# Explaining heterogeneity in medial entorhinal cortex with task-driven neural networks

Aran Nayebi
*Assistant Professor*
*Machine Learning Department*
*Neuroscience Institute (core faculty), Robotics Institute (by courtesy)*
*Carnegie Mellon University*

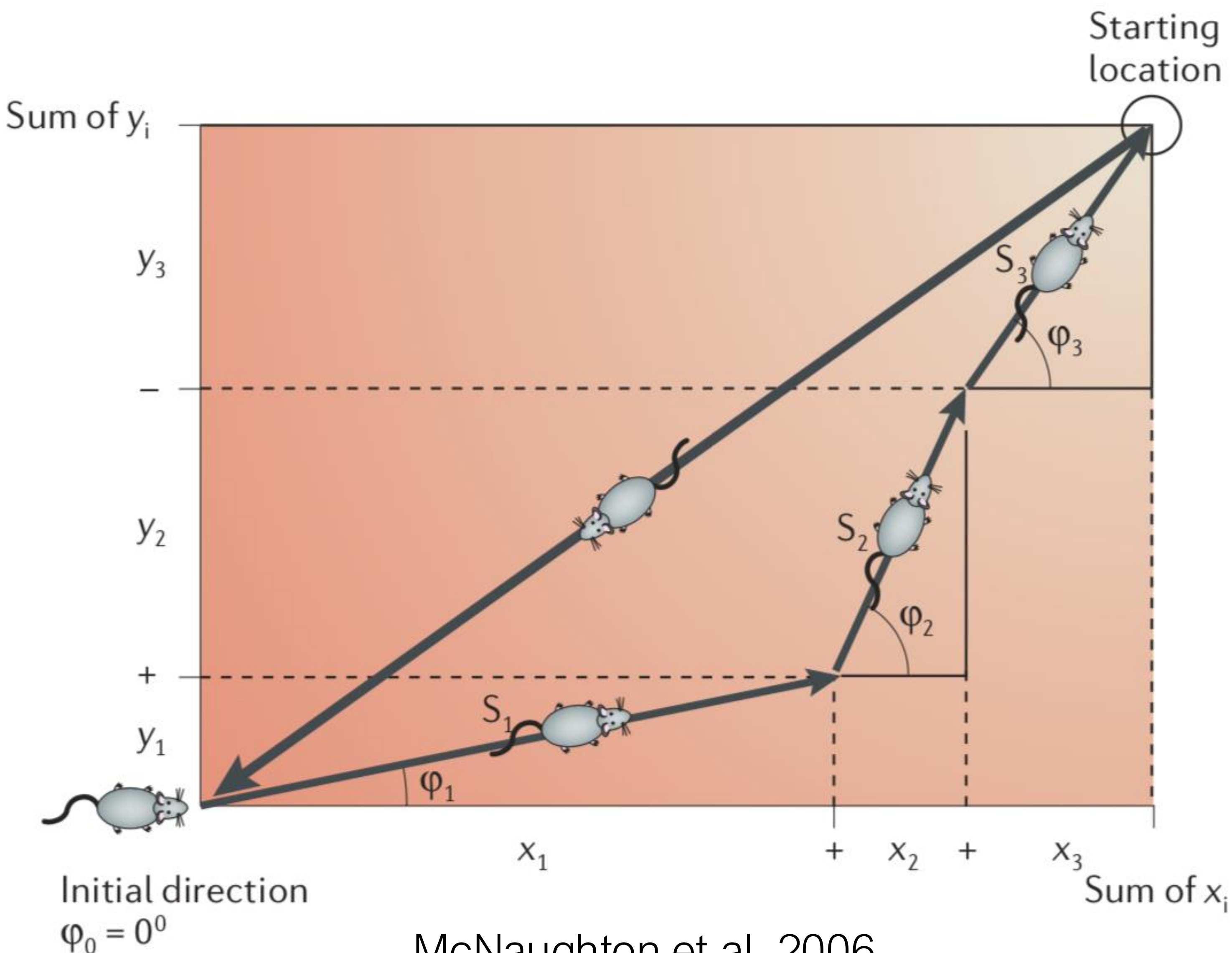
FENS 2025 Regional Meeting
*2025.06.18*

MACHINE LEARNING DEPARTMENT

Carnegie Mellon University Neuroscience Institute

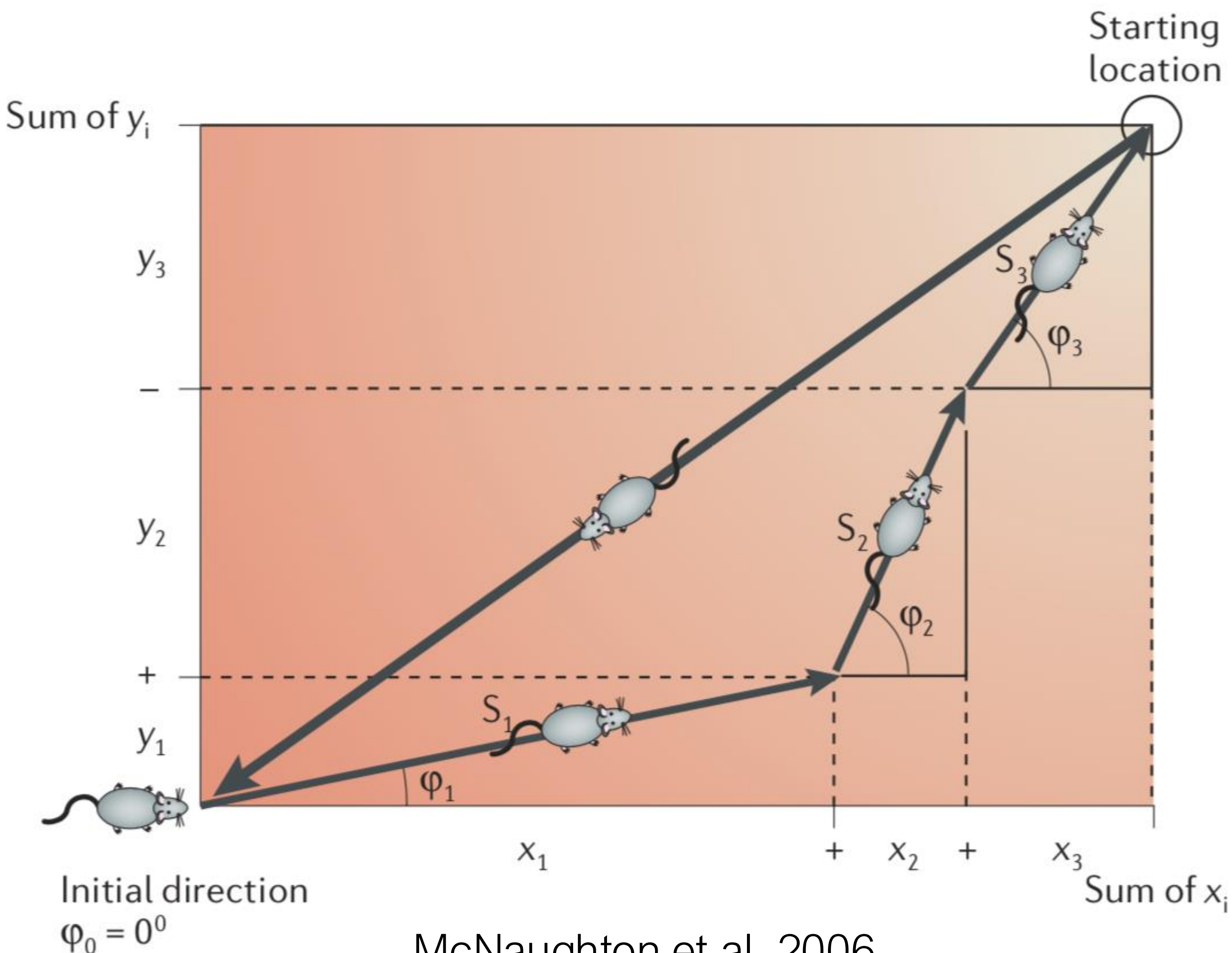
THE ROBOTICS INSTITUTE

# Hippocampal-Entorhinal Spatial Map



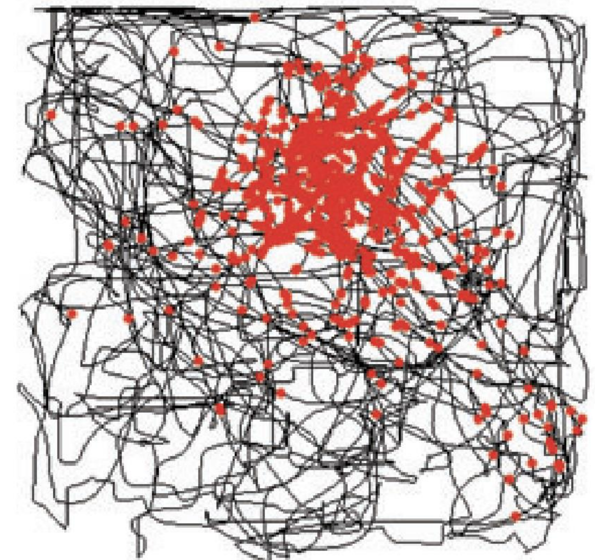

McNaughton et al. 2006

# Hippocampal-Entorhinal Spatial Map

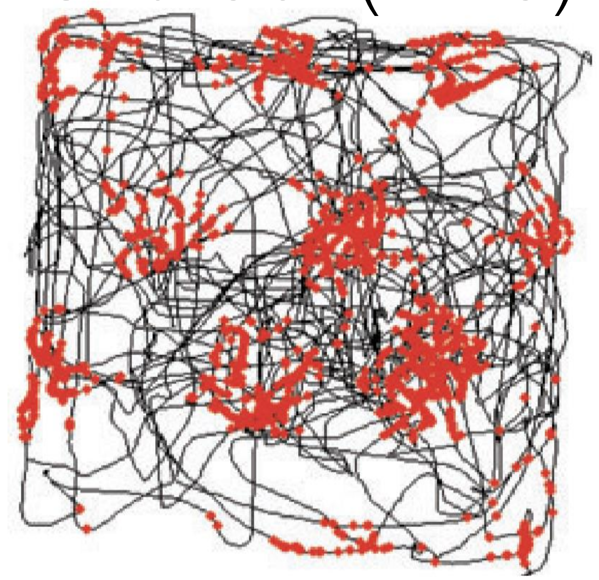



McNaughton et al. 2006

Place Cell (HPC)

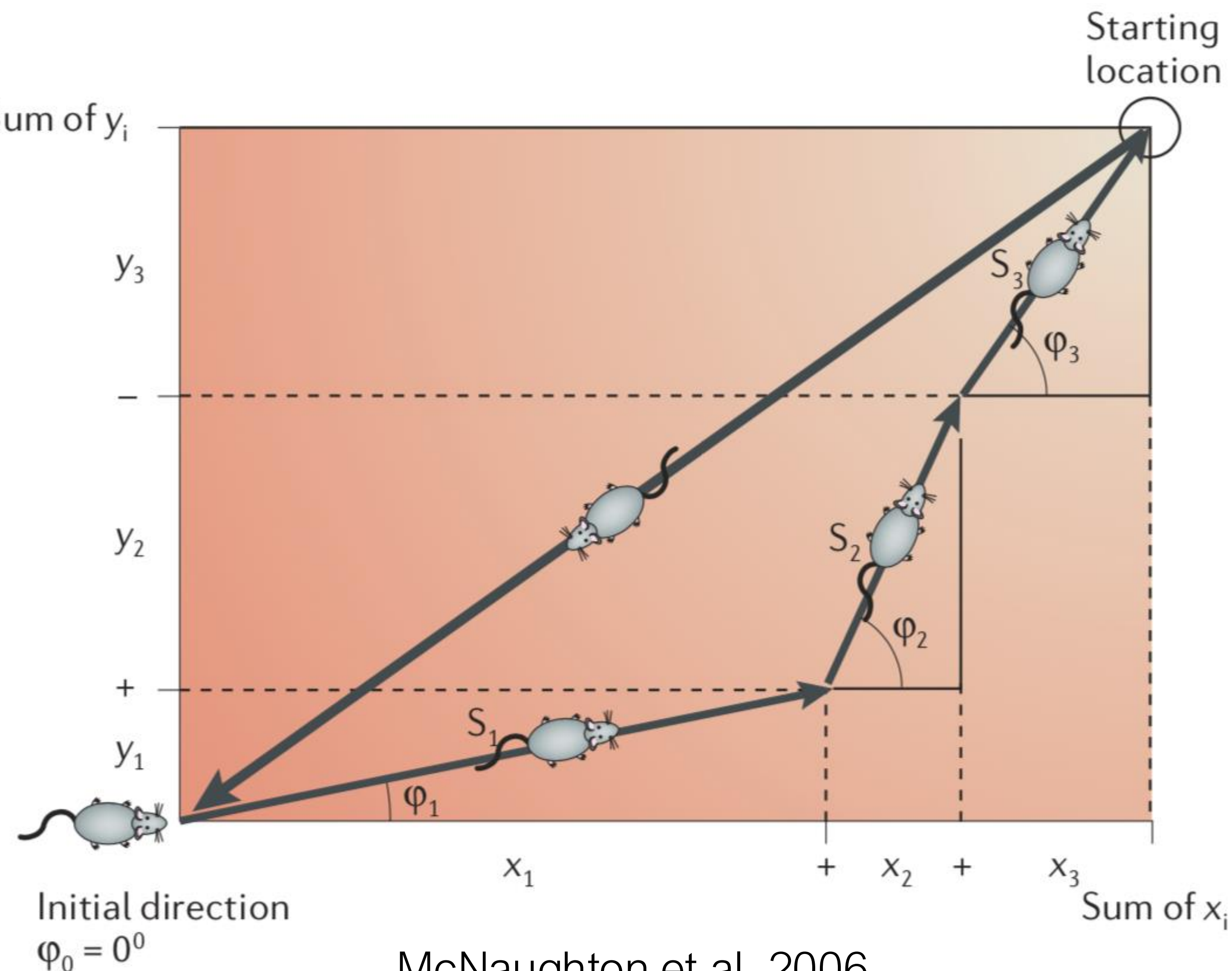


Grid Cell (MEC)



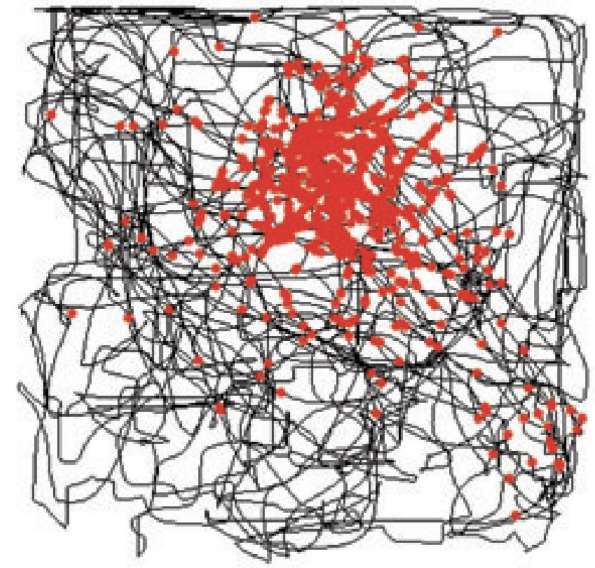
Moser et al. 2008

# Hippocampal-Entorhinal Spatial Map

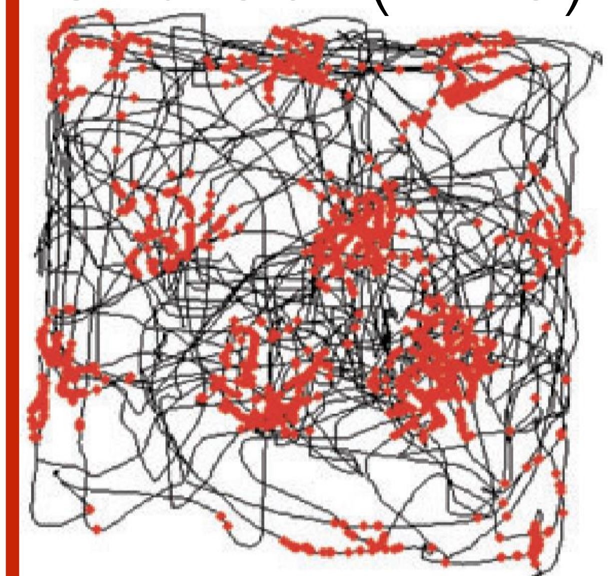



McNaughton et al. 2006

Place Cell (HPC)



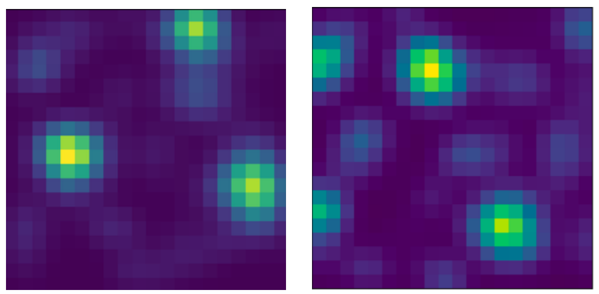
Grid Cell (MEC)



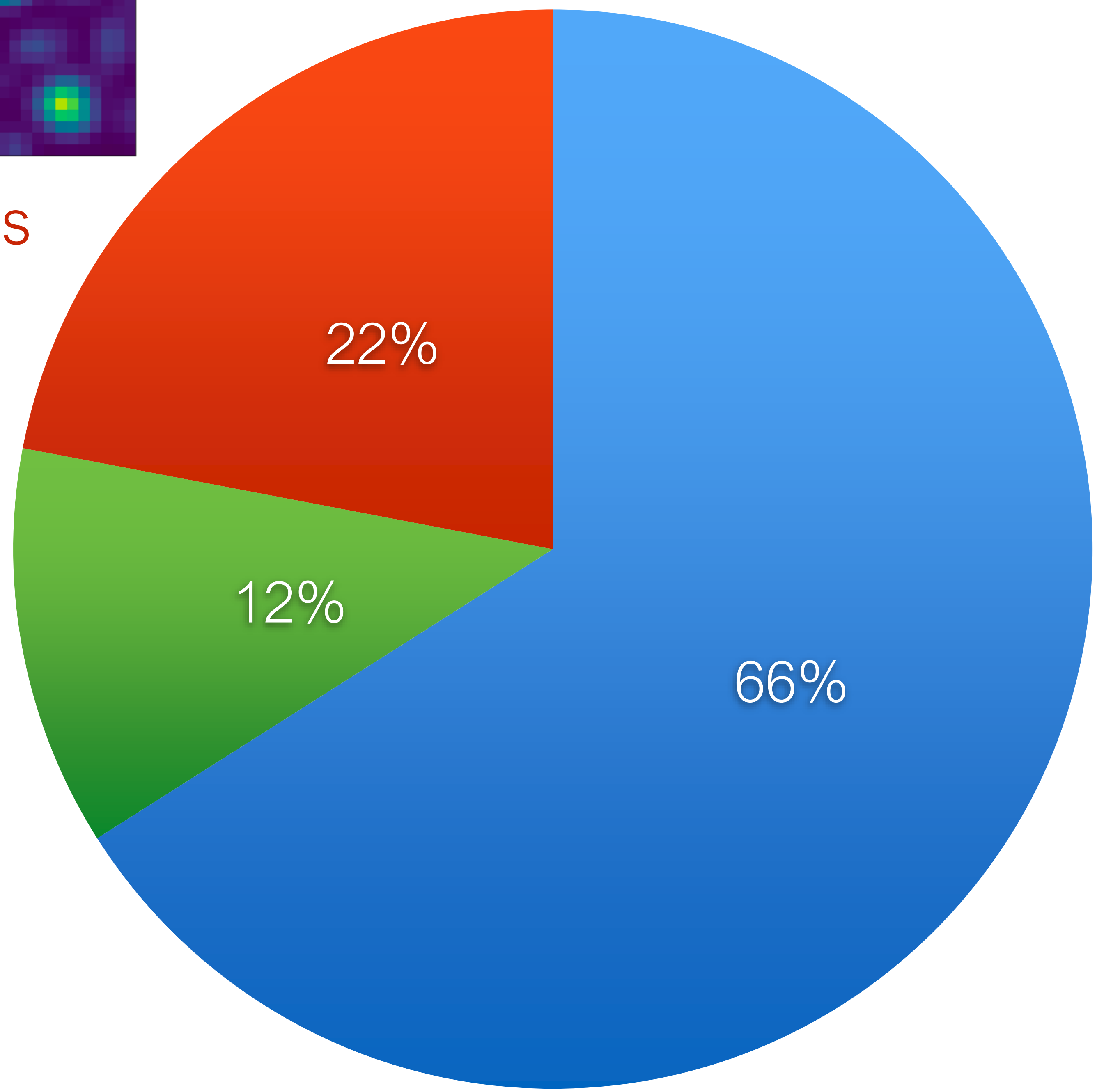
Moser et al. 2008

# Accounting for heterogeneous code?

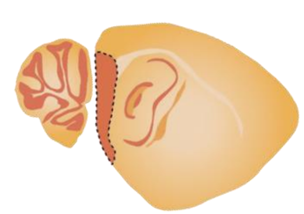
# Accounting for heterogeneous code?



Grid Cells



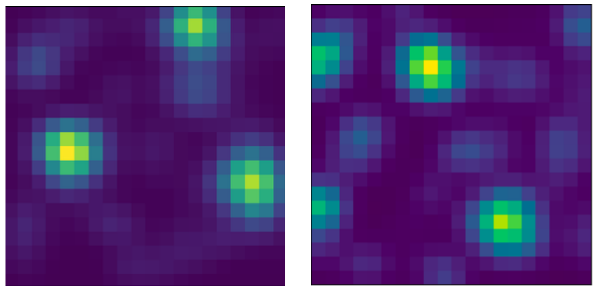

Data from: *Mallory et al. 2021*



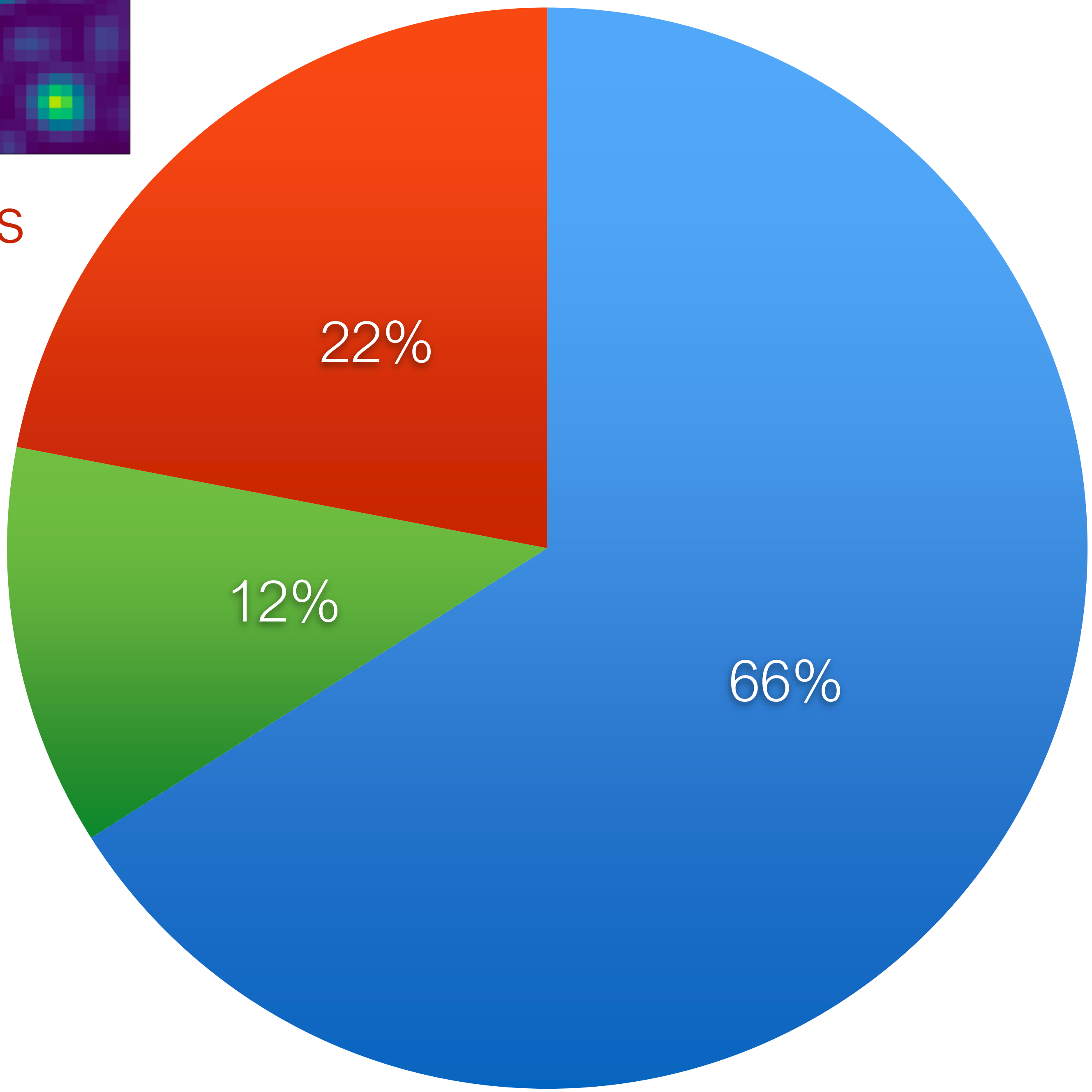
Caitlin Mallory

# Accounting for heterogeneous code?

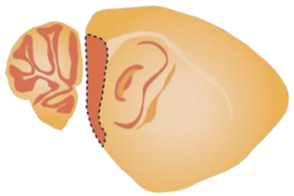
More like ~2-3%!



Grid Cells



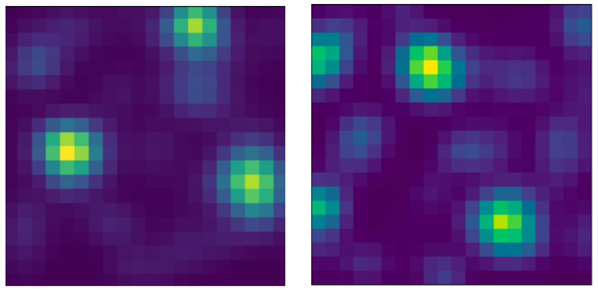

Data from: *Mallory et al. 2021*



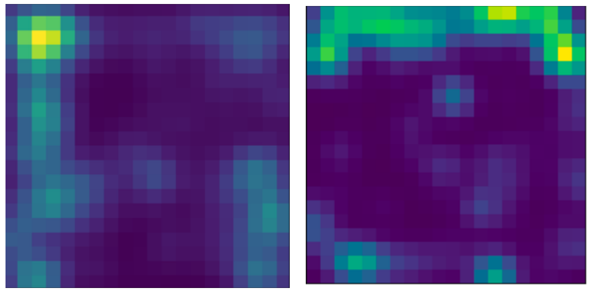
Caitlin Mallory

# Accounting for heterogeneous code?

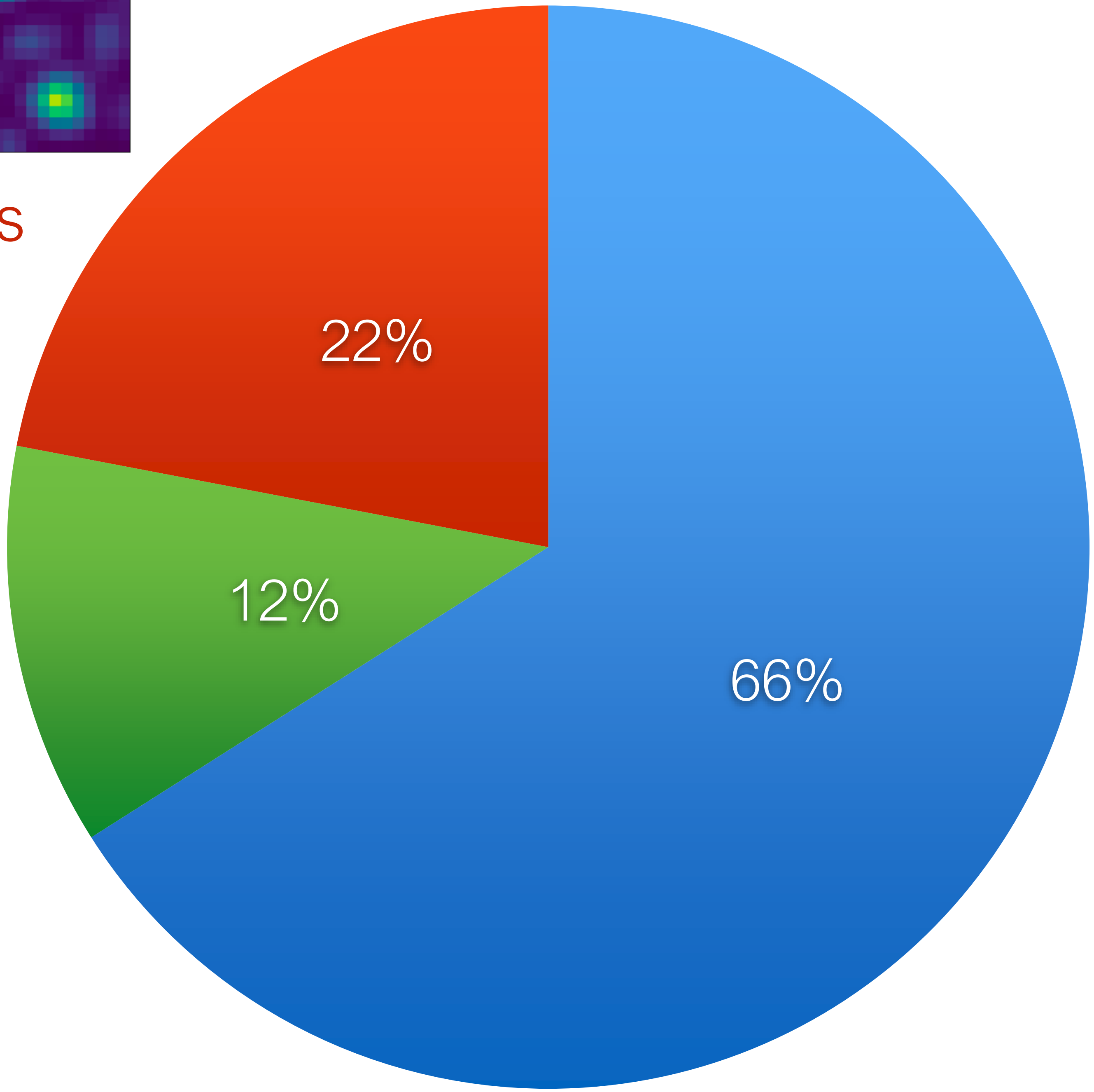
More like ~2-3%!



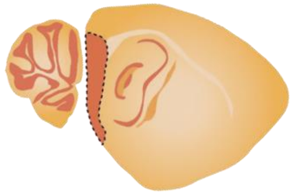
Grid Cells



Border Cells



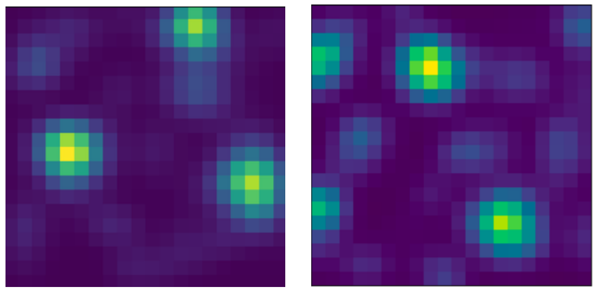

Data from: *Mallory et al. 2021*



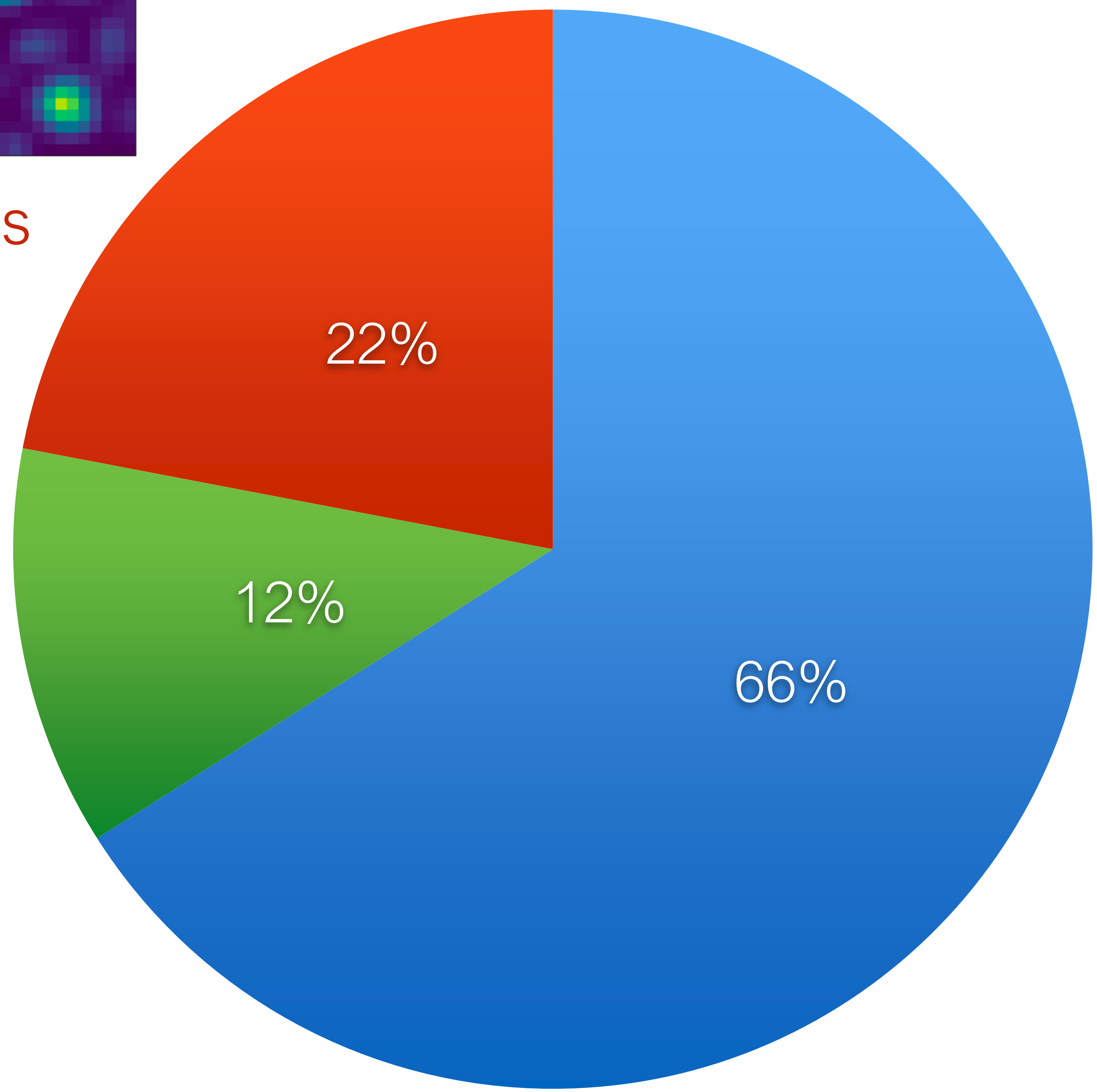
Caitlin Mallory

# Accounting for heterogeneous code?

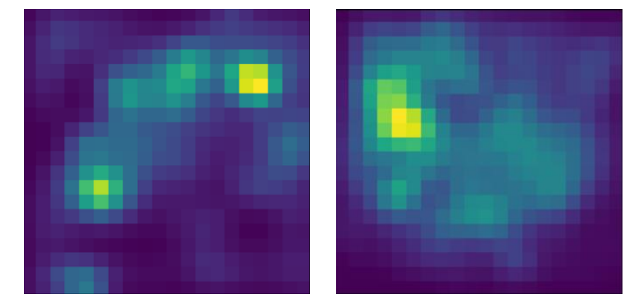
More like ~2-3%!

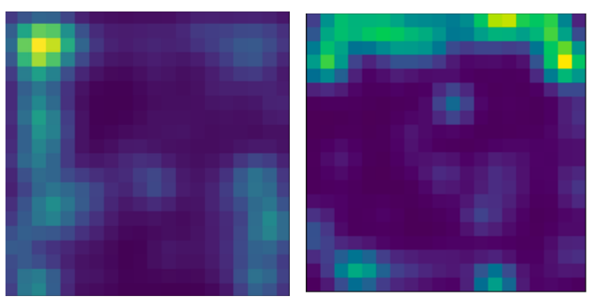


Grid Cells



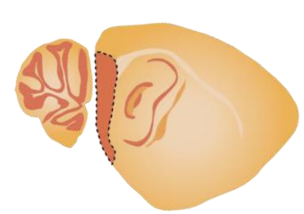

Heterogeneous Cells





Border Cells

Data from: *Mallory et al. 2021*



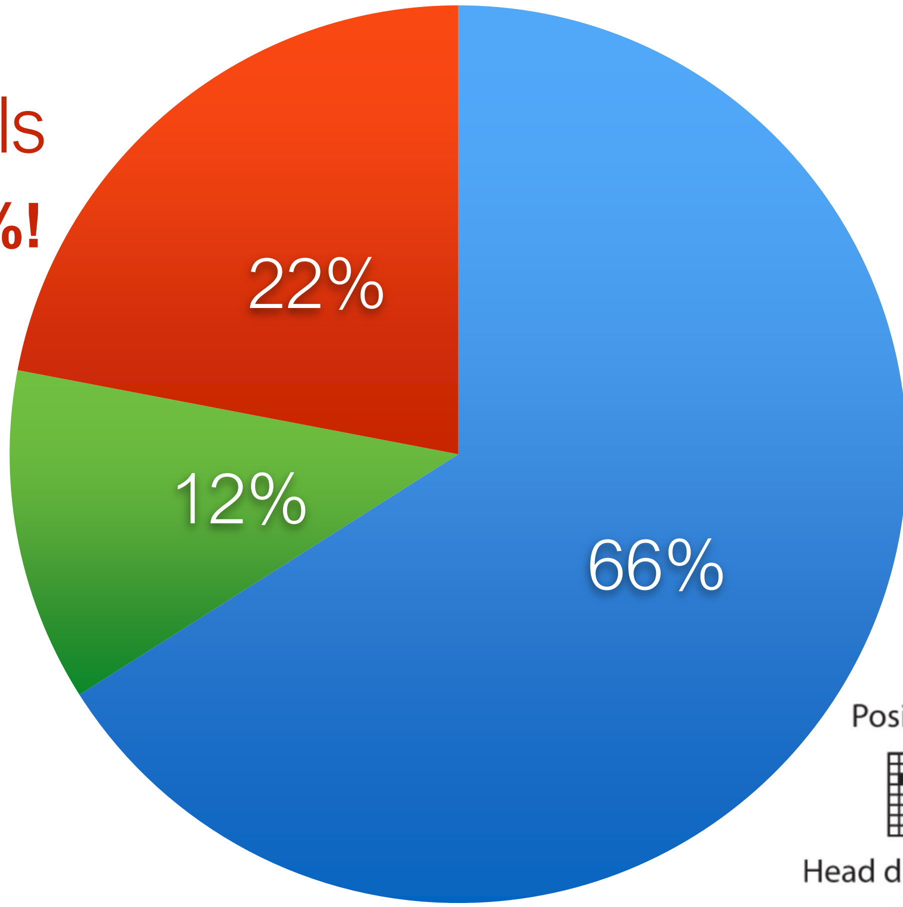
Caitlin Mallory

# Accounting for heterogeneous code?

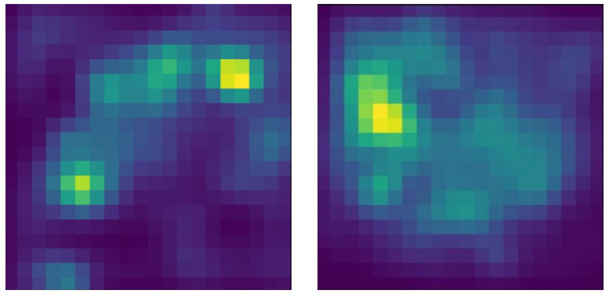
Grid Cells

**More like ~2-3%!**

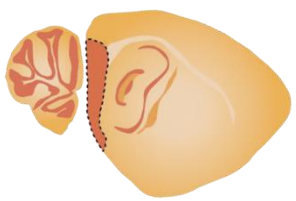
Border Cells

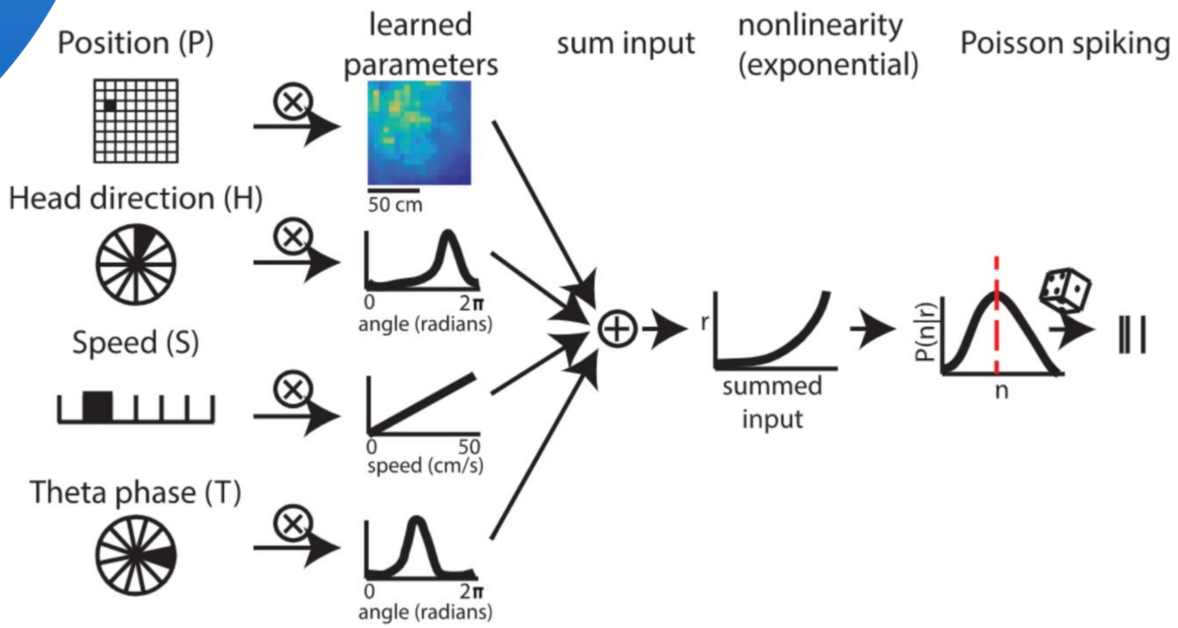


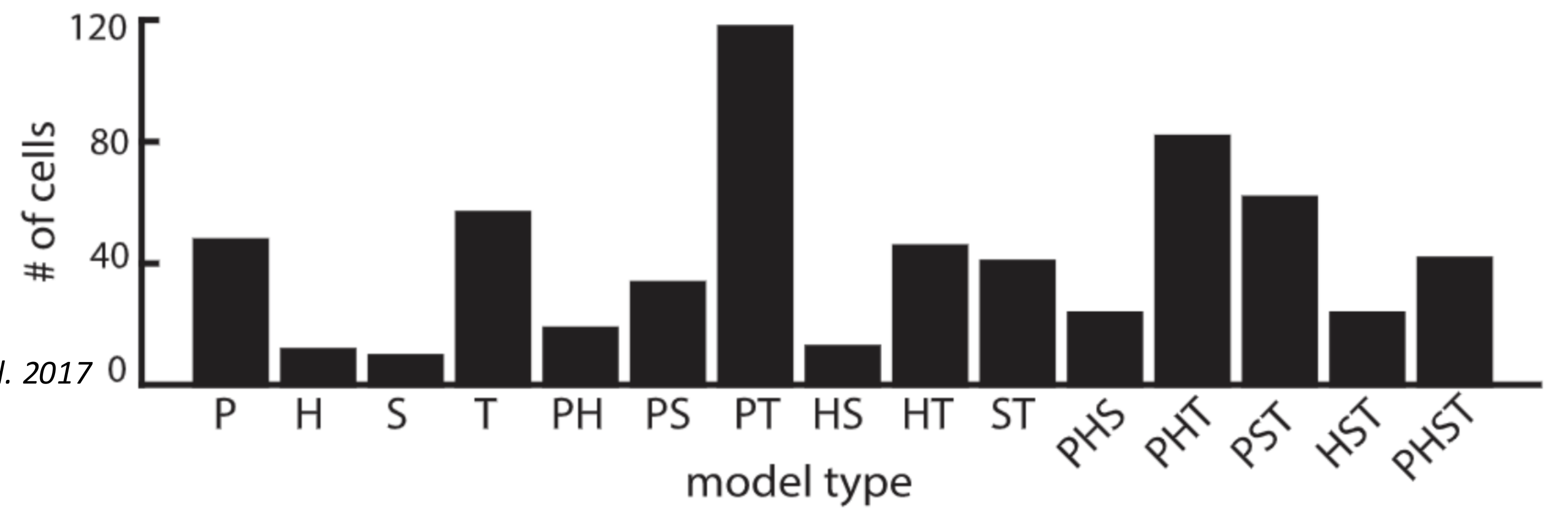

Heterogeneous Cells



Data from: *Mallory et al. 2021*





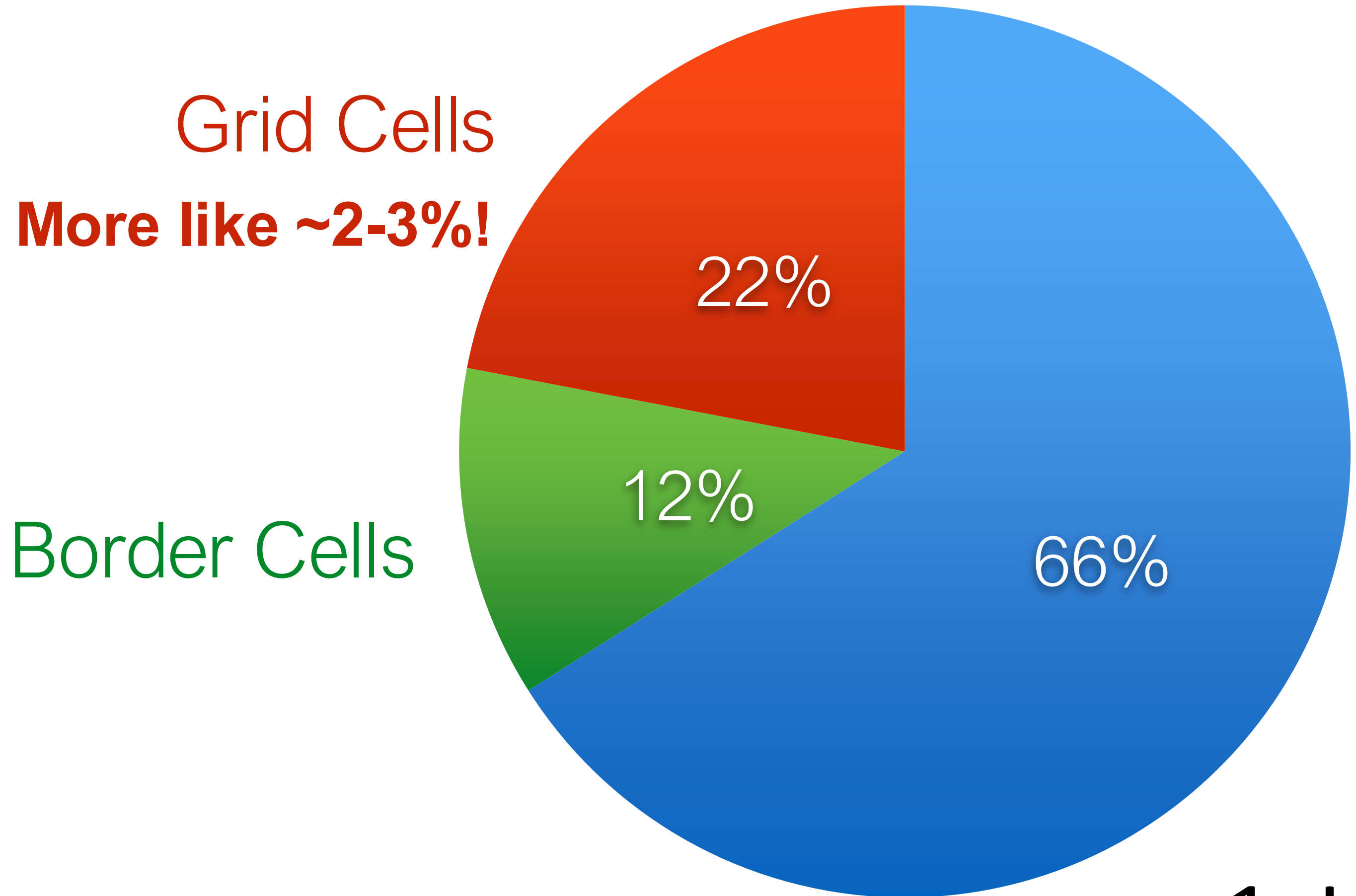

*Hardcastle et al. 2017*

Kiah Hardcastle

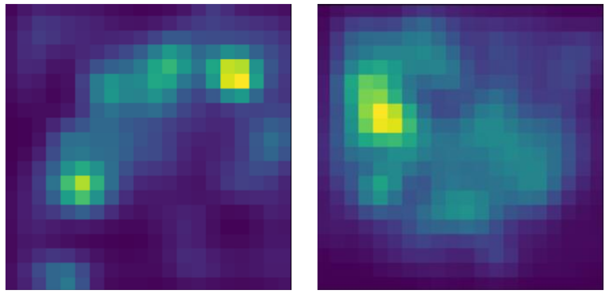
Surya Ganguli

Lisa Giocomo

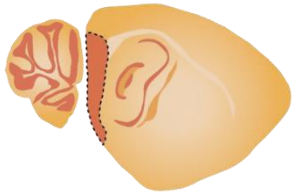
# Accounting for heterogeneous code?




Heterogeneous Cells



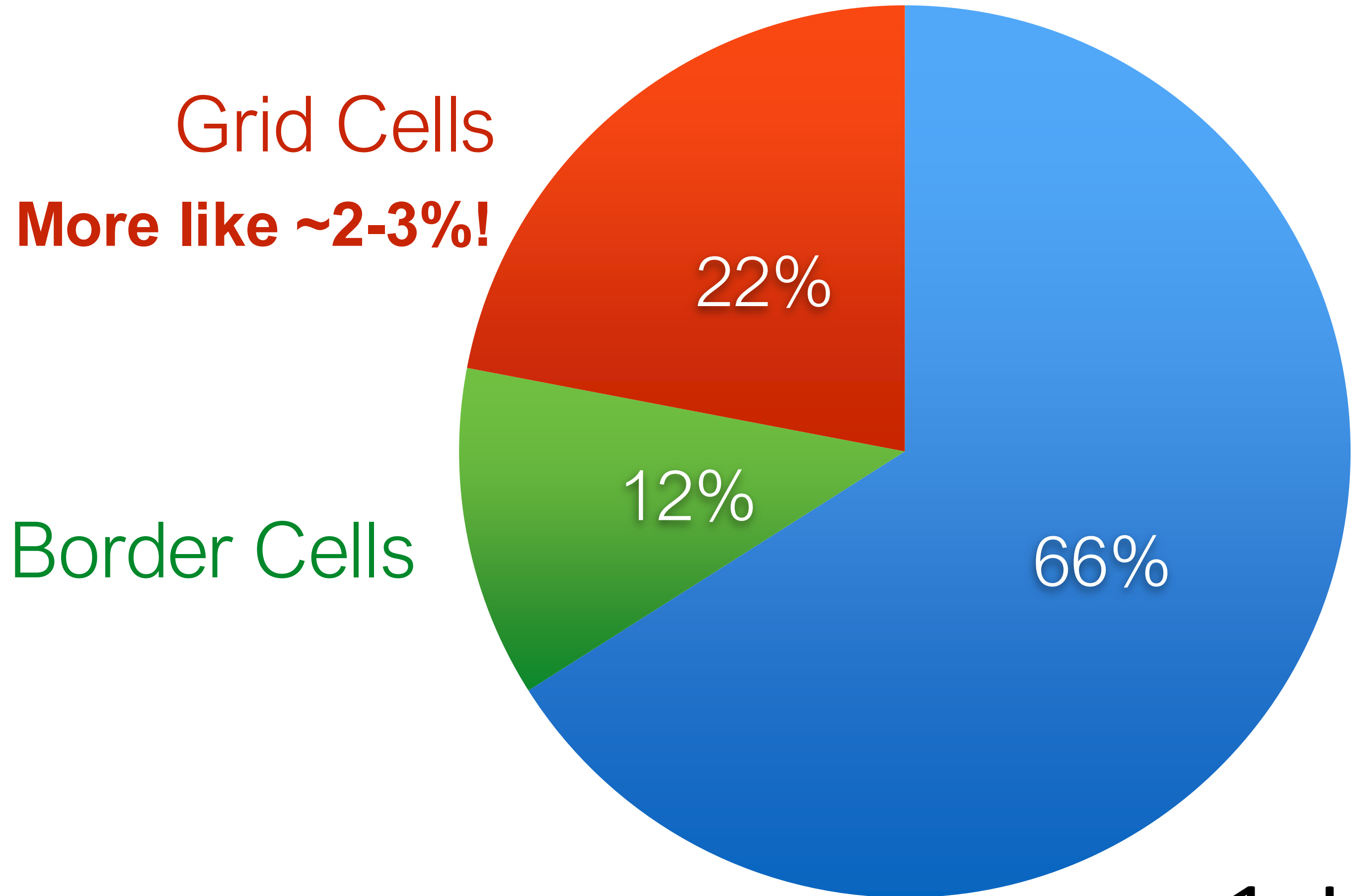
Data from: *Mallory et al. 2021*



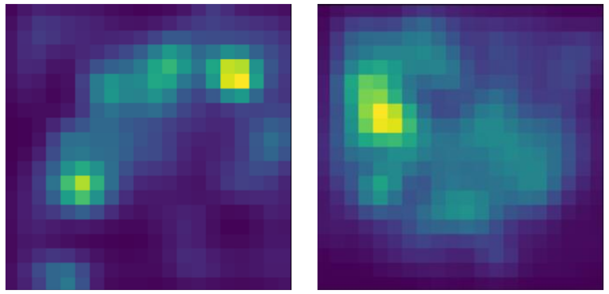
<u>Neurobiological Puzzle(s):</u>

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

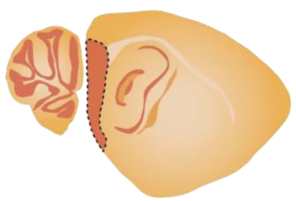
# Accounting for heterogeneous code?




Heterogeneous Cells



Data from: *Mallory et al. 2021*

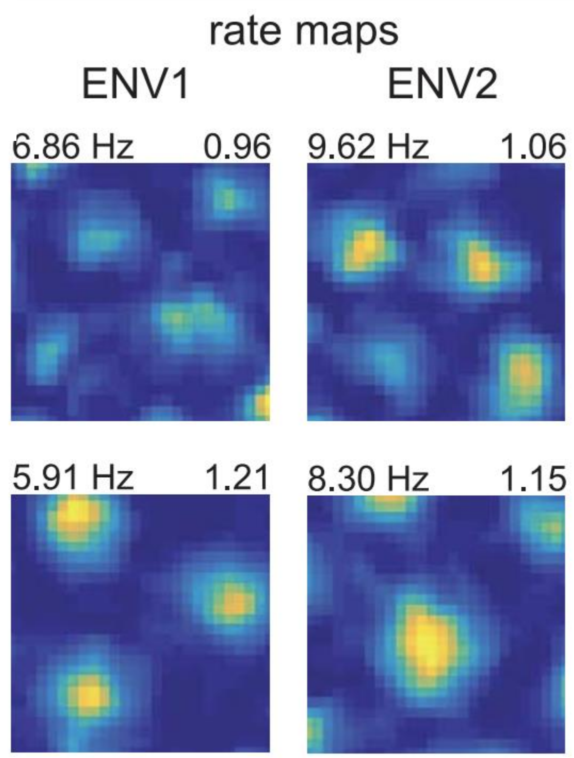


<u>Neurobiological Puzzle(s):</u>

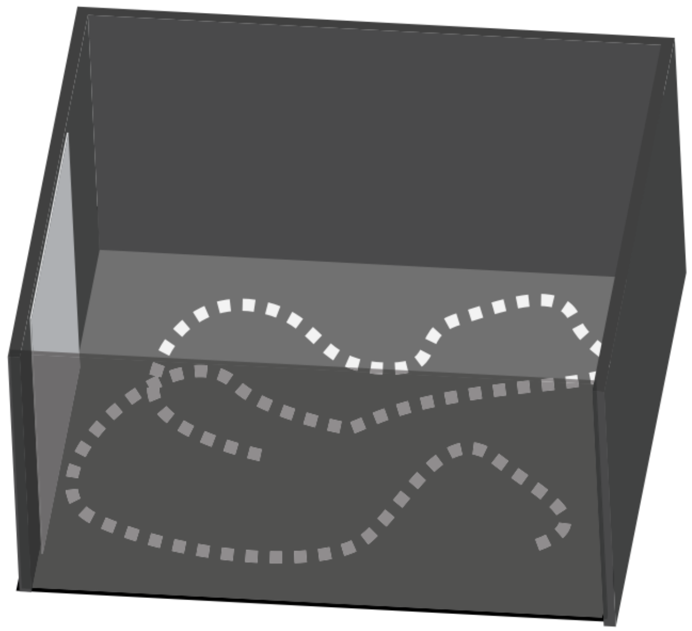
1. How might we characterize what these heterogeneous cells do?

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

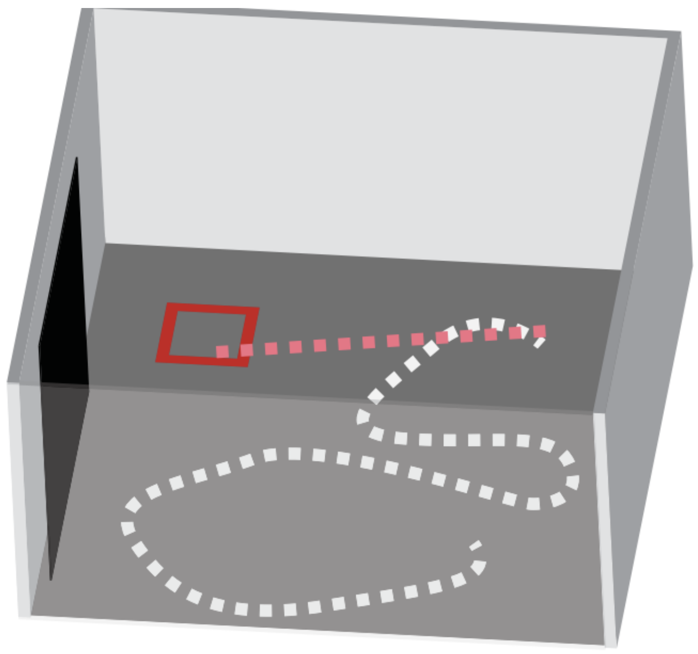
# Accounting for heterogeneous code in the presence of rewards?




Butler*, Hardcastle*, Giocomo 2019



free foraging (ENV1)

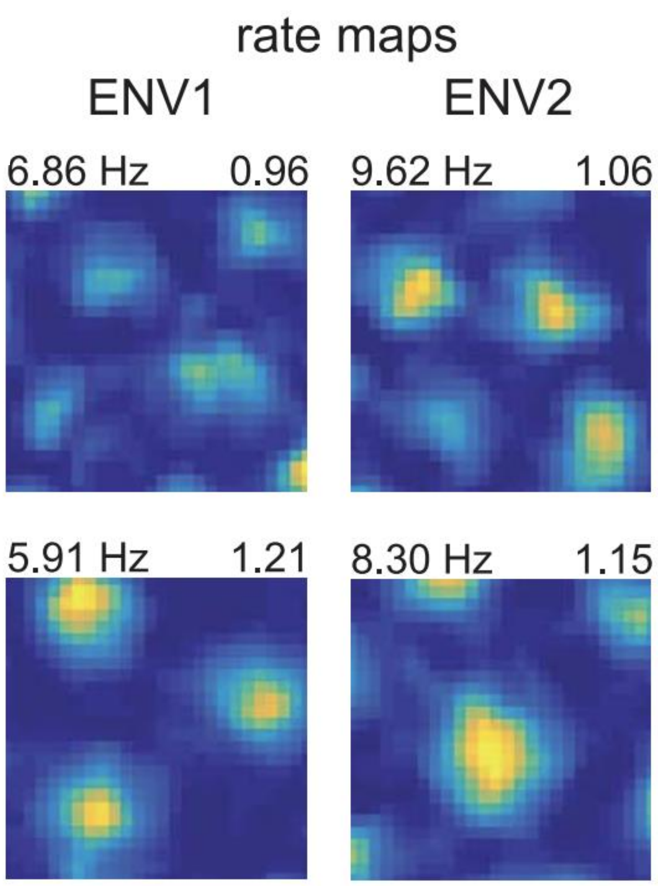


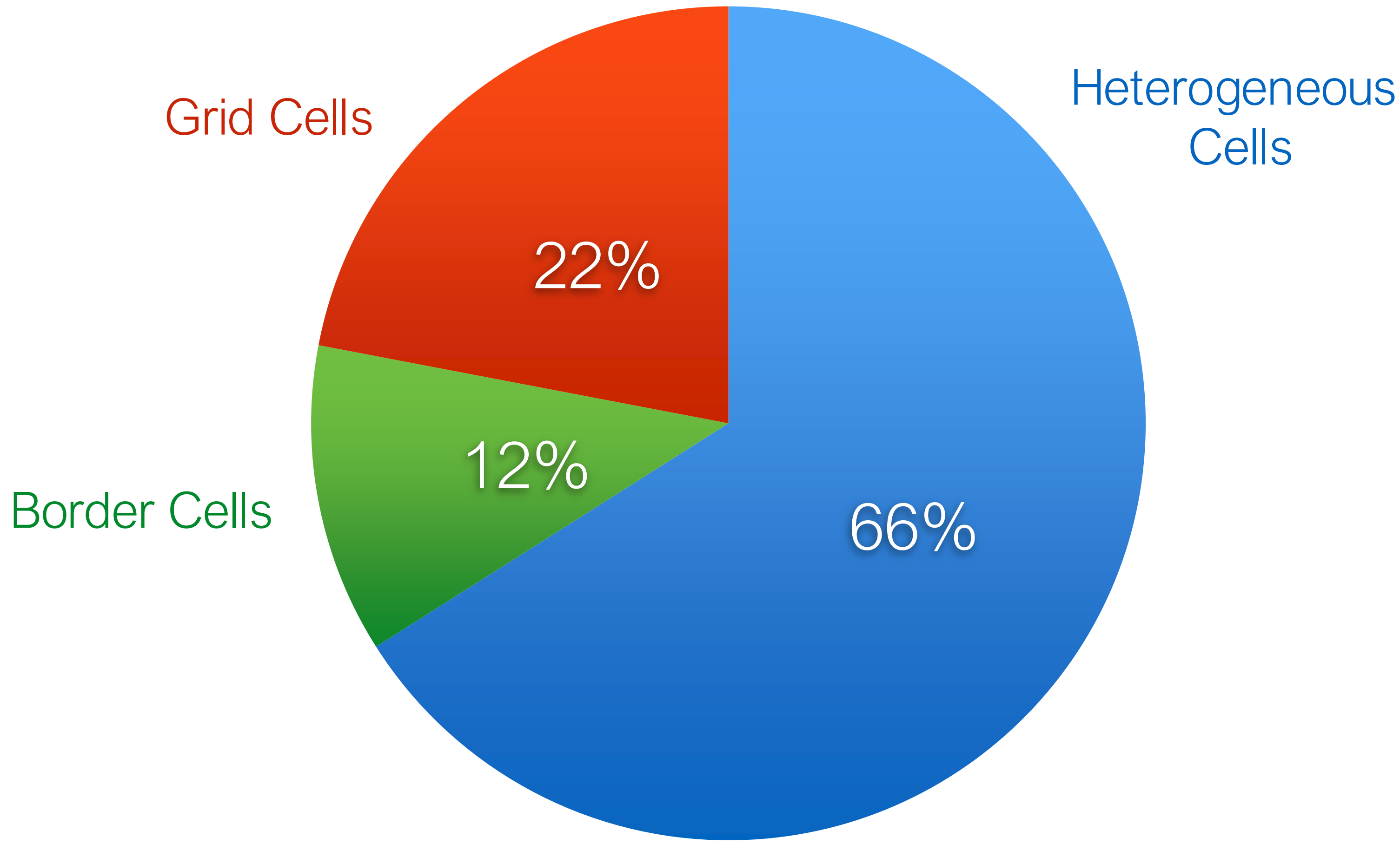
spatial task (ENV2)

3. In fact, MEC remaps in the presence of rewards… so what describes the <u>joint</u> interaction between these heterogeneous cells and reward?

# Taking a modeling approach

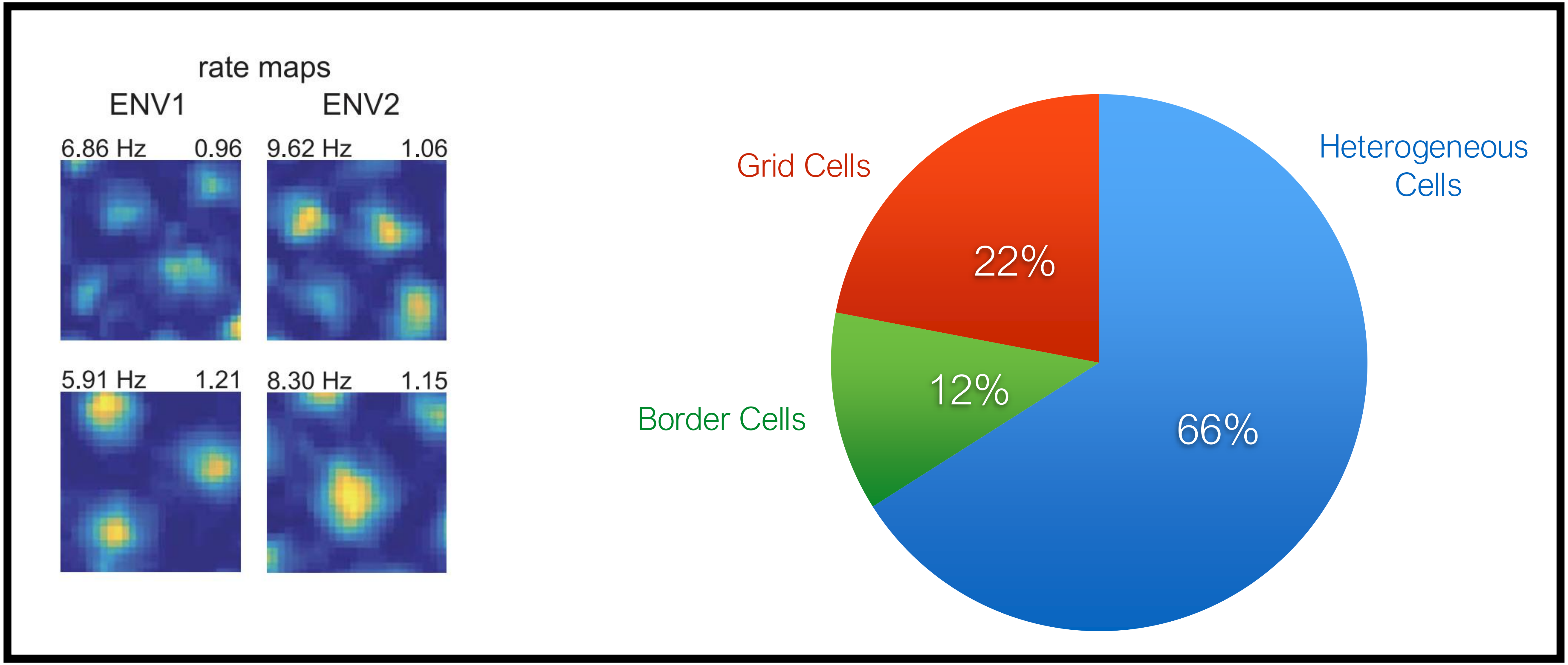
It would be useful to have a **unified** model that can simultaneously explain different types of neural responses in MEC.



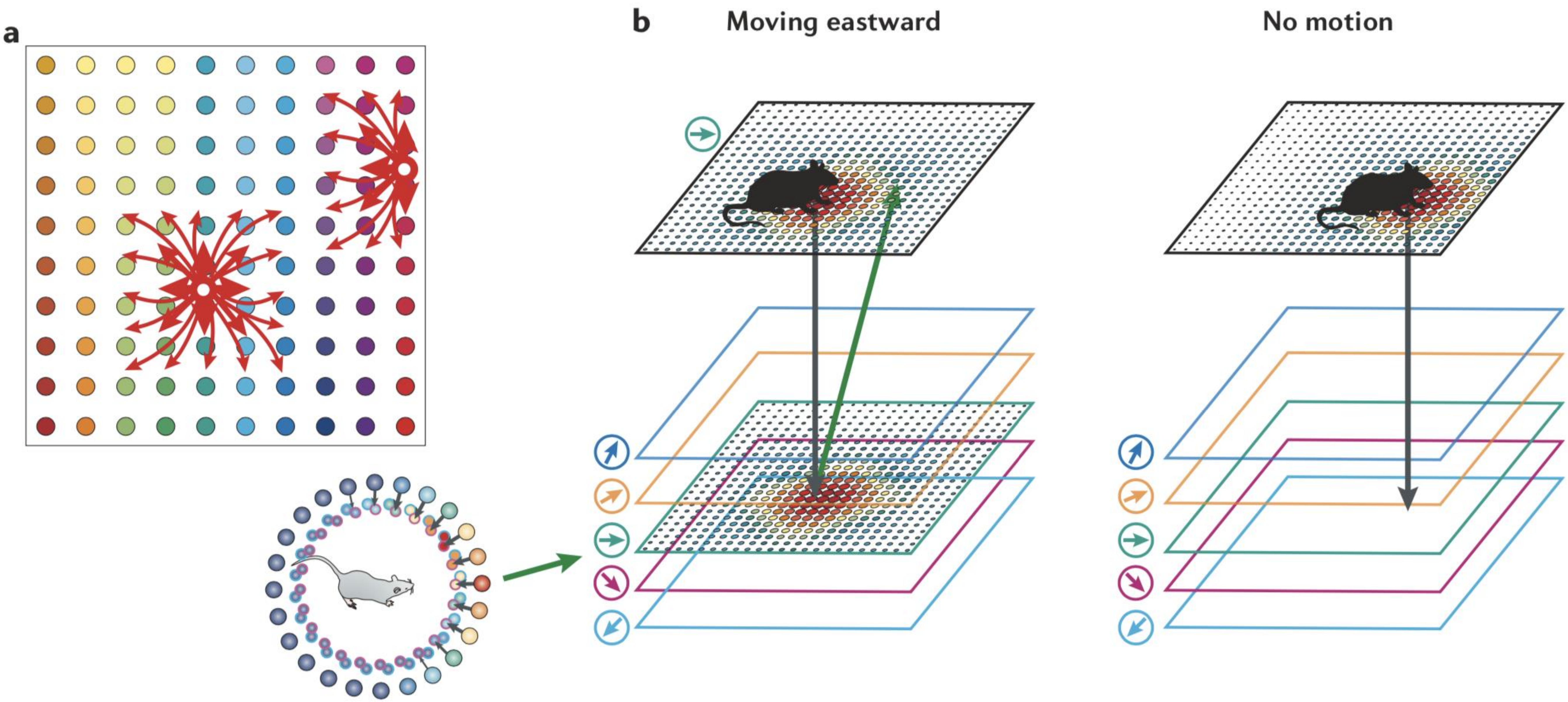
# Taking a modeling approach

It would be useful to have a **unified** model that can simultaneously explain different types of neural responses in MEC.



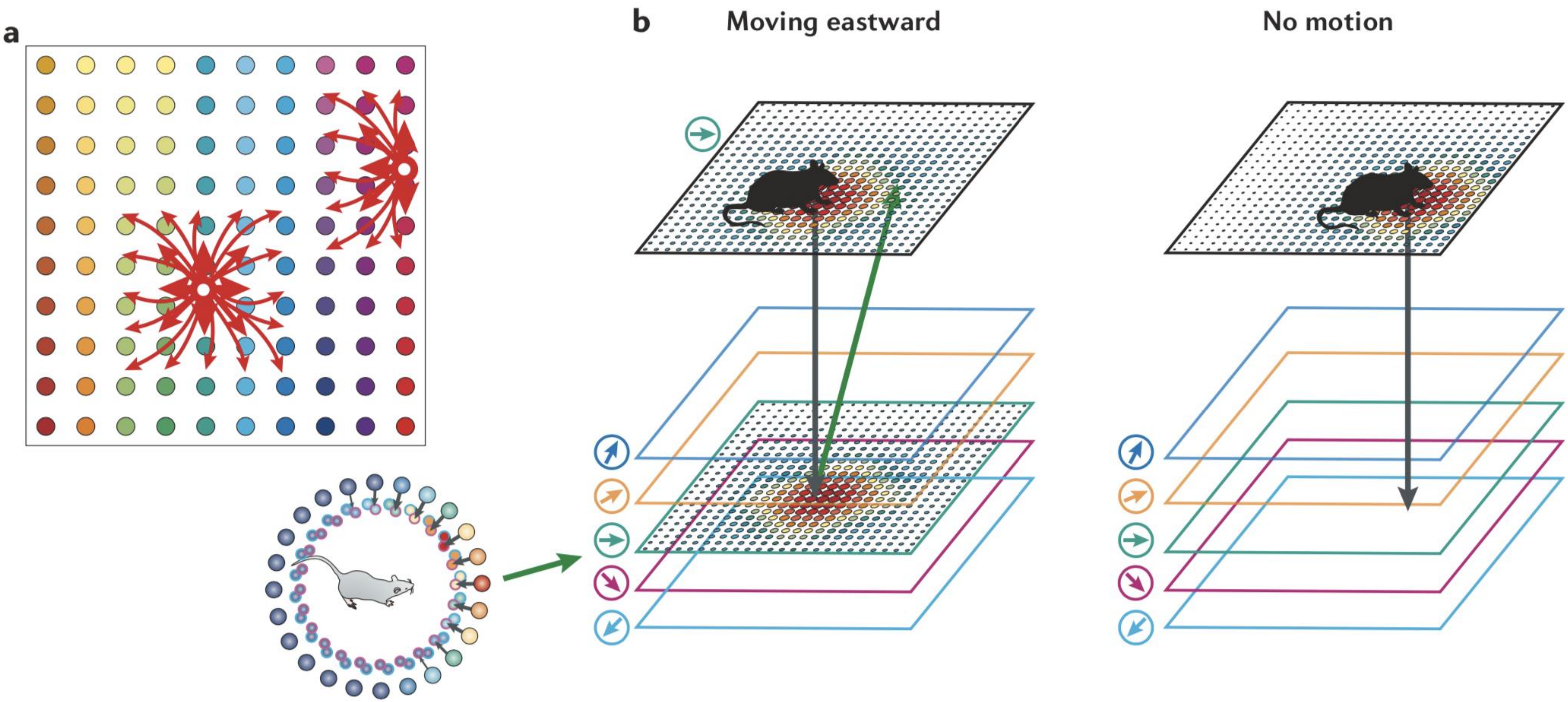

Where do we begin?

# “Hand-Tuned” Attractor Models - 2D Case




McNaughton et al. 2006

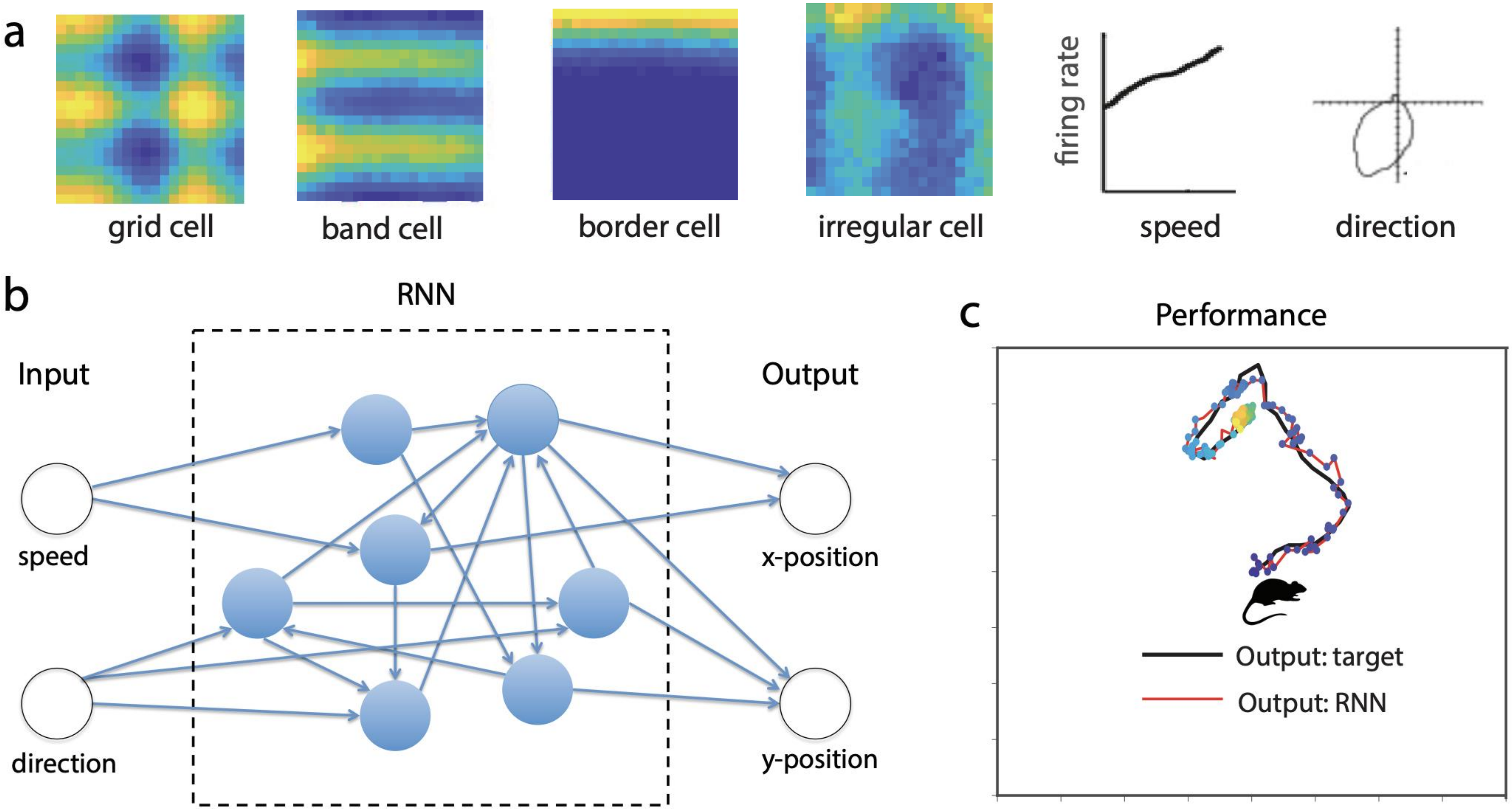
# “Hand-Tuned” Attractor Models - 2D Case




McNaughton et al. 2006

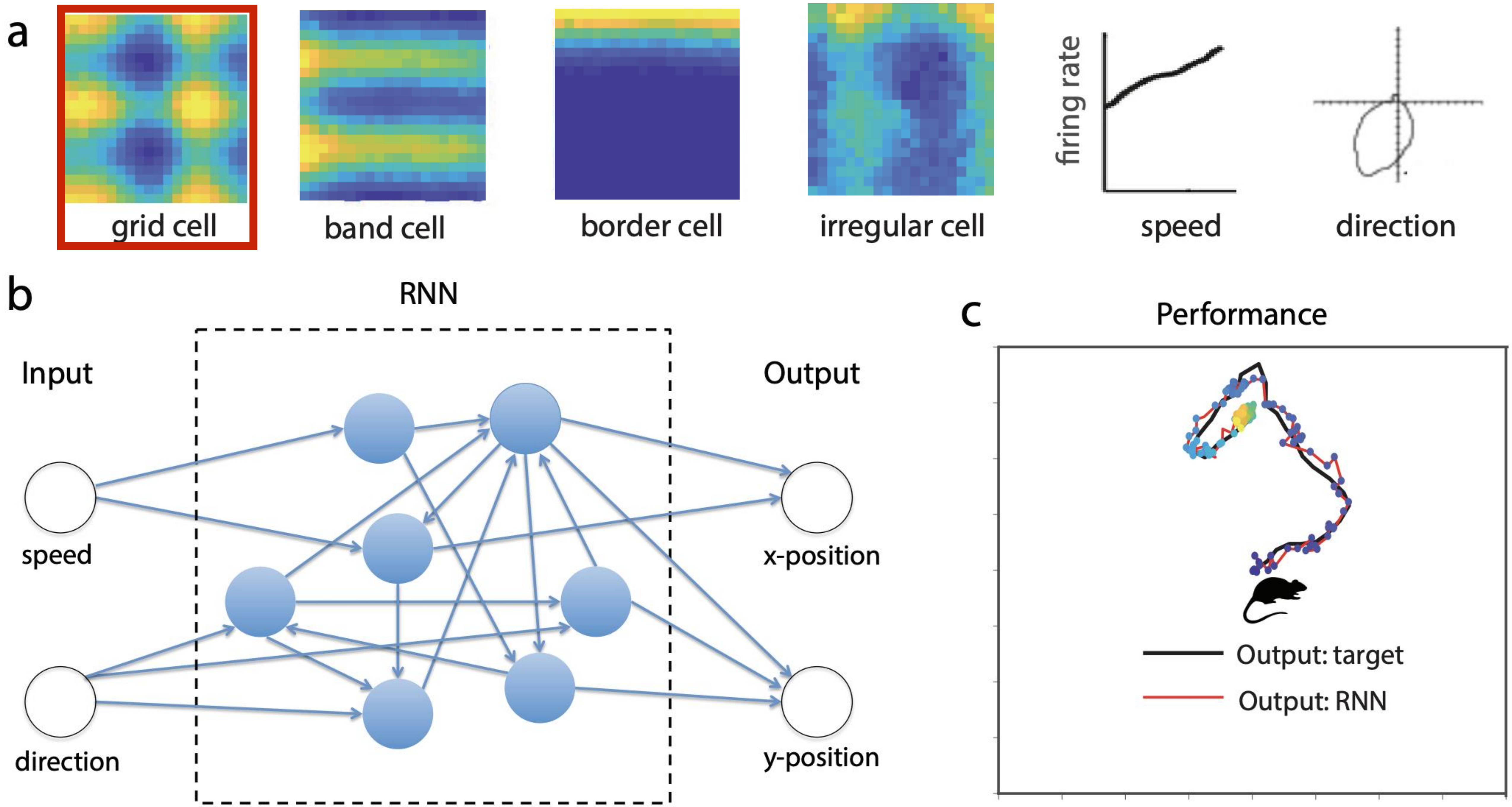
But these hand tuned models capture the properties of stereotypical cell-type classes

# But more recently there are neural network models that “develop” these cells…



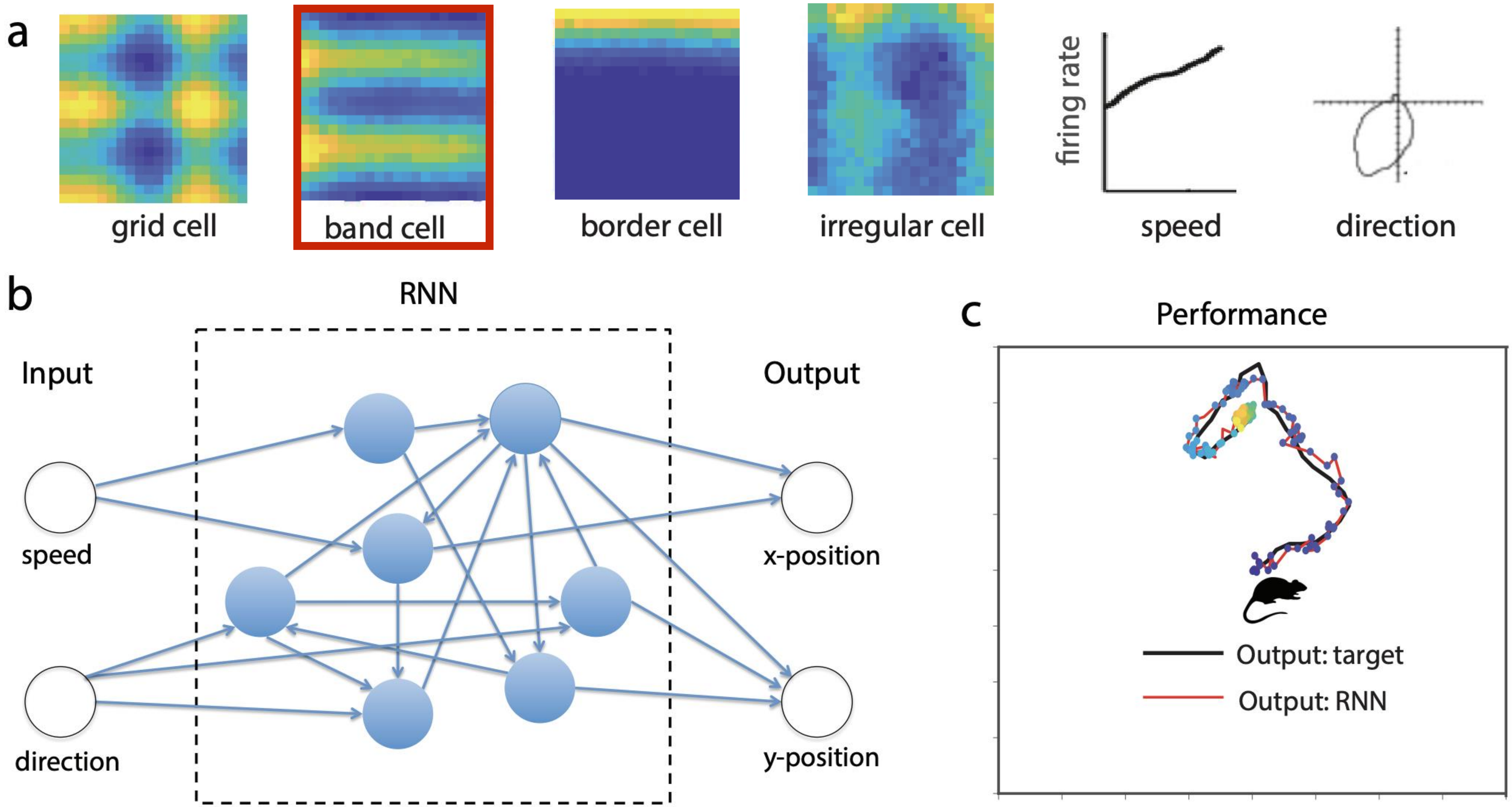

Cueva* & Wei* 2018

# But more recently there are neural network models that “develop” these cells…



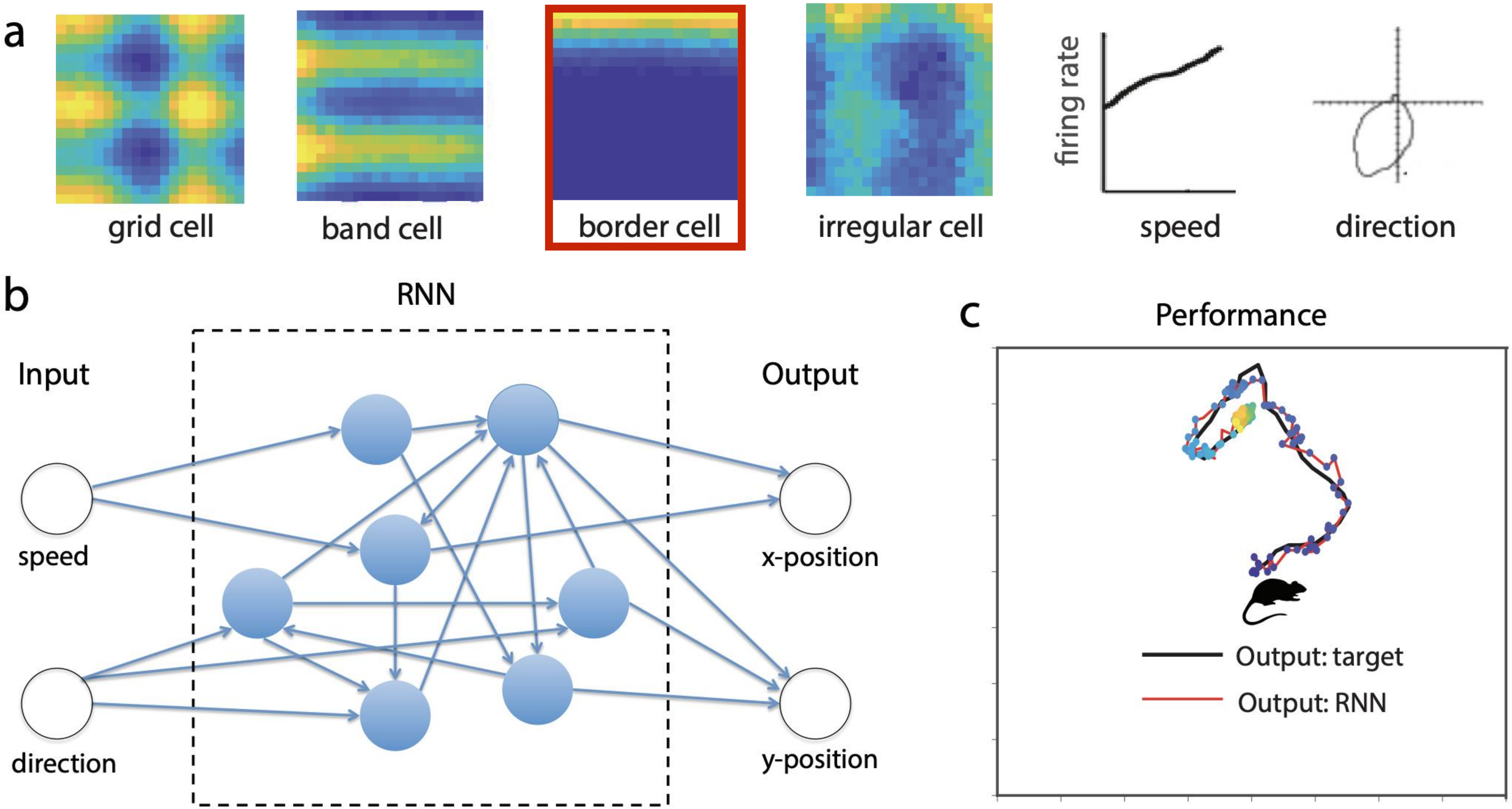

Cueva* & Wei* 2018

# But more recently there are neural network models that "develop" these cells…



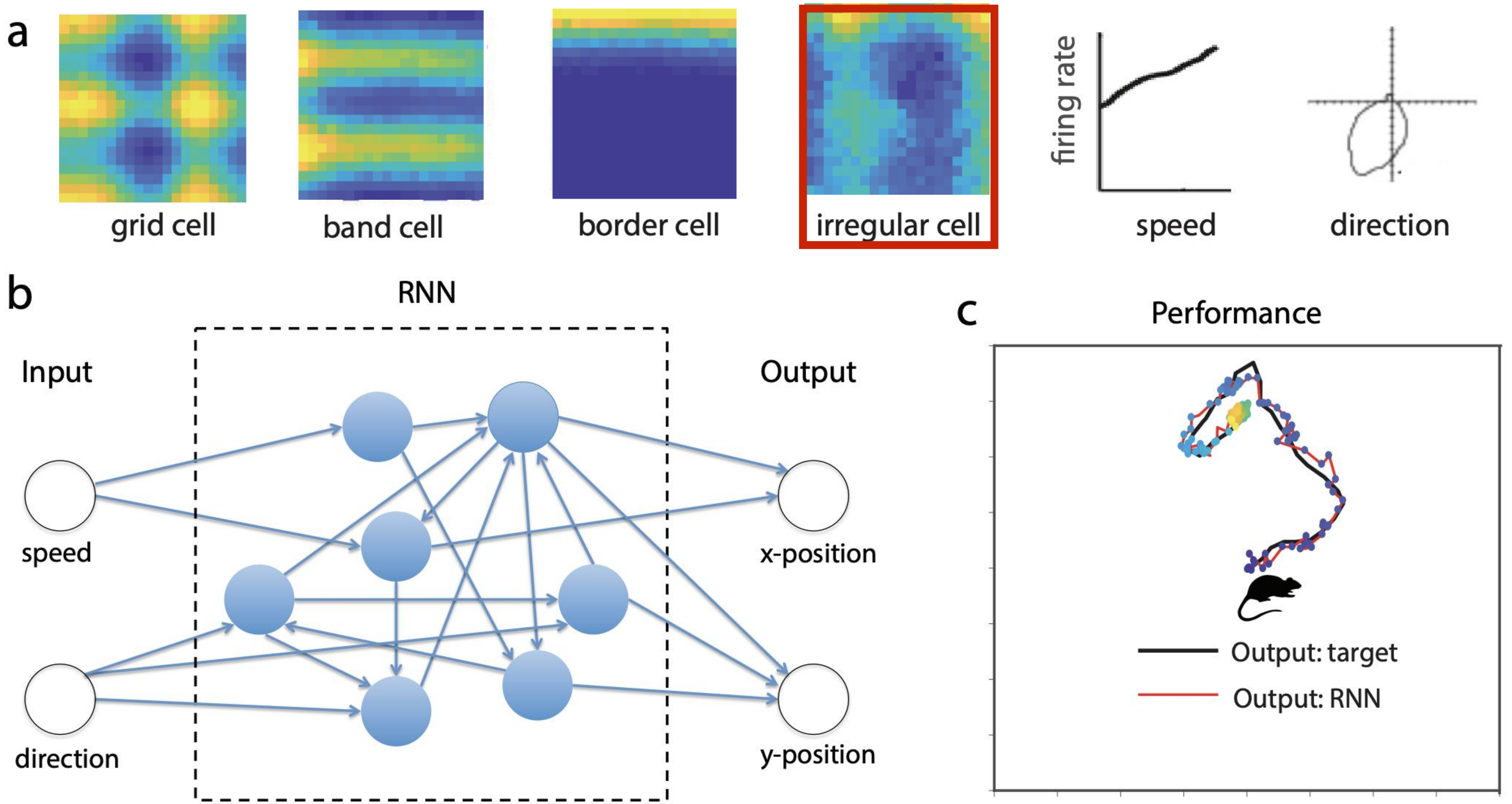

Cueva* & Wei* 2018

# But more recently there are neural network models that "develop" these cells…




Cueva* & Wei* 2018

# But more recently there are neural network models that "develop" these cells…

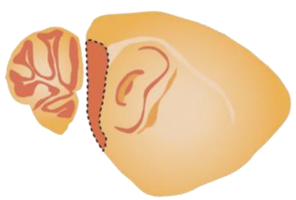


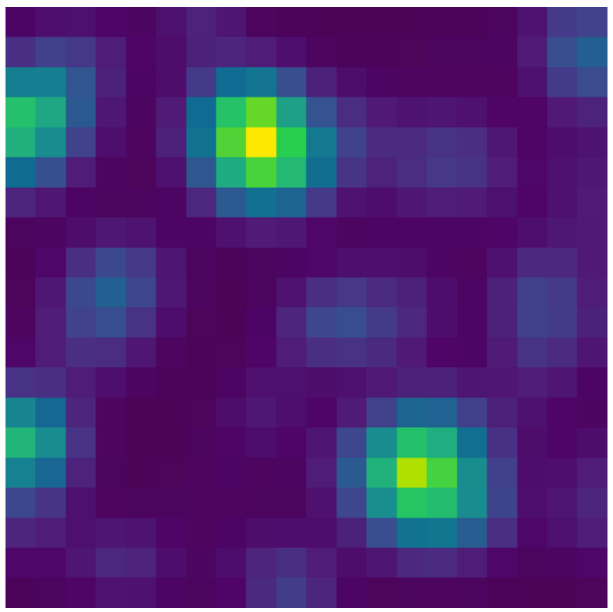

Cueva* & Wei* 2018

# Task-Driven Approach

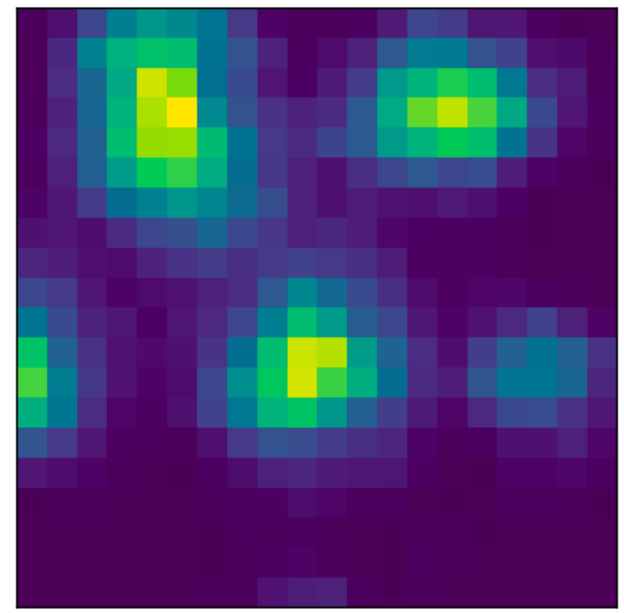
But are they a good ***quantitative*** model of these responses?

MEC Grid Cell

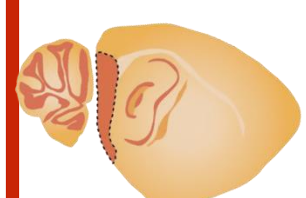


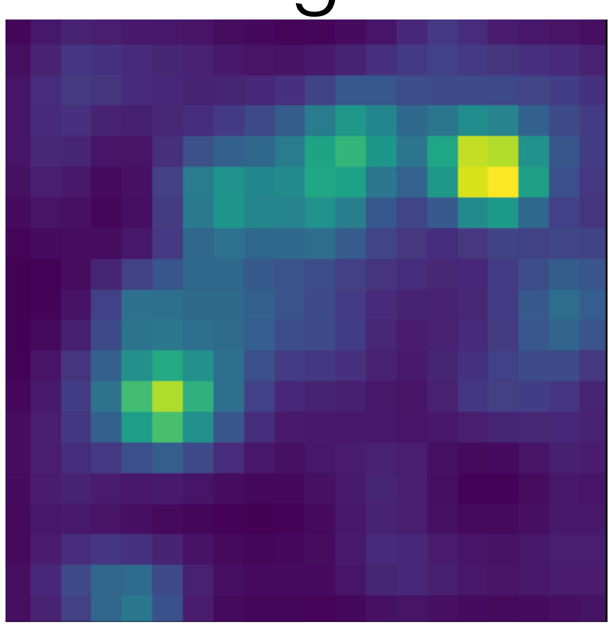


Model Grid Cell



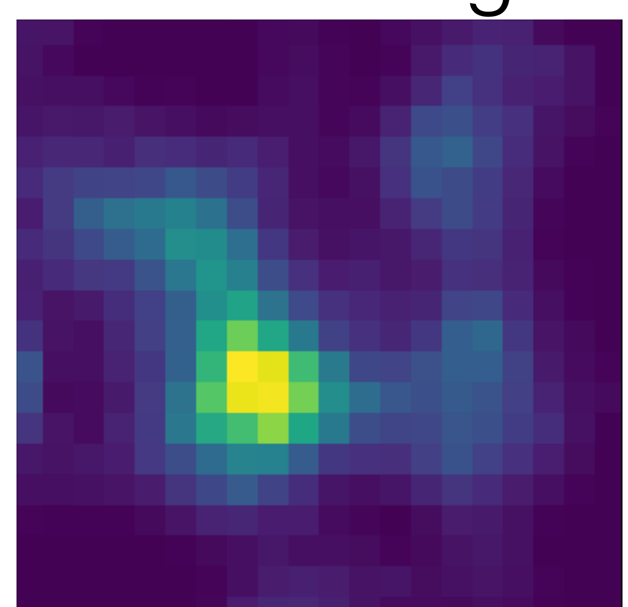
MEC Heterogeneous Cell





?

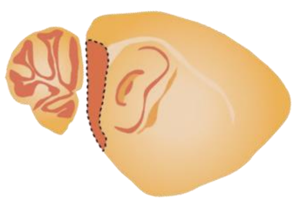
Model Heterogeneous Cell

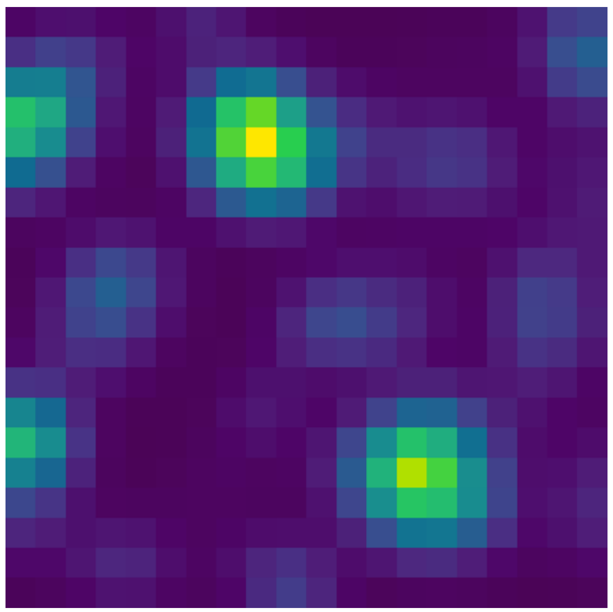


# Task-Driven Approach

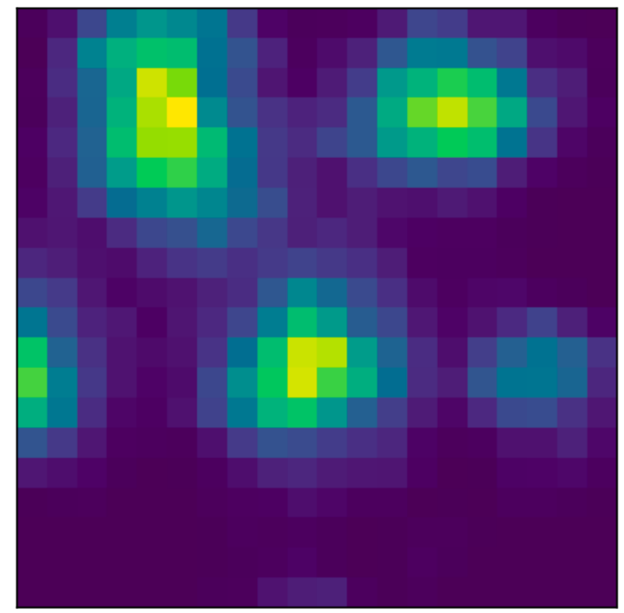
But are they a good ***quantitative*** model of these responses?

MEC Grid Cell

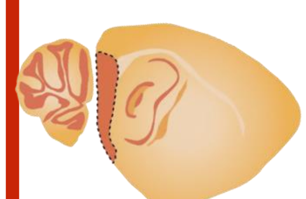


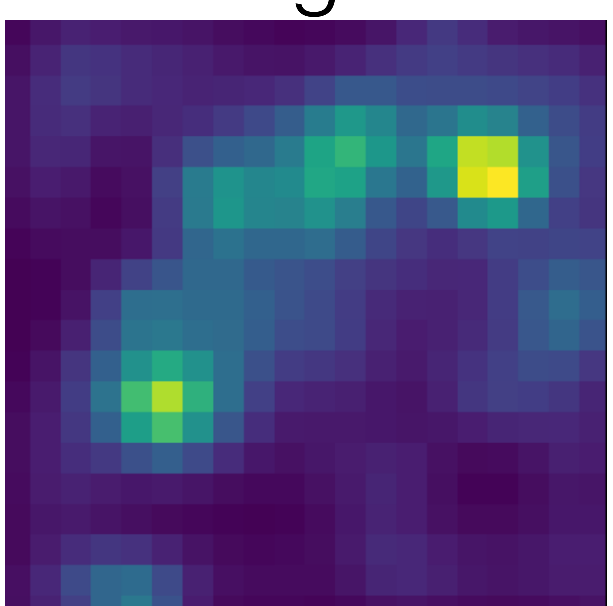


Model Grid Cell



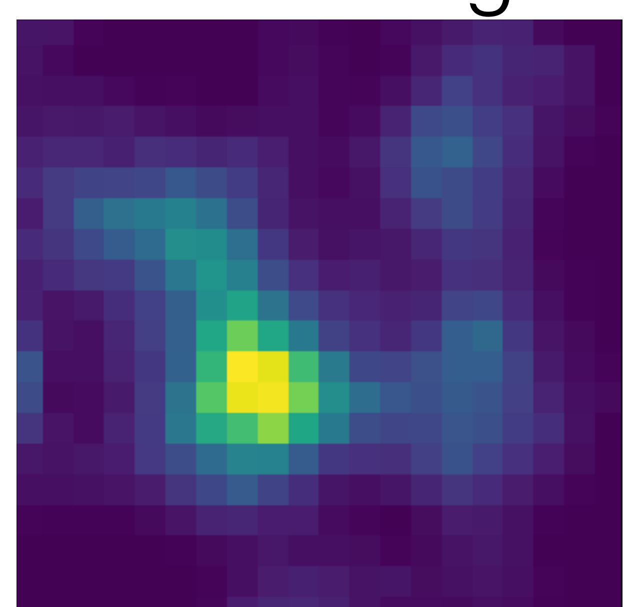
MEC Heterogeneous Cell





**Not all models are equal!**

Model Heterogeneous Cell



# Task-Driven Modeling: Four Components

## Task-Optimization (ML)

**1.**

***A*** = *architecture class*

**2.**

***T*** = *task loss*

**3.**

***D*** = *dataset*

**4.**

***L*** = *learning rule*

# Task-Driven Modeling: Four Components

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

<u>Neurobiology</u>

1.

***A*** = *architecture class*

2.

***T*** = *task loss*

3.

***D*** = *dataset*

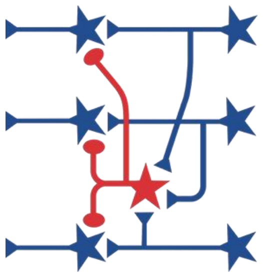
4.

***L*** = *learning rule*

# Task-Driven 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*

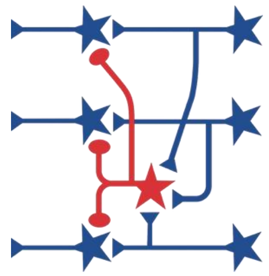
# Task-Driven Modeling: Four Components

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

<u>Neurobiology</u>

**1.**

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



**2.**

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

**3.**

***D*** = *dataset*

**4.**

***L*** = *learning rule*

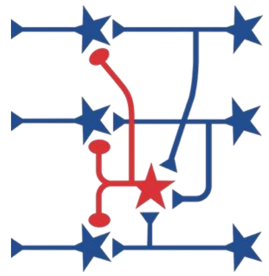
# Task-Driven Modeling: Four Components

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

<u>Neurobiology</u>

**1.**

| |
|---|
| ***A*** = *architecture class* **= *circuit neuroanatomy*** |

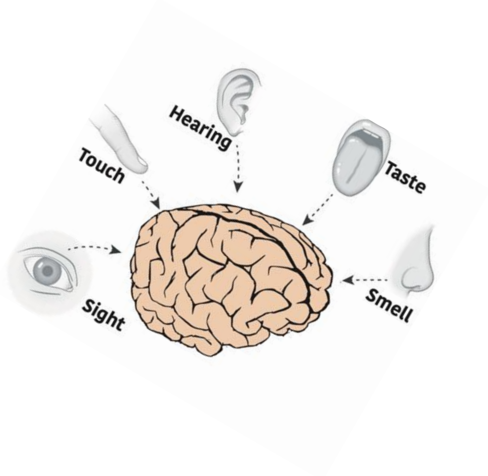


**2.**

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

**3.**

| |
|---|
| ***D*** = *dataset* **= *environment*** |




**4.**

| |
|---|
| ***L*** = *learning rule* |

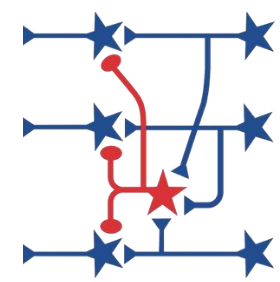
# Task-Driven Modeling: Four Components

## Task-Optimization (ML)

## Neurobiology

**1.**

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

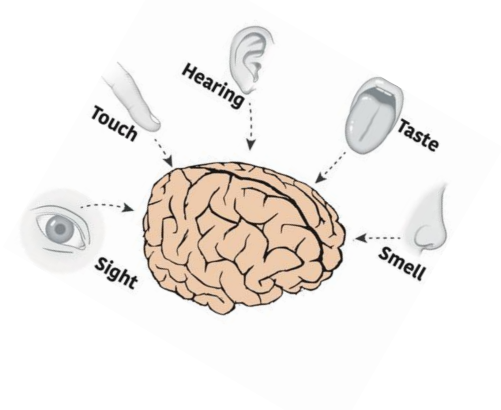


**2.**

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

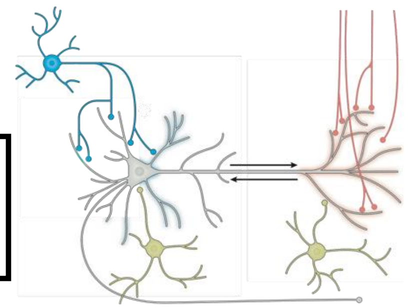
**3.**

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




**4.**

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



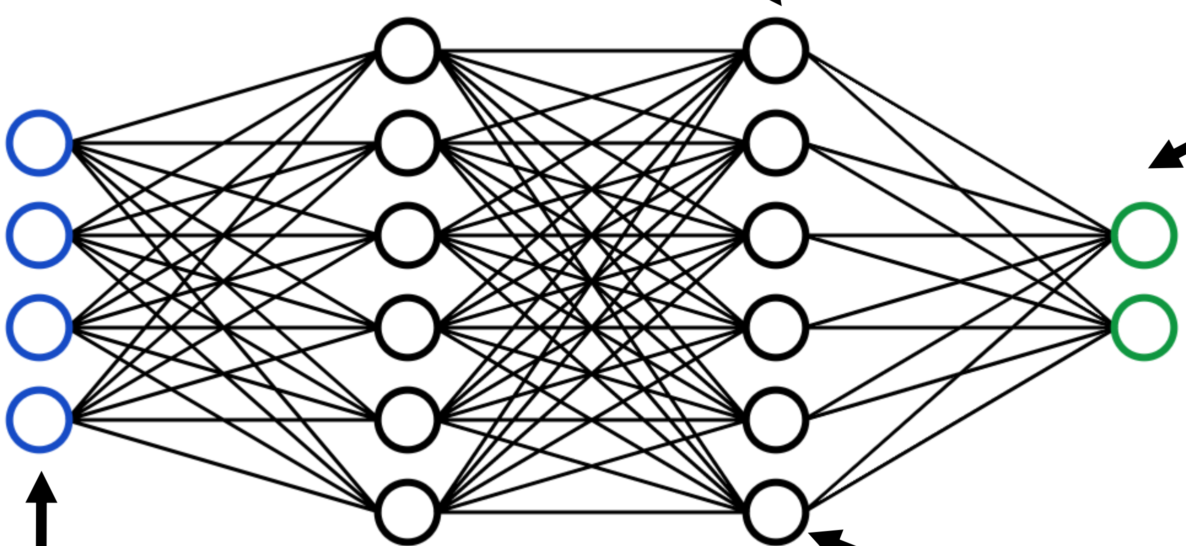
# Task-Driven Modeling: Four Components

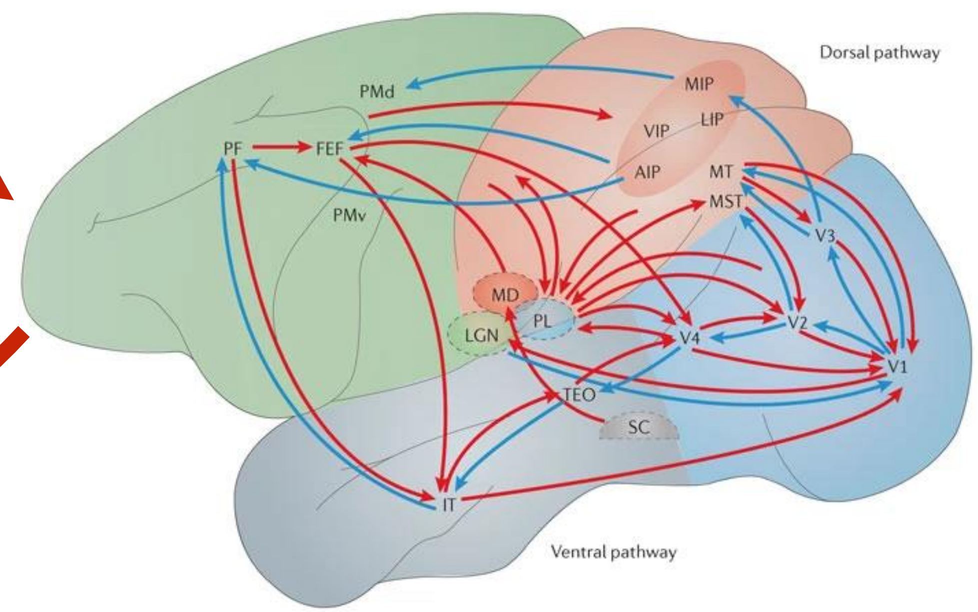
***L*** = *learning rule*

***"Natural selection + plasticity"***

***T*** = *task loss*

***"Ecological niche/behavior"***






***"Environment"***

***D*** = *data stream*

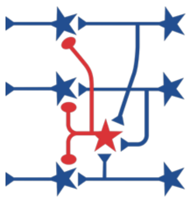
***"Circuit"***

***A*** = *architecture class*

# Task-Driven Approach

***A*** = *architecture class*

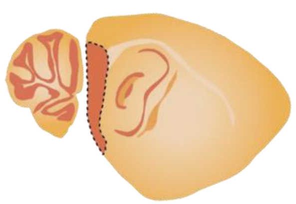
**1.** *"Circuit"*

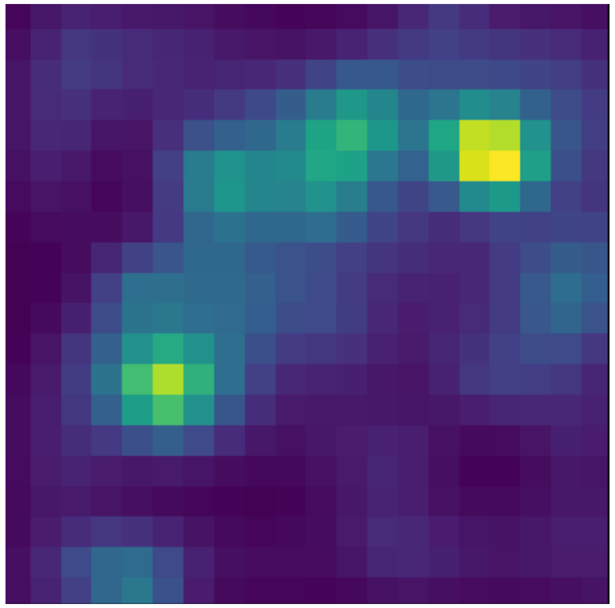


***T*** = *task loss*

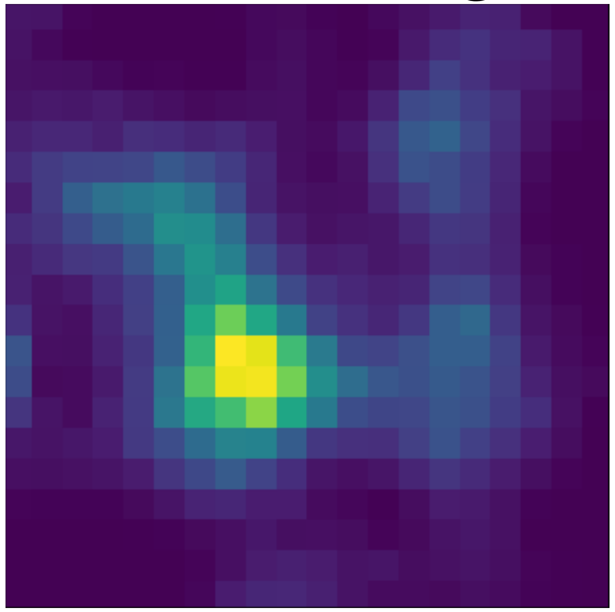
**3.** *"Ecological niche/behavior"*

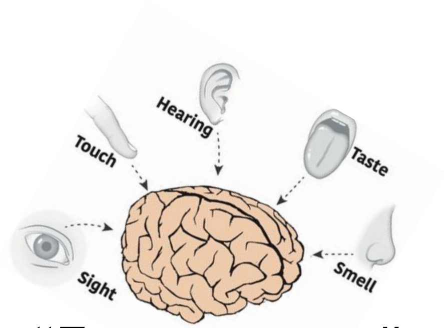
MEC Heterogeneous Cell





Model Heterogeneous Cell






**2.** *"Environment"*

***D*** = *data stream*

# Task-Driven Approach

***A*** = *architecture class*

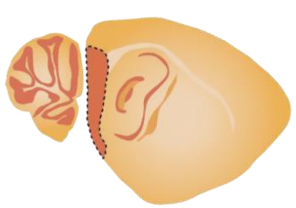
**1.** *"Circuit"*

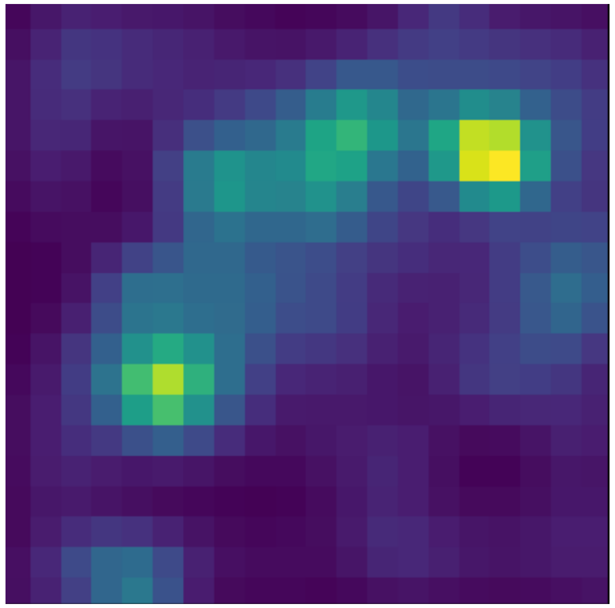


***T*** = *task loss*

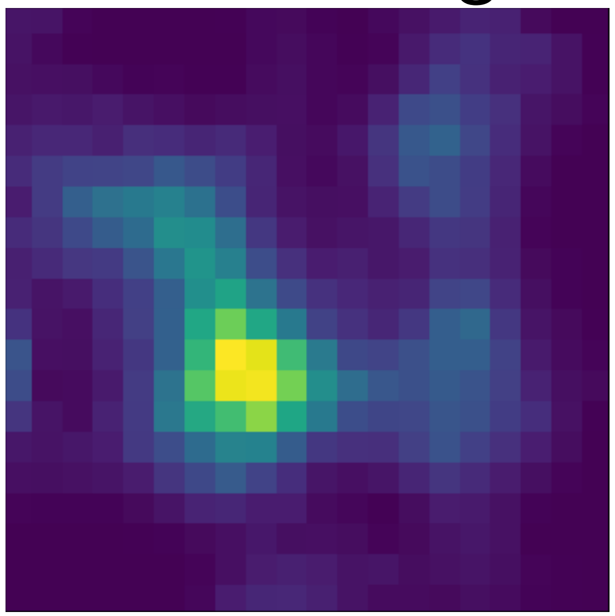
**3.** *"Ecological niche/behavior"*

MEC Heterogeneous Cell





Model Heterogeneous Cell

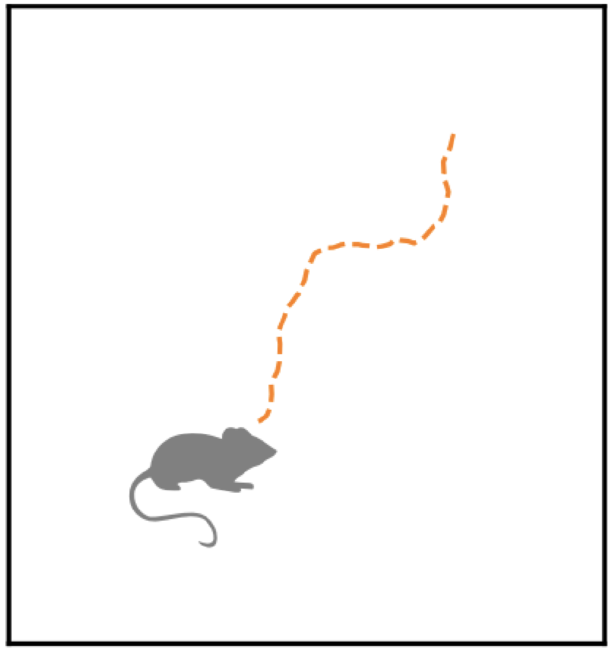



**2.** *"Environment"*

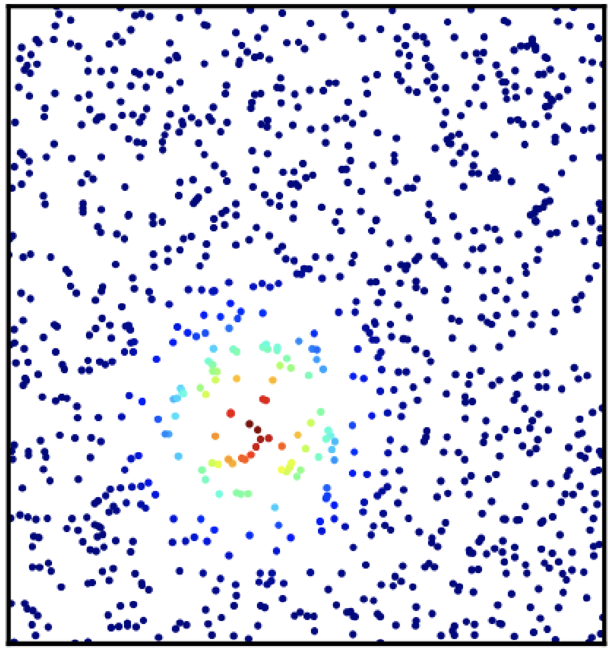
***D*** = *data stream*

# A spectrum of tasks

Simulated trajectory



Place cell centers

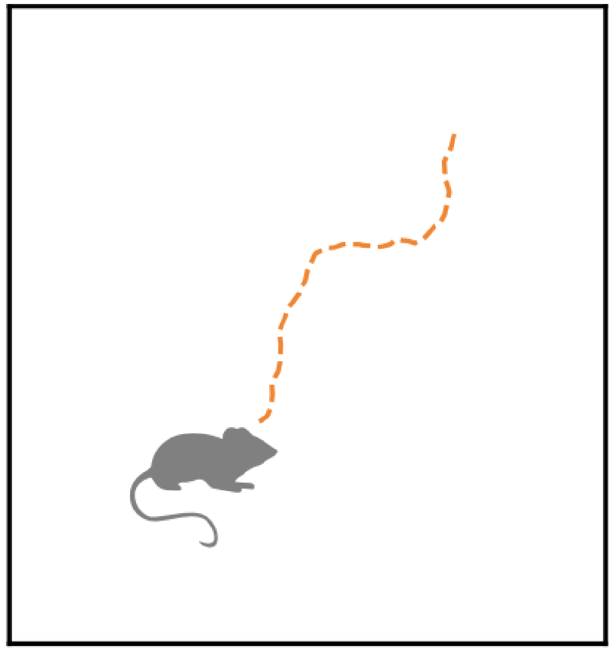


*Banino*, Barry* et al. 2018*

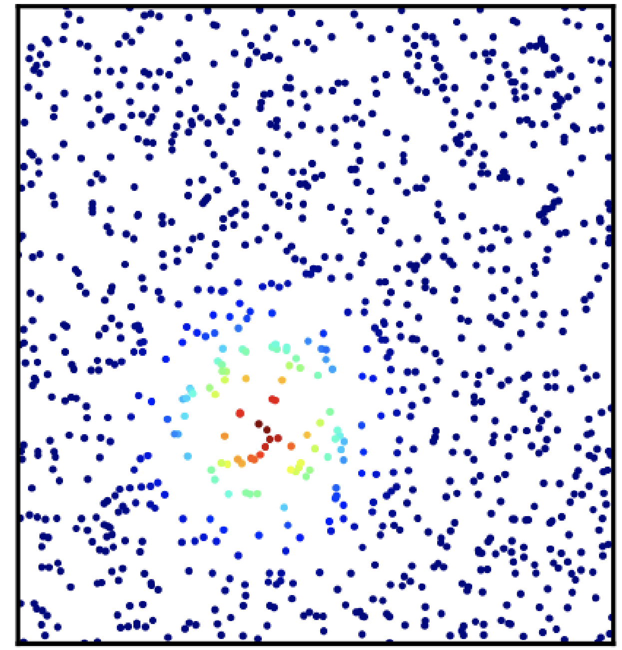
*Sorscher*, Mel* et al. 2019*

# A spectrum of tasks

Simulated trajectory



Place cell centers



*Banino*, Barry* et al. 2018*

*Sorscher*, Mel* et al. 2019*

Velocity → MEC → Place Cells → Position (x,y)

# A spectrum of tasks

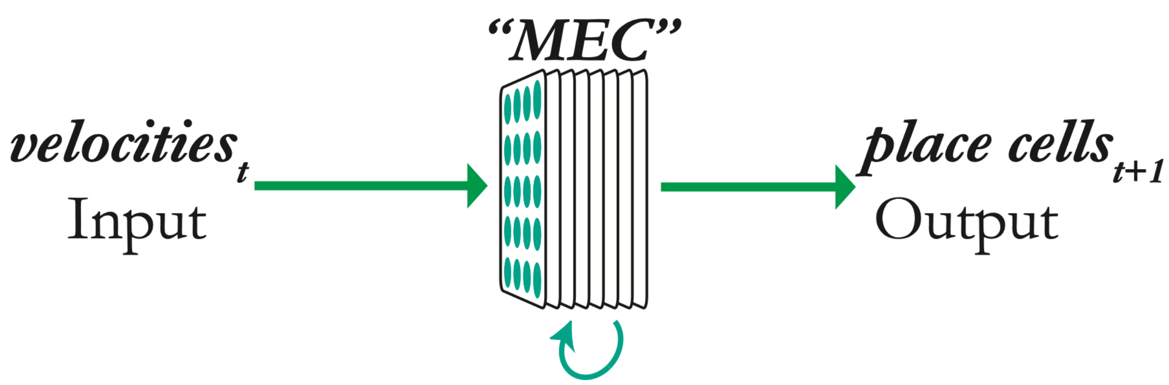
Simplest “model”

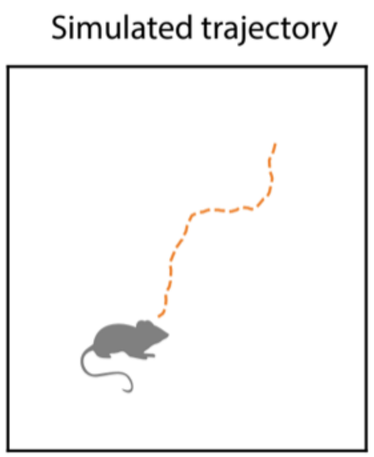
$\boxed{\text{Velocity}} \longrightarrow \text{MEC} \longrightarrow \text{Place Cells} \longrightarrow \text{Position } (x,y)$

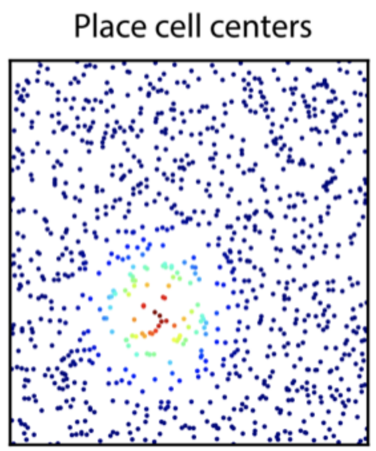
# A spectrum of tasks

$$\mathcal{L}(\hat{p}, p) := -\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N_p} p_i^t \log \hat{p}_i^t$$

*Banino*, Barry* et al. 2018*



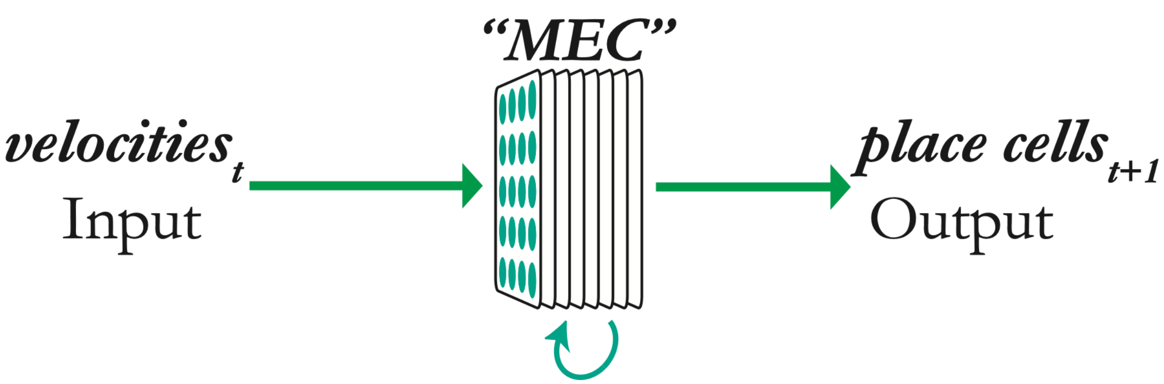




Velocity → MEC → Place Cells → Position (x,y)

# A spectrum of tasks

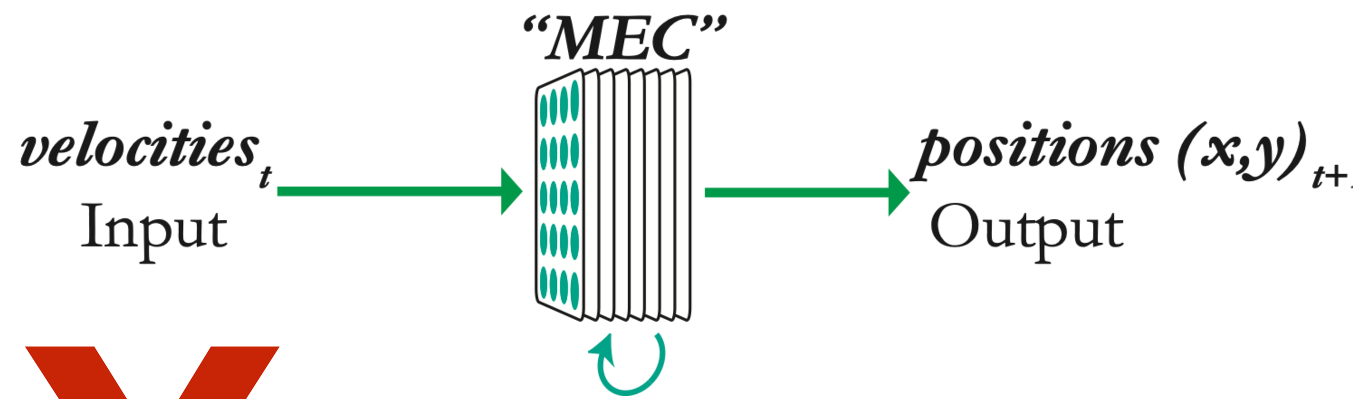
*Banino*, Barry* et al. 2018*

$$\mathcal{L}(\hat{p}, p) := -\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N_p} p_i^t \log \hat{p}_i^t$$




*Cueva* & Wei* 2018*

$$\mathcal{L}(\hat{p}, p) := \frac{1}{2}\frac{1}{T}\sum_{t=1}^{T}\left(\left(p_x^t - \hat{p}_x^t\right)^2 + \left(p_y^t - \hat{p}_y^t\right)^2\right)$$

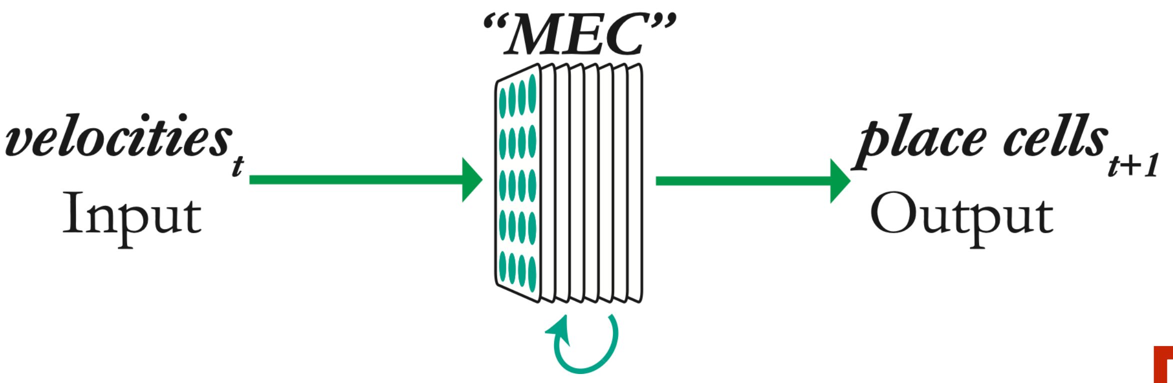



Velocity → MEC → Place Cells → Position (x,y)

# A spectrum of tasks

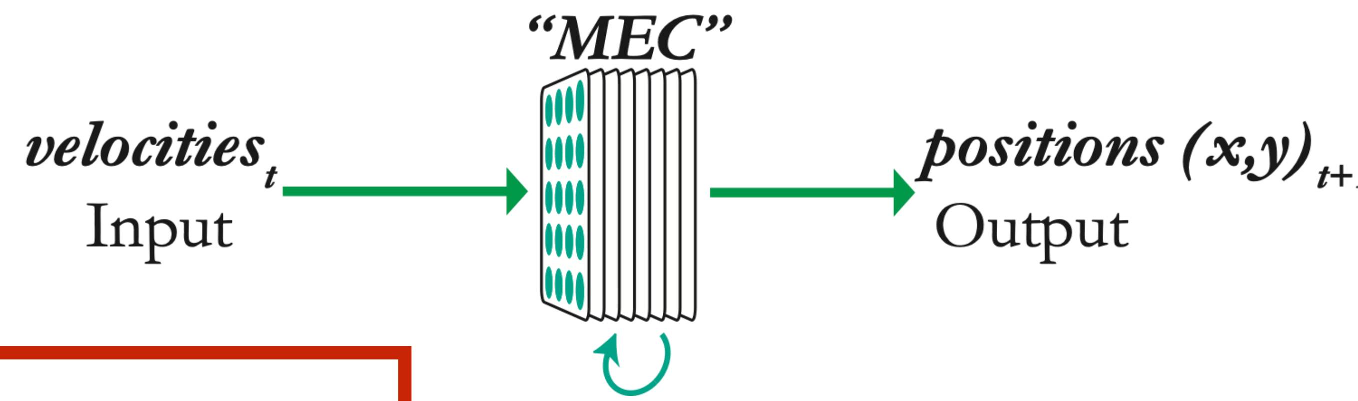
$$\mathcal{L}(\hat{p}, p) := -\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N_p} p_i^t \log \hat{p}_i^t$$

*Banino*, Barry* et al. 2018*




$$\mathcal{L}(\hat{p}, p) := \frac{1}{2}\frac{1}{T}\sum_{t=1}^{T}\left(\left(p_x^t - \hat{p}_x^t\right)^2 + \left(p_y^t - \hat{p}_y^t\right)^2\right)$$

*Cueva* & Wei* 2018*



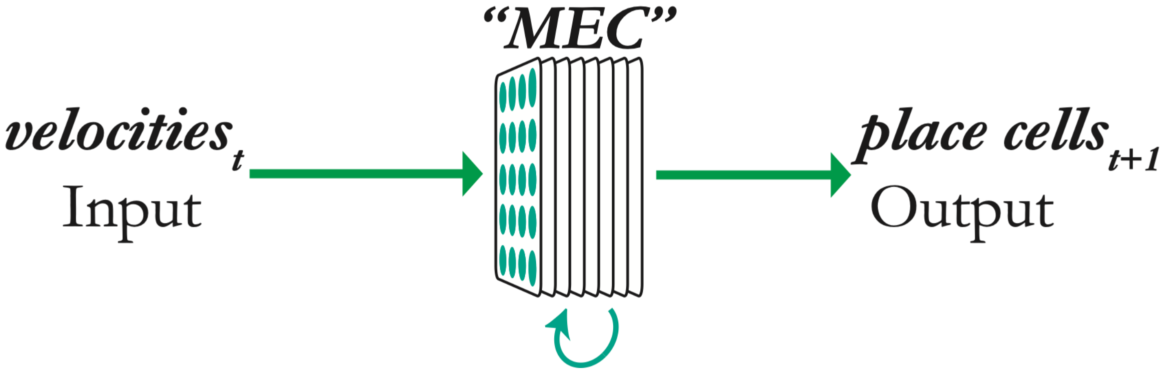

Velocity → MEC → $\boxed{\text{Place Cells}}$ → Position (x,y)

Output-based models

# A spectrum of tasks

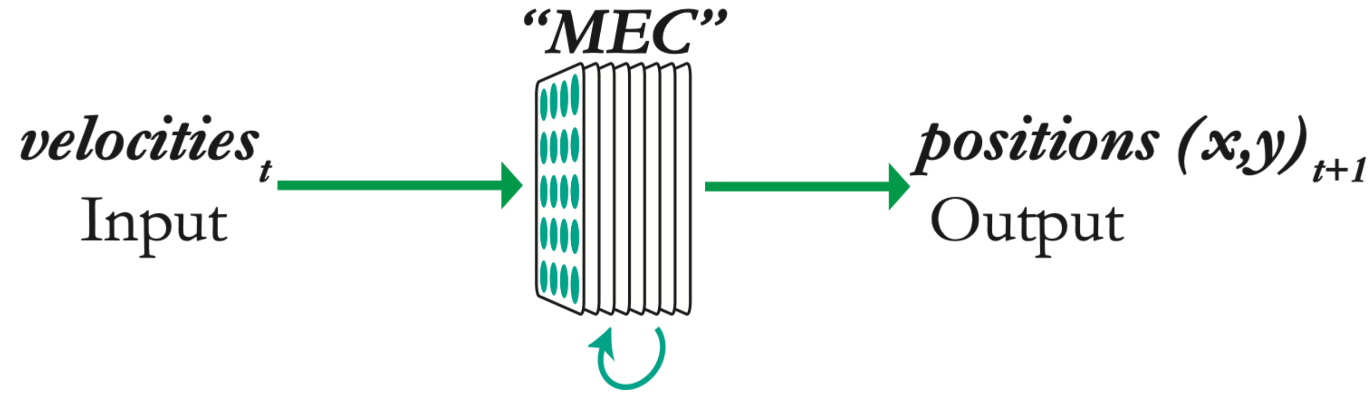
*Banino*, Barry* et al. 2018*

$$\mathcal{L}(\hat{p}, p) := -\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N_p} p_i^t \log \hat{p}_i^t$$



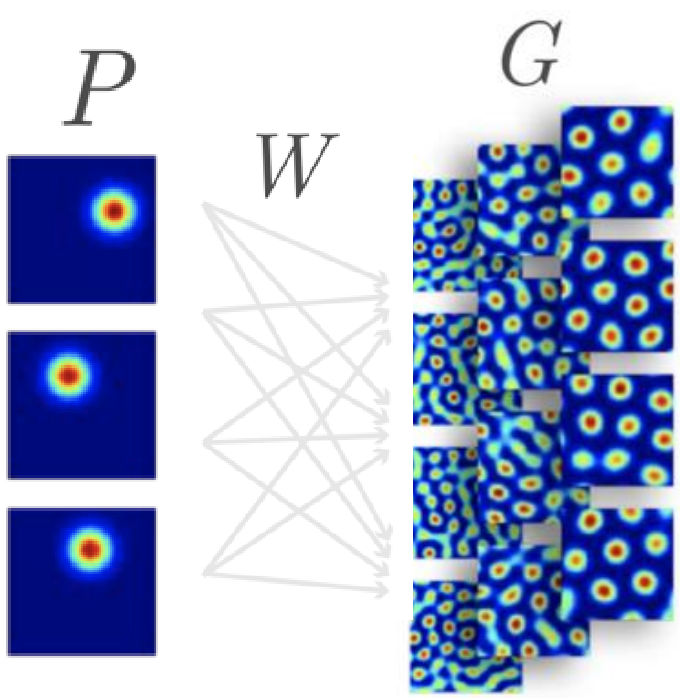

*Cueva* & Wei* 2018*

$$\mathcal{L}(\hat{p}, p) := \frac{1}{2}\frac{1}{T}\sum_{t=1}^{T}\left(\left(p_x^t - \hat{p}_x^t\right)^2 + \left(p_y^t - \hat{p}_y^t\right)^2\right)$$




Velocity ⟶ MEC ⟶ Place Cells ⟶ Position (x,y)

**NMF**
**(Place Cell Input)**

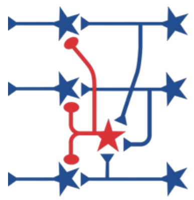



*Dordek et al. 2016*

# Task-Driven Approach

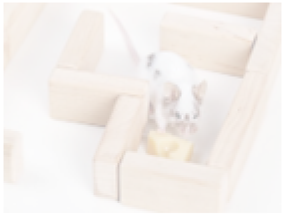
***A*** *= architecture class*

**1.** *"Circuit"*

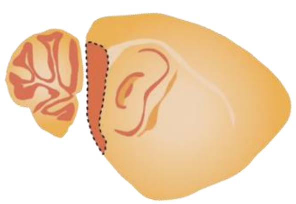


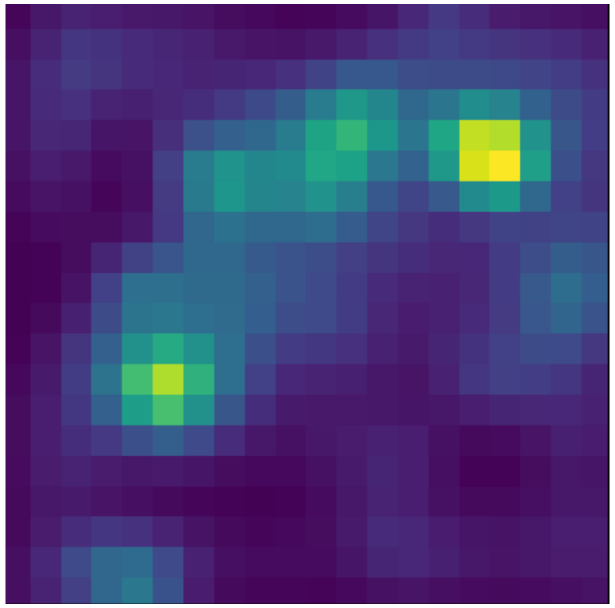
**3.** *"Ecological niche/behavior"*

***T*** *= task loss*

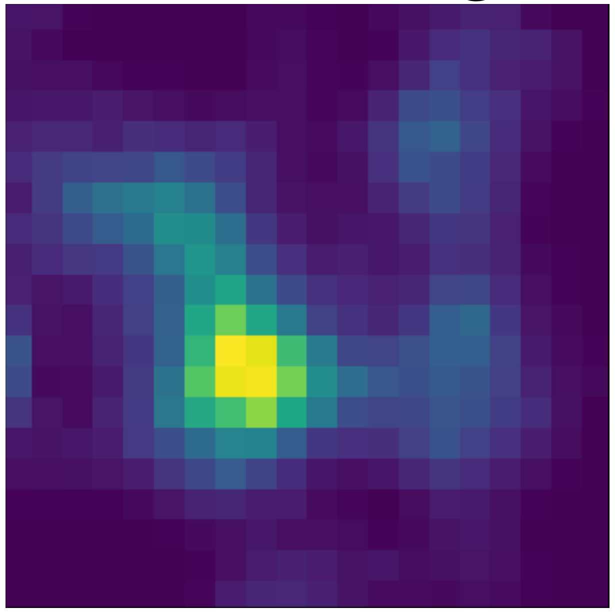


MEC Heterogeneous Cell





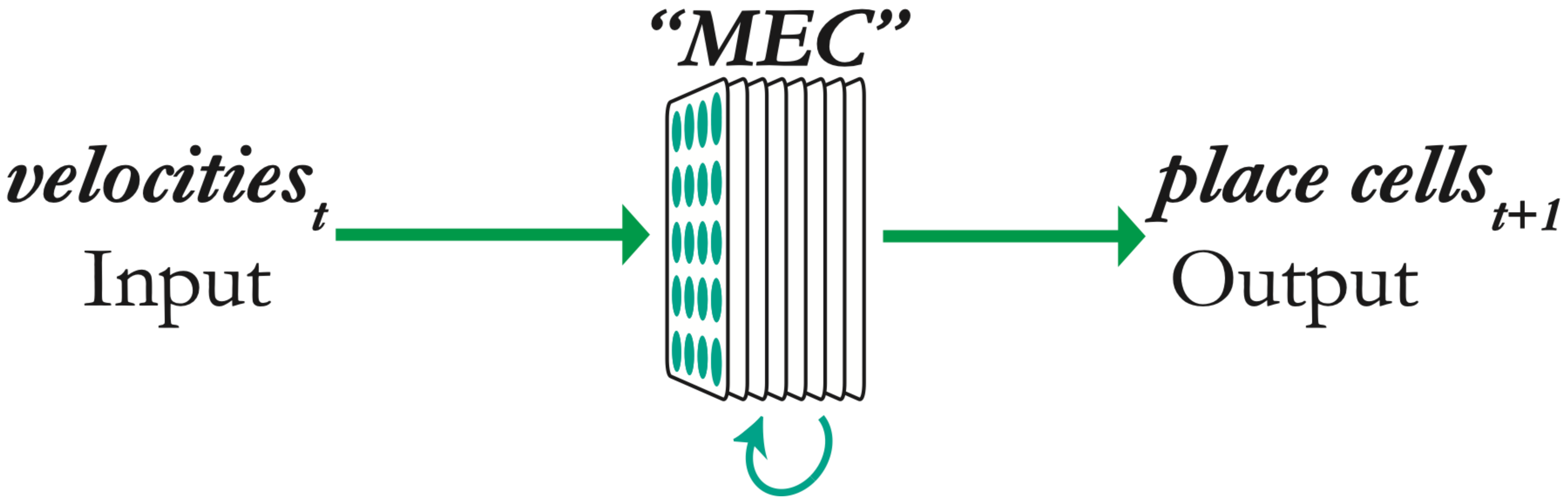
Model Heterogeneous Cell



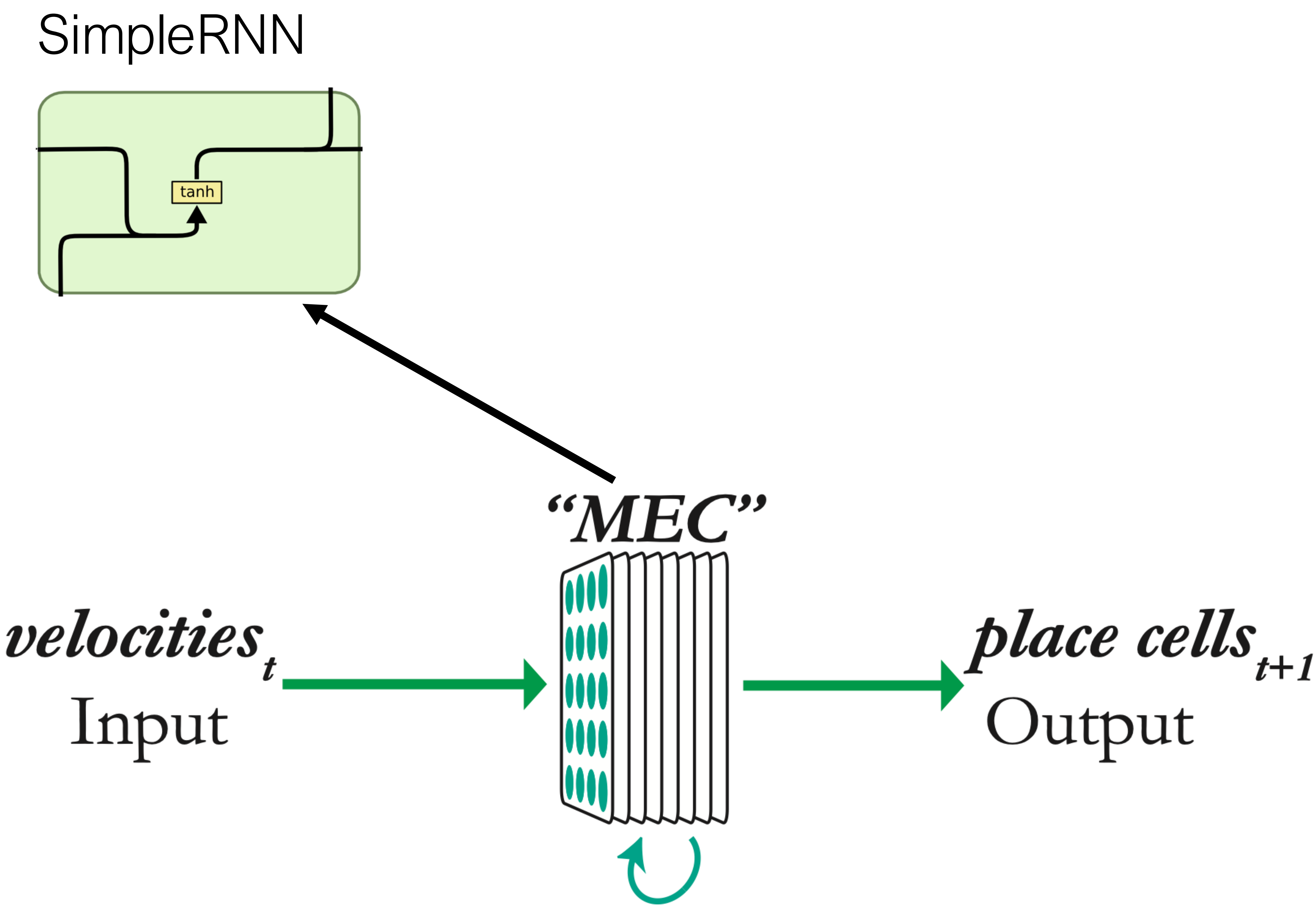

**2.** *"Environment"*

***D*** *= data stream*

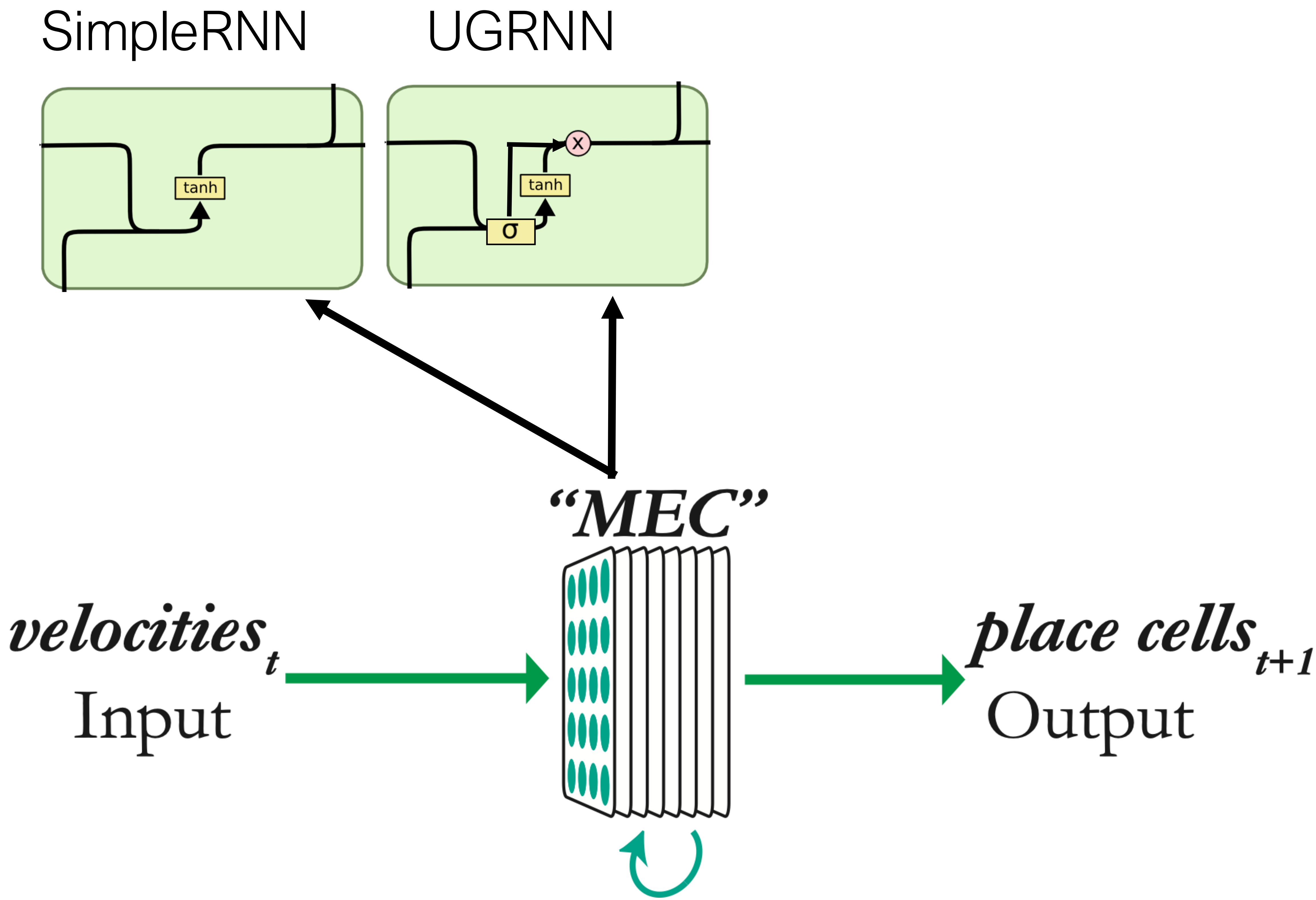
# A spectrum of circuits



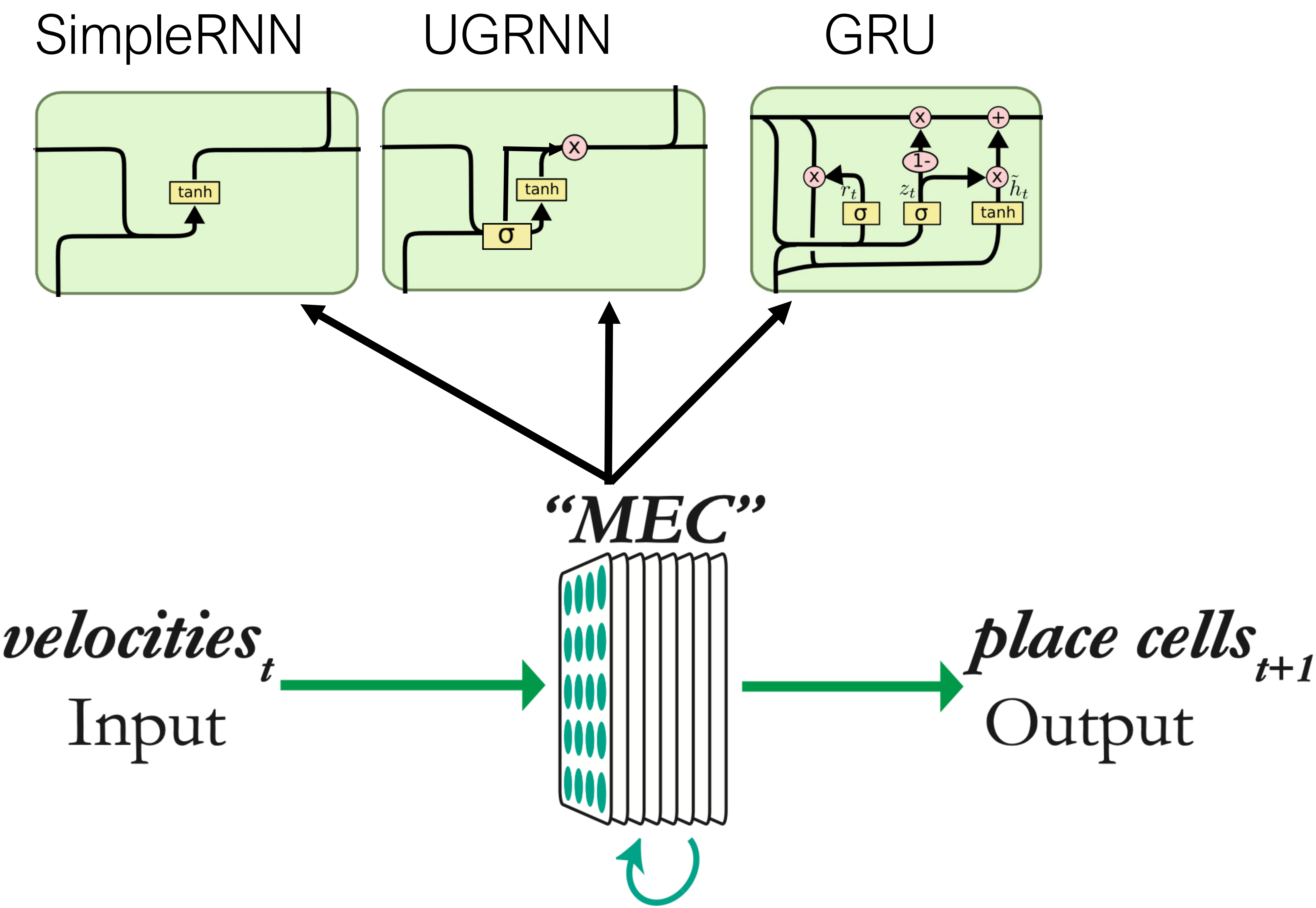
# A spectrum of circuits



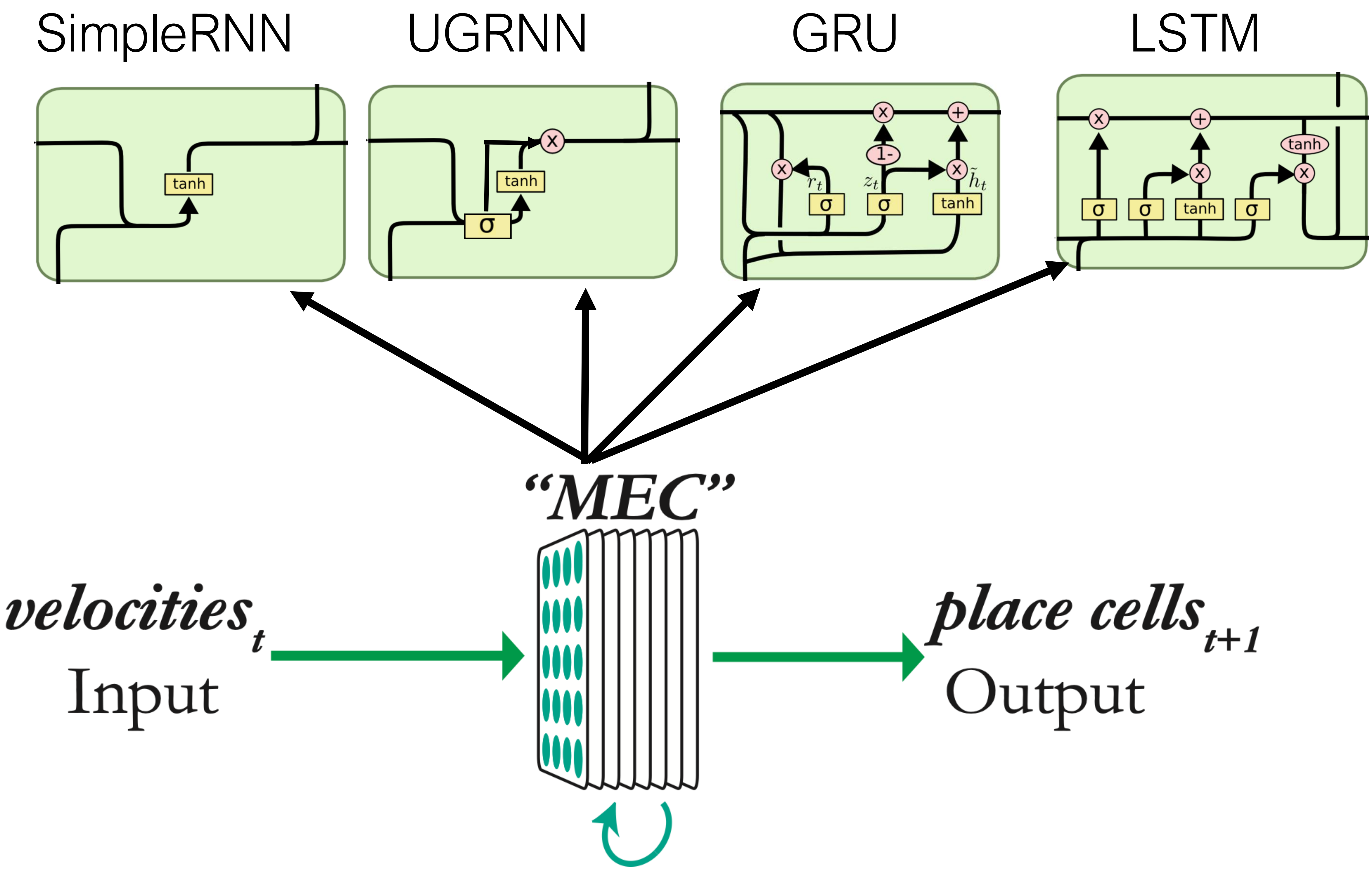
# A spectrum of circuits — learnable modulation ("gating")



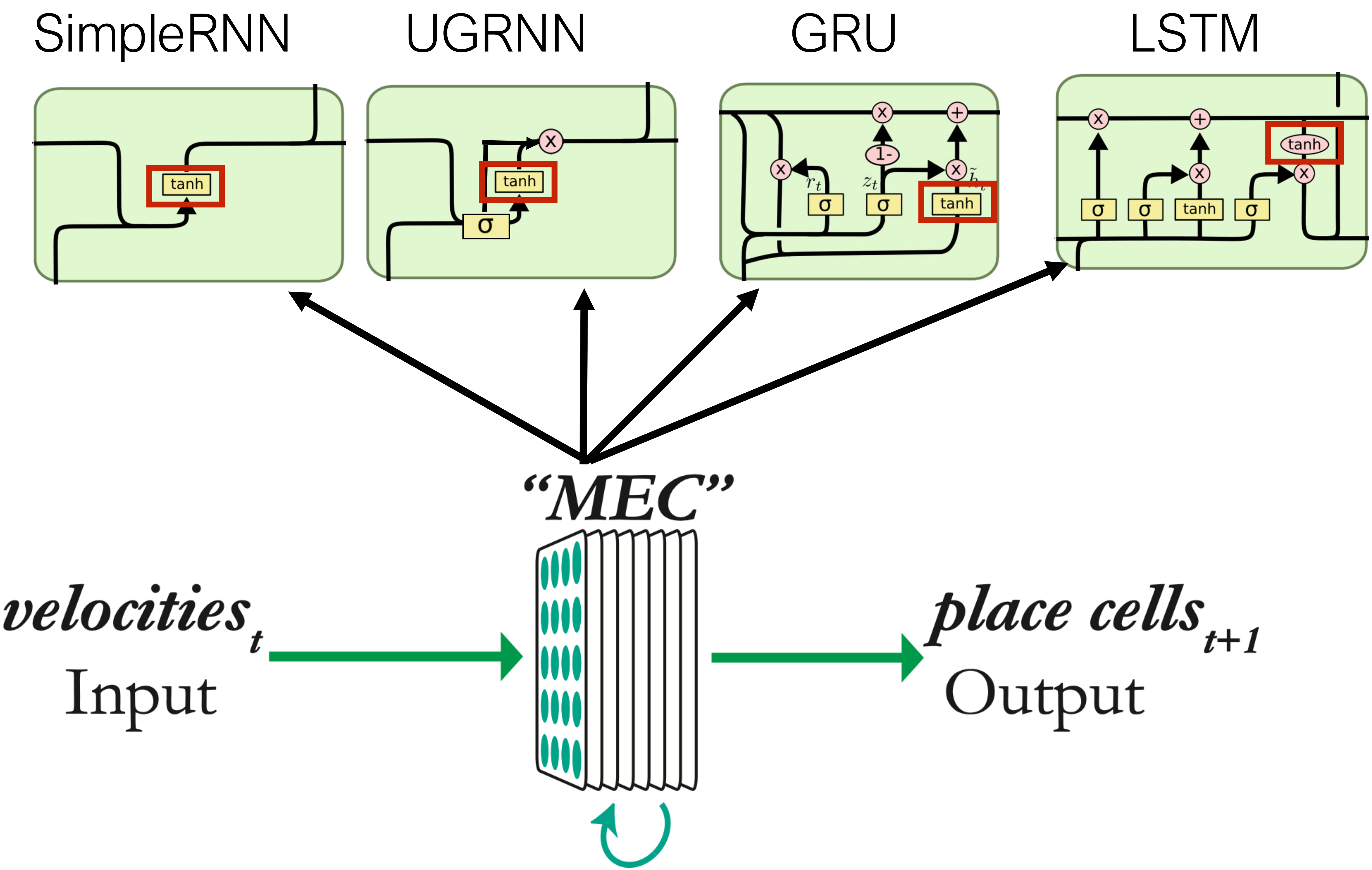
# A spectrum of circuits — learnable modulation (“gating”)



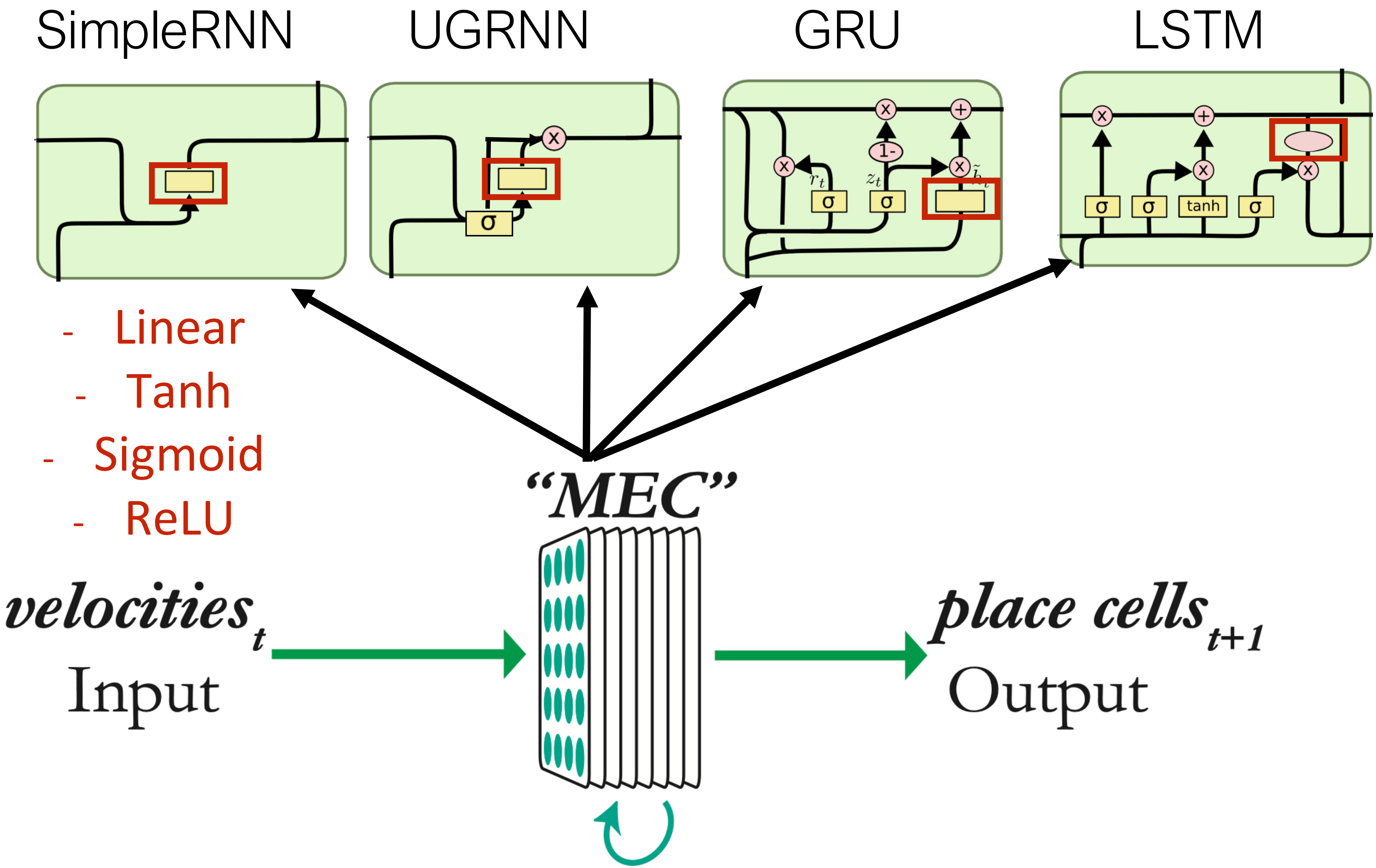
# A spectrum of circuits — learnable modulation (“gating”)



# A spectrum of circuits — output nonlinearity



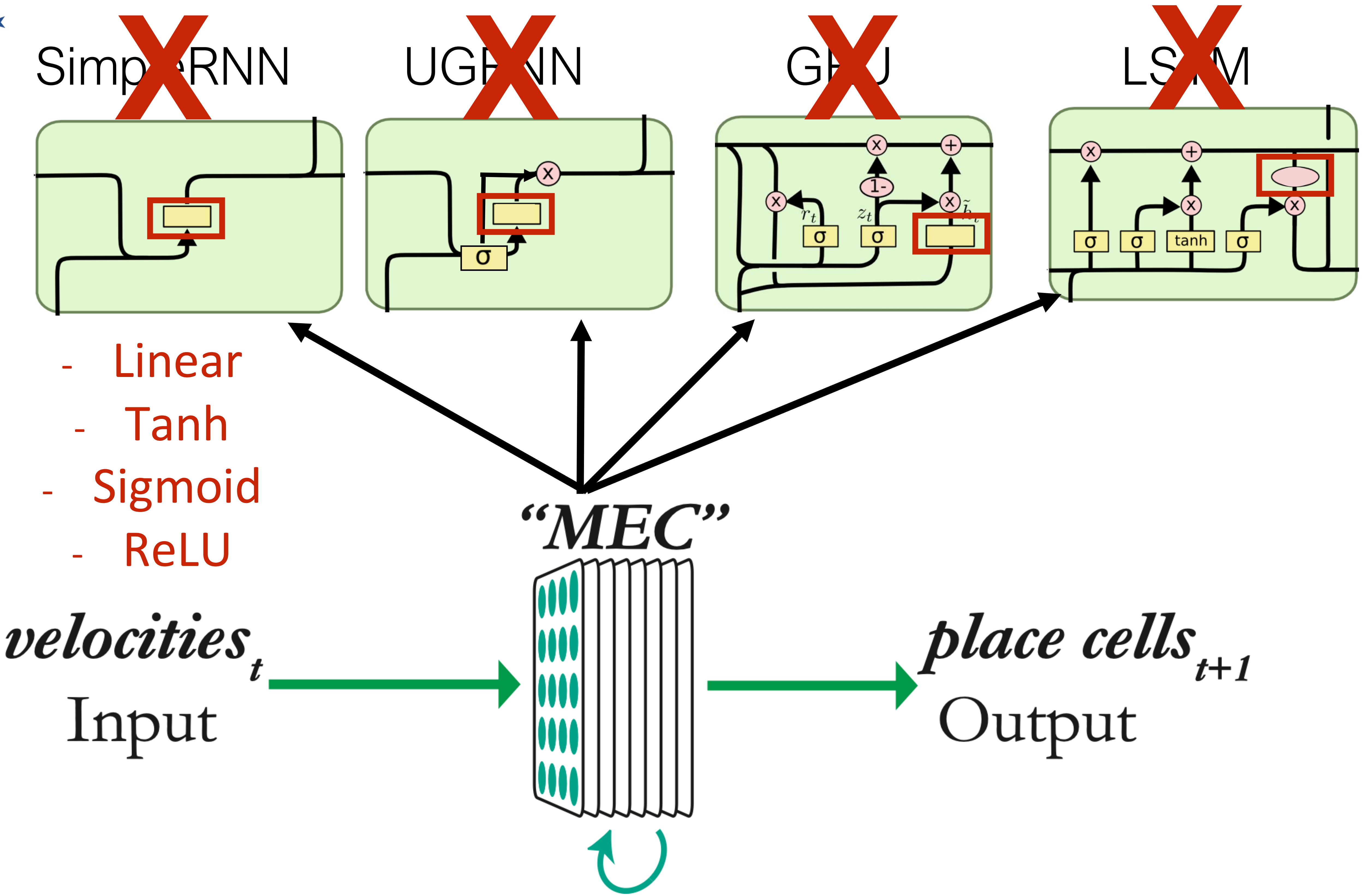
# A spectrum of circuits — output nonlinearity




- Linear
- Tanh
- Sigmoid
- ReLU

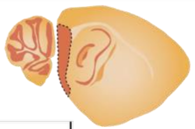
# A spectrum of circuits — output nonlinearity

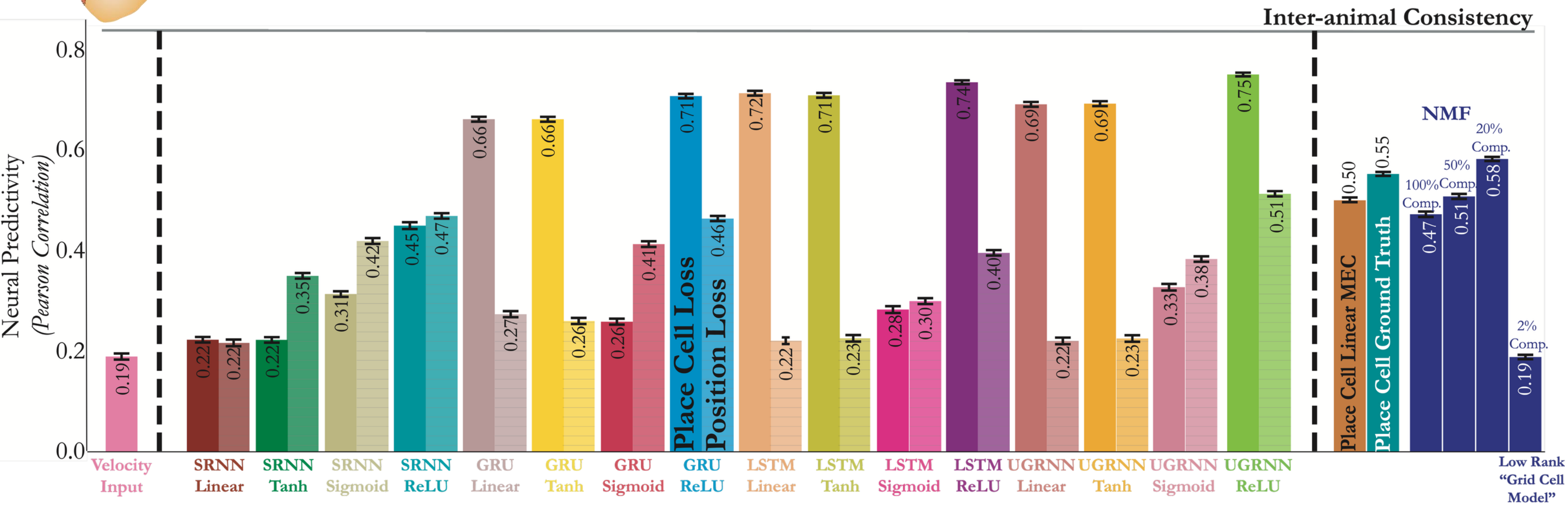
**Circuit busting!**




- Linear
- Tanh
- Sigmoid
- ReLU

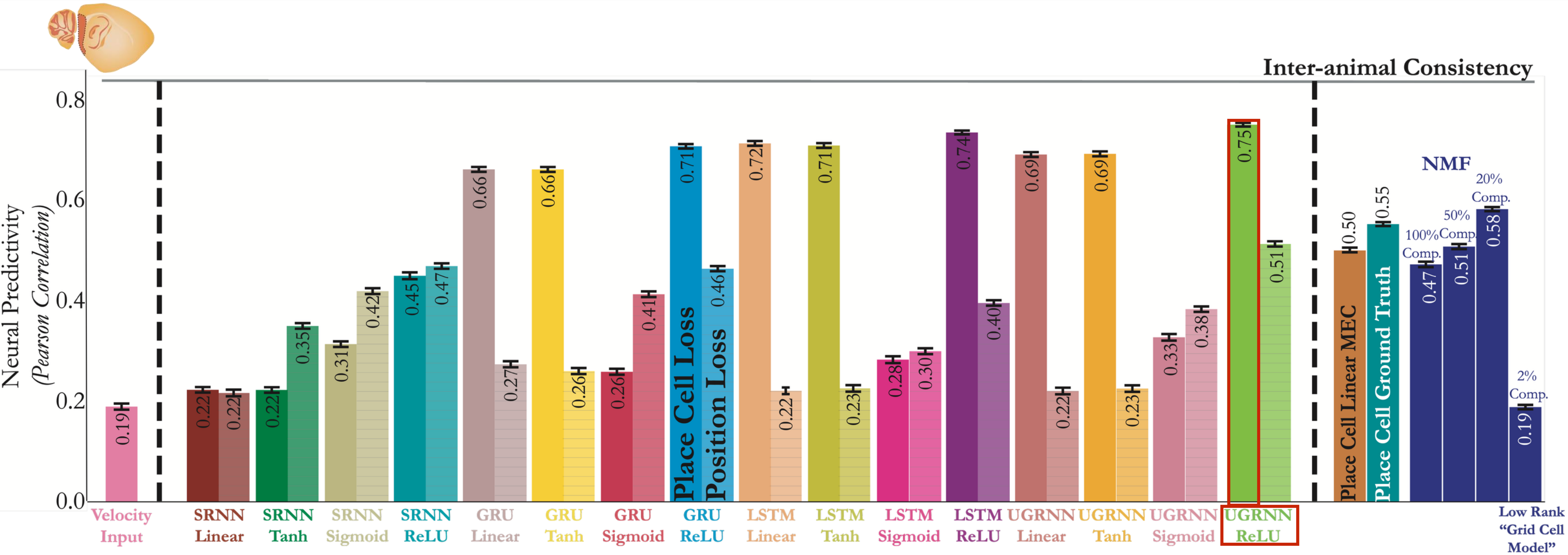
# Benchmarking models with the same transform as between animals



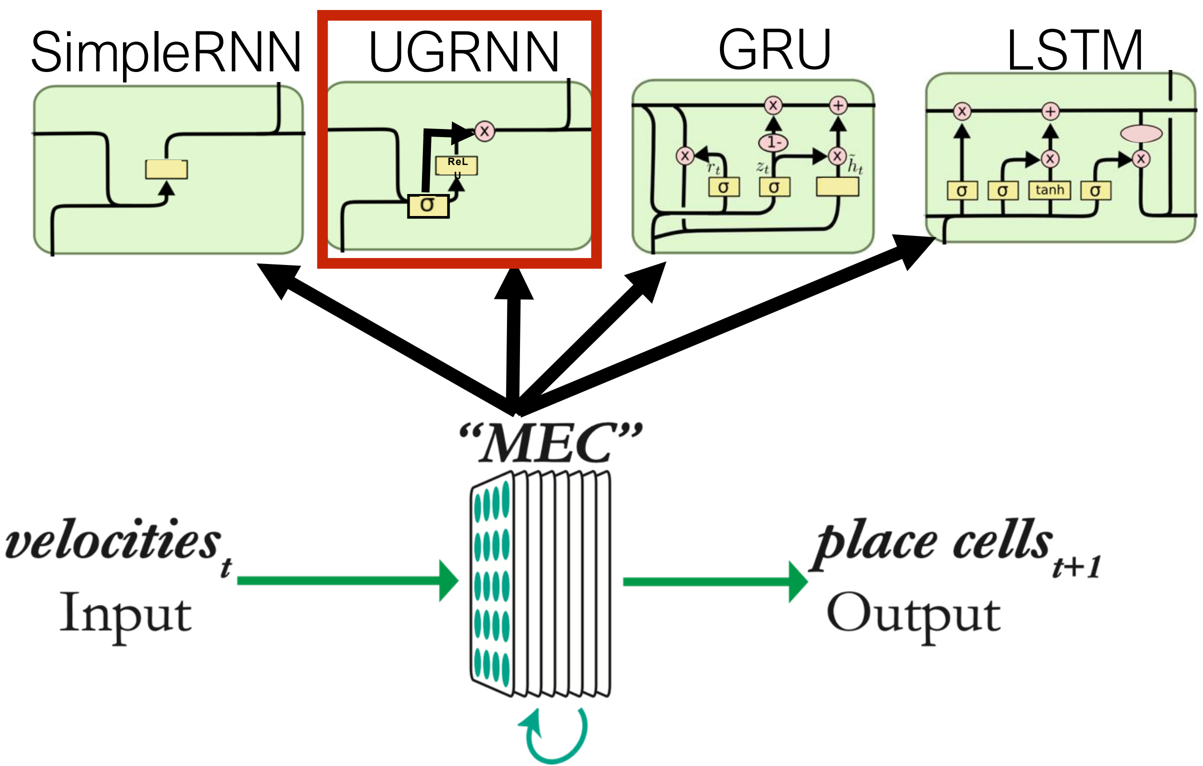



Caitlin Mallory

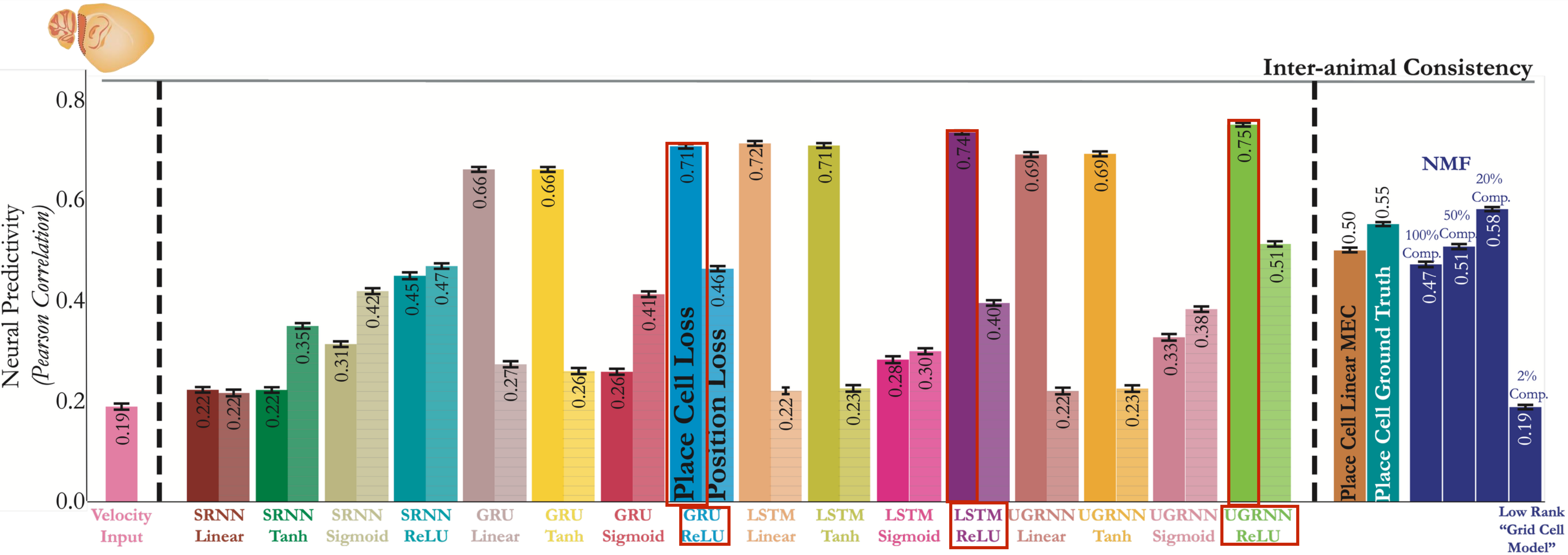
# Task-optimized navigational models best predict the *entire* MEC population




Best task-optimized models explain almost all of the neural variability

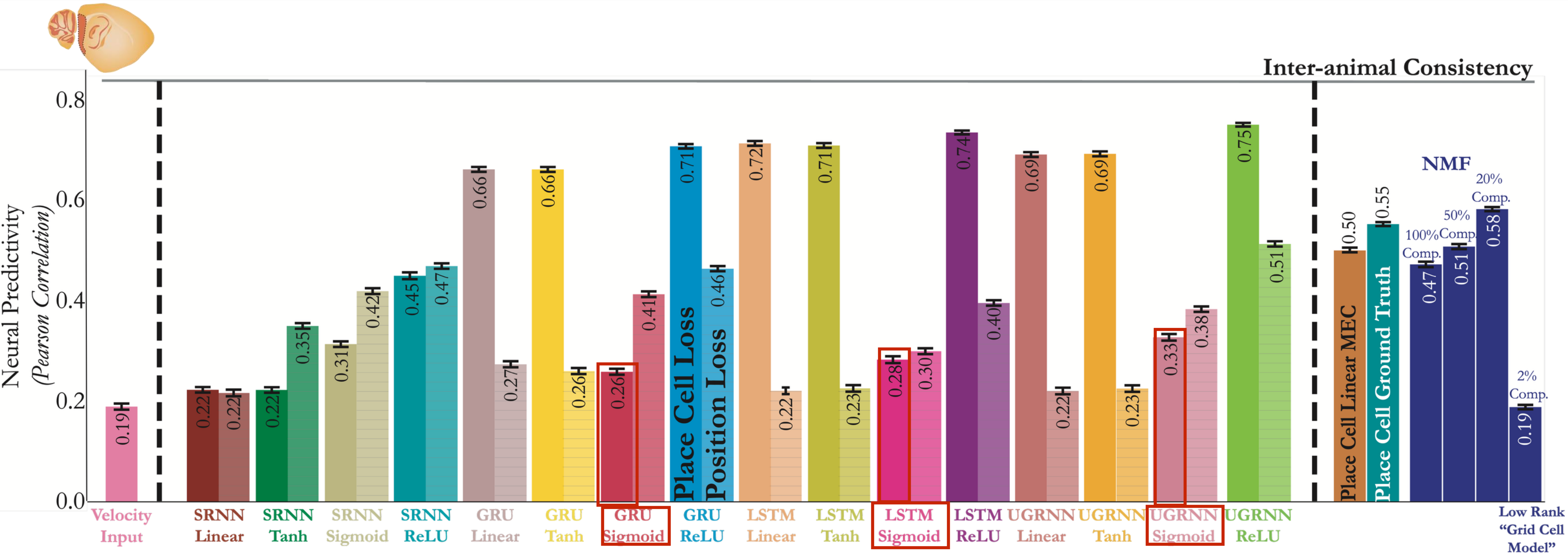


# Nonlinearity type affects generalization



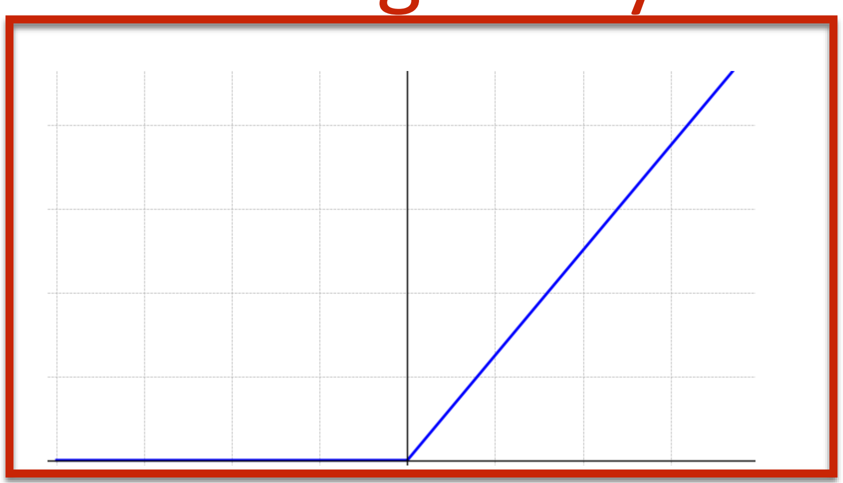

Nonnegativity constraint + gating aids in generalization across environments

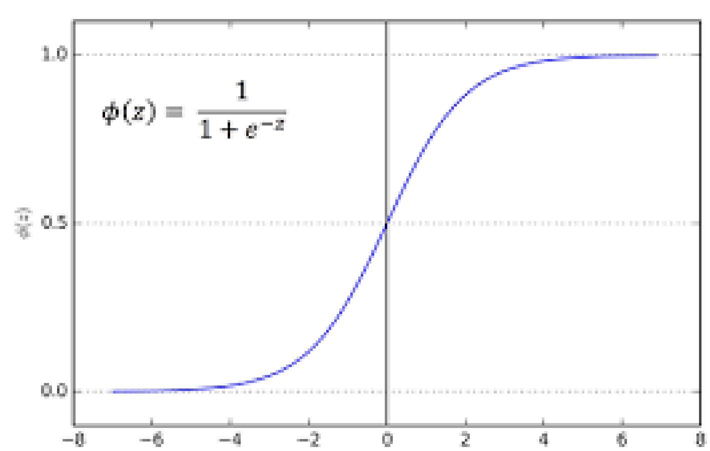
# Nonlinearity type affects generalization



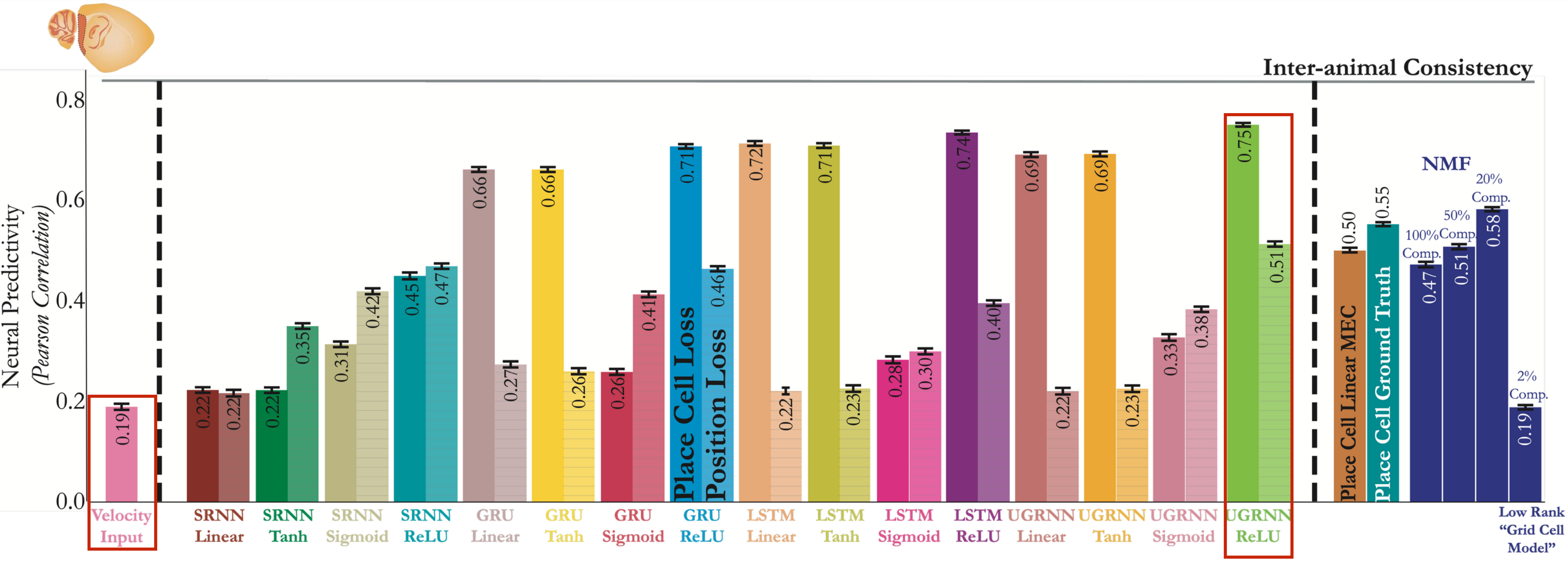

Nonnegativity constraint + gating aids in generalization across environments

But this nonnegativity constraint must *not* saturate either!



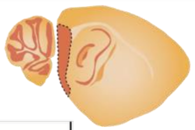


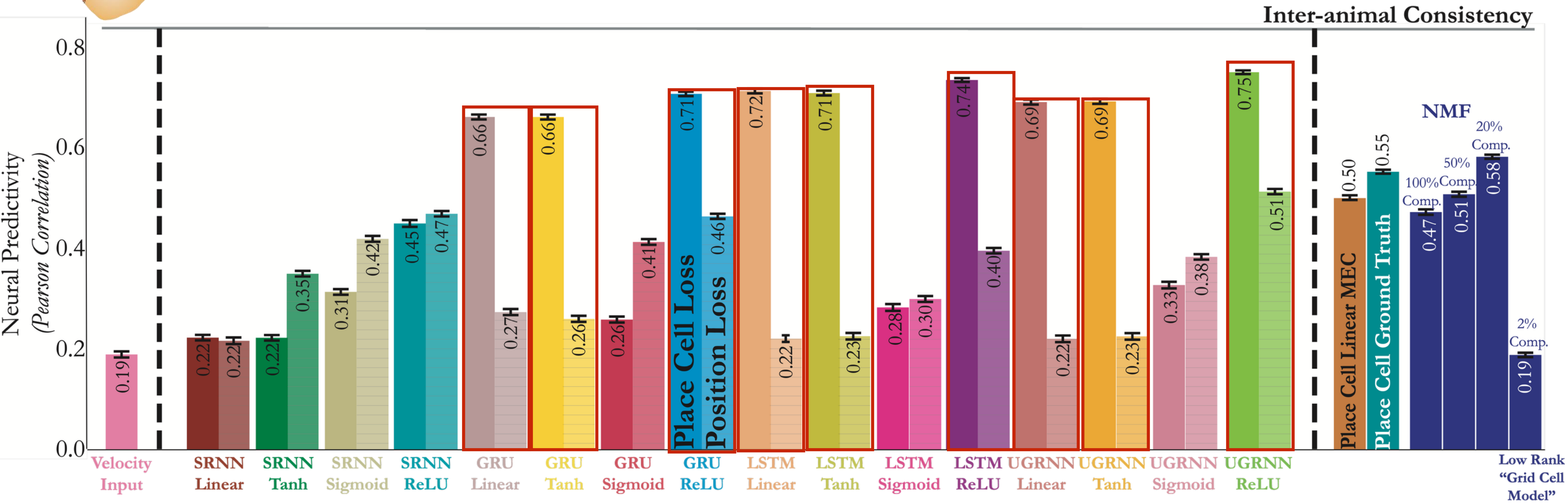
# Model input is a poor predictor of population




Models add a lot of predictive power to their inputs

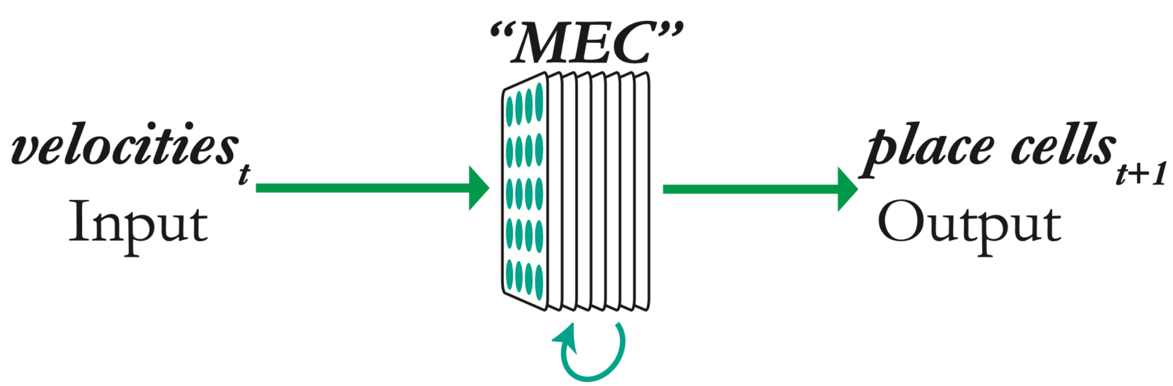
# Direct path integration *fails* to generalize



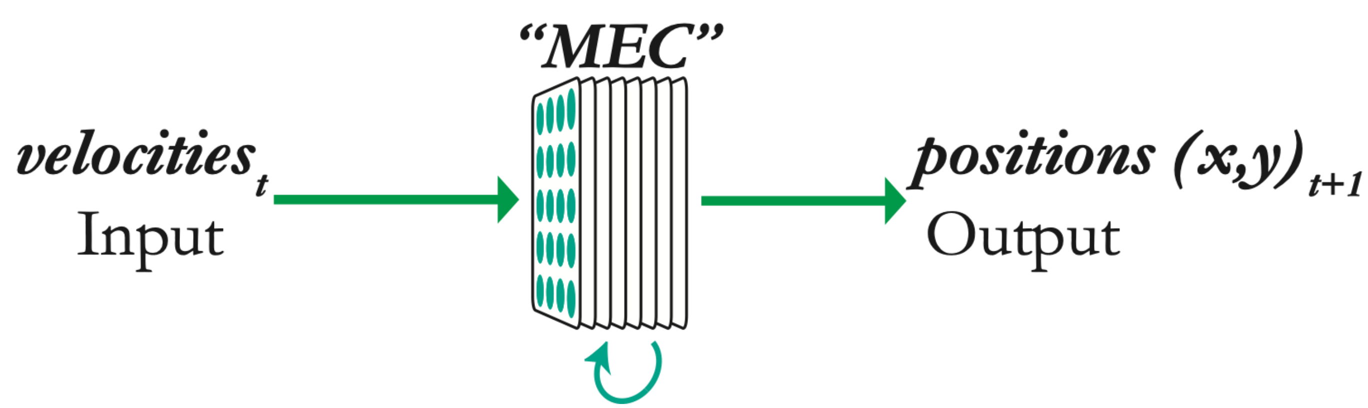



Output place cell supervision provides better generalization over direct supervision of position (path integration)

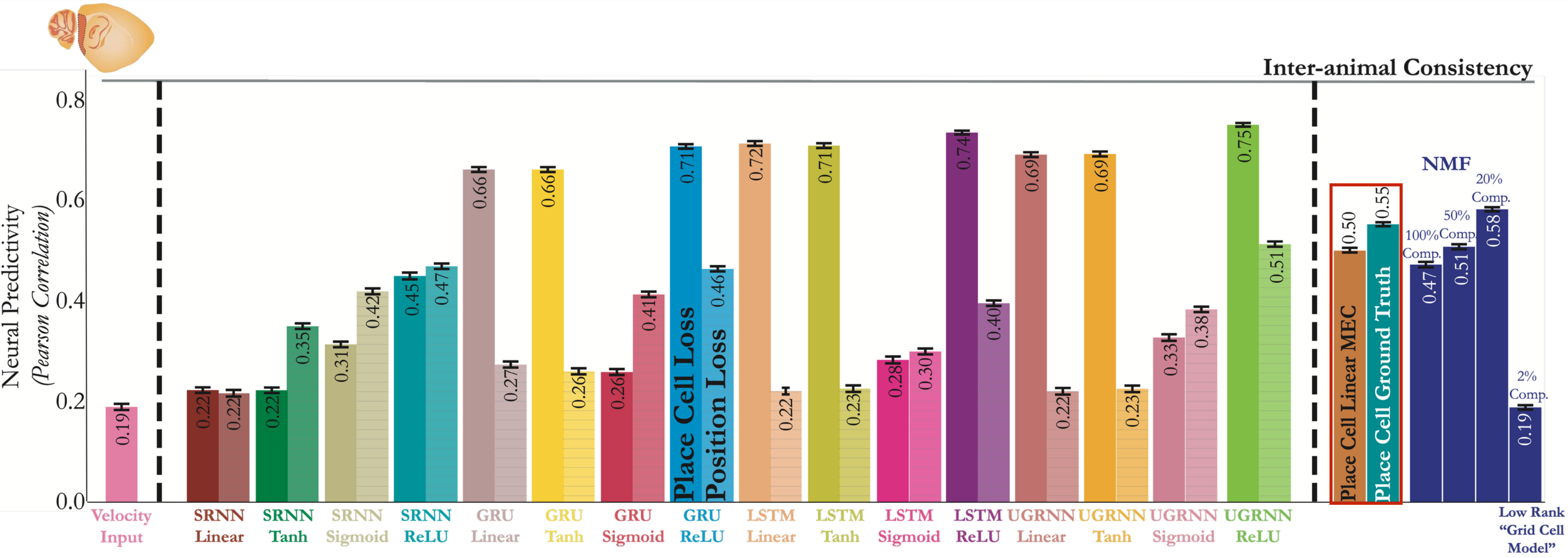
$$\mathcal{L}(\hat{p}, p) := -\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N_p} p_i^t \log \hat{p}_i^t$$




$$\mathcal{L}(\hat{p}, p) := \frac{1}{2}\frac{1}{T}\sum_{t=1}^{T}\left(\left(p_x^t - \hat{p}_x^t\right)^2 + \left(p_y^t - \hat{p}_y^t\right)^2\right)$$



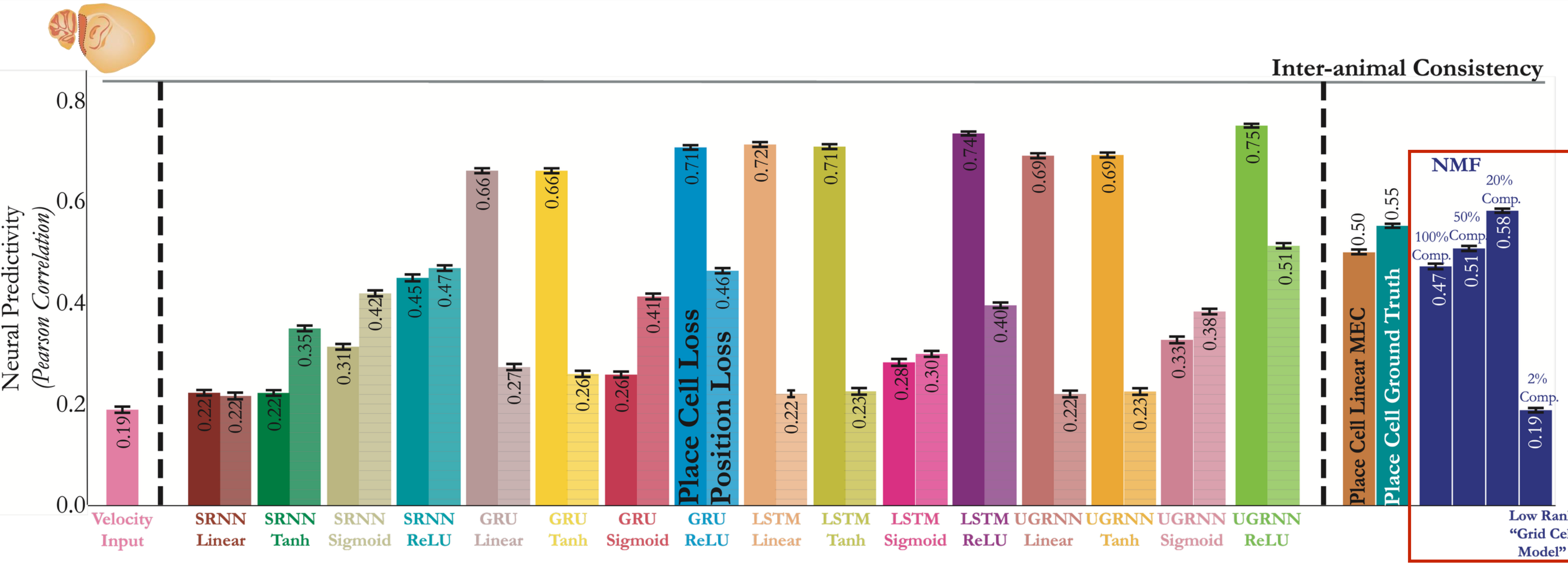
# Place cells alone are a poor predictor




But place cells alone are *not* a good predictor of MEC (good!)

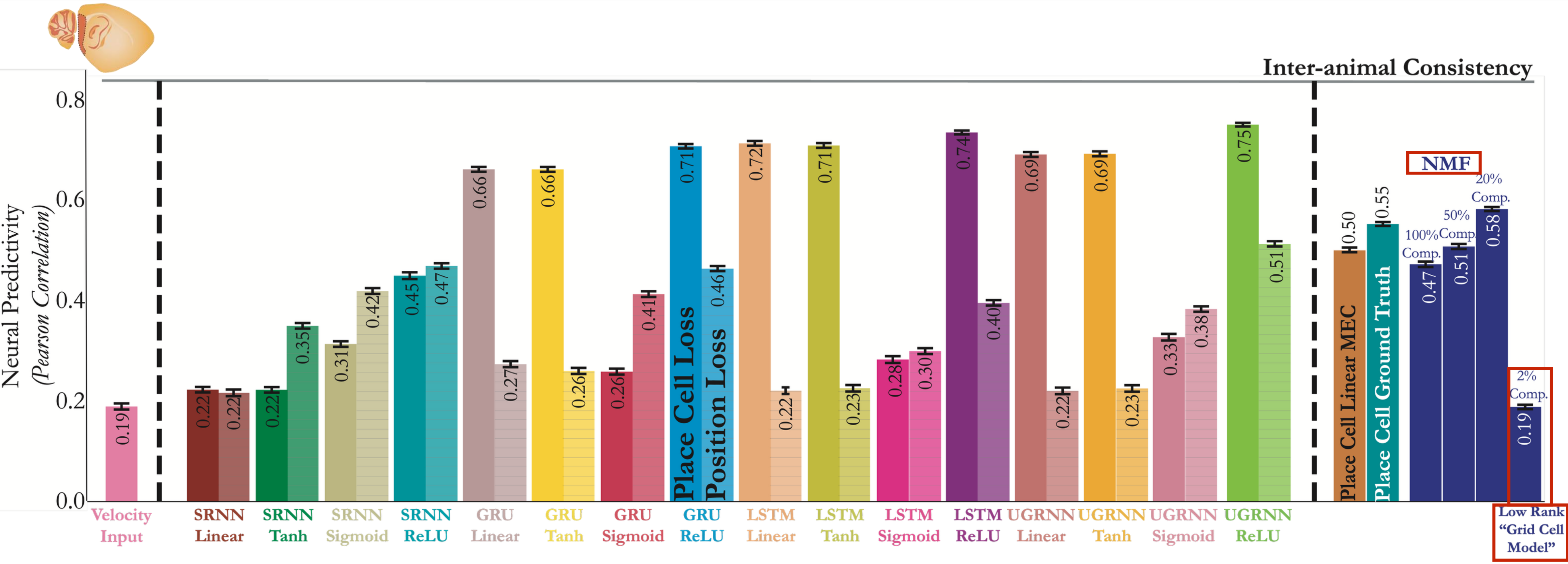
You actually need to integrate them across time!

# NMF is also a poor predictor



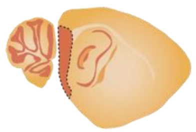

Dimensionality reduction on place cells is *not* a good predictor of MEC either

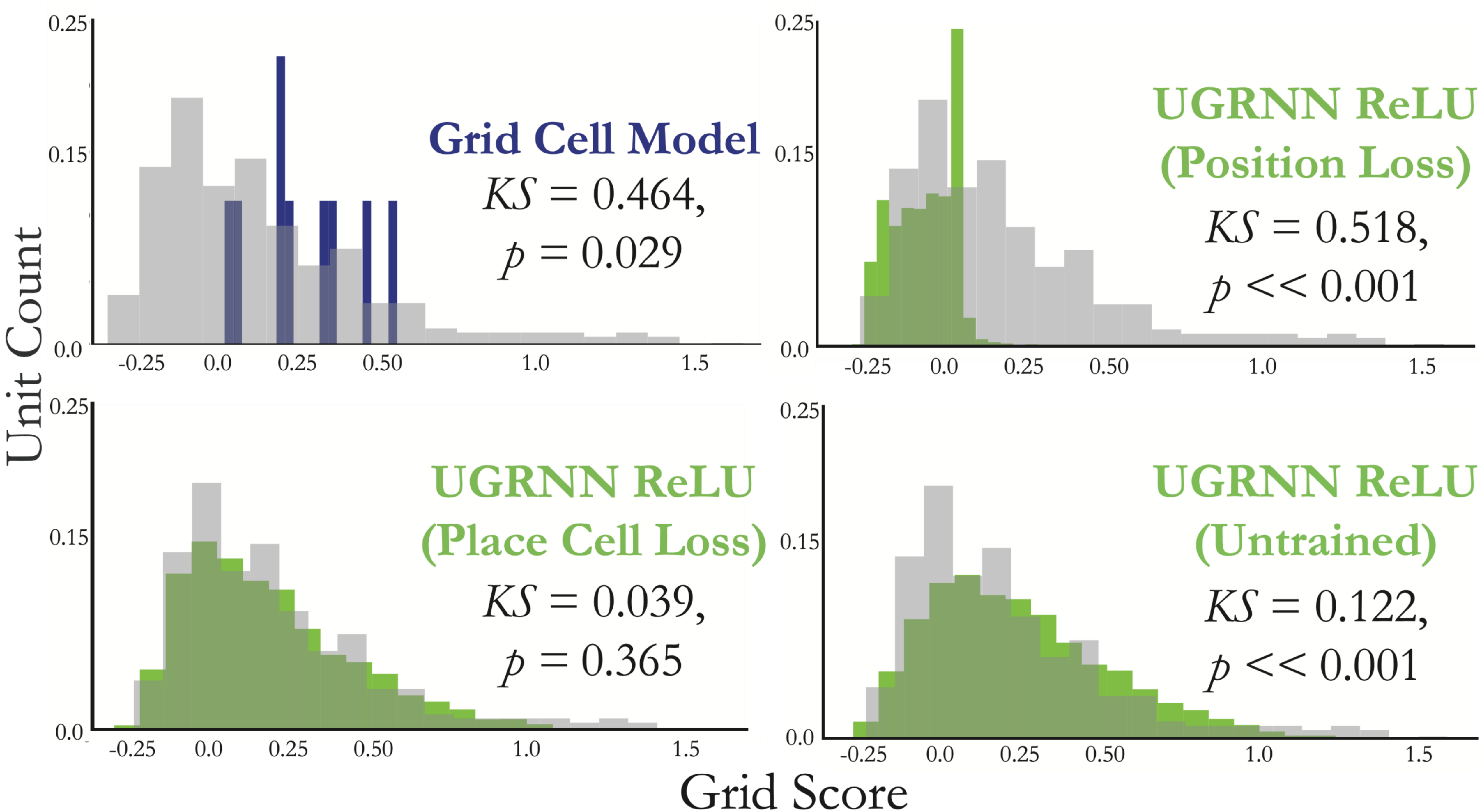
# Grid cell oriented NMF is a poor predictor




Grid cell oriented model is an especially *poor* predictor!

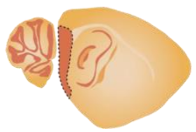
Grid score distribution does not require any parameter fitting

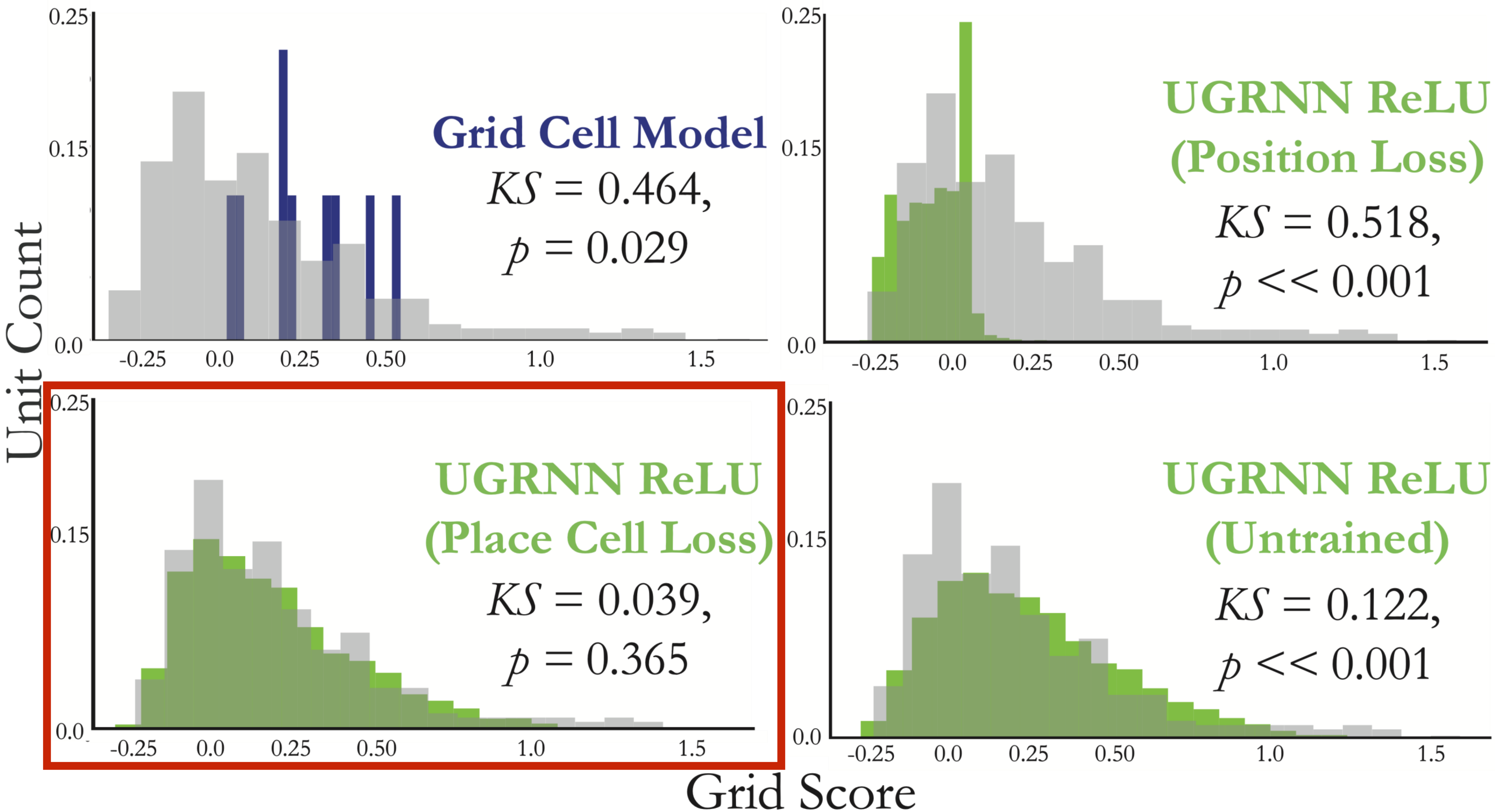




# Task-optimized navigational models best predict the ***entire*** MEC population

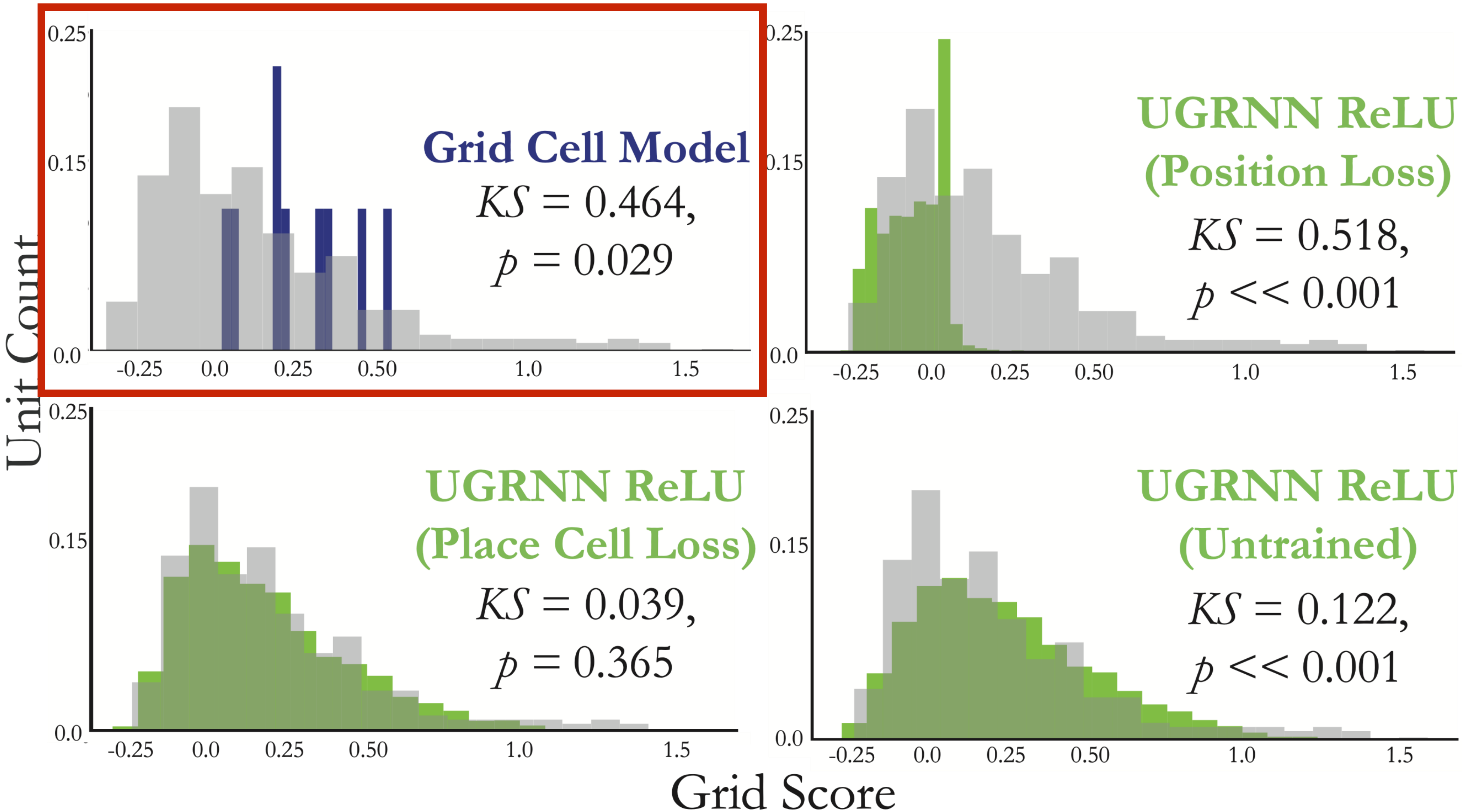
Best model class in terms of neural predictivity also matches grid score distribution in its own synthetic population





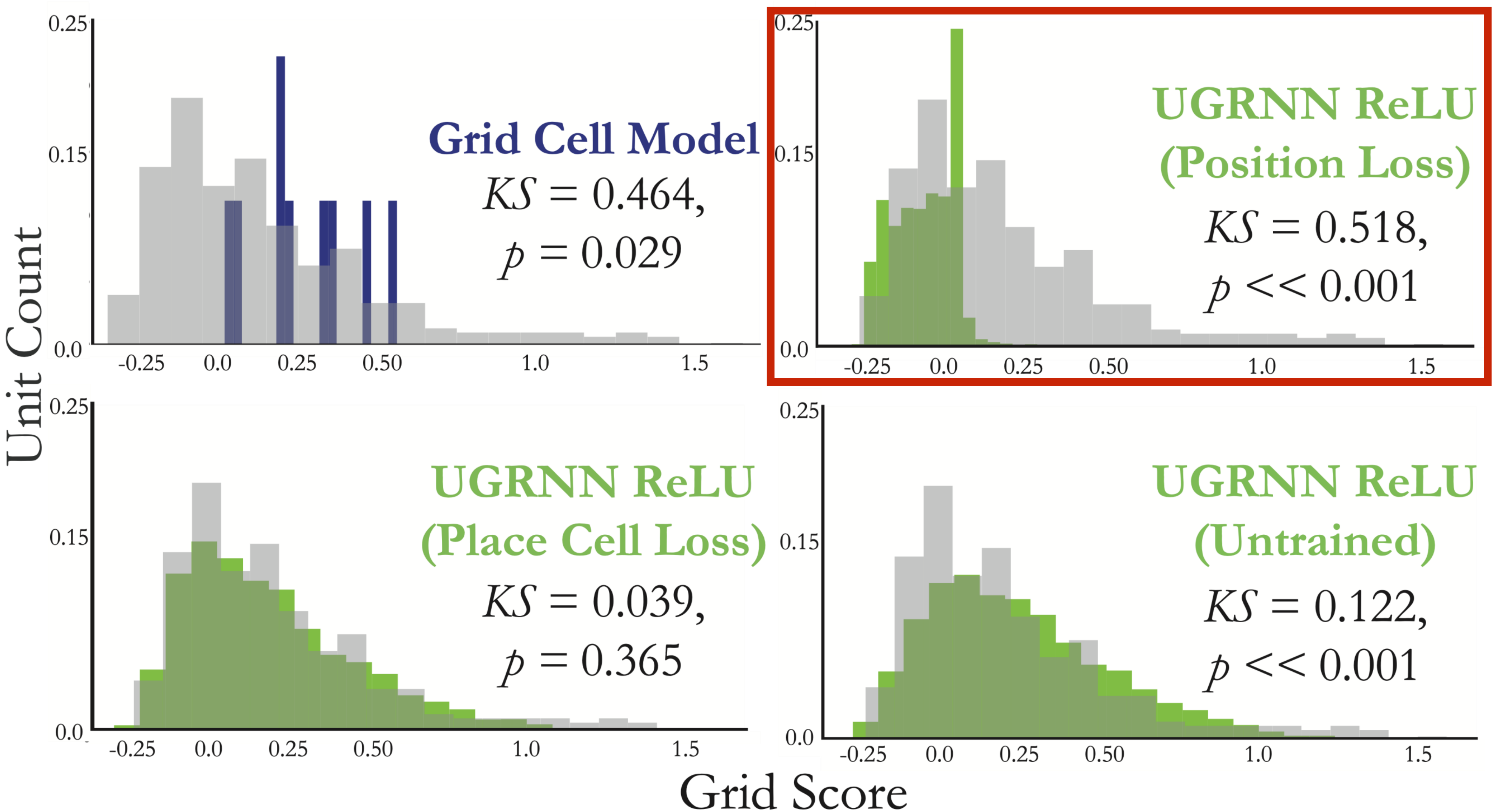
# Task-optimized navigational models best predict the ***entire*** MEC population

## Low-rank model is too biased towards grid-like units

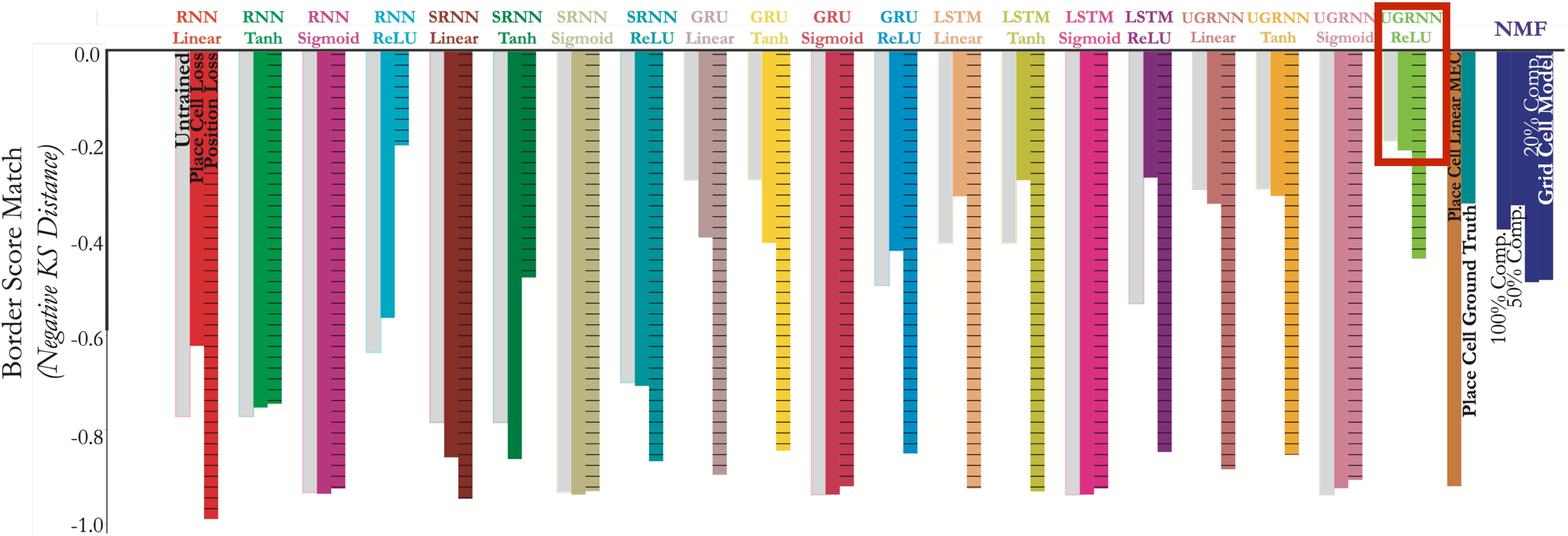


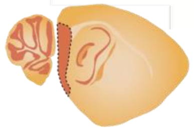
# Task-optimized navigational models best predict the ***entire*** MEC population

## Without place cell integration, the model is too biased towards *non* grid-like units



# More fine-grained unit matching metrics



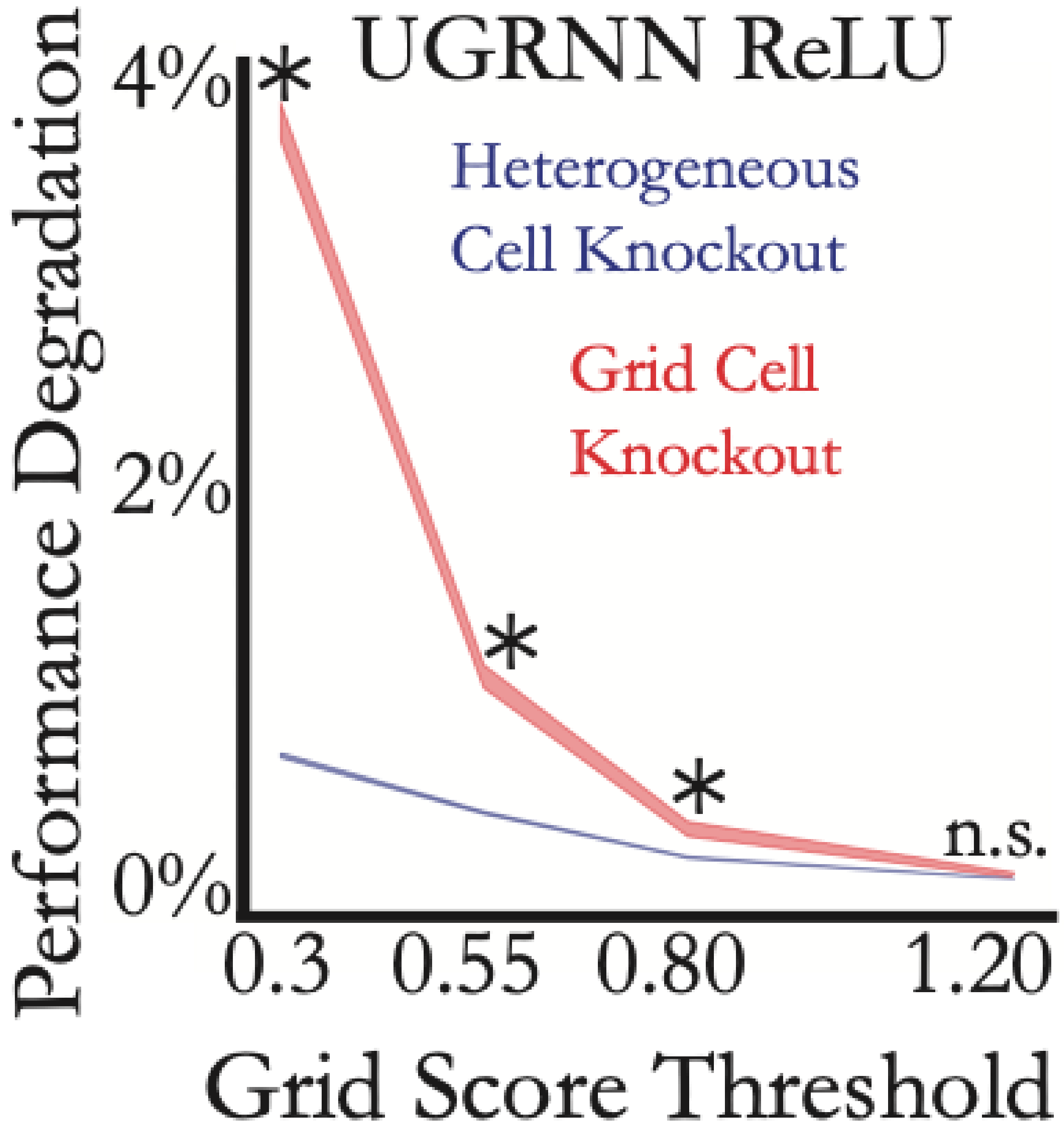
Best model also matches the data's ***border*** score distribution in its own synthetic population

# Knockout experiments

Given that we have a model that exhibits close similarity to MEC, we can use it to generate predictions for experiments that are very difficult to do

# Knockout experiments

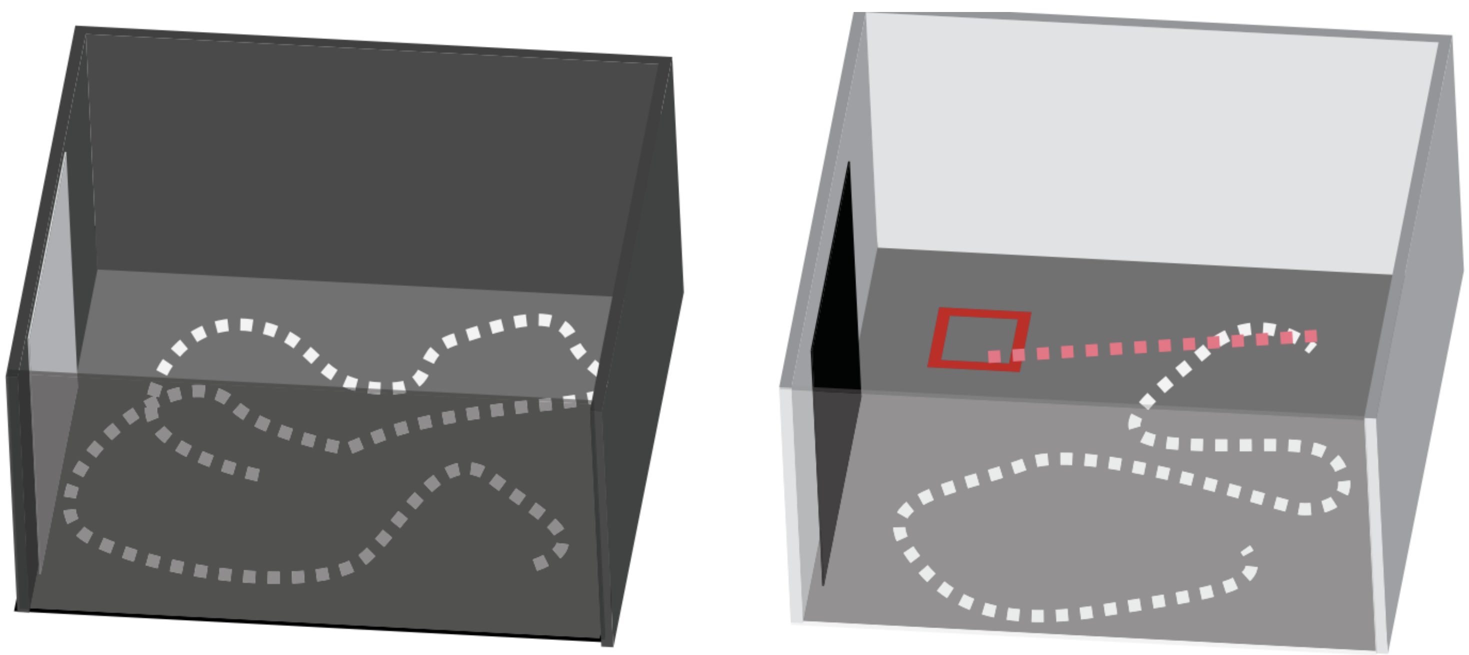
Given that we have a model that exhibits close similarity to MEC, we can use it to generate predictions for experiments that are very difficult to do



# Remapping in the presence of reward

## Remembered reward locations restructure entorhinal spatial maps

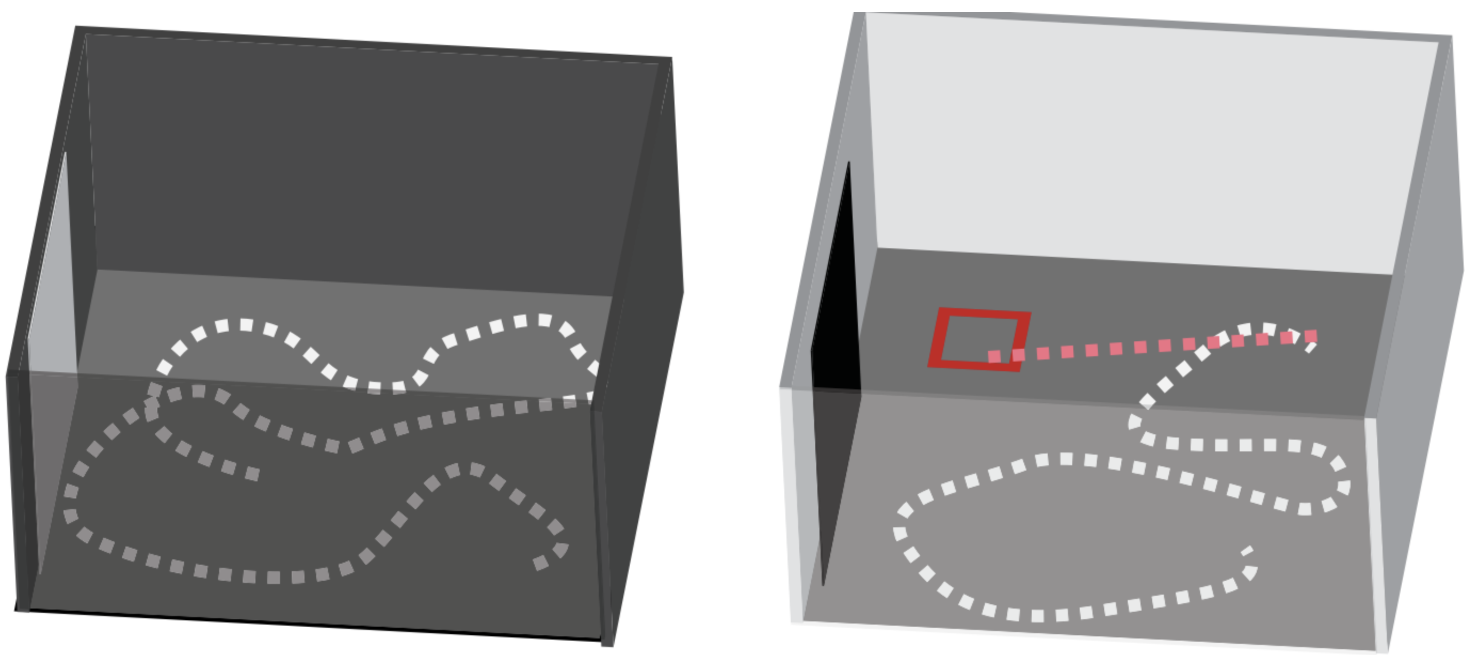
William N. Butler*, Kiah Hardcastle*, Lisa M. Giocomo†



free foraging (ENV1)    spatial task (ENV2)

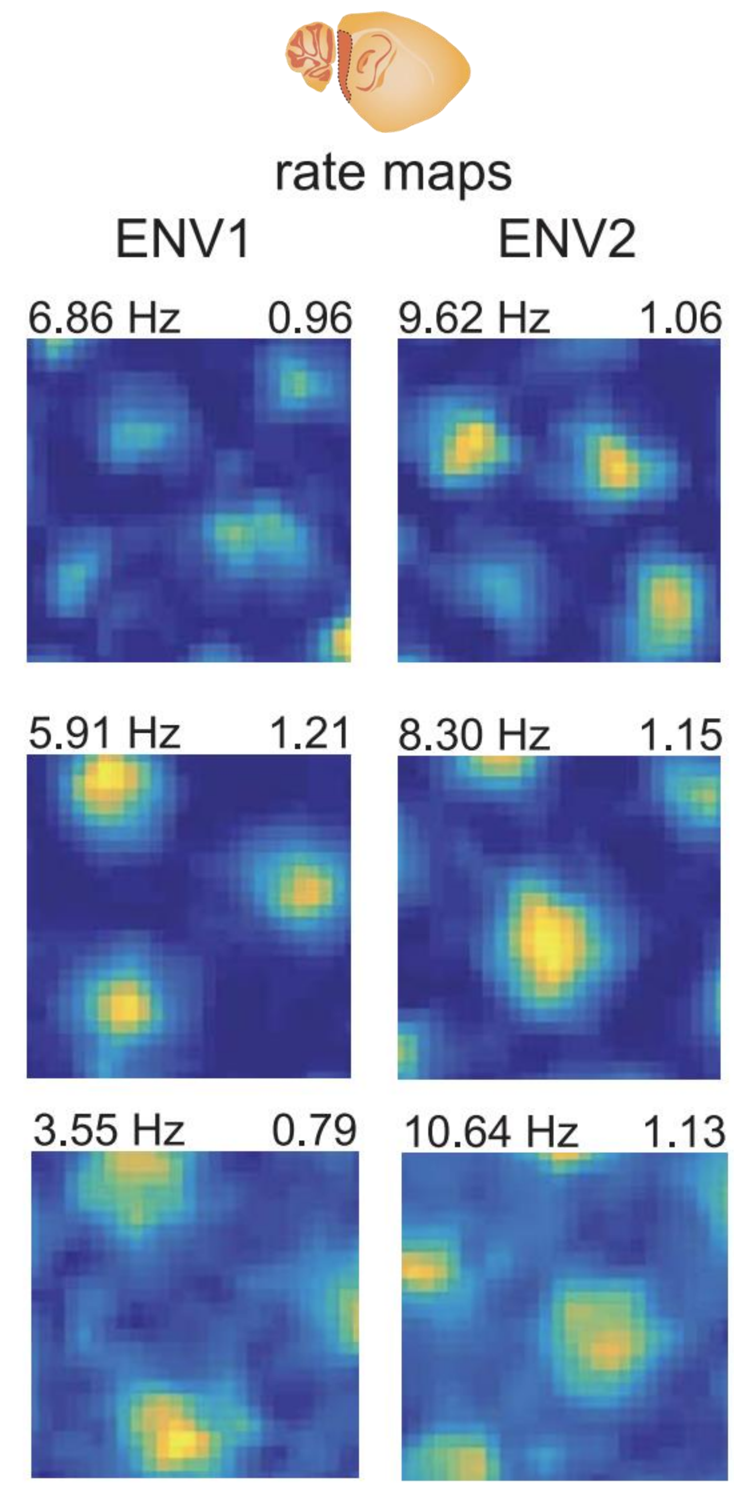
# Remembered reward locations restructure entorhinal spatial maps

William N. Butler*, Kiah Hardcastle*, Lisa M. Giocomo†

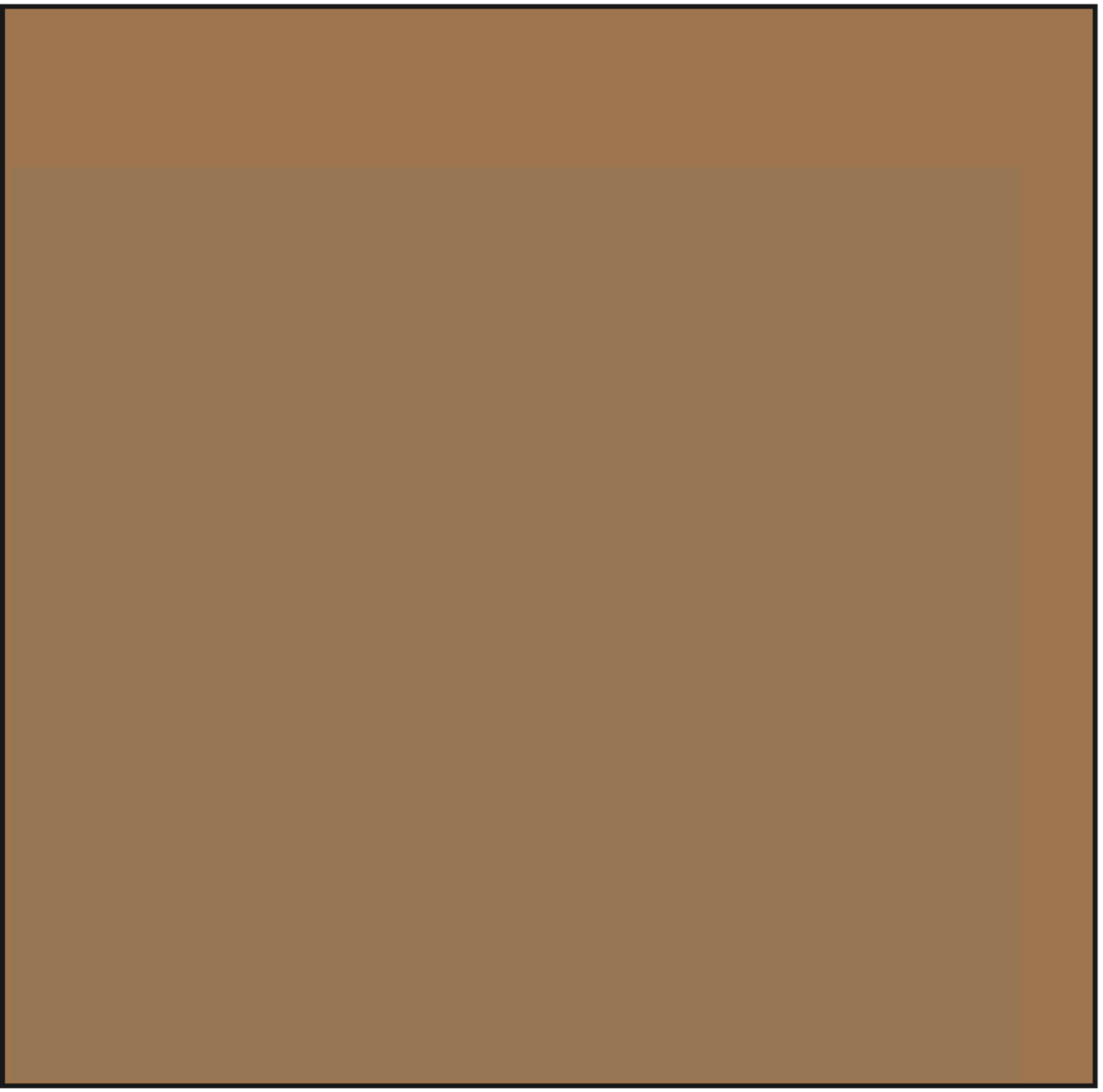


free foraging (ENV1)

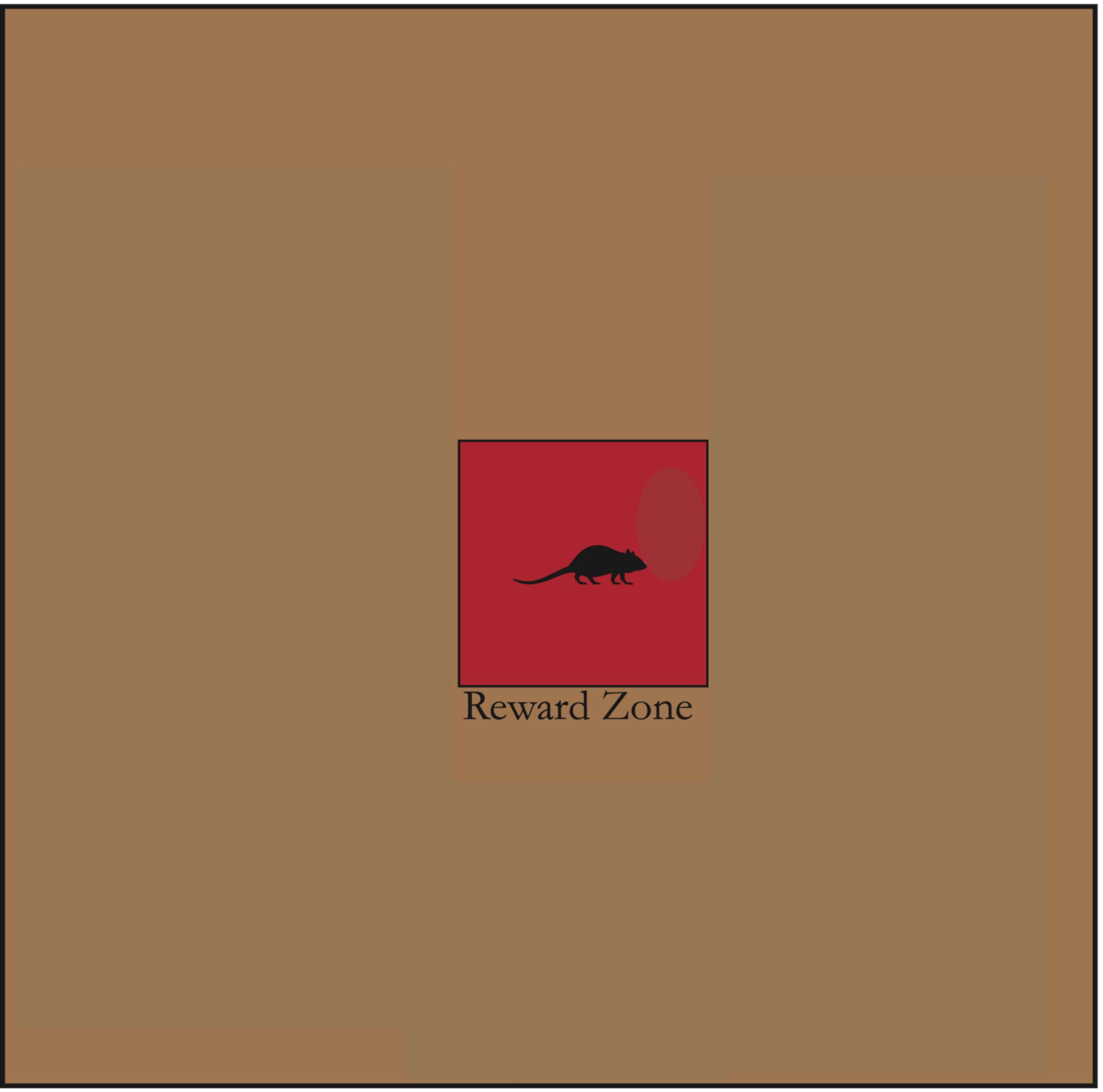
spatial task (ENV2)



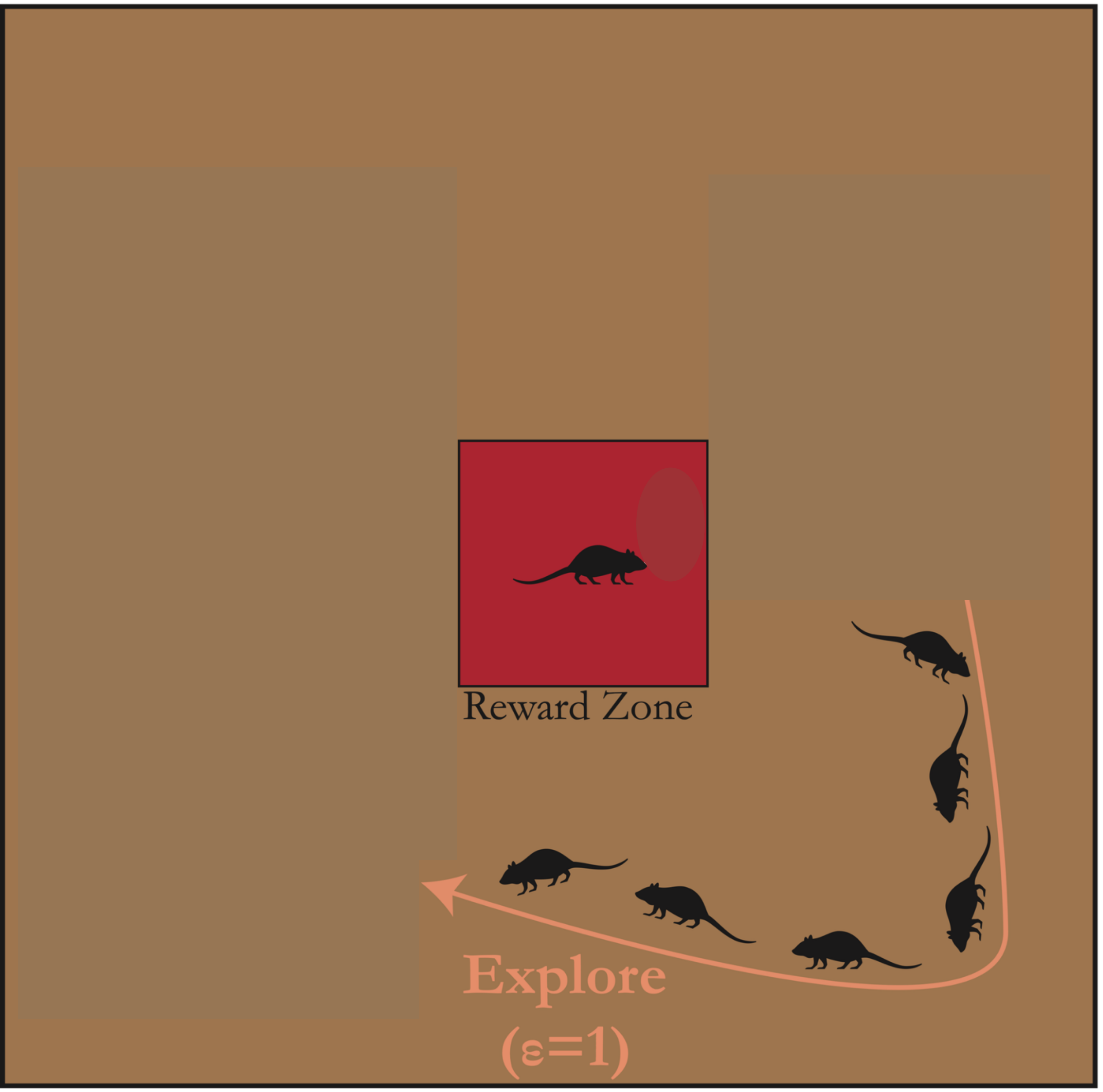
# Modeling rewards



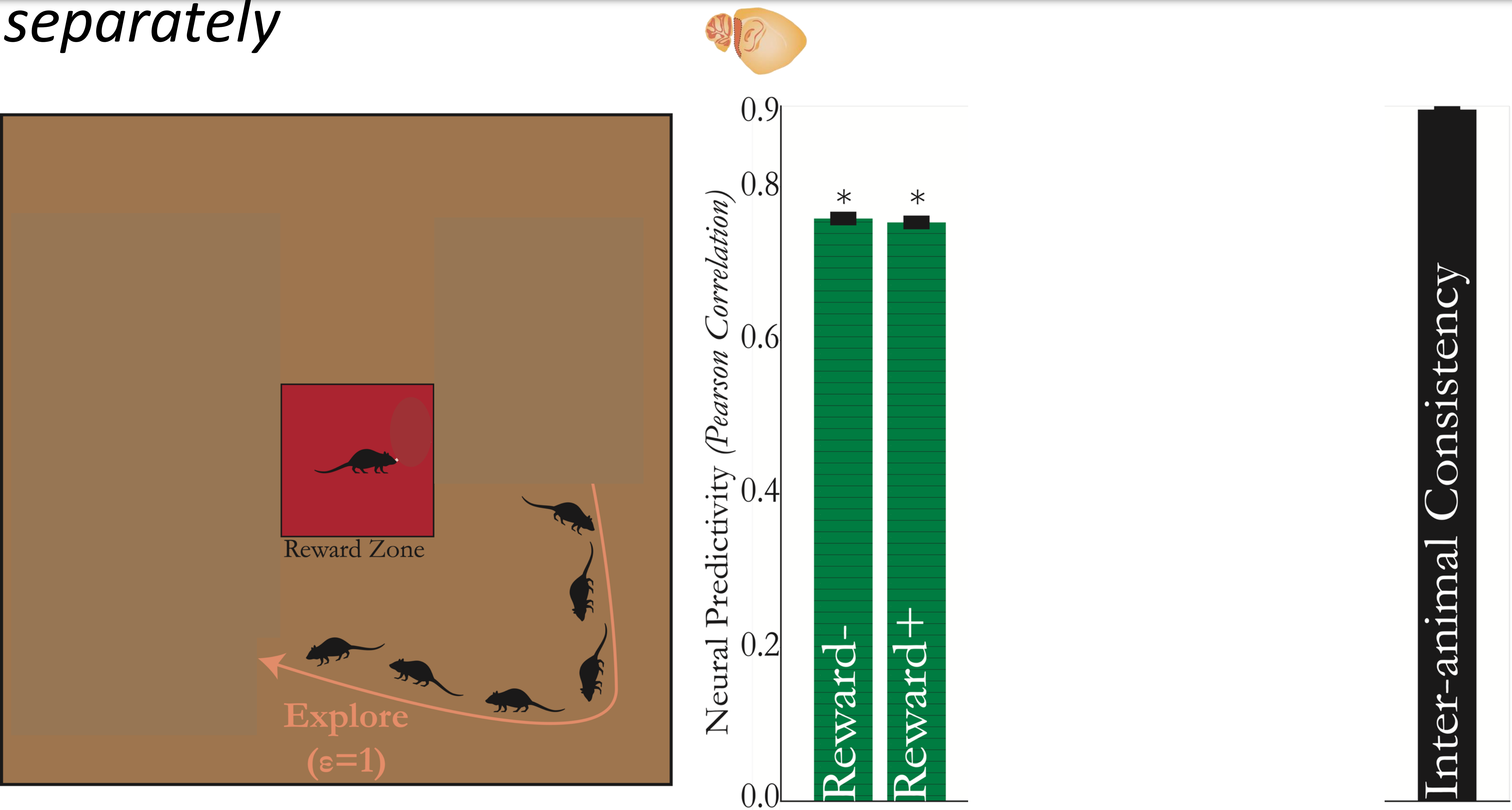
# Modeling rewards



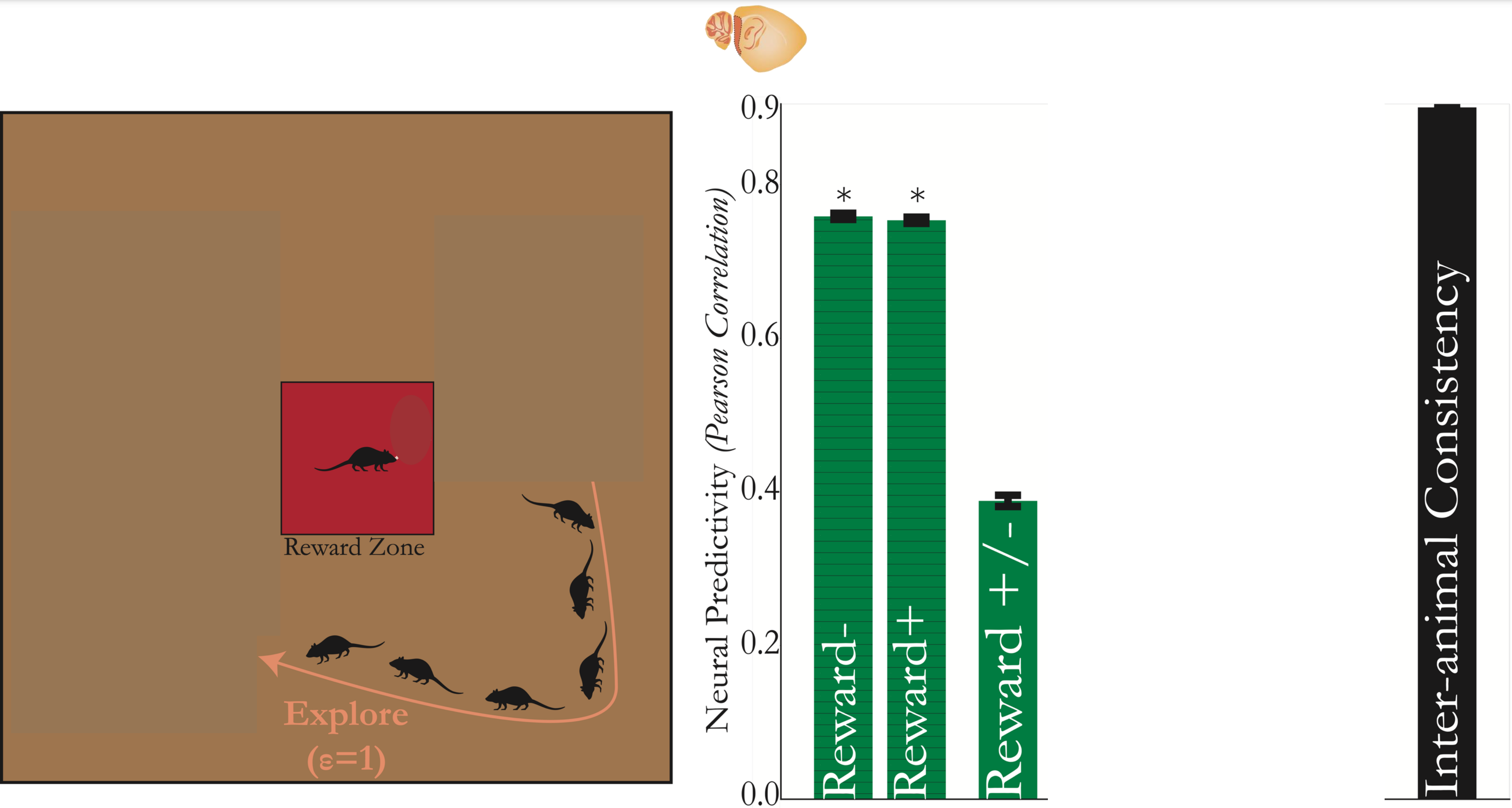
# Modeling rewards - What we have done previously



# Exploration only model captures each condition *separately*

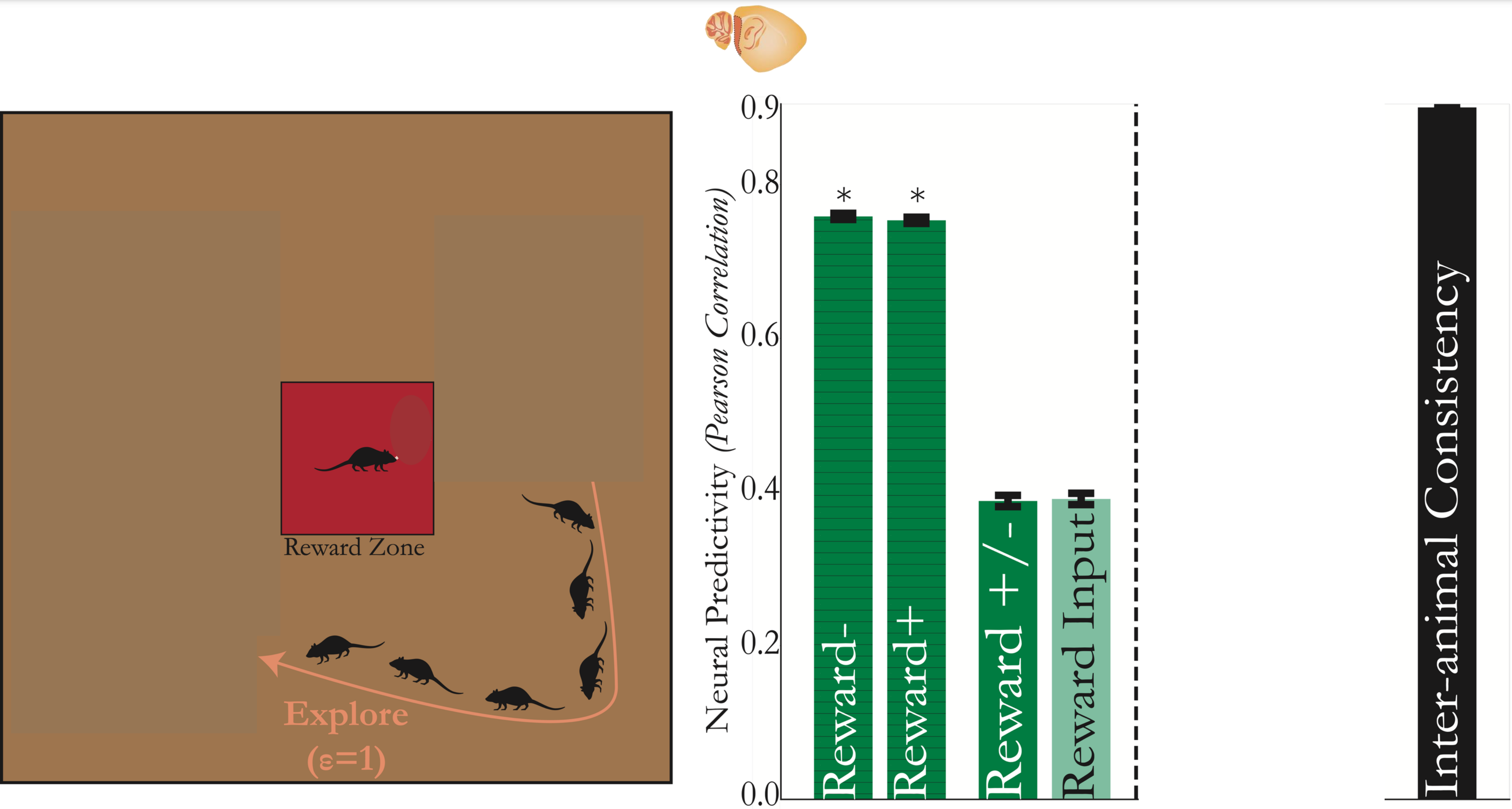


# Exploration only model fails to capture remapping




Failure of pure exploration!

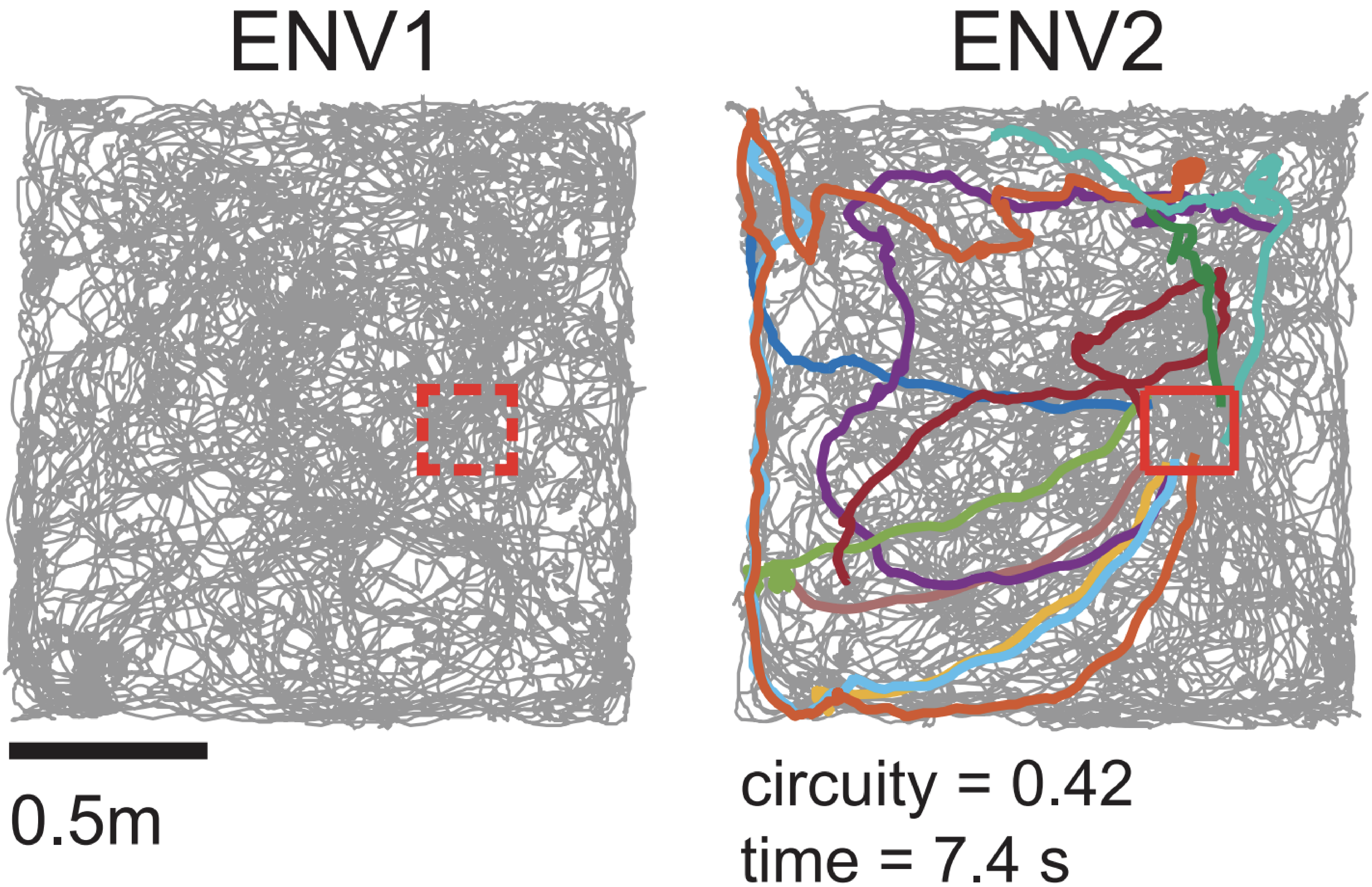
# Reward must be extrinsically modeled



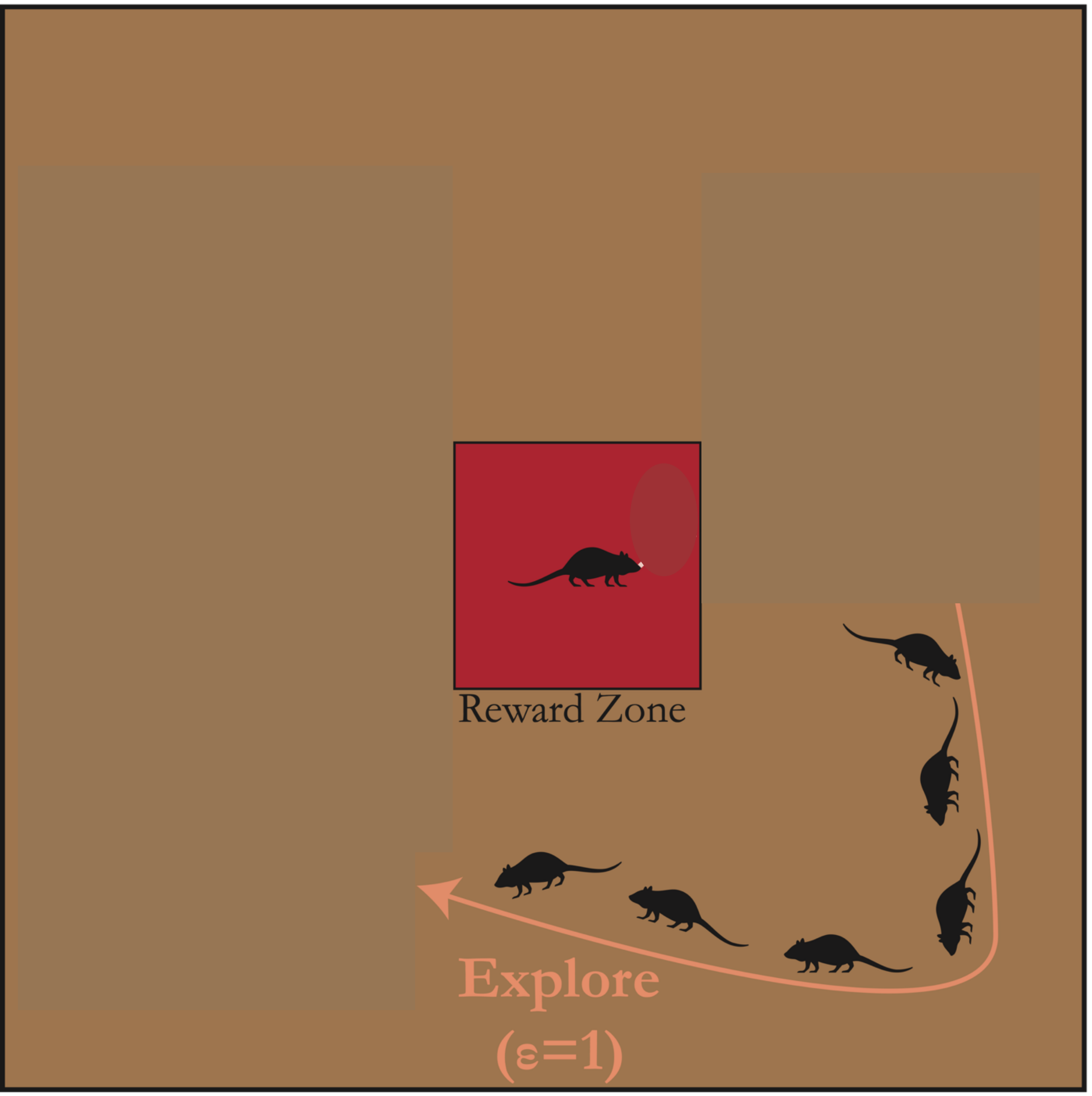

Simply augmenting inputs does not help either

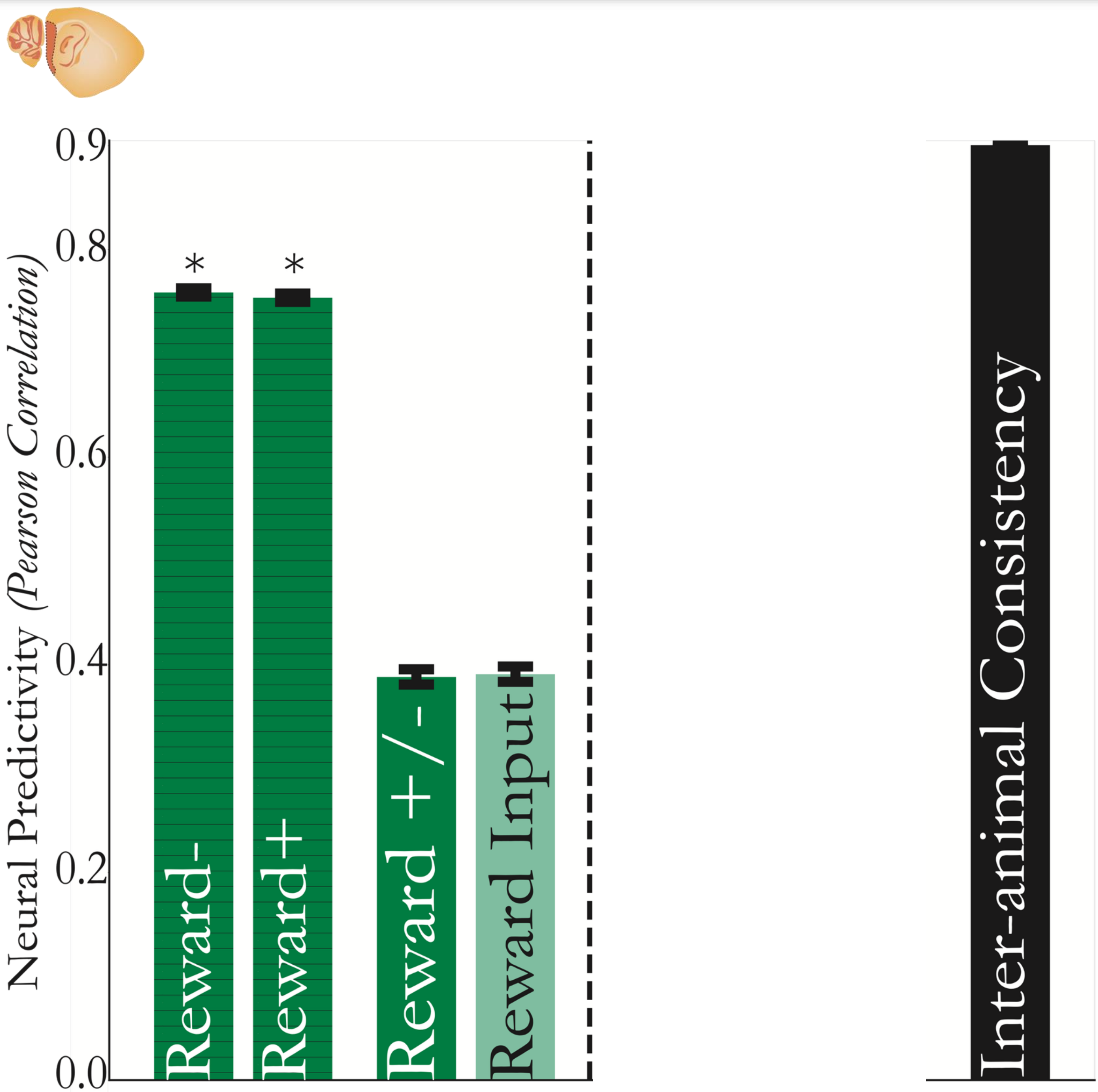
# Inspiration from animal behavior — rapid, direct paths

Animals tend to take rapid, direct paths to reward zone

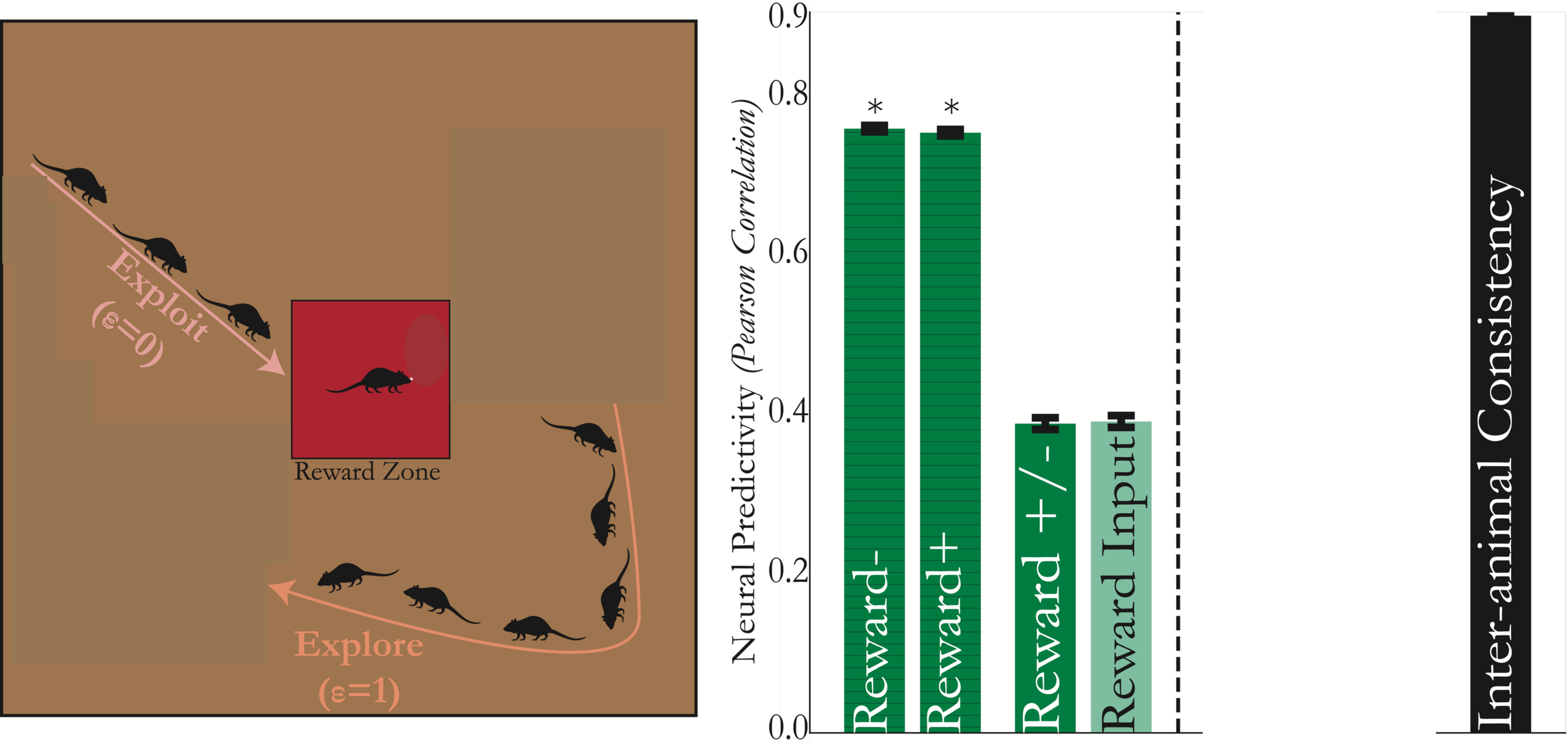


# Reward must be extrinsically modeled

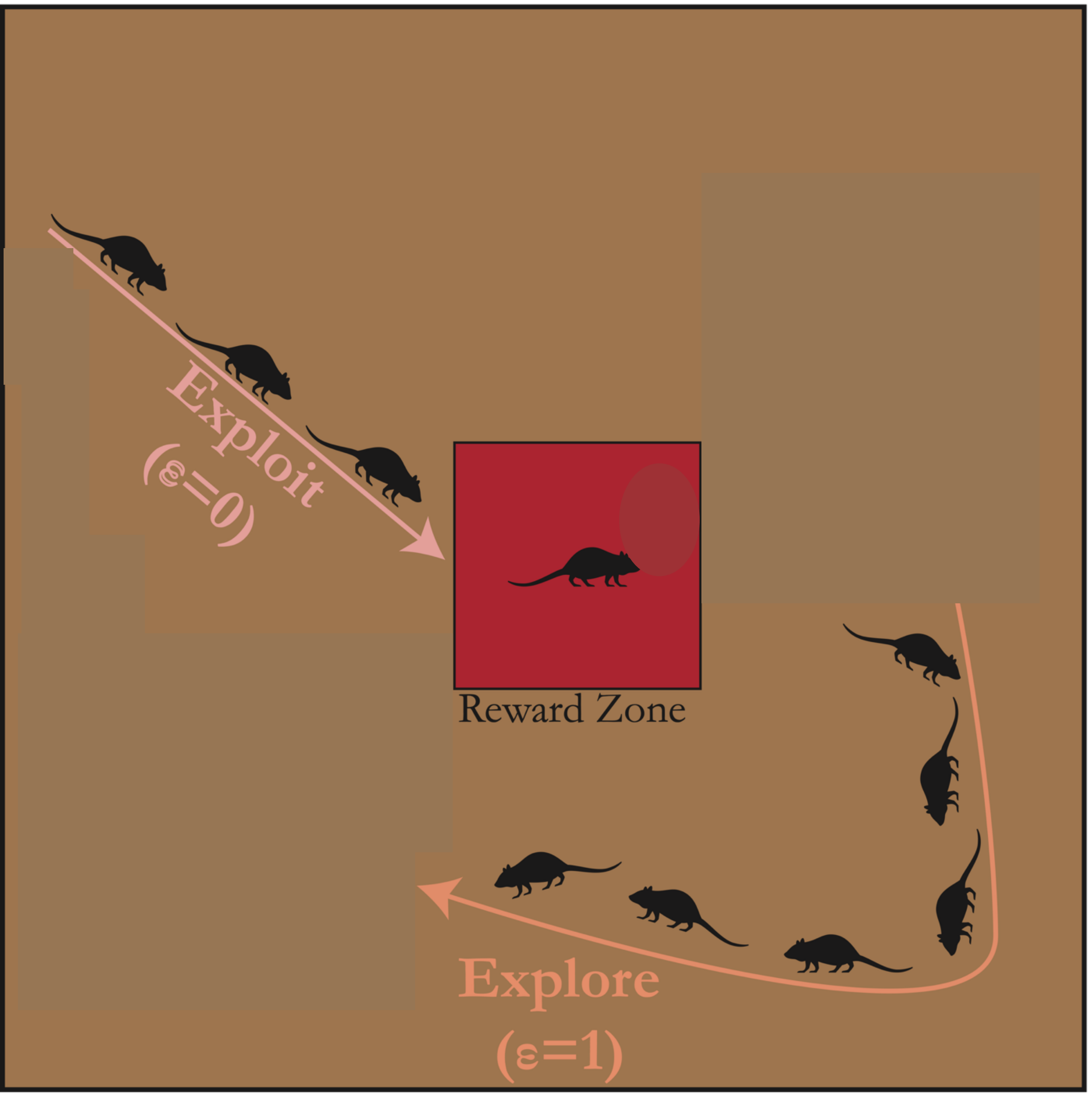


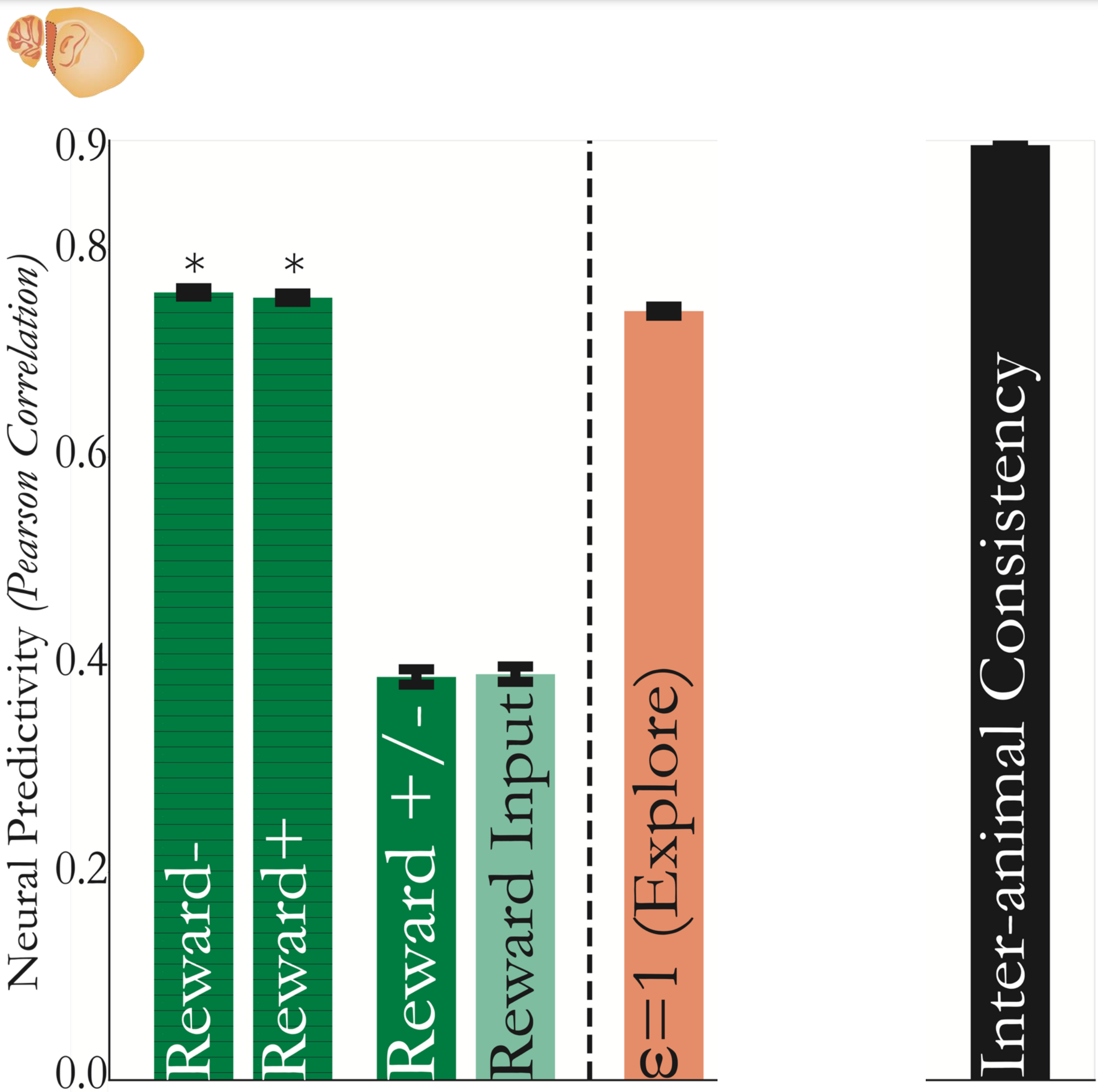



# Modeling rewards as *biased* path integration

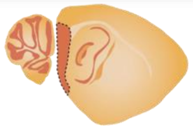


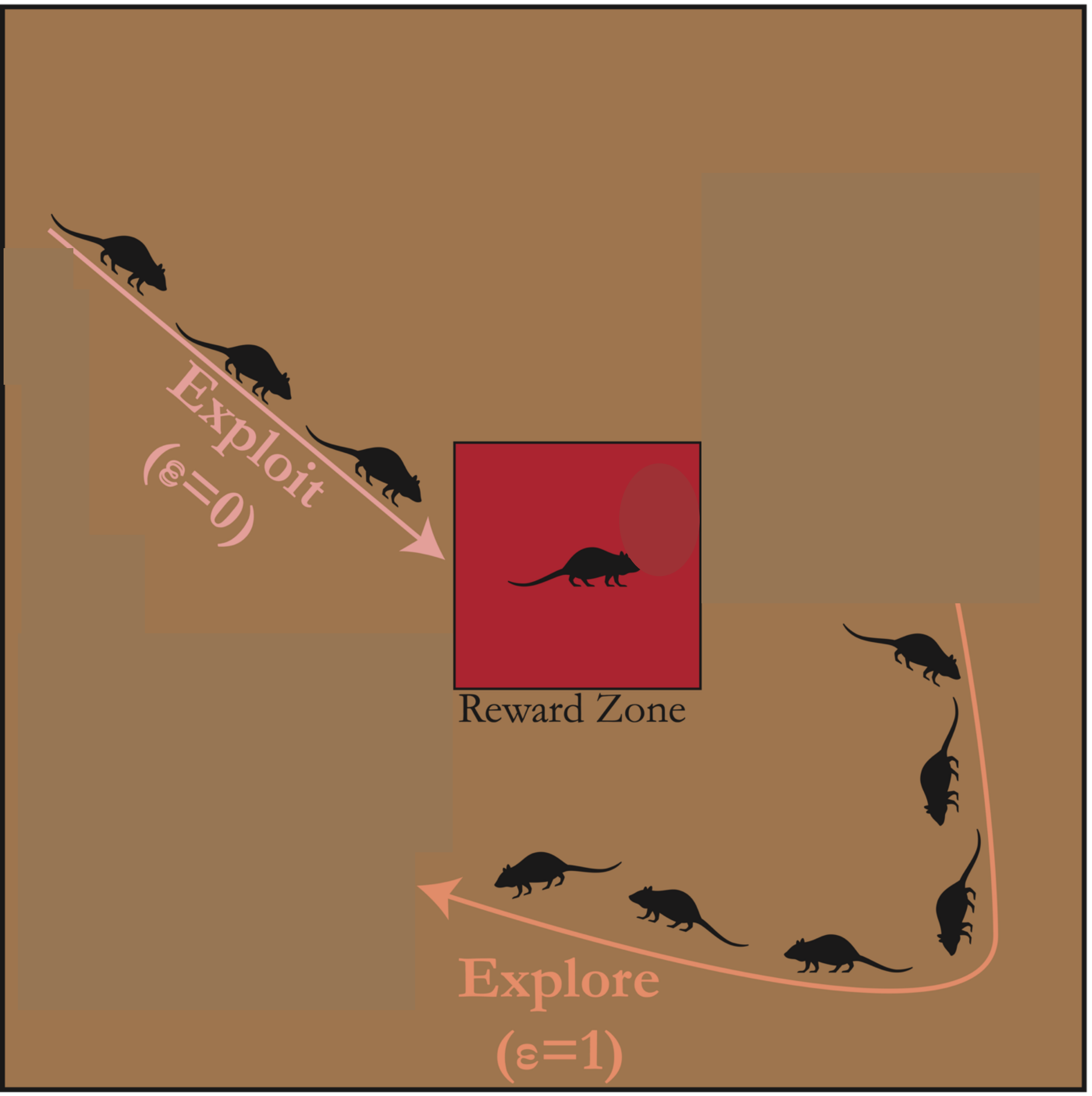
# Modeling rewards as *biased* path integration

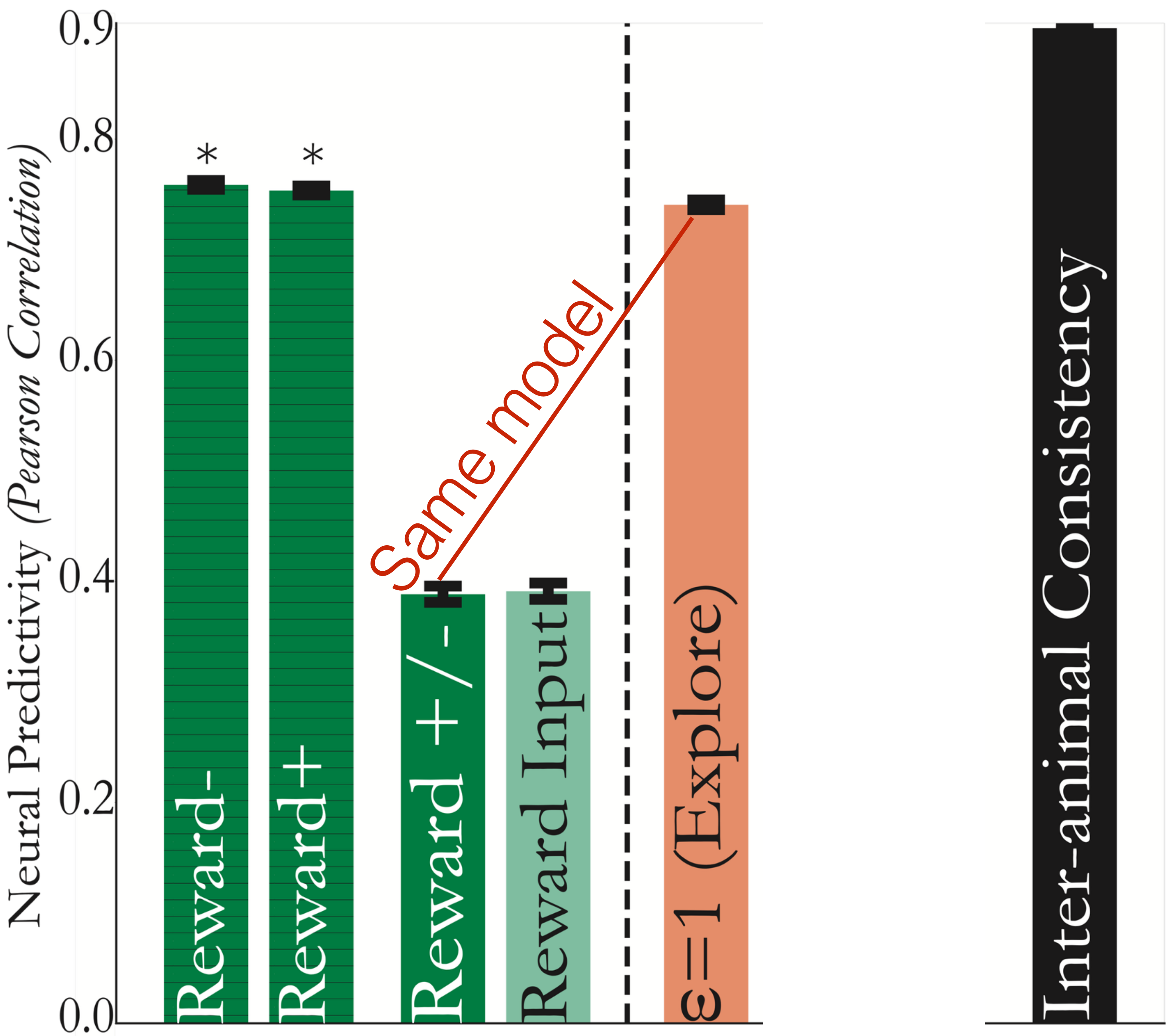


# Modeling rewards as *biased* path integration

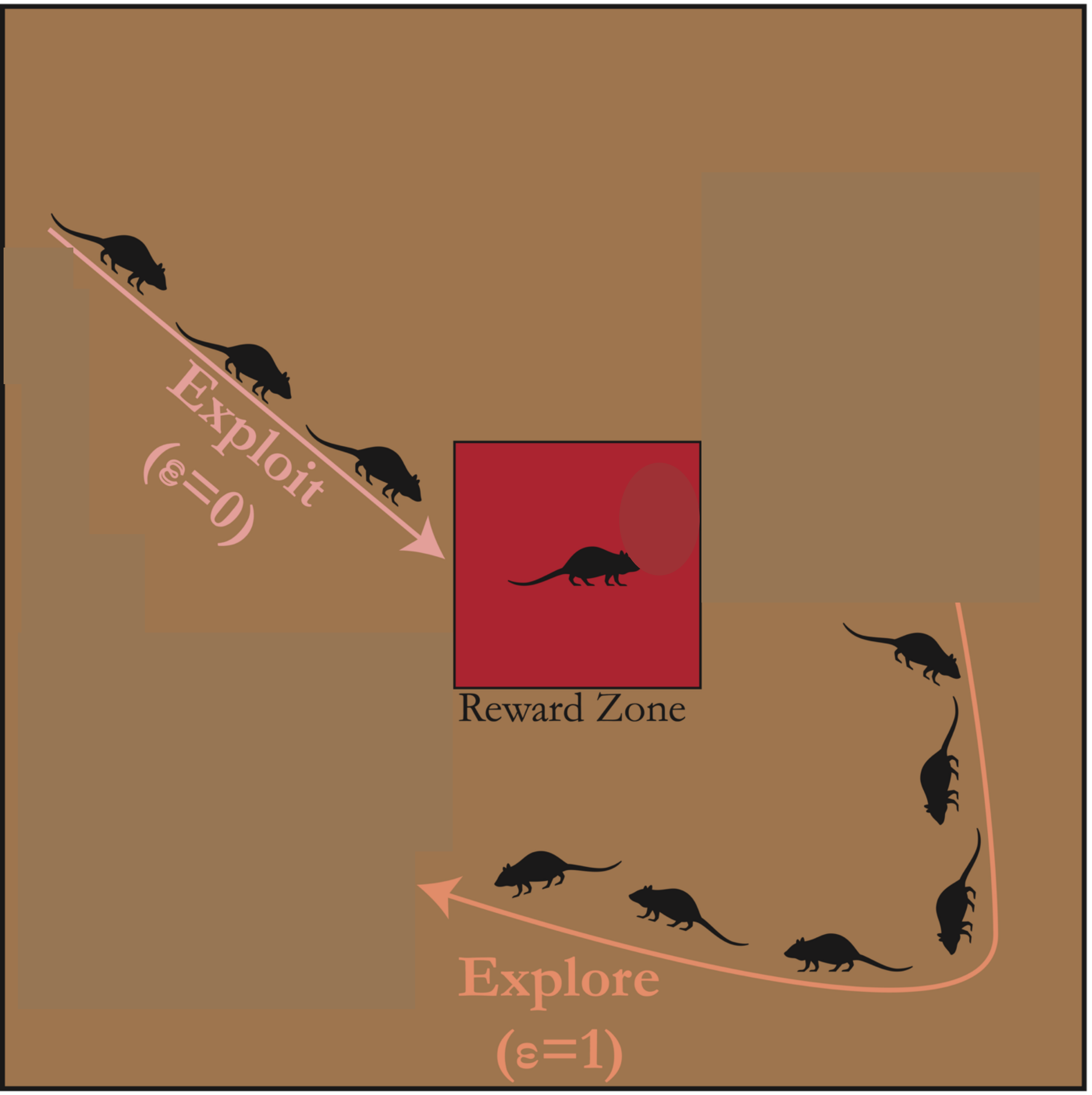


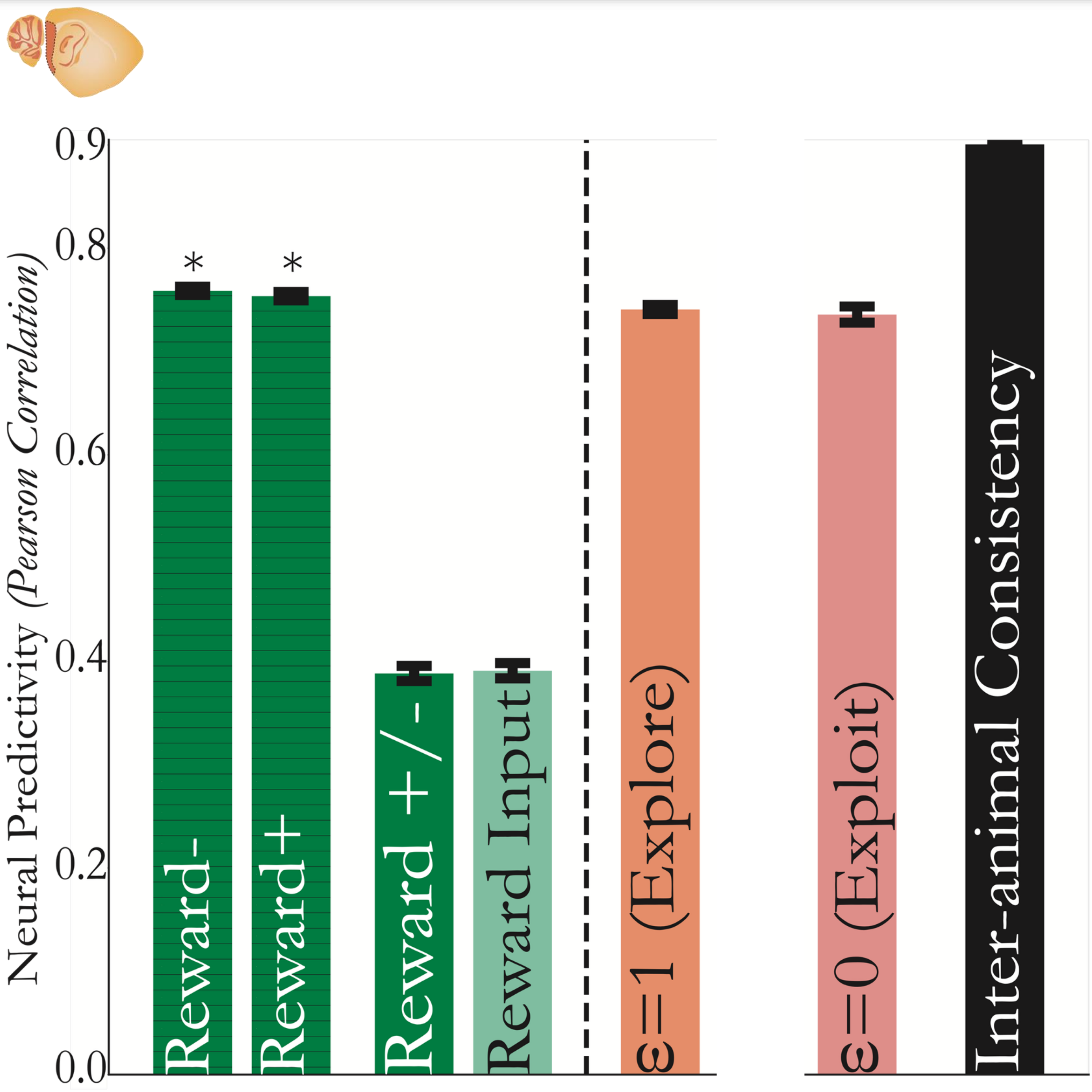


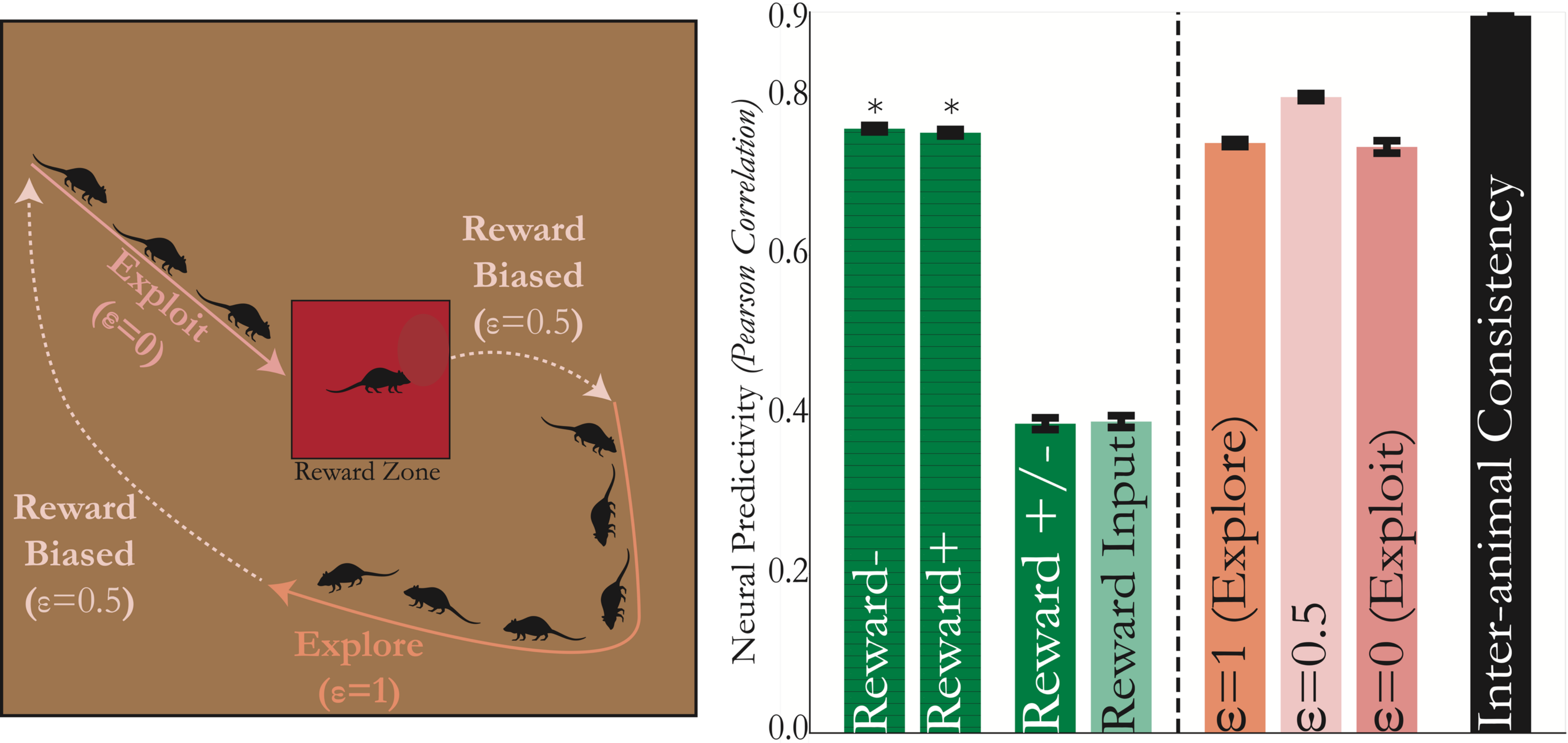

Reward remapping strongly input driven!

# Pure exploitation isn't any better



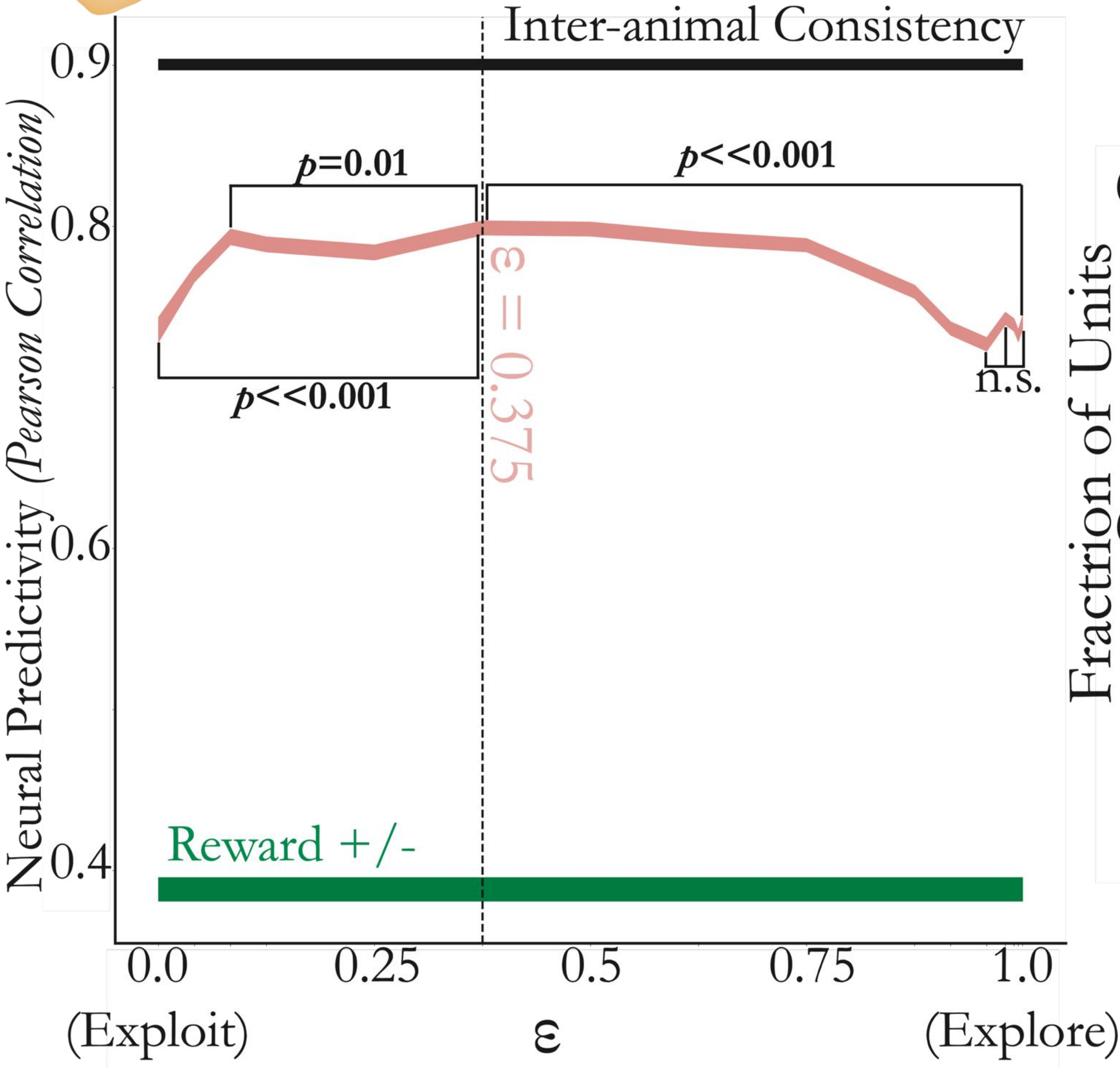
*Reward-biased* path integration captures remapping of responses in the presence of reward

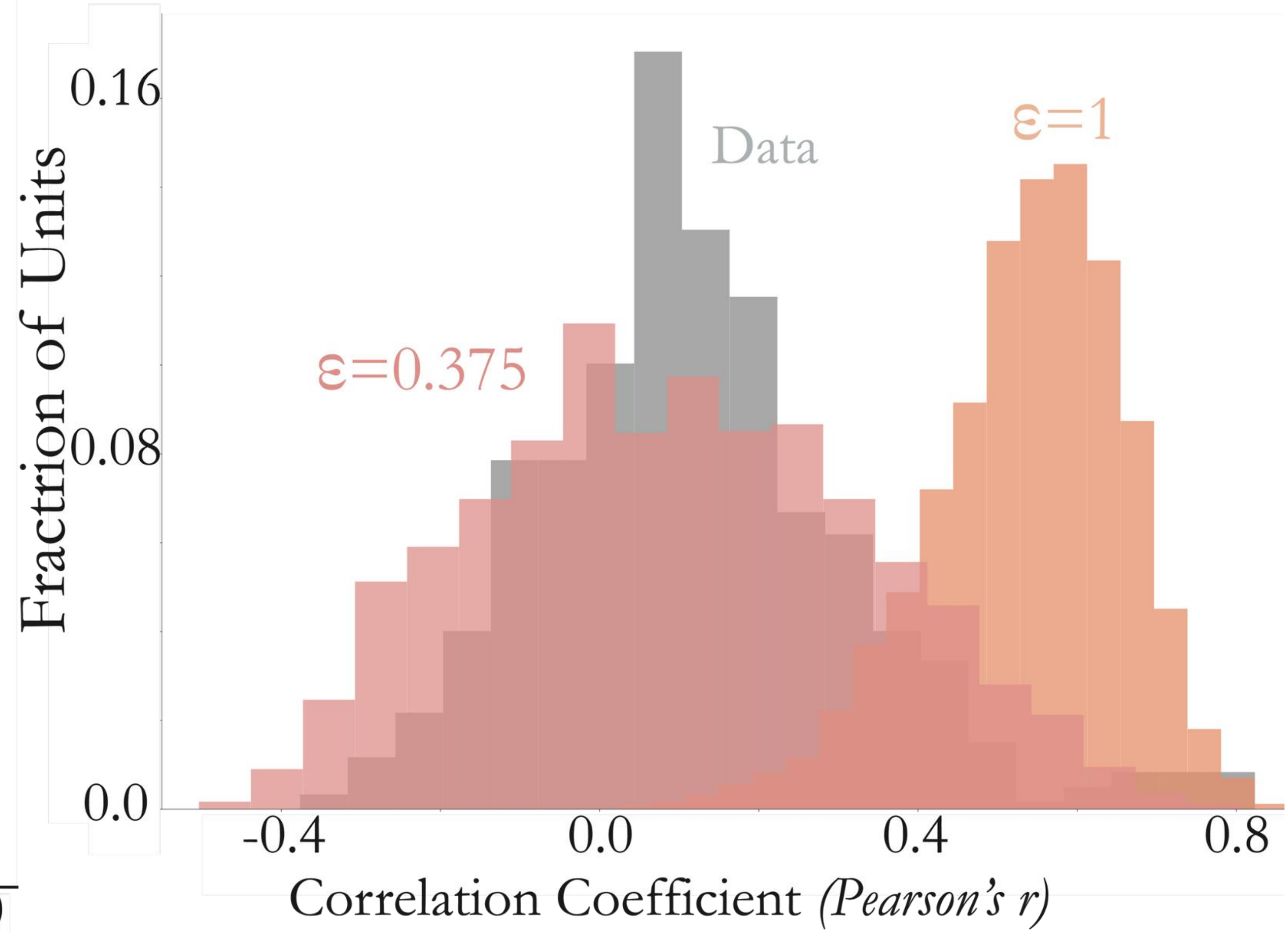



Reward-biased path integrator best captures remapping

*Reward-biased* path integration captures remapping of responses in the presence of reward

Slight bias to exploitation preferred



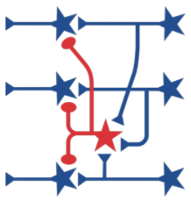




Reward-biased path integrator best captures remapping

# Takeaways

***A*** = *architecture class*

**1.** *"Circuit"*



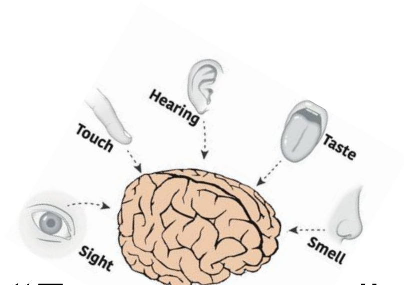
***T*** = *task loss*

**3.** *"Ecological niche/behavior"*

<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?




**2.** *"Environment"*

***D*** = *data stream*

# Takeaways

***A*** *= architecture class*

**1.** *"Circuit"*

gating + nonnegativity

***T*** *= task loss*

**3.** *"Ecological niche/behavior"*

place cell integration

~~path integration~~

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

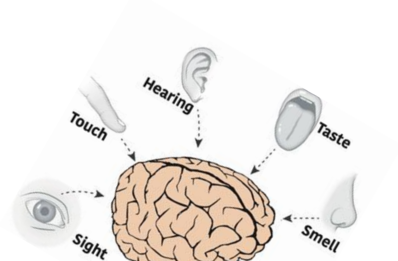
1. How might we characterize what these heterogeneous cells do?

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

<u>*Partial* Resolution:</u>

1. <u>Characterization:</u> Close to perfect neural predictivity with the above constraints — more complex environments are needed!

2. <u>Functional Role:</u> Grid cells are *not* more functionally relevant for navigation! Both heterogeneous and grid cells arise *jointly* through task optimization.




*"Environment"*

**2.**

***D*** *= data stream*

# Next Steps

What about heterogeneity in hippocampus (HPC)? Is there a single end-to-end **unsupervised** model that gets both the heterogeneity in MEC and HPC?

# Next Steps

What about heterogeneity in hippocampus (HPC)? Is there a single end-to-end **unsupervised** model that gets both the heterogeneity in MEC and HPC?

Can we further **rule out** models by making much more complex, realistic environments with more complicated spatial structure, reward structure, and visual cues?

# Next Steps

What about heterogeneity in hippocampus (HPC)? Is there a single end-to-end **unsupervised** model that gets both the heterogeneity in MEC and HPC?

Can we further **rule out** models by making much more complex, realistic environments with more complicated spatial structure, reward structure, and visual cues?

Does the animal even do memory/navigation tasks that today's best AI models are unable to?

# Next Steps

What about heterogeneity in hippocampus (HPC)? Is there a single end-to-end **unsupervised** model that gets both the heterogeneity in MEC and HPC?

Can we further **rule out** models by making much more complex, realistic environments with more complicated spatial structure, reward structure, and visual cues?

Does the animal even do memory/navigation tasks that today's best AI models are unable to?

a. If not, then: hypothesis: some existing AI model can be morphed to be a completely correct model of rodent MEC

# Next Steps

What about heterogeneity in hippocampus (HPC)? Is there a single end-to-end **unsupervised** model that gets both the heterogeneity in MEC and HPC?

Can we further **rule out** models by making much more complex, realistic environments with more complicated spatial structure, reward structure, and visual cues?

Does the animal even do memory/navigation tasks that today's best AI models are unable to?

a. If not, then: hypothesis: some existing AI model can be morphed to be a completely correct model of rodent MEC

b. If so, then: hypothesis: if we build a new and improved AI model that solves the task, it will be an improved model of rodent MEC in the rich condition

# Next Steps

Likely critical for the next generation of AI agents to be able to deal with continuous real world inputs, rather than limited context windows!

Can we further **rule out** models by making much more complex, realistic environments with more complicated spatial structure, reward structure, and visual cues?

Does the animal even do memory/navigation tasks that today's best AI models are unable to?

a. If not, then: hypothesis: some existing AI model can be morphed to be a completely correct model of rodent MEC

b. If so, then: hypothesis: if we build a new and improved AI model that solves the task, it will be an improved model of rodent MEC in the rich condition

# Acknowledgments

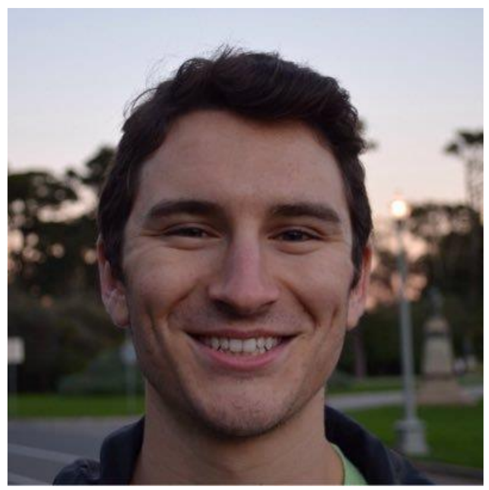

Alexander Attinger

Malcolm G. Campbell

Kiah Hardcastle

Isabel I.C. Low

Caitlin S. Mallory



Gabriel C. Mel

Ben Sorscher

Alex H. Williams

Surya Ganguli

Lisa M. Giocomo

Daniel L.K. Yamins



**Contact:**

anayebi@cs.cmu.edu

@aran_nayebi