The Helmholtz Hippocampus: A biologically plausible generative model of the hippocampal formation

Tom George[1], Caswell Barry[1,2], Kimberly Stachenfeld[3,4], Claudia Clopath[1,5], Tomoki Fukai[6]

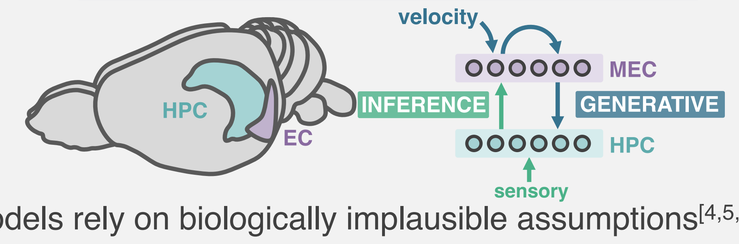
[1]Sainsbury Wellcome Centre, UCL, London [2]Department of Cell & Developmental Biology, UCL, London [3]Columbia University, New York [4]Google DeepMind, London UK [5]Imperial College, London [6]Okinawa Institute of Science and Technology, Japan

1 THE DUAL ROLE OF THE HIPPOCAMPUS

The hippocampal formation (HPC & MEC) has two main roles in navigation:

1. INFER self-location
place cells, grid cells etc.

2. GENERATE trajectories offline
replay[1], memory consolidation[2], planning[3]



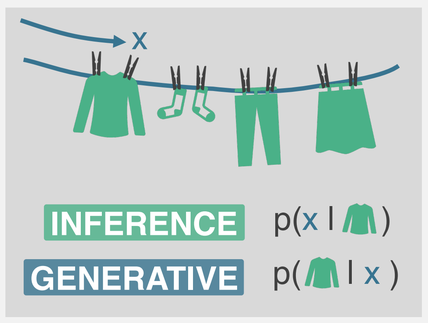
Existing models rely on biologically implausible assumptions[4,5,6].

- Can a **biologically plausible** model account for these **inferential** and **generative** capacities.
- What would the **architecture**, **dynamics** and **learning rules** look like?
- What might it teach us about generative models in the brain?

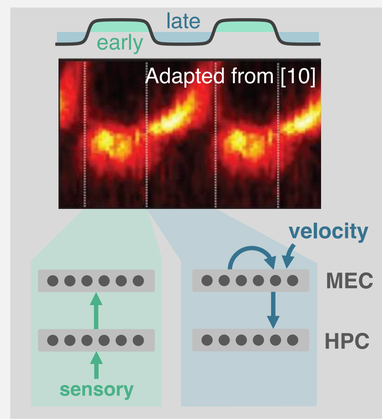
TL;DR?
See box 6

2 BACKGROUND

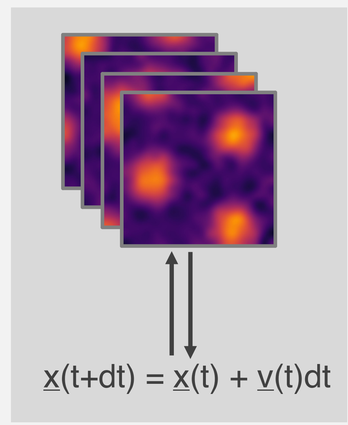
1. MEC learns a sensory-agnostic map of space. HPC binds sensory stimuli onto positions in this space[7,5].



2. During behaviour, **inference** and **generation** occur in distinct **early** and **late** phase of the 5-10 Hz theta rhythm[8,9].

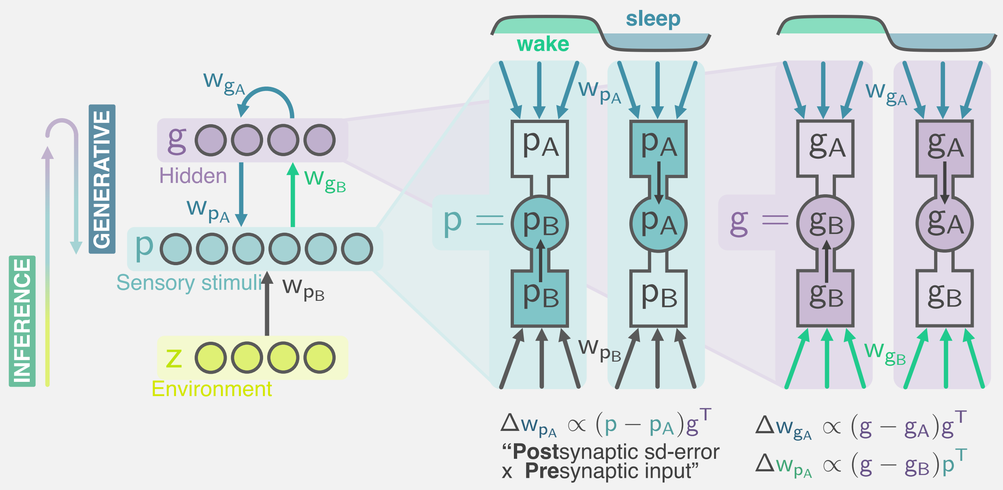


3. Path integration – a "**generative**" process – is a fundamental pillar of MEC function[11,4].



3 THE HELMHOLTZ HIPPOCAMPUS

- A hierarchical network receives a continuous stream of sensory inputs.
- **Inference** and **generative** pathways arrive at distinct **basal** and **apical** dendritic compartments.
- LFP theta-rhythm (5-10 Hz) gates which compartment drives the soma, thus which way information flows.



- **Inference** and **generative sampling** occur in alternating 5-10 Hz **wake-sleep** phases.
- **Hebbian learning:** minimizes the somatic "prediction errors"[12] until $p_B \approx p_A$ and $g_B \approx g_A$

4 THEORETICAL INTERPRETATION

Helmholtz machines[13] (aka Boltzmann machines) learn latent models by matching inference and generative distributions in alternating learning phases.

HPC / sensory MEC / latent

$$p_{inf}(p, g) = p_w(g|p)p(p)$$
$$p_{gen}(p, g) = p_w(p|g)p_w(g)$$
$$\mathcal{L}(w) = D_{KL}(p_{inf}||p_{gen})$$

…our **Hebbian learning rule** is derived from this **loss** function[14].

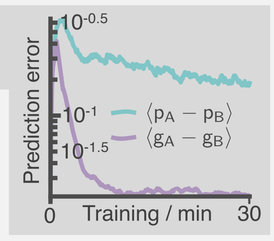
"What I cannot [**generate**] I do not **understand**"

early theta phases ⇔ "wake"
late theta phases ⇔ "sleep"

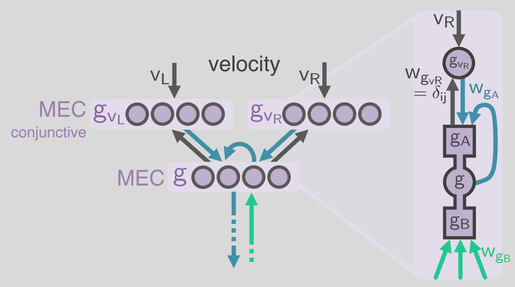
iPad

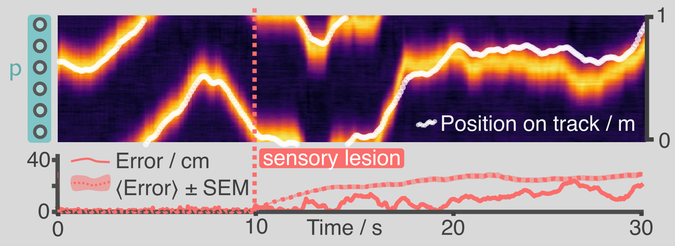
5 RESULTS

1. **COMPRESSION:** HPC learns to efficiently **compress mixed, high-dimensional stimuli** deriving from a small number of independent latents.

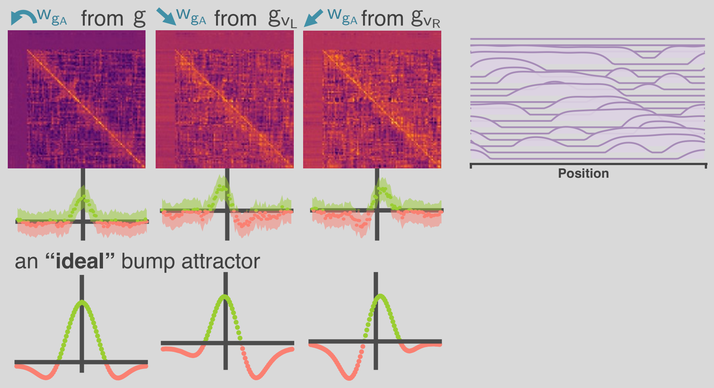


2. **PATH INTEGRATION** is learnt by the model in a 1D spatial navigation task[15]
   i. Initially un-tuned conjunctive inputs provide velocity signal to MEC, HPC receives spatially-selective "place cell" inputs.
   ii. When sensory input is lesioned, top-down generative input from MEC accurately maintains position estimate by "integrating" its velocity inputs.

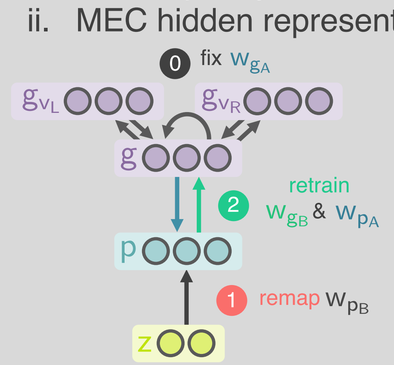


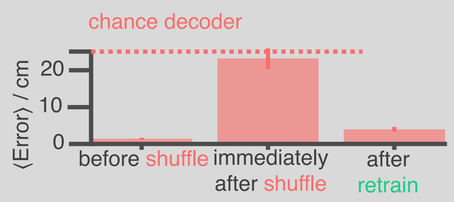


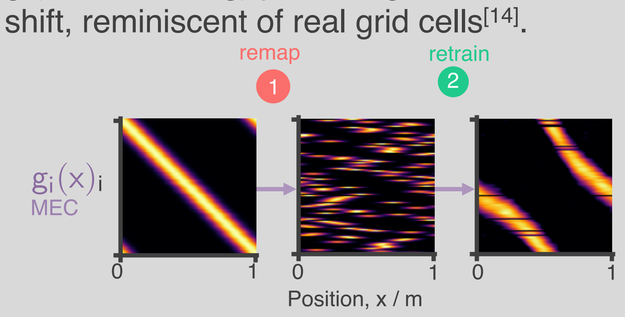
3. **RING-ATTRACTOR:** Weights in the hidden layer reveal a stable calibrated ring-attractor. This theoretically justified[16] solution is hard to achieve with local learning rules[17,18].



4. **REMAPPING AND TRANSFER LEARNING:** We tested "remapping" by **fixing the MEC weights** and **shuffling sensory input**
   i. Model quickly **remastered new environment** by transferring (not relearning) path integration .
   ii. MEC hidden representations reformed with constant phase shift, reminiscent of real grid cells[14].







6 CONCLUSIONS

- Hippocampal **architecture**, **dynamics** and **learning rules** are highly similar to those of a **Helmholtz machine**. HPC maps sensory input to MEC which learns hidden latent structure.
- **LFP Theta oscillations** rapidly gate information flow to generate **"wake-sleep"** learning phases with exclusively **local learning rules**.
- Various navigational function, including self-location, path-integration and transfer learning, can thence be explained **without backpropagation.**

NeurIPS 2023 paper:

[1] Lee & Wilson (2002), Memory of sequential experience in the hippocampus during slow wave sleep **[2]** Carr et al. (2011), Hippocampal replay in the awake state: a potential substrate for memory consolidation and retrieval **[3] Spiers and Maguire (2006)**, Thoughts, behaviour, and brain dynamics during navigation in the real world **[4] Banino et al. (2018)**, Vector-based navigation using grid-like representations in artificial agents **[5] Whittington et al. (2020)**, The Tolman-Eichenbaum Machine: Unifying Space and Relational Memory through Generalization in the Hippocampal Formation **[6] Sorscher et al. (2022)**, A unified theory for the computational and mechanistic origins of grid cells **[7] Bellmund et al. (2018)**, Navigating cognition: Spatial codes for human thinking **[8] Hasselmo et al. (2002)**, A Proposed Function for Hippocampal Theta Rhythm: Separate Phases of Encoding and Retrieval Enhance Reversal of Prior Learning **[9] Sanders and Lisman (2015)**, Grid Cells and Place Cells: An Integrated View of their Navigational and Memory Function **[10] Wang et al. (2020)**, Alternating sequences of future and past behavior encoded within hippocampal theta oscillations **[11] Dorrell et al. (2023)**, Actionable Neural Representations: Grid Cells from Minimal Constraints **[12] Urbanczik and Senn (2014)**, Learning by the Dendritic Prediction of Somatic Spiking **[13] Dayan et al. (1995)**, The Helmholtz Machine **[14] Bredenberg et al. (2021)**, Impression learning: Online representation learning with synaptic plasticity **[15] George et al. (2024)**, RatInABox, a toolkit for modelling locomotion and neuronal activity in continuous environments **[16] Zhang (1996)**, Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory **[17] Burak and Fiete (2009)**, Accurate path integration in continuous attractor network models of grid cells **[18] Vafidis et al. (2022)**, Learning accurate path integration in ring attractor models of the head direction system **[19] Fyhn et al. (2007),** Hippocampal remapping and grid realignment in entorhinal cortex.

Sainsbury Wellcome Centre
tom.george.20@ucl.ac.uk