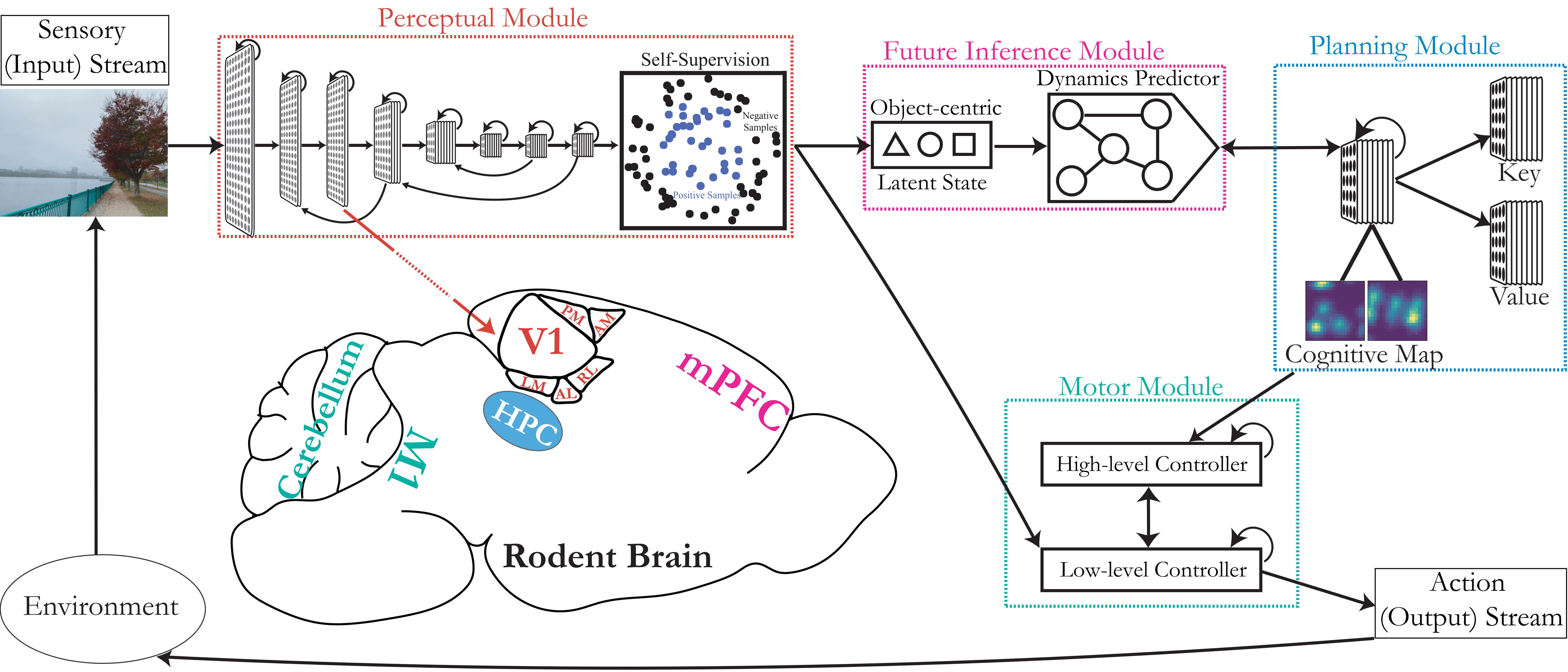


Practical Outcomes:

1. Open-source benchmarks
2. Predict outcomes for new experiments
3. Systems-level understanding of an organism
   - Hard-wired intrinsic goals
   - Identifying biological learning rules
   - Species-conserved algorithms across modules

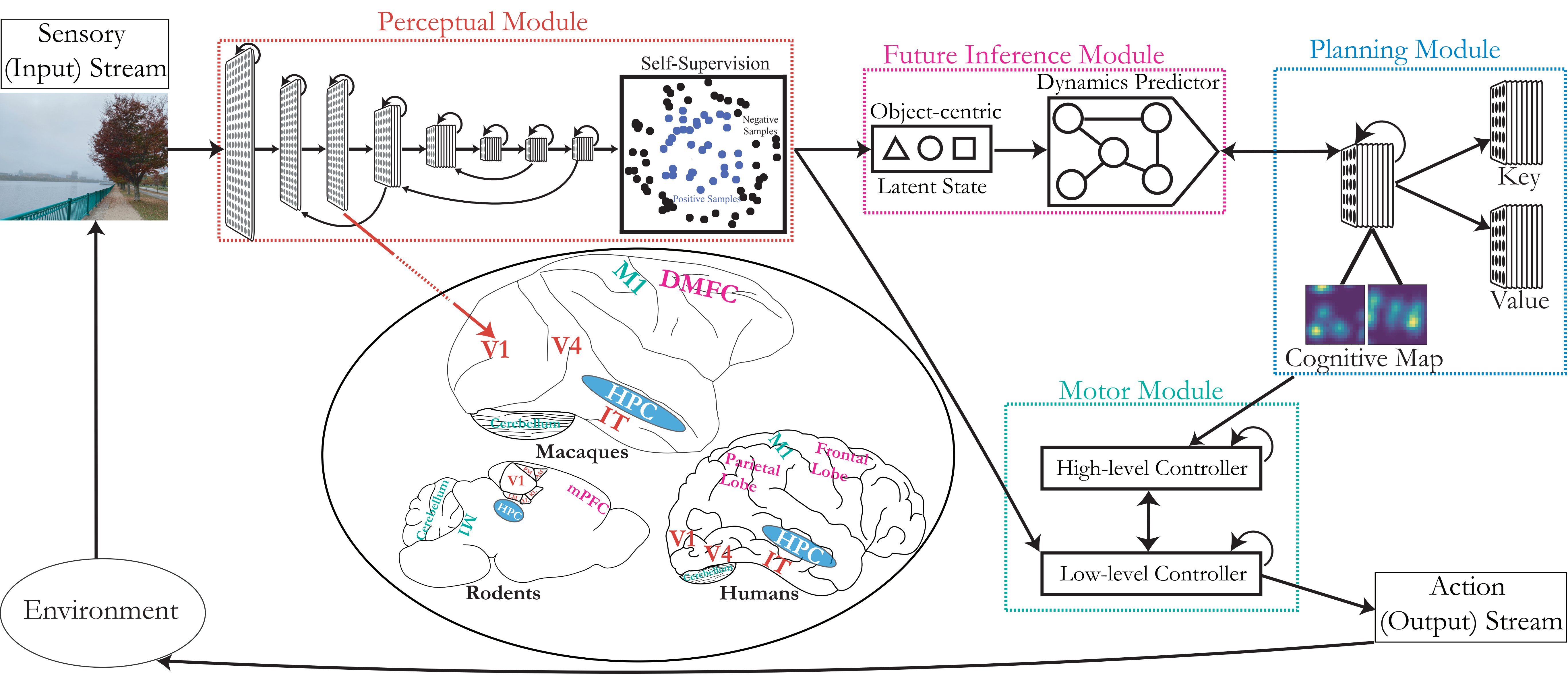
Next Steps: Applying the Embodied Agent

How does the brain *represent*, *predict*, *plan*, and enable *action*?

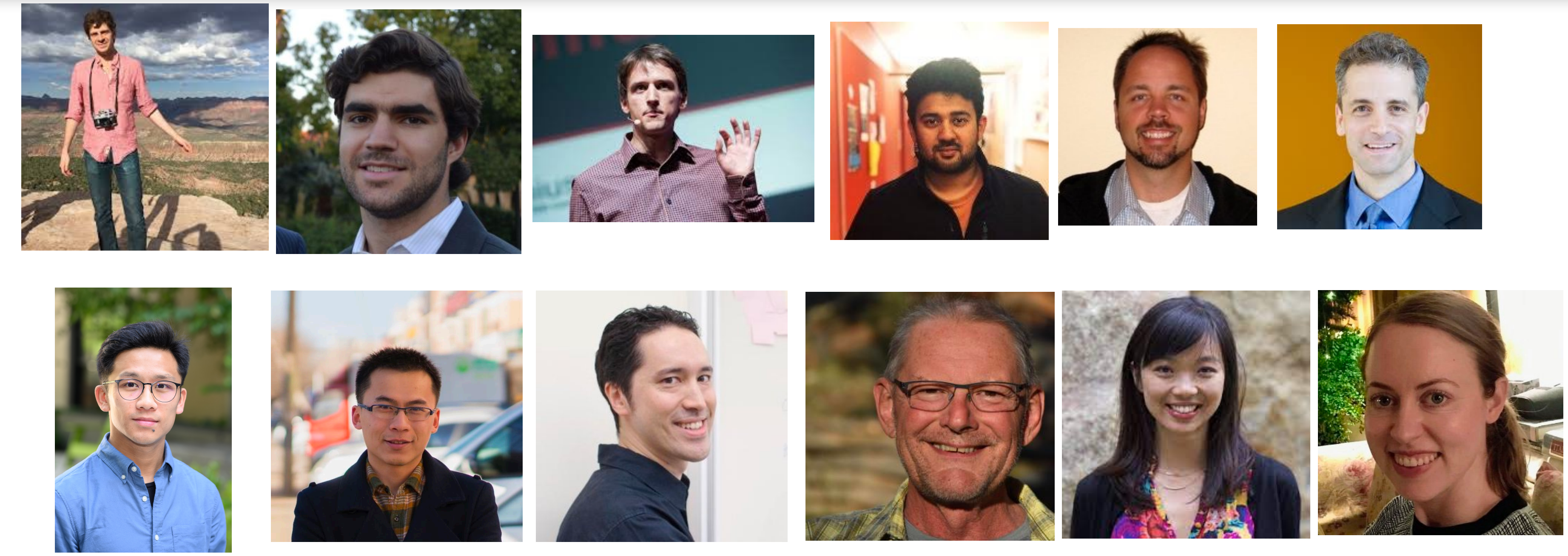


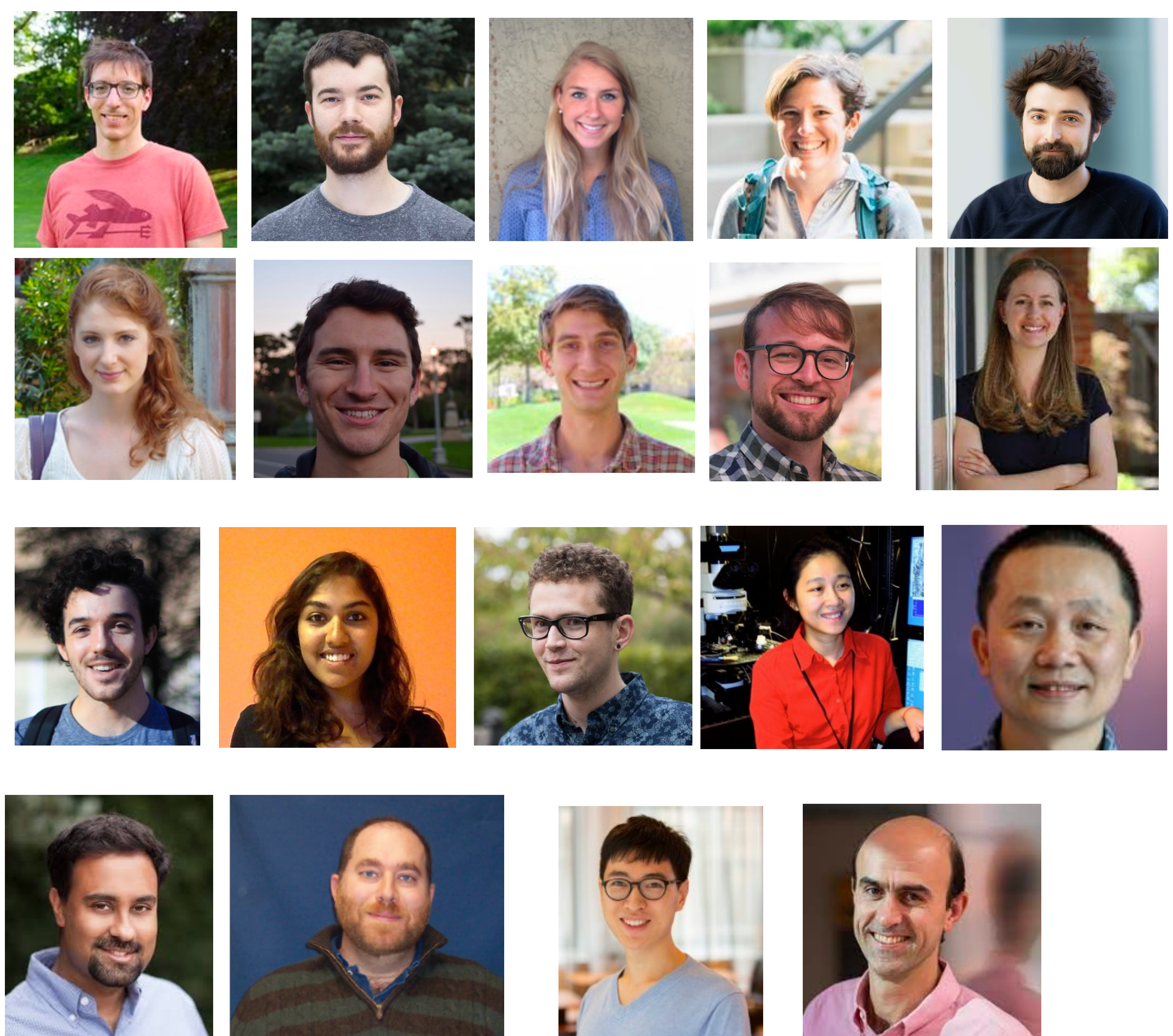
Long-Term Outcome: Artificial **Organisms**

How does the brain *represent*, *predict*, *plan*, and enable *action*?



Acknowledgements





Contact:

anayebi@mit.edu

@aran_nayebi

Funding:

K. Lisa Yang ICoN Postdoctoral Fellowship, McGovern Institute, MIT

Stanford Neurosciences PhD Program

Stanford Mind, Brain, Computation and Technology Training Program, Wu Tsai Neurosciences Institute