Hippocampus Is What Happens while You're Busy Making Other Plans

Kiyohito Iigaya[1,*] and John P. O'Doherty[1,*]
[1]Division of the Humanities and Social Sciences, California Institute of Technology, 1200 E. California Blvd., Pasadena, CA 91125, USA
*Correspondence: kiigaya@caltech.edu (K.I.), jdoherty@caltech.edu (J.P.O.)
https://doi.org/10.1016/j.neuron.2019.04.027

In this issue of *Neuron*, Vikbladh et al. (2019) provide evidence to suggest that the human hippocampus, long known to support spatial memory, also plays a causal role in model-based planning.

Think of your commute in the morning. If it is to the same place that you worked for the past 20 years, your commute has become habitual and you are less likely to plan your exact path to your workplace before you leave. However, if on one day your route is unexpectedly blocked by roadwork, you may need to plan more explicitly about which alternative route to take.

These two decision-making strategies are captured in machine-learning algorithms called model-free and model-based reinforcement learning (Daw et al., 2011). A model-free agent makes decisions without explicit prospection, solely relying on the overall success rates of past actions. In contrast, a model-based agent envisions how to get to a desired goal using an internal model of the world that is learned and stored in the agent's memory (model-based planning). Previous studies have suggested that humans utilize both strategies (Daw et al., 2011).

We currently have a very limited understanding of how the human brain performs model-based planning, though there are reasonable grounds to suspect that the hippocampus might be involved. In rodents, sequential activations of hippocampal place cells occur when animals pause and look about while navigating a maze as if they were planning, a behavior called vicarious trial and error (Redish, 2016). A recent muscimol inactivation study showed the rodent hippocampus as a causal neural substrate for planning (Miller et al., 2017). However, in humans, we lack causal evidence about the contribution of hippocampus to model-based planning.

In this issue of *Neuron*, Vikbladh et al. (2019) addressed the lack of evidence in humans by administering two behavioral tasks to patients who had undergone unilateral anterior temporal lobectomy (ATL) for the treatment of intractable epilepsy. These patients are missing a part of their hippocampi on one side, as well as adjacent medial temporal lobe structures.

In the first task, Vikbladh et al. (2019) tested whether the patients were impaired in model-based planning using a well-studied behavioral task (Daw et al., 2011). On each trial, participants chose from one of two spaceships, each of which had a unique probability to take the participants to one of the two possible planets. Upon arriving on a planet, participants were then asked to choose one alien from the two that were always present on each planet. The chosen alien would give participants a reward or no reward, whose probability changed slowly over time.

Vikbladh et al. (2019) used computational modeling to dissect participants' behavior, as model-based and model-free computational strategies are known to predict distinct behaviors in this task (Daw et al., 2011). A model-free agent makes decisions based on the overall reward rate of each spaceship. Such an agent is more likely to simply choose a previously rewarded spaceship, ignoring subsequent planets and aliens. In contrast, a model-based agent aims to reach the alien that is currently giving out the most rewards and would thus choose a spaceship that is most likely to take them to the desired planet on which that alien resides. Consequently, a model-based agent does not necessarily repeat the exact same choice on the next trial even if it was rewarded on the previous one (Daw et al., 2011).

Consistent with previous studies (Daw et al., 2011), participants with intact hippocampi showed a mixture of model-based and model-free strategies in their choices (Vikbladh et al., 2019). This was also the case in the ATL patients; however, those patients relied more heavily on a model-free strategy and less on a model-based one unlike the comparison group (Vikbladh et al., 2019). Further analyses suggested that the effect was mainly driven by those patients who had right hippocampal lesions, as opposed to those with left hippocampal lesions. In fact, the size of the right hippocampal lesion across patients was negatively correlated with an individual's tendency to use model-based planning overall.

Vikbladh et al. (2019) also showed, using a virtual spatial navigation task, that boundary-based place memory is significantly impaired in the ATL patients, replicating the results of previous hippocampal lesion studies (e.g., Guderian et al., 2015). They then compared task behavior between this spatial memory task and the spaceship choice task. Individuals with intact hippocampi showed a positive correlation between the spatial memory and the tendency to engage in model-based planning, while ATL patients did not (Vikbladh et al., 2019). This suggests that the hippocampus might serve as a common substrate for both spatial memory and model-based planning (Figure 1).

One caveat of the findings, however, is that it is challenging to localize the impairment to the hippocampus per se, because the ATL lesion encompasses multiple structures including parahippocampal cortex and amygdala. Vikbladh et al. (2019) did control for overall lesion

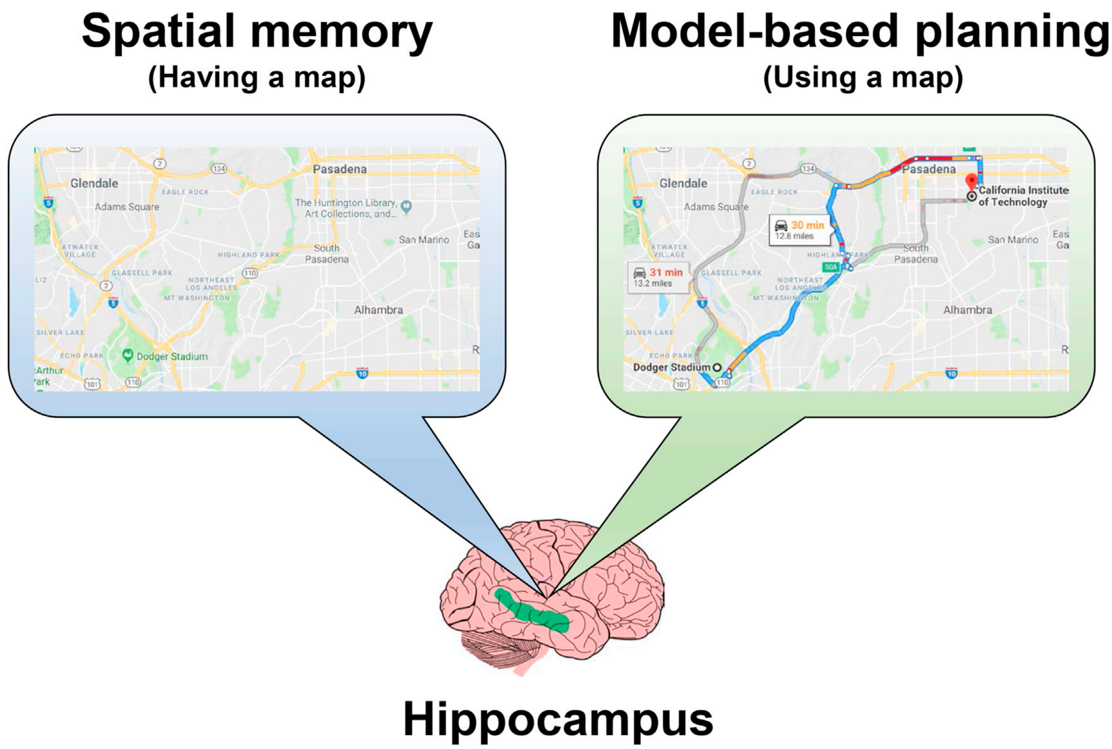

Figure 1. Schematic Illustration of the Findings in Vikbladh et al.
The hippocampus, a subcortical structure in the medial temporal lobe, plays a central role in both spatial memory (having a map) and model-based planning (using the map). Map data: © 2019 Google.

size while still showing a correlation with hippocampal lesion extent, thereby partly mitigating this concern. Patients with epilepsy may also have abnormal tissue in the ATL that could complicate generalization of these conclusions to healthy brains. In spite of these limitations, human lesion studies, such as the present one, are highly valuable because they can illuminate causal roles for human deep brain structures in neural computations in the absence of the precise causal manipulations available to non-human animal researchers.

One open question is whether the present findings can generalize to a context that is not spatial. The cover story of spaceships and planets might have encouraged participants to utilize a spatial planning strategy (e.g., by thinking about which planet to visit, though it is also debatable how the hippocampus plays a role in this). However, not all plans need to be spatial. In fact, a previous study showed that animals with hippocampal lesions can show intact, goal-directed (presumably also model-based) performance involving lever pressing, which likely does not depend strongly on spatial planning (Corbit and Balleine, 2000). Nevertheless, the human hippocampal complex has been suggested to represent even non-spatial information in the form of a cognitive map (Garvert et al., 2017). Thus, the hippocampus could still play a role in model-based planning even in non-spatial situations.

Another key open question is how the neurons in hippocampus perform model-based planning. First, how does the brain combine a map and a goal? In order to select the best path, one has to locate a goal that has the highest payoff. Studies of spatial navigation suggest that hippocampal-prefrontal cortex coupling may play a key role in the capacity for flexible navigation (Preston and Eichenbaum, 2013). Second, how does the brain learn a map and a goal in the first place? Dopaminergic systems may couple with hippocampus to achieve this end (Shohamy and Wagner, 2008). Lastly, the hippocampus has been suggested to play an essential role in another (third) reinforcement learning strategy called episodic control (Lengyel and Dayan, 2008), whereby decisions are guided by the memories of past episodes instead of a model. Further studies should address whether the hippocampus differentially contributes to multiple reinforcement learning strategies.

In sum, Vikbladh et al. (2019) show that hippocampus may be causally involved in model-based planning in humans. Further work will be needed to shed light on the computational mechanisms by which this process is mediated.

REFERENCES

Corbit, L.H., and Balleine, B.W. (2000). The role of the hippocampus in instrumental conditioning. J. Neurosci. *20*, 4233–4239.

Daw, N.D., Gershman, S.J., Seymour, B., Dayan, P., and Dolan, R.J. (2011). Model-based influences on humans' choices and striatal prediction errors. Neuron *69*, 1204–1215.

Garvert, M.M., Dolan, R.J., and Behrens, T.E. (2017). A map of abstract relational knowledge in the human hippocampal-entorhinal cortex. eLife *6*, e17086.

Guderian, S., Dzieciol, A.M., Gadian, D.G., Jentschke, S., Doeller, C.F., Burgess, N., Mishkin, M., and Vargha-Khadem, F. (2015). Hippocampal volume reduction in humans predicts impaired allocentric spatial memory in virtual-reality navigation. J. Neurosci. *35*, 14123–14131.

Lengyel, M., and Dayan, P. (2008). Hippocampal contributions to control: the third way. Adv. Neural Inf. Process. Syst. *20*, 889–896.

Miller, K.J., Botvinick, M.M., and Brody, C.D. (2017). Dorsal hippocampus contributes to model-based planning. Nat. Neurosci. *20*, 1269–1276.

Preston, A.R., and Eichenbaum, H. (2013). Interplay of hippocampus and prefrontal cortex in memory. Curr. Biol. *23*, R764–R773.

Redish, A.D. (2016). Vicarious trial and error. Nat. Rev. Neurosci. *17*, 147–159.

Shohamy, D., and Wagner, A.D. (2008). Integrating memories in the human brain: hippocampal-midbrain encoding of overlapping events. Neuron *60*, 378–389.

Vikbladh, O.M., Meager, M.R., King, J., Blackmon, K., Devinsky, O., Shohamy, D., Burgess, N., and Daw, N.D. (2019). Hippocampal contributions to model-based planning and spatial memory. Neuron *102*, this issue, 683–693.