Modelling hippocampal and striatal contributions to reward-based navigation

Jesse P. Geerts[1,2], Fabian Chersi[2], Kimberly L. Stachenfeld[3], Neil Burgess[2]
[1]Sainsbury Wellcome Centre for Neural Circuits and Behaviour, University College London, 25 Howland Street, London W1T 4JG, UK
[2]Institute for Cognitive Neuroscience, University College London, 17 Queen Square, London WC1N 3AZ, UK
[3]DeepMind, 6 Pancras Square, London N1C 4AG, UK

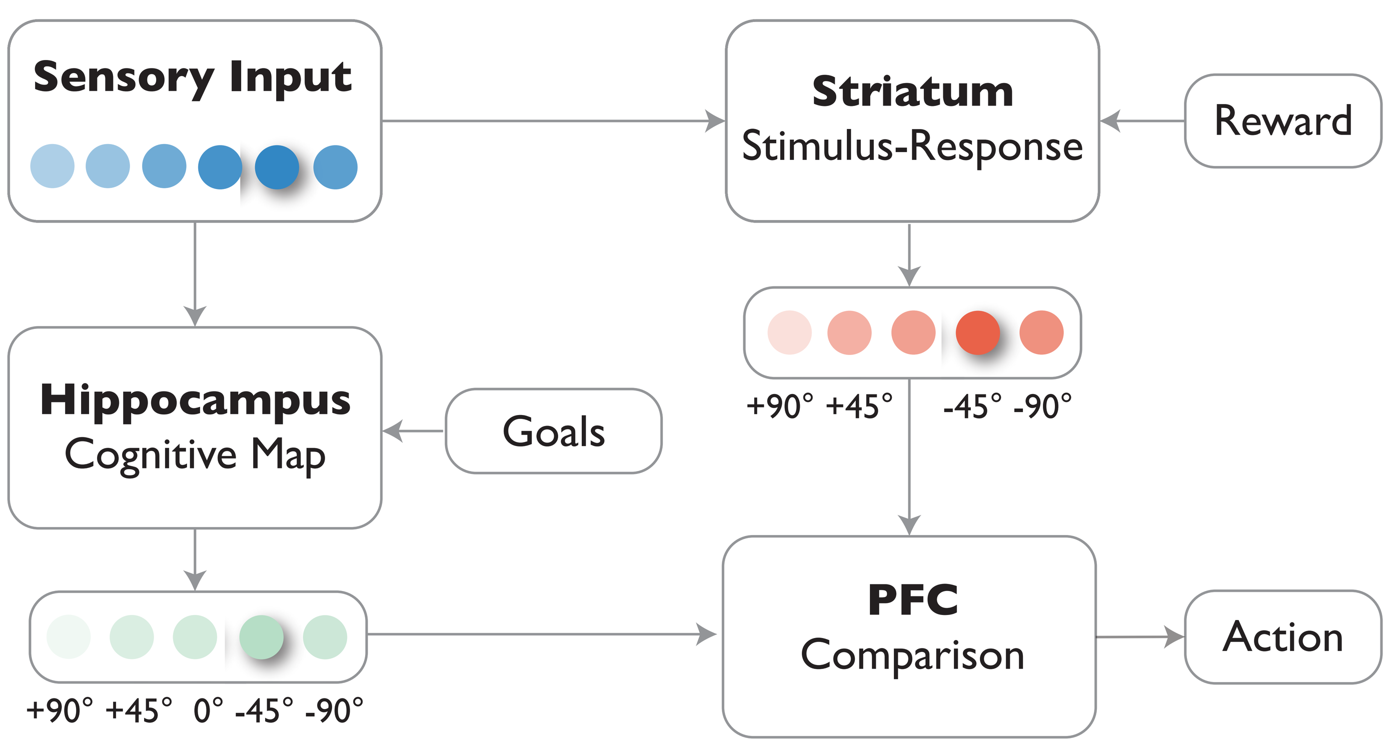
UCL

I. Introduction

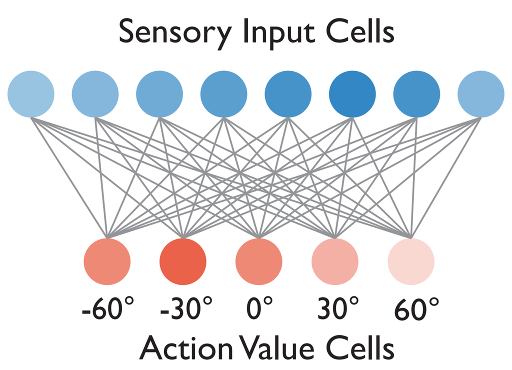
- There are many ways to find a goal location, and animals have been shown to use distinct strategies (Chersi & Burgess, 2015).
- One strategy, called **response learning**, involves executing a learnt sequence of actions, depending on *current sensory cues and past actions*. Another strategy, which we call **place learning**, uses a *a cognitive map*
- Previous studies in rodents (Packard & McGaugh, 1996; Pearce et al., 1998) and humans (Doeller & Burgess, 2008) have shown that these strategies depend on different brain areas. While the striatum underlies response learning, place learning is supported by the hippocampus.
- An open question that remains is when animals choose for a place strategy versus a response strategy
- Here, we introduce a model that aims to capture these effects. Our model consists of a striatum learning stimulus-response associations using model-free RL, a hippocampus that uses a Hebbian learning rule to learn the weights to a goal cell, and a model medial prefrontal cortex (PFC) that arbitrates between these two
- We use our model to simulate data from a set of experiments probing response learning and place learning in the Morris Water Maze and the Plus Maze

II. Model Architecture

i. Overview



ii. Striatal system underlying stimulus-response learning



- Transformed sensory inputs indicate the relative angle to the landmarks (see also Dollé et al., 2017)
- These connect to action value neurons coding for the value of each egocentric heading direction

$$Q_{s,a} = v_a^{striatum} = \phi\left[\sum_{i=1}^{N} v_i^{sensory} w_{i,a}\right]$$

- These value tracking neurons allow us to compute the temporal difference (TD) prediction error δ

$$\delta_t = r_{t+1} + \gamma \max_{a'} Q(s_{t+1}, a') - Q_{s_t,a_t}$$

- The weights between the sensory and value neurons are updated using this prediction error:

$$\Delta w_{i,a} \propto \alpha \delta_t e_{i,a}$$

- Here, α is the learning rate and $e_{i,a}$ is the eligibility trace of the weight. The trace is updated as follows, with trace decay parameter λ:

$$e_{i,a}(t+1) = v_i^{sensory} v_j^{striatum} + \lambda e_{i,a}(t)$$

iii. Hippocampal system underlying incidental learning

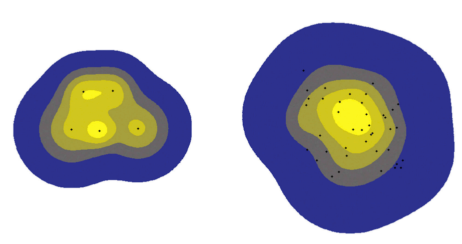
- The hippocampus was modelled as a set of place cells with Gaussian receptive fields, and a goal cell (Gauthier & Tank, 2017):

$$v^G = \sum_{i=1}^{N} z_i v_i$$

- Weights between them are learned using one-shot Hebbian learning when the goal is reached, with learning rate η:

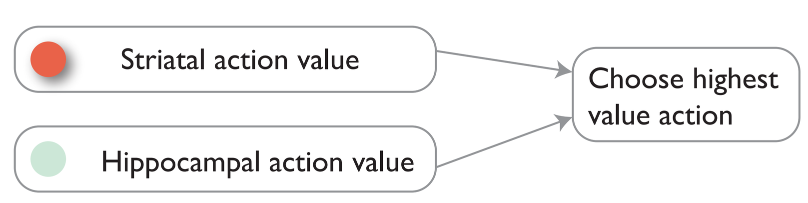
$$\Delta z_i = \eta v_i^{PC} v^G$$

- The goal cell firing rate map constitutes a global value function that can be used to navigate to the goal, when maximising its slope at each time step (Chersi & Burgess, 2015)



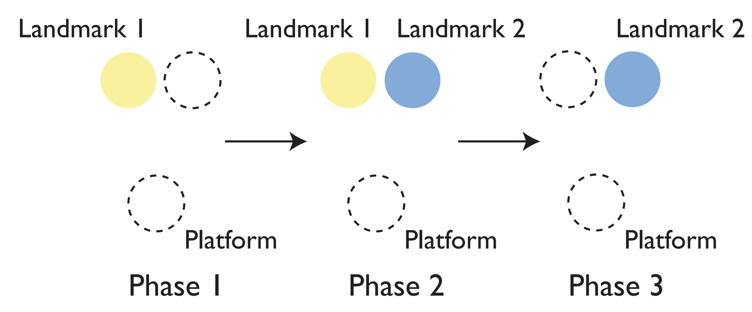
iv. Prefrontal cortex selects action with highest value

- Both the striatal and hippocampal systems result in a proposed action, the values of which are compared to make a final choice

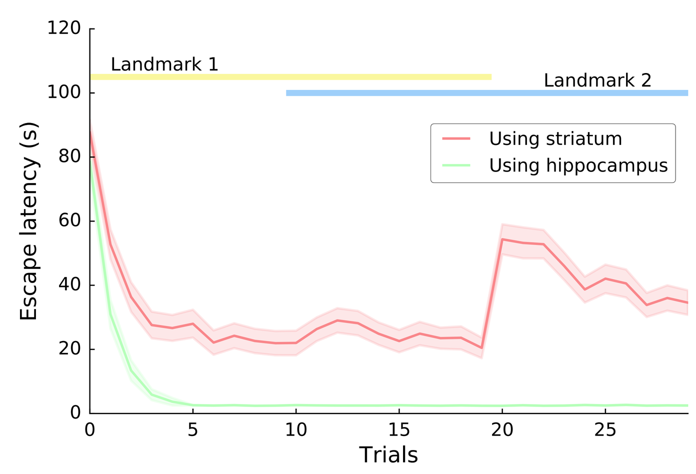


III. Striatum but not hippocampus is sensitive to spatial blocking

- We tested spatial blocking (Rescorla & Wagner, 1972) in the water maze: agents learnt to navigate to a hidden platform close to an intra-maze landmark
- Consistently with experimental data in humans (Doeller & Burgess, 2008), learning about one landmark blocked learning about a second landmark for agents using the striatal system based on prediction error learning

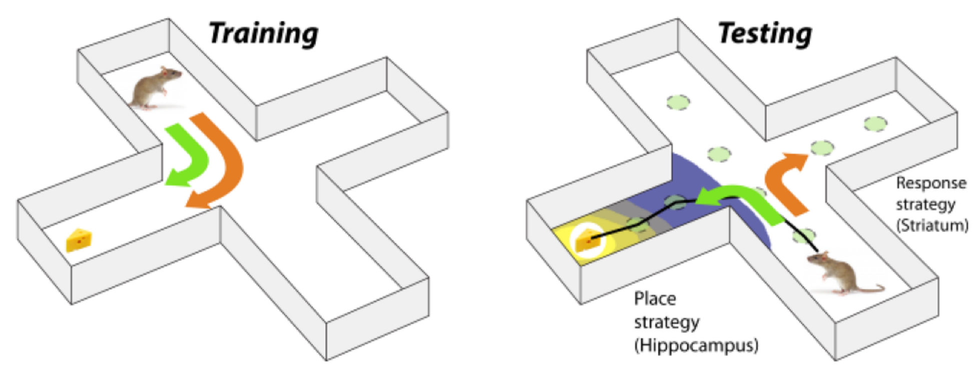


- In contrast, agents using the hippocampal system to navigate did not show the blocking effect

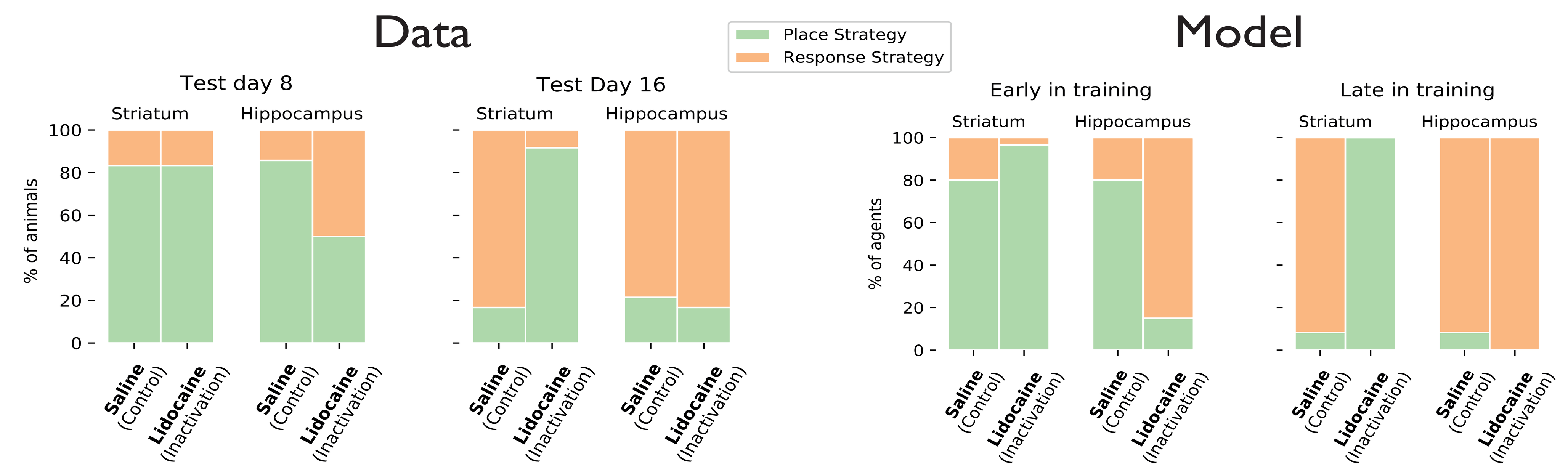


IV. Animals switch to response strategy on the Plus Maze

- Animals were trained to navigate on the Plus Maze (Packard & McGaugh, 1996)
- During training, animals learnt to approach a consistently baited goal arm, always starting from the same start box
- On day 8 (early) and day 16 (late) animals performed a probe trial, starting from the opposite start box



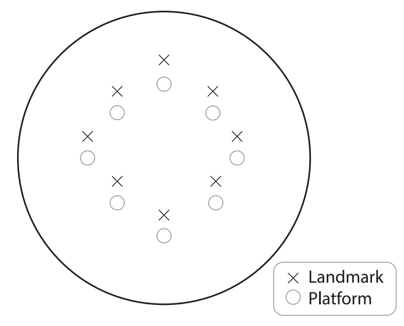
- A **place strategy** was defined as going to the place were the food was during training. A **response strategy** was defined as making the same turning response as during training
- We modelled lidocaine inactivation as turning off the striatal and hippocampal parts of the model, respectively

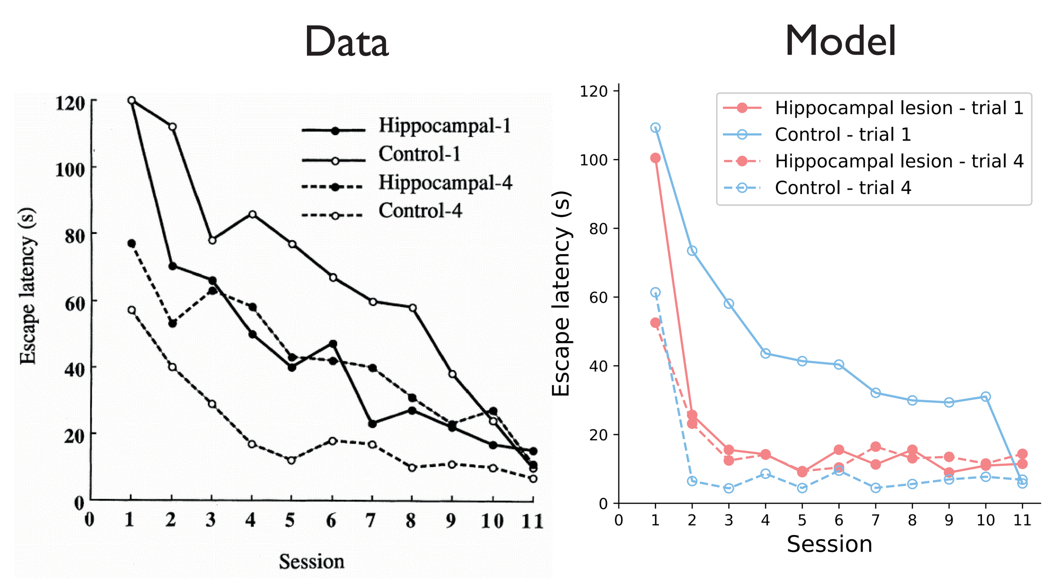


- Early in training, most animals follow a place strategy, which can be partially reversed by inactivating hippocampus
- Later in training, animals switch to a response strategy. However, this can be reverted by inactivating the striatum
- Our model captures both these effects: inactivating hippocampus
- Conversely, inactivating the striatum caused a switch towards place strategies

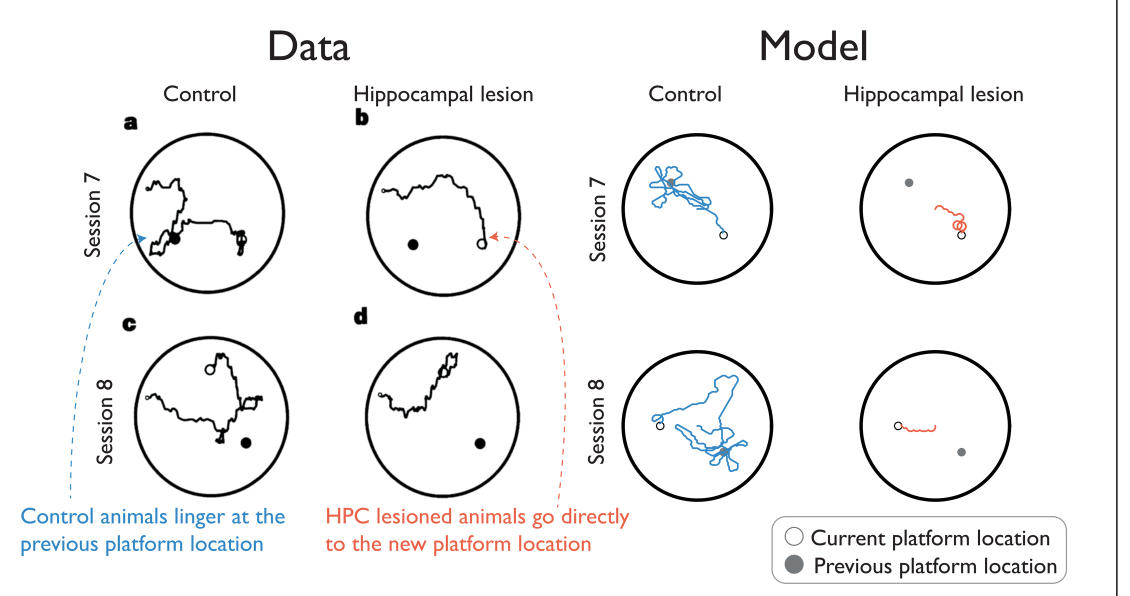
V. Effects of hippocampal lesions on water maze performance

- Pearce et al. (1998) trained animals to navigate in a water maze with intra-maze landmarks. The landmark was always 20 cm north of the platform, but the landmark and platform pair were moved each session to one of 8 different locations
- Hippocampal lesions impair within-session learning, but over sessions the task is still learnt
- Crucially, animals with hippocampal lesions performed better than control animals on the first trial after the platform and landmark moved





Escape latency in the water maze for hippocampal lesioned and control animals (left) and agents (right) on trial 1 (solid lines) and 4 (dashed line) of each session.



Example trajectories from the first trials of sessions 7 and 8. Animals and agents using a hippocampal strategy tend to wander around the previous platform location

VI. Conclusions and Directions

- We simulated hippocampal and striatal contributions to spatial learning in the Morris Water Maze and the Plus Maze, using a model relying on model-free RL (striatum) and Hebbian learning (hippocampus)
- Using this model, we were able to explain spatial blocking (Doeller & Burgess, 2008), a gradual switch to response strategies (Packard & McGaugh, 1996) and the effects of hippocampal lesions in a water maze with changing reward locations (Pearce et al., 1998)
- Our framework is not limited to the spatial domain, as RL can operate on any Markovian state representation, and hippocampus has been shown to represent non-spatial variables (Aronov et al, 2017). In the near future we will apply our model to non-spatial learning tasks that probe model-based RL (Daw et al., 2011; Doll et al., 2015), which has been shown to involve hippocampus (Miller et al., 2017)

D. Aronov, R. Nevers, D. W. Tank, Nature. 543, 719–722 (2017).
F. Chersi, N. Burgess, Neuron. 88, 64–77 (2015).
1. N. D. Daw, S. J. Gershman, B. Seymour, P. Dayan, R. J. Dolan, Neuron. 69, 1204–1215 (2011).
1. C. F. Doeller, J. A. King, N. Burgess, 105, 5915–5920 (2008).
1. B. B. Doll, K. D. Duncan, D. A. Simon, D. Shohamy, N. D. Daw, Nat. Neurosci. 18, 767–772 (2015).
1. L. Dollé, R. Chavarriaga, A. Guillot, M. Khamassi, PLoS Comput Biol (2018).
1. K. J. Miller, M. M. Botvinick, C. D. Brody, Nat. Neurosci. 20, 1269–1276 (2017).
1. M. G. Packard, J. L. McGaugh, Neurobiol. Learn. Mem. 72, 65–72 (1996).
1. J. M. Pearce, A. D. L. Roberts, M. Good, Nature. 62, 1997–1999 (1998).
1. R. A. Rescorla, A. R. Wagner, Class. Cond. II Curr. Res. theory. 2, 64–99 (1972).

GATSBY
wellcome trust
Sainsbury Wellcome Centre