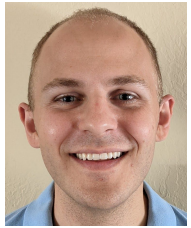
# No Free Lunch from Deep Learning in Neuroscience

Rylan Schaeffer

Joint work with Mikail Khona & Ila Fiete

NeurIPS 2022



@RylanSchaeffer
@KhonaMikail

# The claimed promise of deep networks in neuroscience

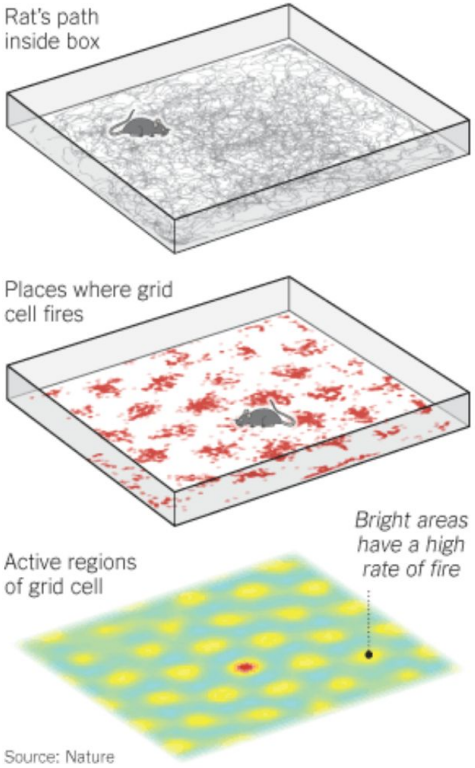
- Deep networks in neuroscience: not just a tool for data analysis, but a possible model of the brain
- 2 claimed promises for what deep learning can offer scientifically:
  - Deep networks can shed light on the brain's optimization problems
  - Deep networks can yield novel predictions about neural phenomena

In this paper, we ask: For grid cells, does deep learning deliver on either of these claims?

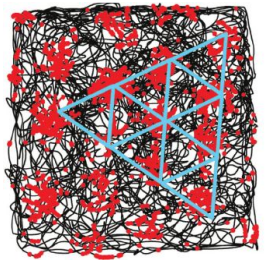
**We studied** recent deep learning models of hippocampus and medial entorhinal cortex

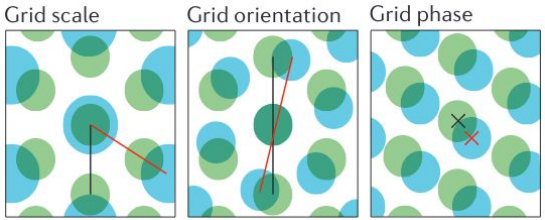
**We showed** deep learning models of neural circuits may tell us *less* about fundamental scientific truths and *more* about programmers' particular implementation choices, and as a result might be more post hoc than predictive.

# Background: Grid Cells in Medial Entorhinal Cortex (MEC)




- Firing pattern of single grid cell forms a regular triangular lattice in physical space
- Grid cells display different scales, orientations and phases
- Grid cells display only a few different periods (scales)





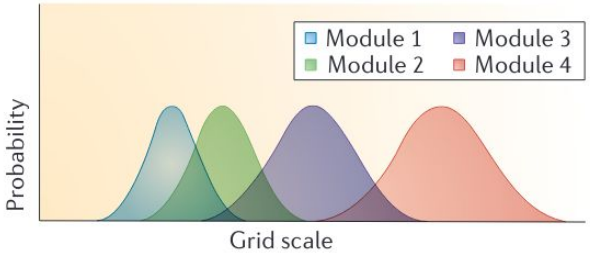




Figure sources:
https://www.nytimes.com/2013/04/30/science/how-grid-cells-in-brain-help-map-out-space.html
https://www.nature.com/articles/nrn3766.pdf

# What insight(s) have deep learning models claimed to offer?

Claim: "Path integration causes the formation of grid cells"

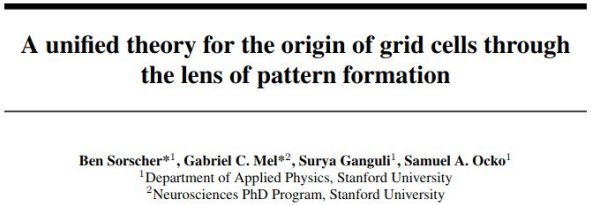
Letter | Published: 09 May 2018

**Vector-based navigation using grid-like representations in artificial agents**

Andrea Banino, Caswell Barry, ... Dharshan Kumaran + Show authors

*Nature* **557**, 429–433 (2018) | Cite this article

*"Our results show that grid-like representations reminiscent of those found in the mammalian entorhinal cortex emerge in a generic network trained to path integrate"*



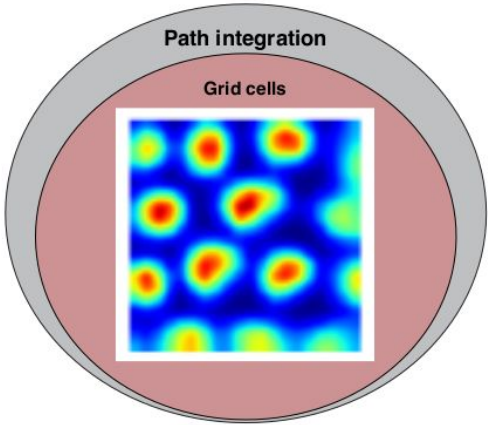
**A unified theory for the origin of grid cells through the lens of pattern formation**

Ben Sorscher*[1], Gabriel C. Mel*[2], Surya Ganguli[1], Samuel A. Ocko[1]
[1]Department of Applied Physics, Stanford University
[2]Neurosciences PhD Program, Stanford University

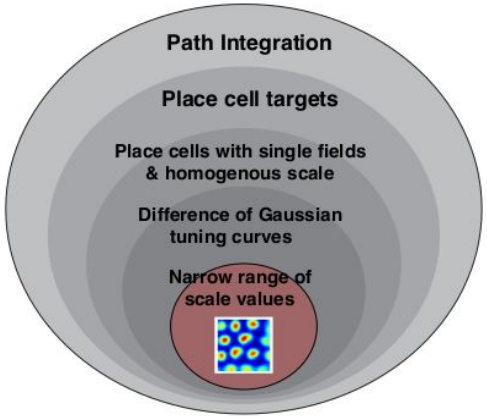
*"Remarkably, in each case the networks learn [...] grid-like representations"*

*"diverse architectures [...] all converge to a grid-like solution"*

# "Path integration causes the formation of grid cells" (?)

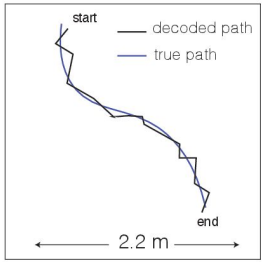


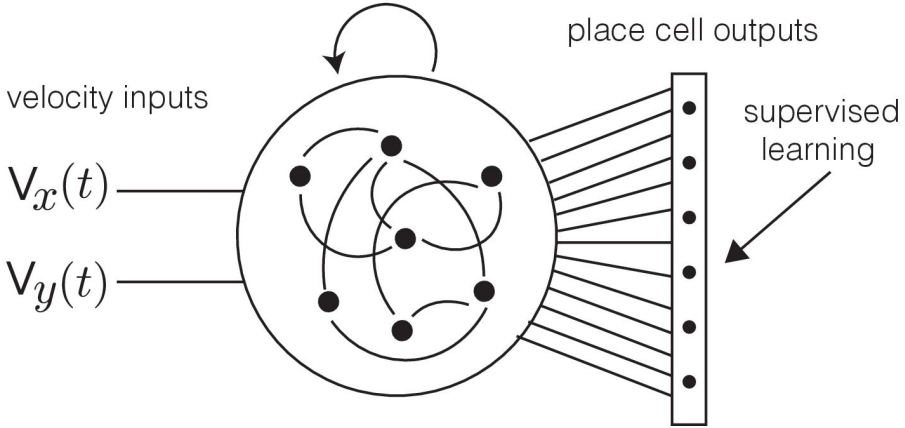

Previous work: presents story that path integration drives the formation of grid cells

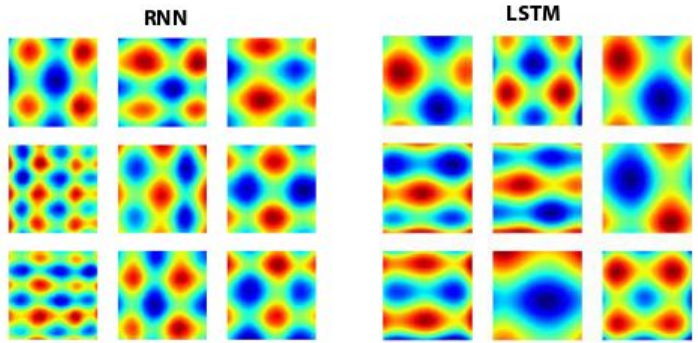


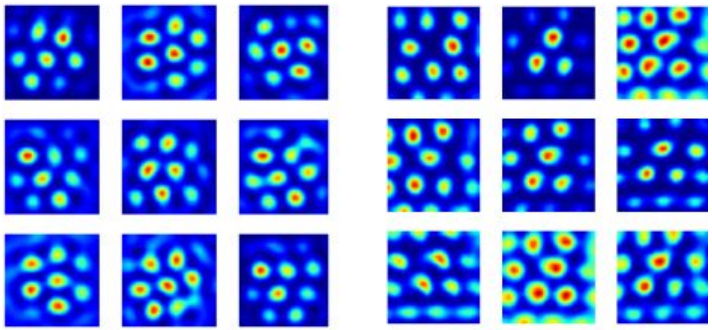

Our work: (1) path integration does not require grid cells, (2) grid-like units emerge in a very small fraction of hyperparameter space, (3) only under biologically invalid implementation choices, and (4) lack key properties of biological grid cells

# Deep learning path integration model setup



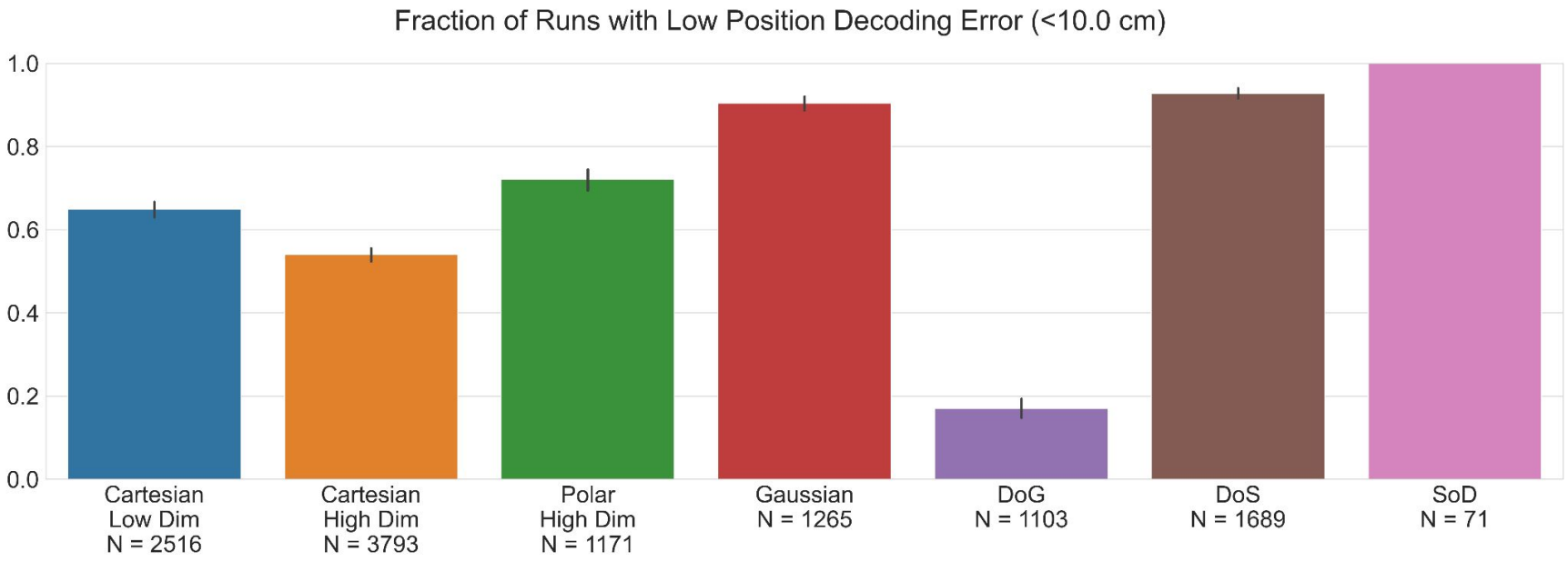



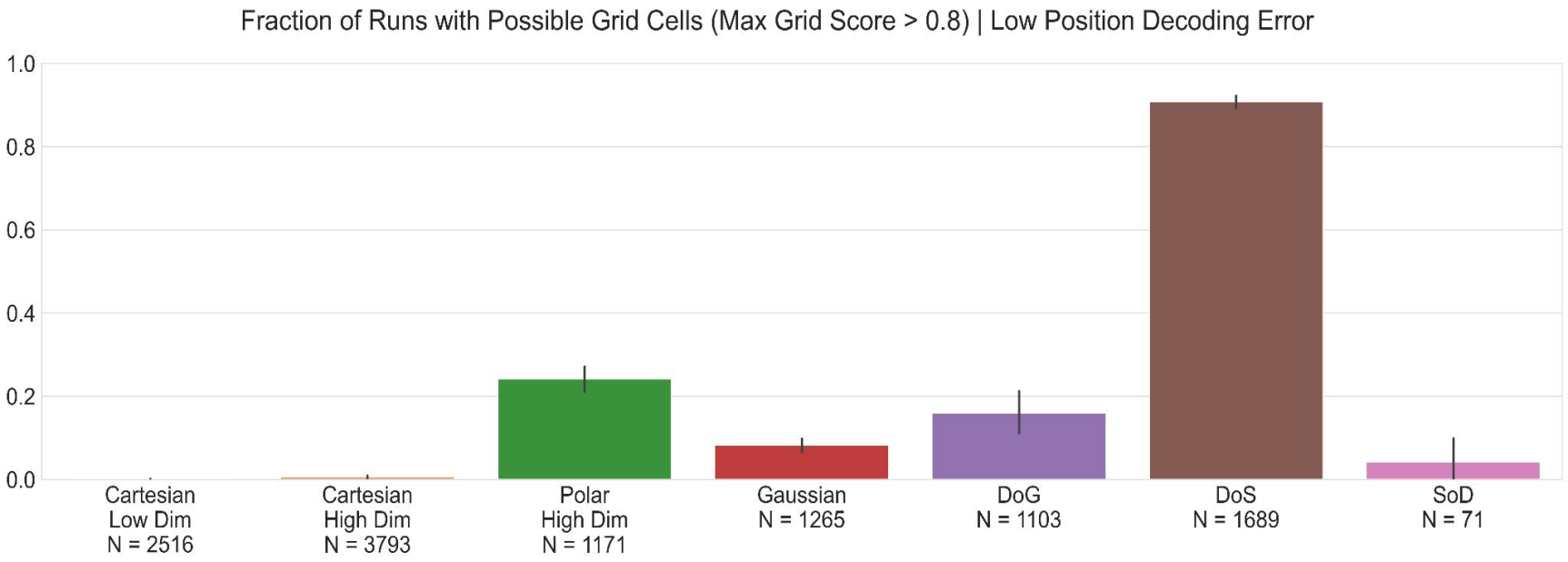



Sorscher et al. NeurIPS 2019

# Result #1: Most networks accurately path integrate…

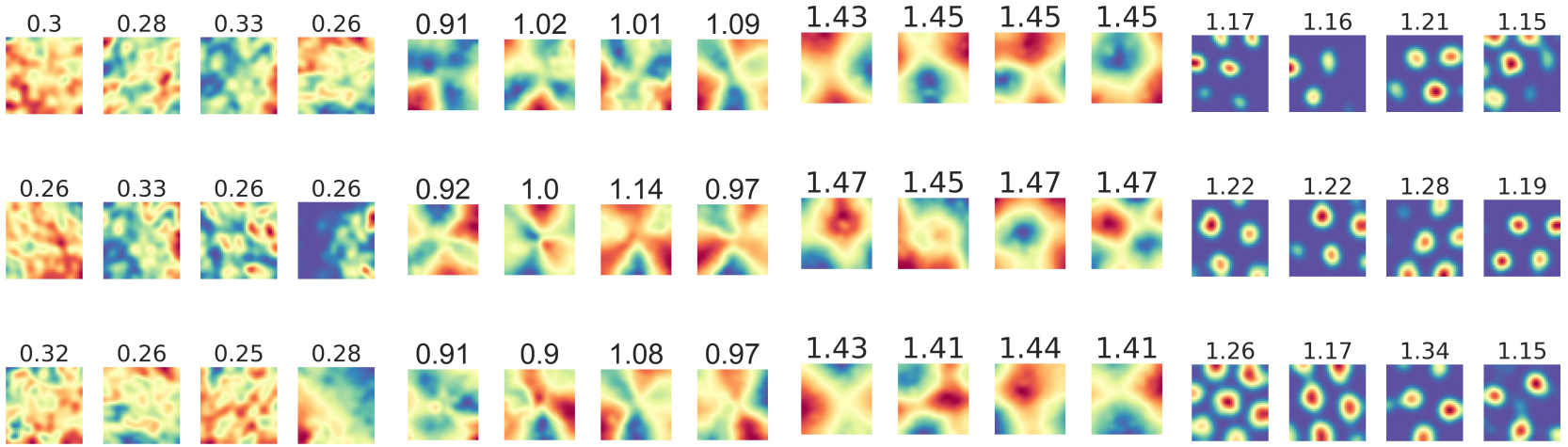


# Result #1: Most networks accurately path integrate, but few learn *possible* grid-like units



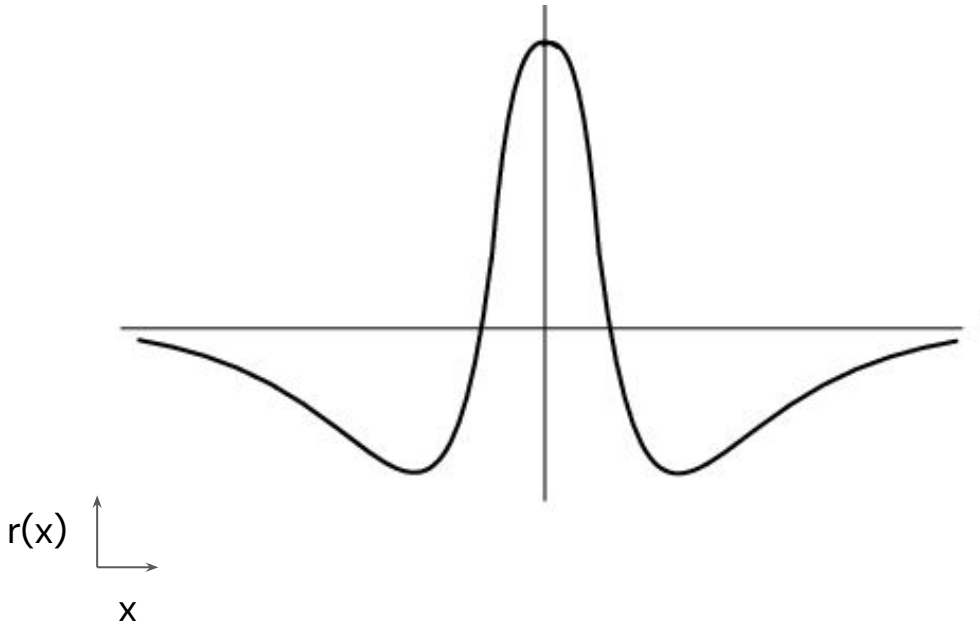

cf. Sorscher et al. 2019: “Remarkably, in each case the networks learn […] grid-like representations”

# Result #2: Grid-like units only emerge under one specific encoding of position



# Emergence of grid-like units requires post-hoc constraints that are biologically invalid for many reasons

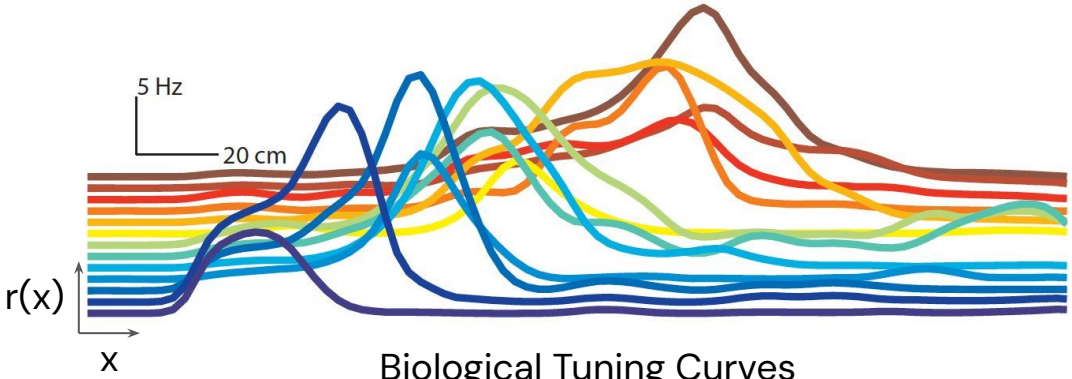
- To obtain grid-like units, artificial place cells must have:
  - single scale across the population
  - single field per cell
  - isotropic tuning curves with...
  - ...a Difference-of-Softmaxes functional form



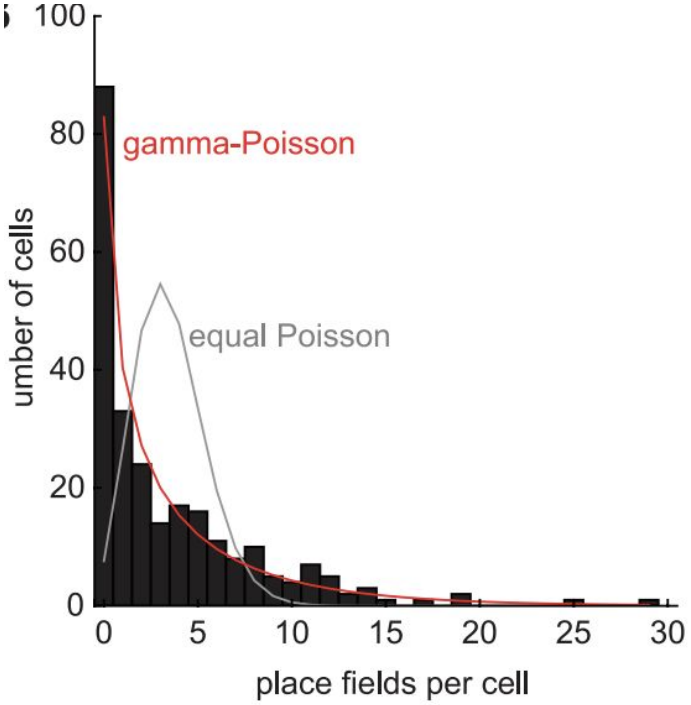

Artificial Tuning Curve

# Emergence of grid-like units requires post-hoc constraints that are biologically invalid for many reasons

- Biological place cells are the opposite
  - heterogeneous scales across the population
  - multiple fields per cell
  - anisotropic tuning curves with…
  - …no clear functional form

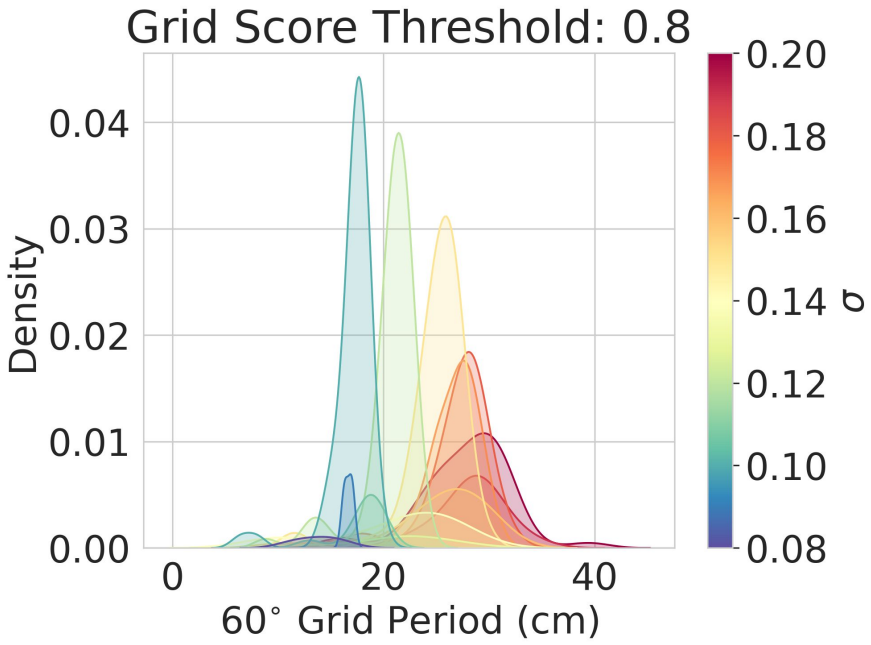


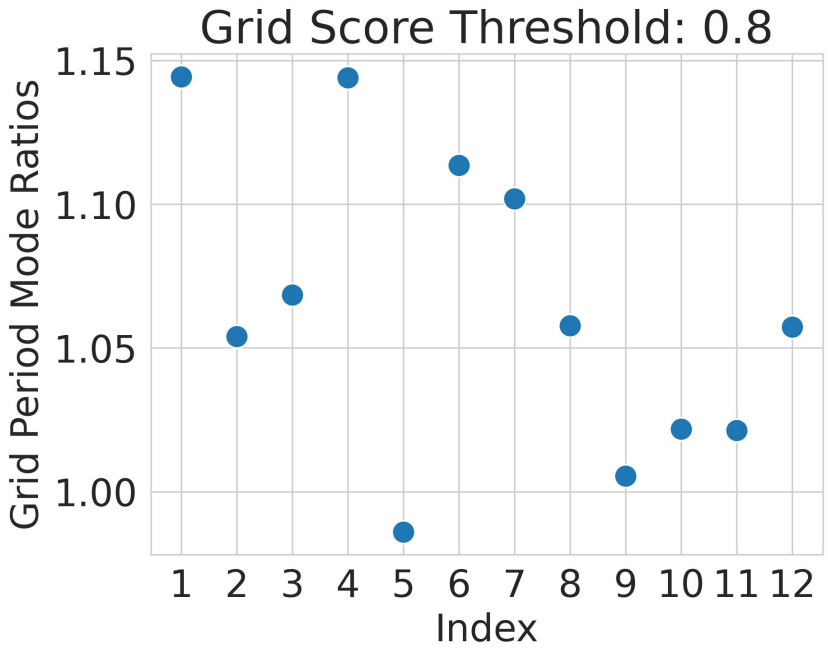

Biological Tuning Curves




Rich et al, Science(2014)

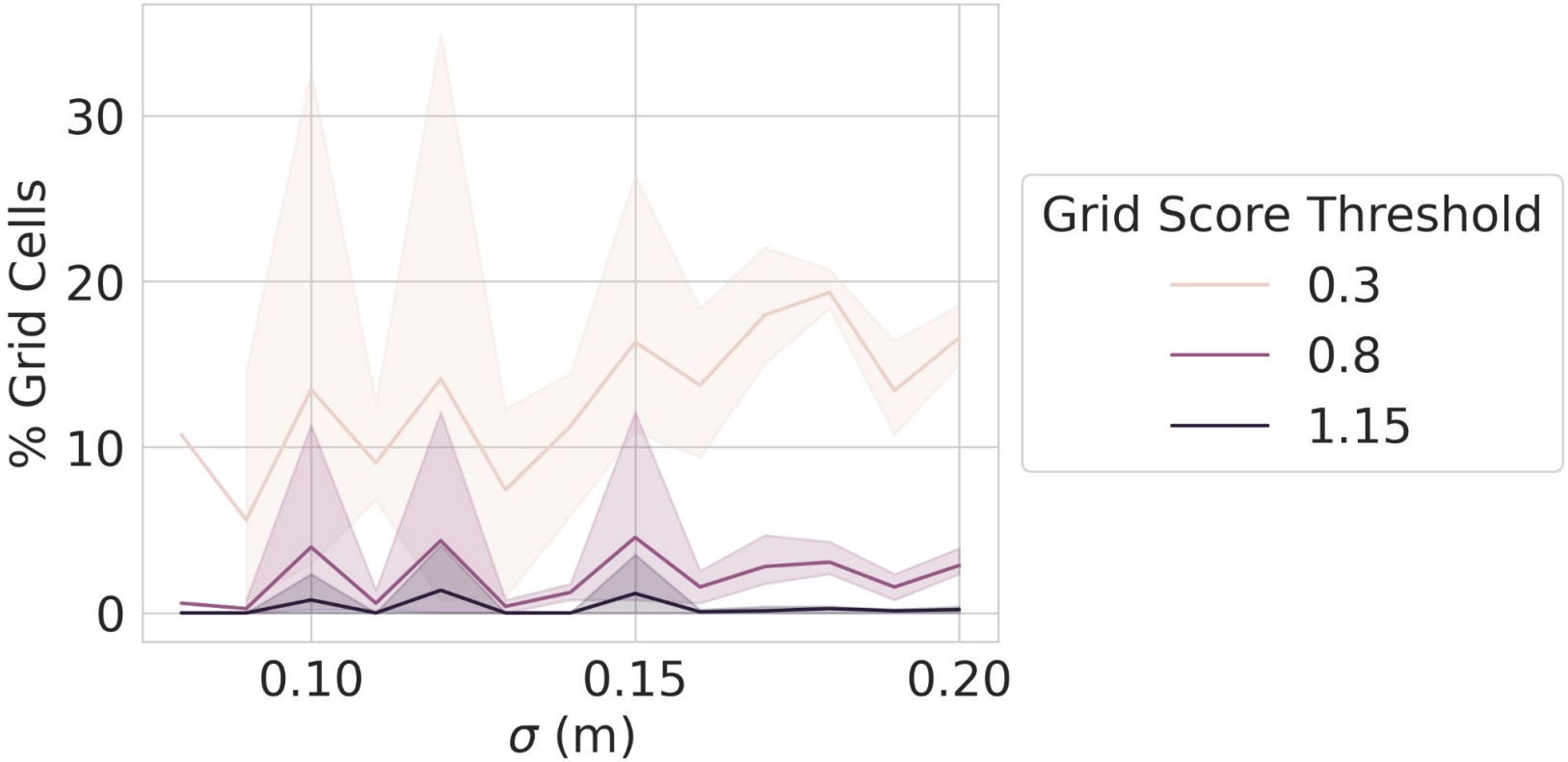
# Result #3: Grid-like units lack key properties of biological grid cells (multiple modules, specific ratios)



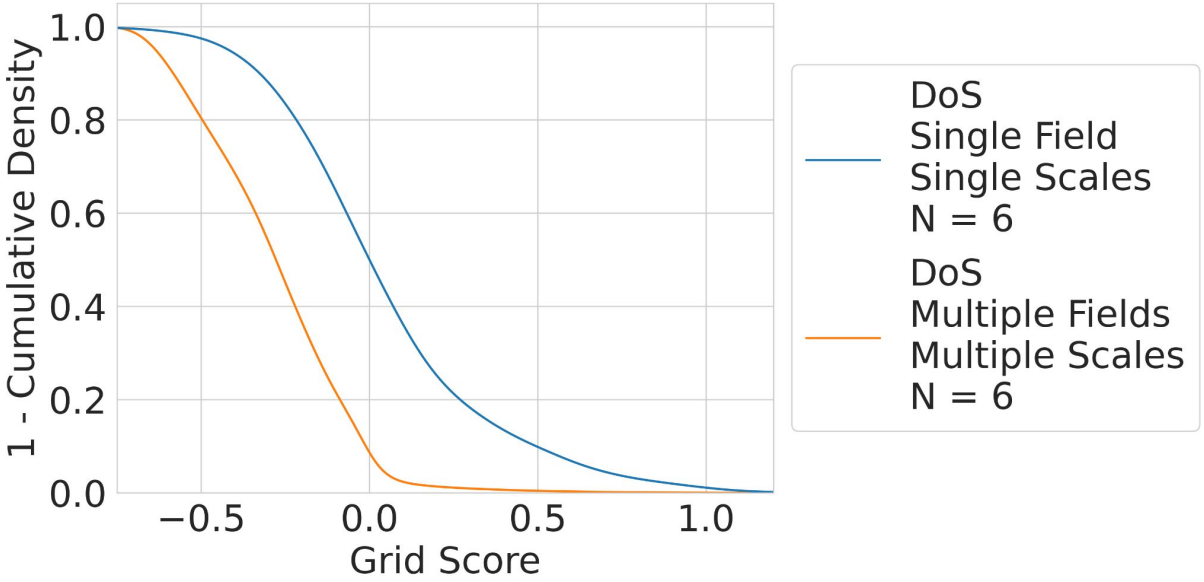



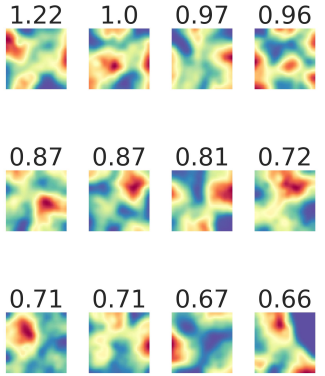

Each network learns 1 periodicity (1 peak) cf. biological grid cells display multiple peaks

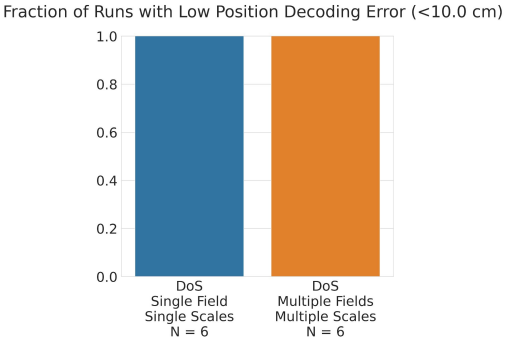
# Result #4: Small perturbations to ideal hyperparameters prevent the formation of grid-like units



# Result #5: Under more biologically plausible conditions, grid-like units do not emerge (without harming performance)



# Result #6: First-principles models can explain all these findings!

$$\dot{r}(x) = -r(x) + g(W \star r)$$

If interaction is translationally invariant: $W(x, x') = W(x - x') = W(\Delta x)$

We can linearize: $\dot{r}(x) \sim -r(x) + f(\Delta x) \star r$

Fourier transform: $\dot{\tilde{r}}(k) = -\tilde{r}(k) + \tilde{f}(k)\tilde{r}(k)$

Formed pattern scale: $k^* = \mathrm{argmax}_k \tilde{f}(k)$

# Result #6: First-principles models can explain all these findings!

For MSE reconstruction error: $||P - W_{\text{readout}} r||^2$

where $\Sigma = PP^T$

Place cell correlation matrix

Sorscher et al(2019): $\dot{r} = -\lambda r + \Sigma r$

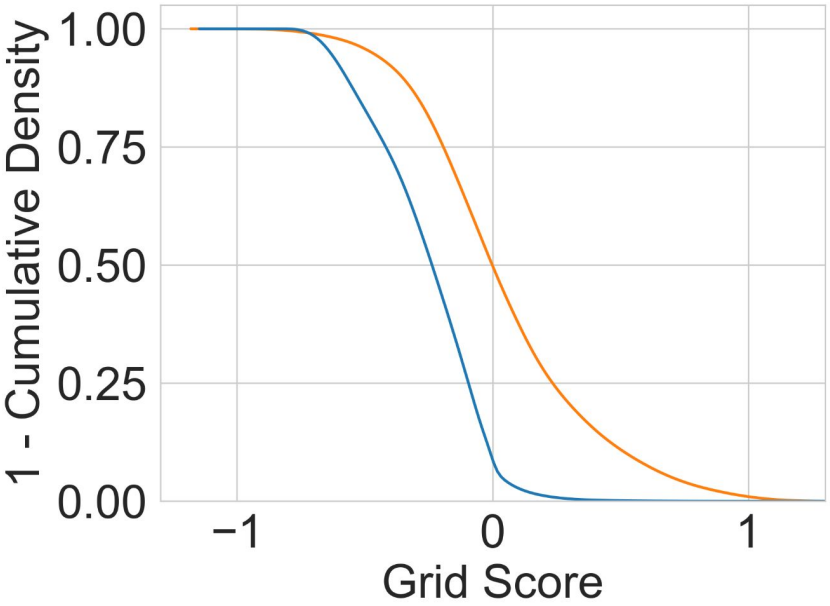
Difference of Gaussians: $W(\Delta x) \equiv f(\Delta x) = \alpha_E \exp\left(-\frac{(\Delta x)^2}{2\sigma_E^2}\right) - \alpha_I \exp\left(-\frac{(\Delta x)^2}{2\sigma_I}\right)$

$$\tilde{f}(k) = \int_{\mathscr{R}} d(\Delta x) f(\Delta x) e^{ik\Delta x} = \alpha_E \sigma_E \exp\left(-\frac{\sigma_E^2 k^2}{2}\right) - \alpha_I \sigma_I \exp\left(-\frac{\sigma_I^2 k^2}{2}\right)$$

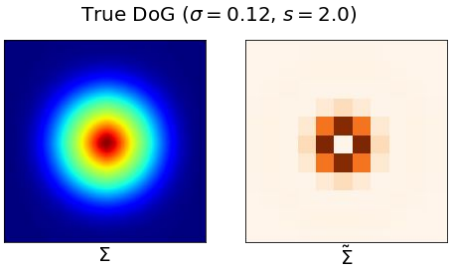
Non-zero maxima = pattern formation! $[k^*]^2 = \frac{2}{\sigma_E^2 - \sigma_I^2} \log\left(\alpha_E \sigma_E^3 / \alpha_I \sigma_I^3\right)$

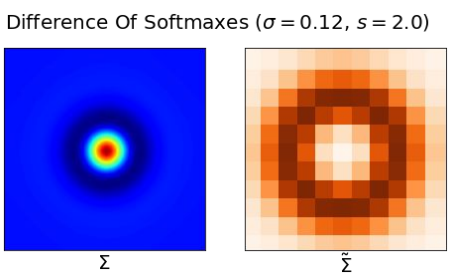
# Result #7: Unmentioned implementation details are necessary for the formation of grid-like units

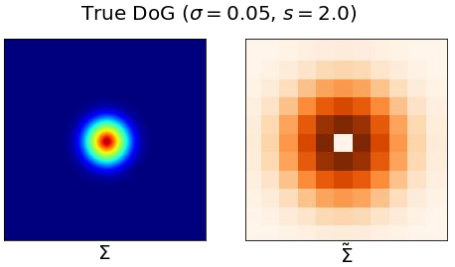
$$p_i(x) = e^{-||x-c_i||^2/2\sigma_1^2} - e^{-||x-c_i||^2/2\sigma_2^2}$$







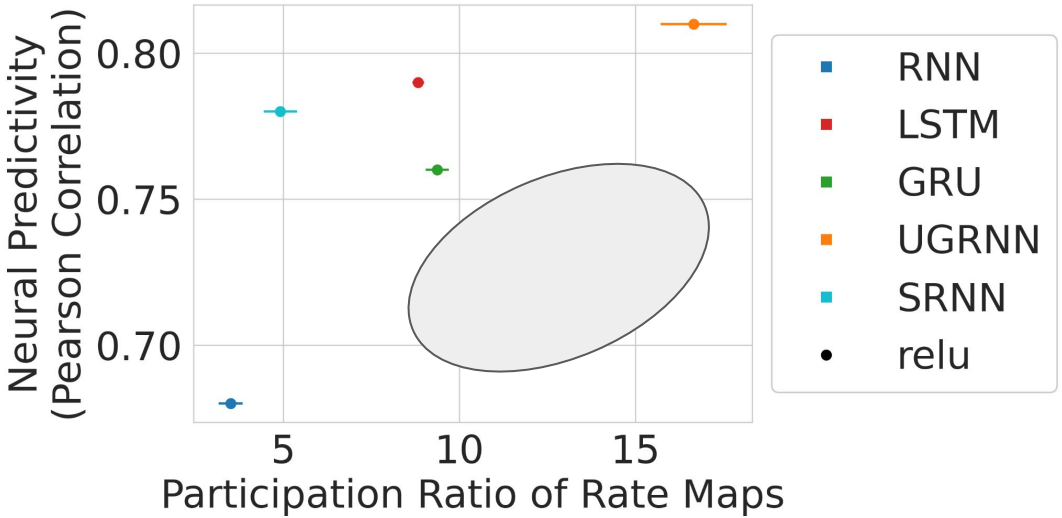



# Paradox: How do networks predict neural activity so well?

- Nayebi et al. NeurIPS 2021 (Spotlight) found that these networks could explain variance in mouse MEC neural activity, comparable to variance explained by other mice's MEC neural activity
- Lent additional support that these networks are good models of MEC-HPC
- However, these networks are inconsistent with (a) key biological features of grid cells (e.g., learning only a single module), and (b) multiple known properties of place cells (e.g. requiring single field, single scale, isotropic Difference-of-Softmax tuning curves)
- Paradox: How do these networks predict mouse MEC neural activity so well?

# Paradox: How do networks predict neural activity so well?

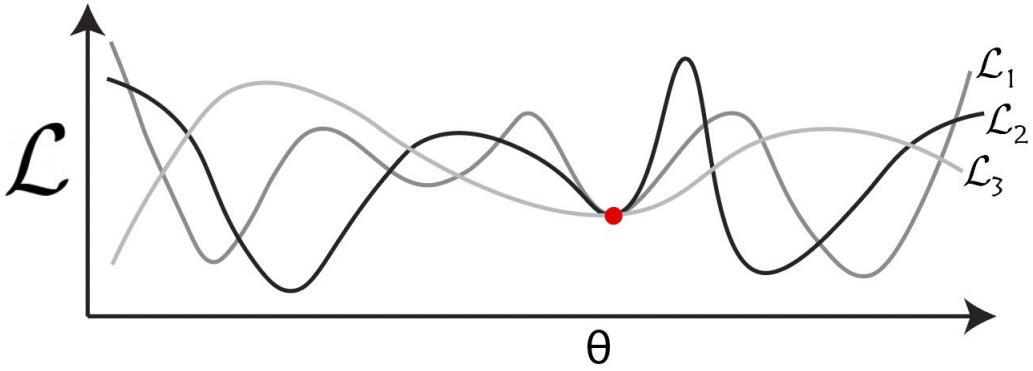
- Hypothesis: Different networks provide different dimensional bases for the regression comparisons, thereby achieving higher (test) correlations
- Data and analysis code were not public, but we have preliminary supporting evidence:




**Q: Why are there no models with high ED, but low brain match?**

# Conclusion: No Free Lunch with Deep Learning Models for Neuroscience

You have found an optimization problem to train a network to replicate the brain's neural tuning. But multiple optimization problems can share an optimum → ***brain could be solving a different optimization problem.***




For an optimization problem correctly matched to one the brain is solving, a model may learn a different optimum that yields different neural tuning → ***model is not a predictive model of tuning.***