No Free Lunch from Deep Learning in Neuroscience: A Case Study through Models of the Entorhinal-Hippocampal Circuit

Rylan Schaeffer[1,3], Mikail Khona[2,3], Ila Rani Fiete[3,4]

[1]Stanford Computer Science
[2]MIT Physics
[3]MIT Brain and Cognitive Sciences
[4]McGovern Institute for Brain Research

2nd AI for Science
An ICML 2022 Workshop

Motivation

- Deep learning is driving renaissances across many scientific disciplines

- Neuroscience is interested in deep learning, not just as a tool, but as a model system of the brain

- The promises of deep learning-based models of brain circuits are twofold:
    1. Make novel predictions about neural phenomena, and/or
    2. Shed light on the brain's optimization problems and solutions

- We show, through the case-study of **Grid Cells (GCs)** in medial-entorhinal cortex, that deep learning models of brain circuits often provide neither

What are grid cells (GCs)?

Published: 19 June 2005

Microstructure of a spatial map in the entorhinal cortex

Torkel Hafting, Marianne Fyhn, Sturla Molden, May-Britt Moser & Edvard I. Moser

Nature **436**, 801–806 (2005) | Cite this article

- Type of neuron located in the mammalian medial entorhinal cortex
- Play a fundamental role in spatial navigation through the process of **path integration**

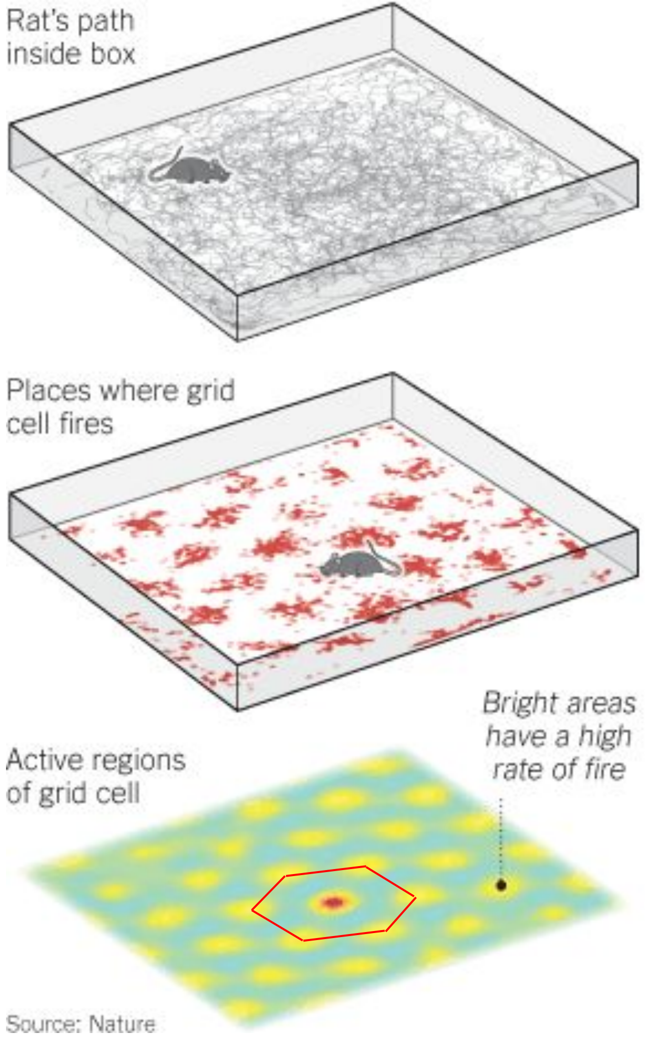
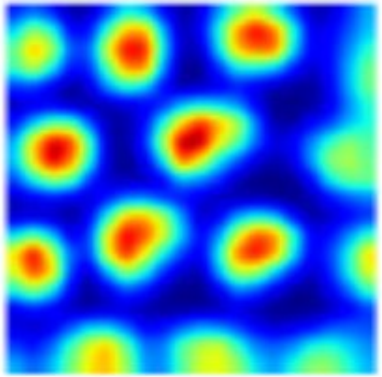


Figure Source: NYT 2013

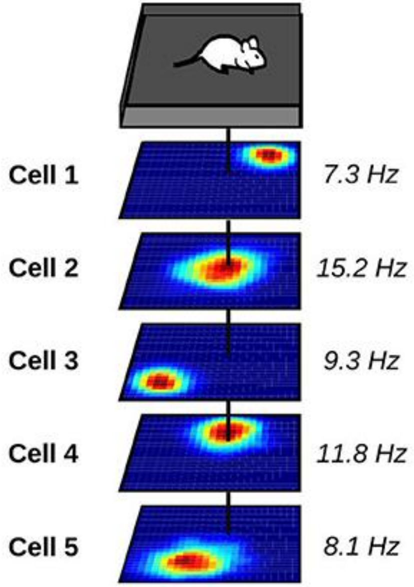
What is path integration and how is it done by the brain?

Path integration is the task of estimating one's absolute spatial position in an environment by adding up velocity estimates

Velocity information

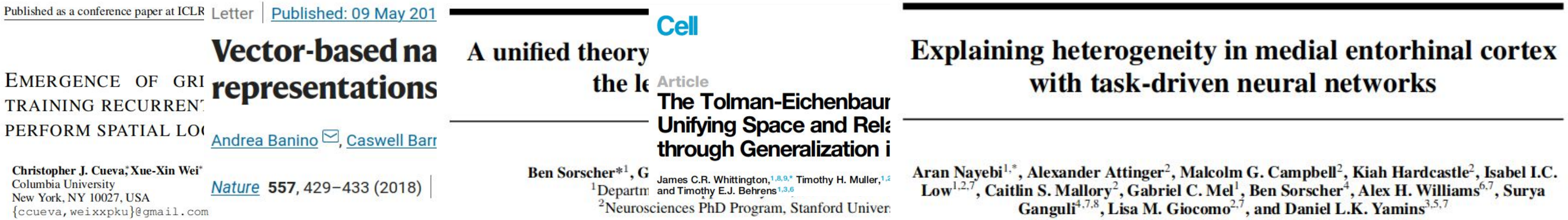


Grid cells in the medial entorhinal cortex

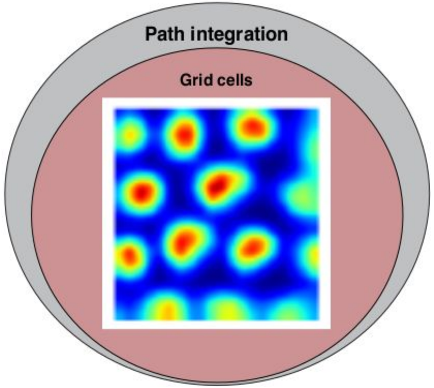


Place cells in the Hippocampus

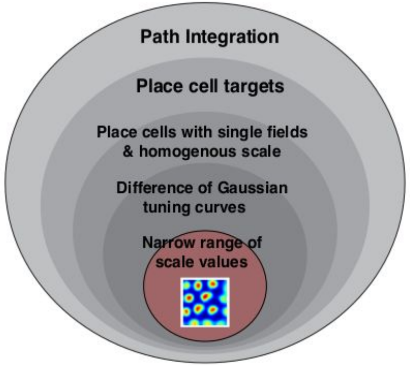
Deep Learning: “In general, path integration yields grid cells”


Published as a conference paper at ICLR
EMERGENCE OF GRI
TRAINING RECURREN
PERFORM SPATIAL LO
Christopher J. Cueva, Xue-Xin Wei
Columbia University
New York, NY 10027, USA
{ccueva,weixxpku}@gmail.com
Letter | Published: 09 May 201
Vector-based na
representations
Andrea Banino, Caswell Barr
Nature 557, 429–433 (2018)
A unified theory
the le
Ben Sorscher[*1], G
[1]Departm
[2]Neurosciences PhD Program, Stanford Univer
Cell
Article
The Tolman-Eichenbaur
Unifying Space and Rela
through Generalization i
James C.R. Whittington,[1,8,9,*] Timothy H. Muller,[1,2]
and Timothy E.J. Behrens[1,3,6]
Explaining heterogeneity in medial entorhinal cortex with task-driven neural networks
Aran Nayebi[1,*], Alexander Attinger[2], Malcolm G. Campbell[2], Kiah Hardcastle[2], Isabel I.C. Low[1,2,7], Caitlin S. Mallory[2], Gabriel C. Mel[1], Ben Sorscher[4], Alex H. Williams[6,7], Surya Ganguli[4,7,8], Lisa M. Giocomo[2,7], and Daniel L.K. Yamins[3,5,7]

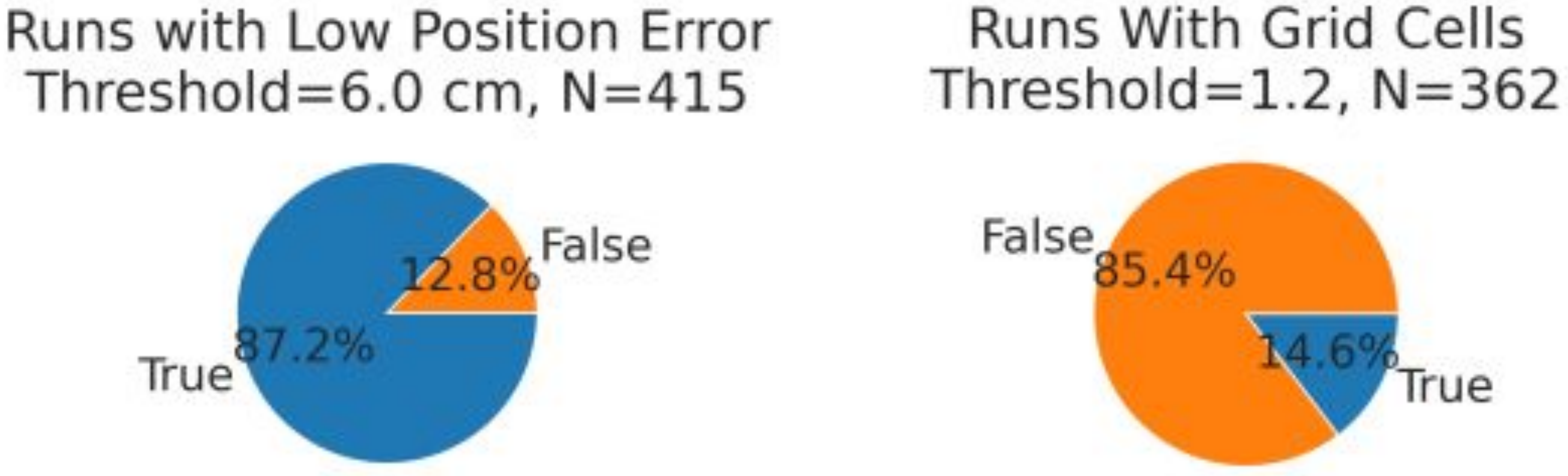
Prior work presents a story that path integration generically drives the formation of grid cells



In this work, we show grid cells only emerge in a small fraction of hyperparameter space, and only when baked in intentionally via specific implementation choices

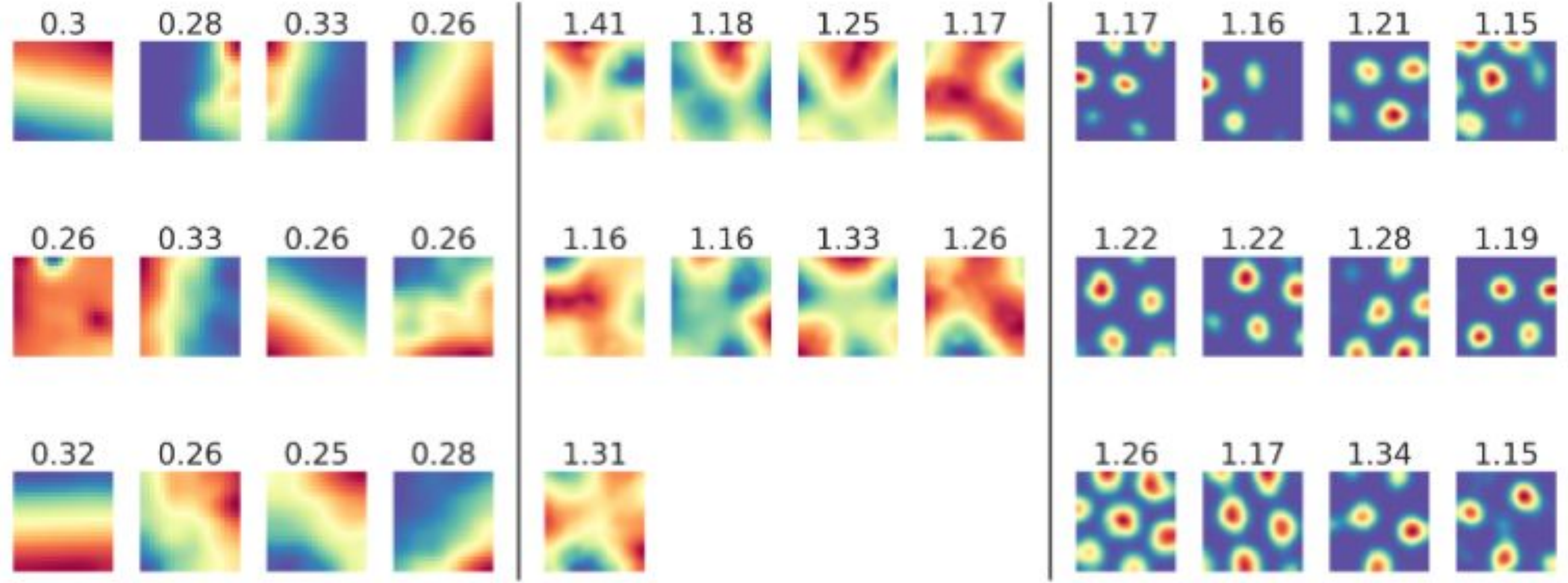


Result 1: Most hyperparameter configurations solve the task of path integration; however, few learn Grid Cells

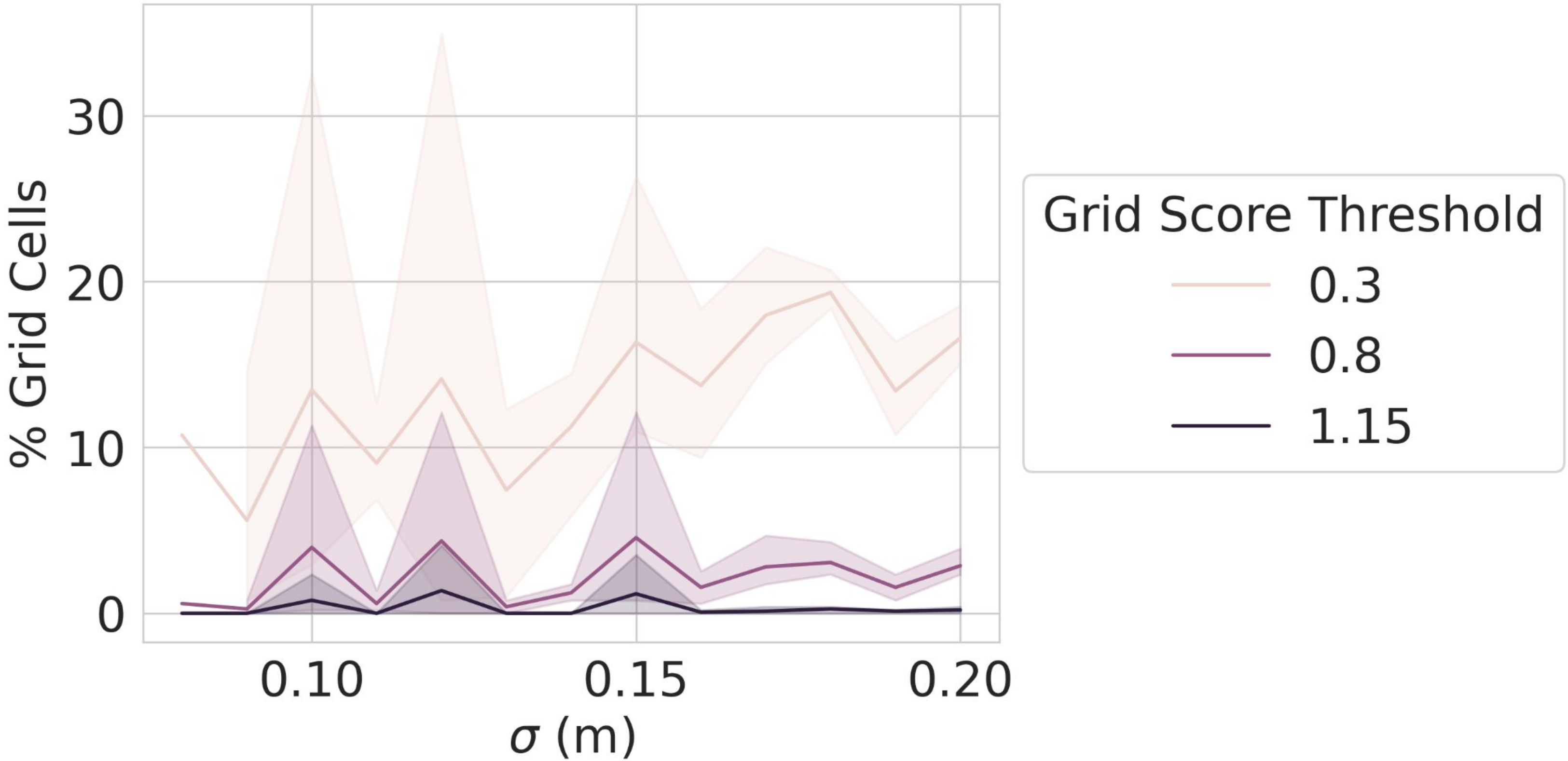


Note: Hyperparameter sweep is biased **towards** grid cell formation

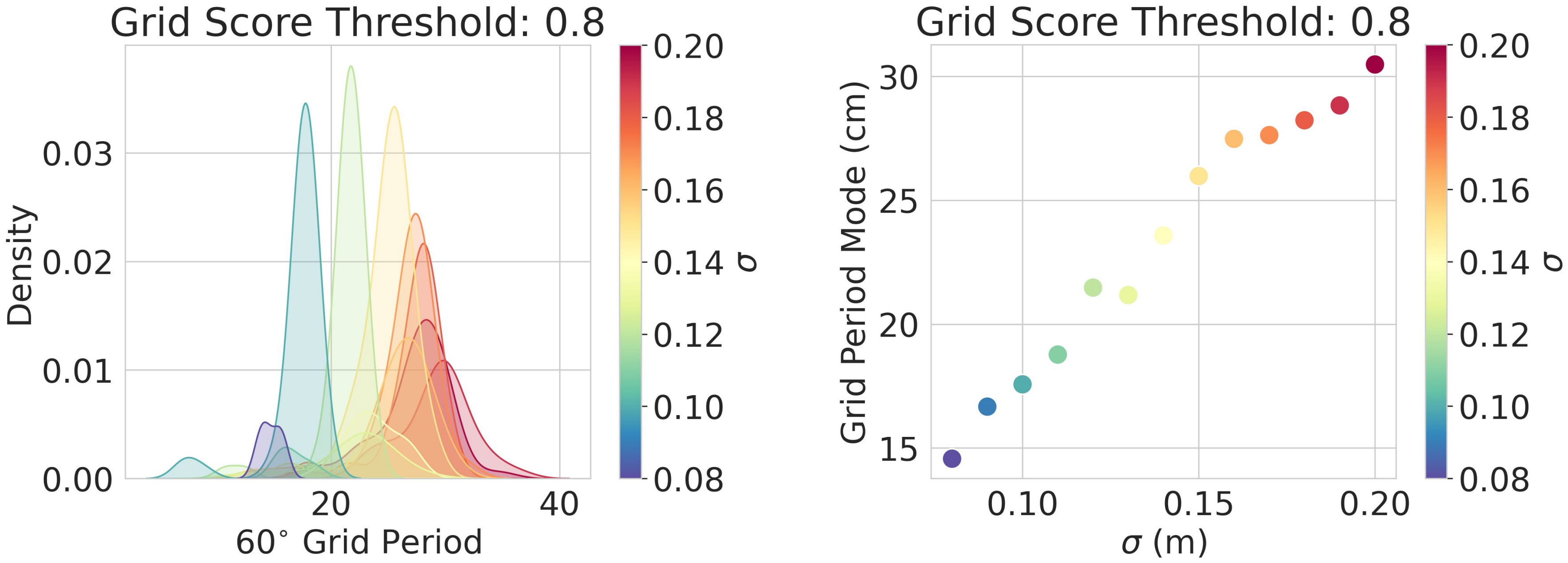
Result 2: Grid Cells only emerge under particular encoding of the (supervised) target i.e. position



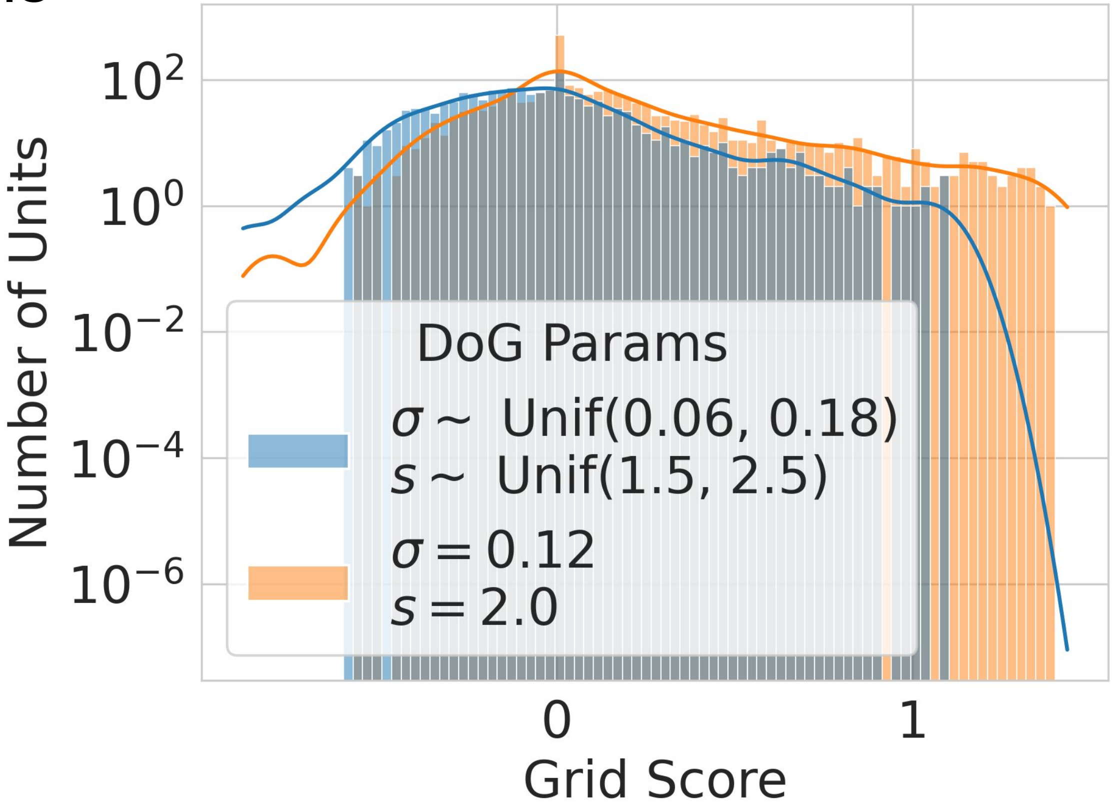
Result 3: Small perturbations to ideal hyperparameters prevent the formation of Grid Cells



Result 4: Unlike brains, path-integrating deep networks do not learn multiple grid modules

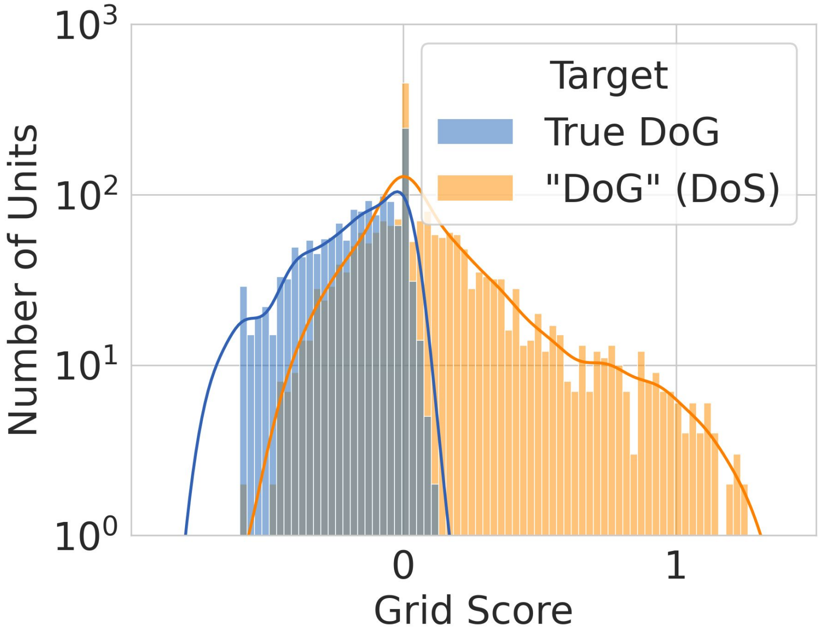


Result 5: Realistic Place Cell targets prevent the formation of Grid Cells

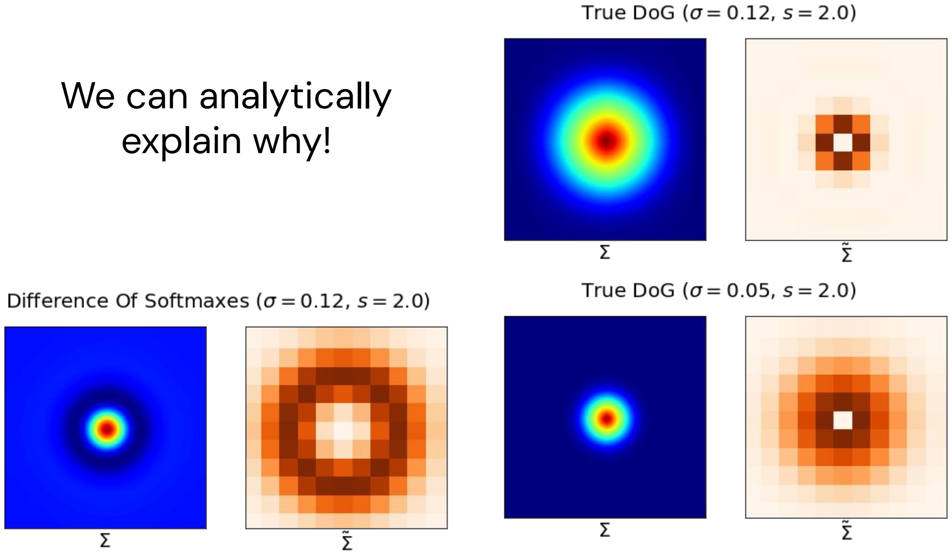


Result 6: Unmentioned implementation details are critical for the formation of Grid Cells

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



We can analytically explain why!



Conclusion: No Free Lunch with Deep Learning Models for Neuroscience

Suppose you want to know what optimization problem the brain is solving, and you have a candidate model that replicates the brain's behavior and neural responses. Because optimization problems can share optima, the brain may solve a different optimization problem than the model.

Even if the brain's optimization problem is correctly identified, a model may learn different optima that yield different behavior or different neural responses.