Predictive representations in hippocampal and prefrontal hierarchies

Katja Schach, Paula Verde Puerto & Franziska Gekeler

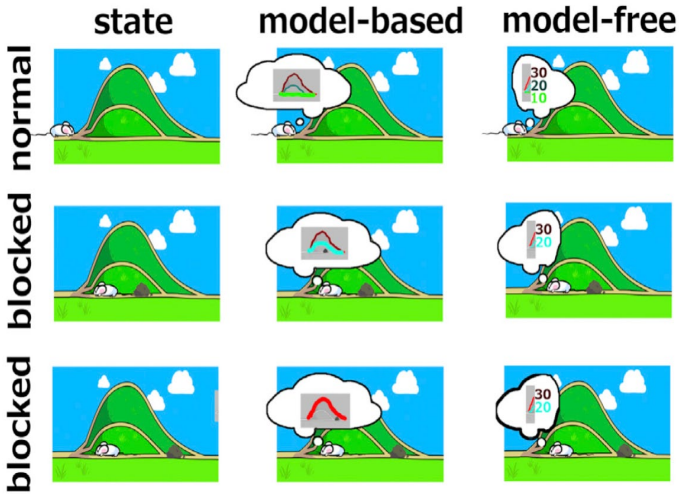
Index

Reinforcement Learning (RL) and Successor Representation (SR)

Limitations for RL

Model-based → too much computational effort

Model-free → not flexible enough



(Dolan & Dayan, 2013. https://doi.org/10.1016/j.neuron.2013.09.007)

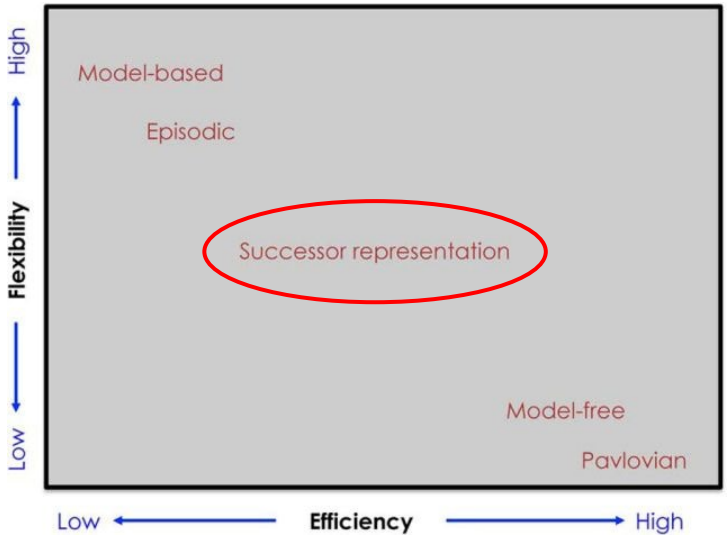
Reinforcement Learning (RL) and Successor Representation (SR)

Limitations for RL

Model-based → too much computational effort

Model-free → not flexible enough

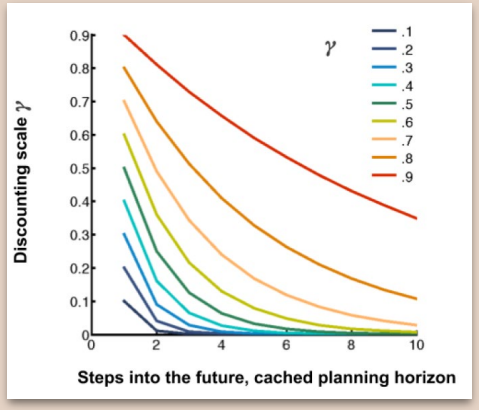
Successor Representation



(Gershman., 2018. https://doi.org/10.1523/JNEUROSCI.0151-18.2018)

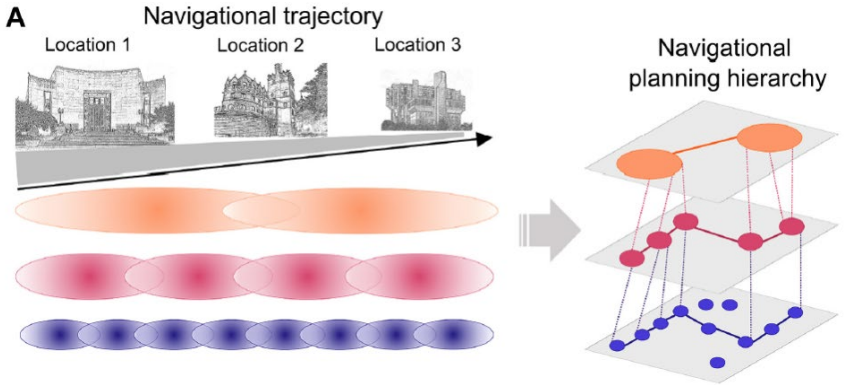
Multi-scale SR

Simultaneous caching of several SR with different γ values

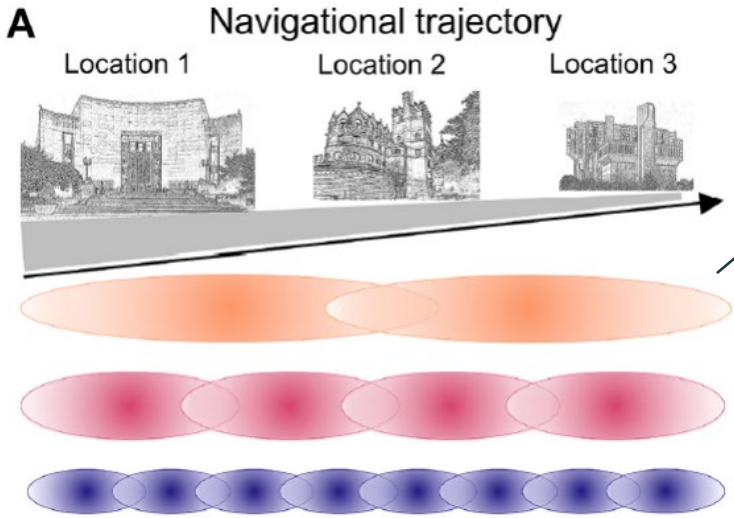


↑ γ → ↑ predictive horizon → ↑ representation

↓ γ → ↓ predictive horizon → ↓ representation



Example: let's go to Stuttgart!



high γ

low γ

Go to Stuttgart

go to the hbf

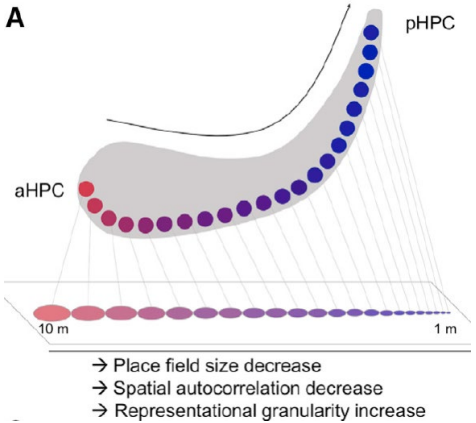
take the TüBus

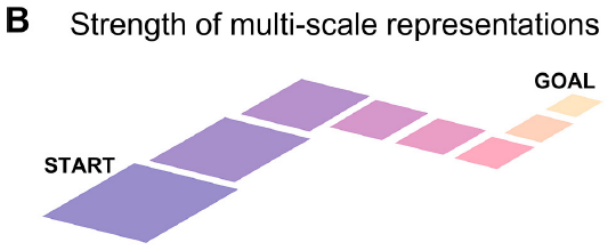
Multi-scale SR

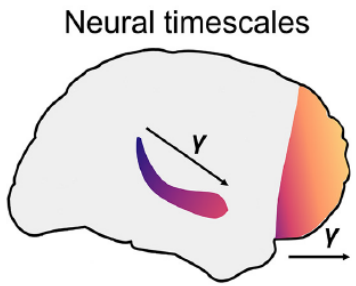
Prefrontal cortex (PFC)

Hippocampus

Multi-scale SR







First hypothesis

Anterior HPC → Longer predictive horizons

Posterior HPC → smaller predictive horizons

Second hypothesis

Anterior PFC → representation of more distant states than posterior PFC

Third hypothesis

antPFC representations

>

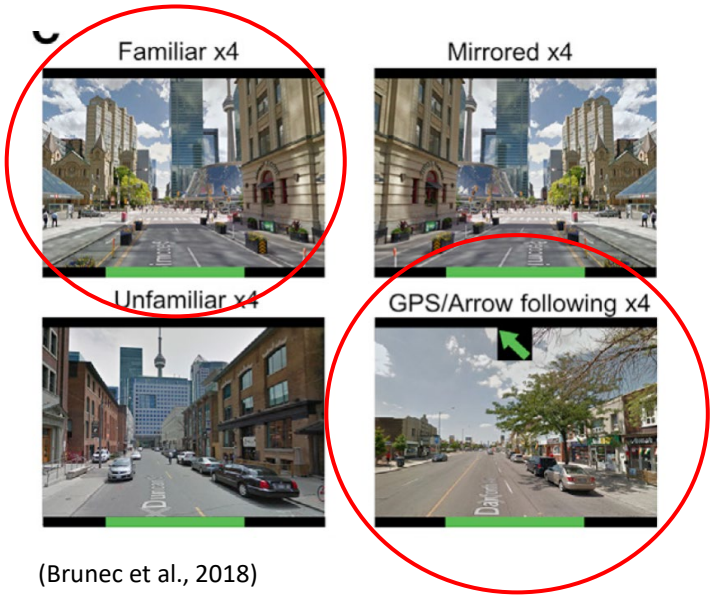
hippocampal representations

Experiment conditions

Multiple Scales of Representation along the Hippocampal Anteroposterior Axis in Humans

Iva K Brunec [1], Buddhika Bellana [2], Jason D Ozubko [3], Vincent Man [4], Jessica Robin [5], Zhong-Xu Liu [5], Cheryl Grady [6], R Shayna Rosenbaum [7], Gordon Winocur [8], Morgan D Barense [9], Morris Moscovitch [10]

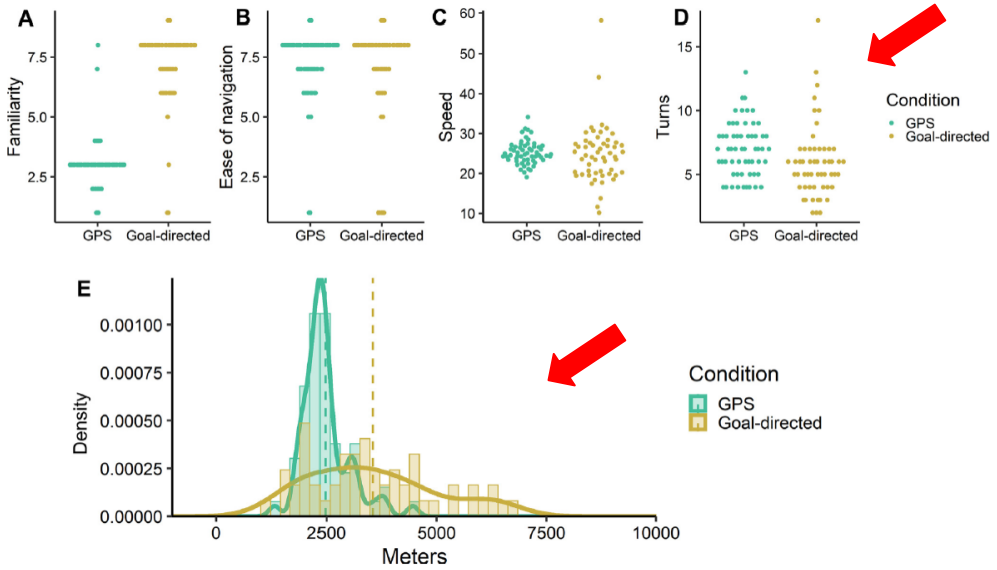
- Navigation of a virtual version of Toronto while undergoing fMRI
- Participants had lived in Toronto for at least 2 years
- 4 conditions: Familiar, Unfamiliar, Mirrored and GPS guided



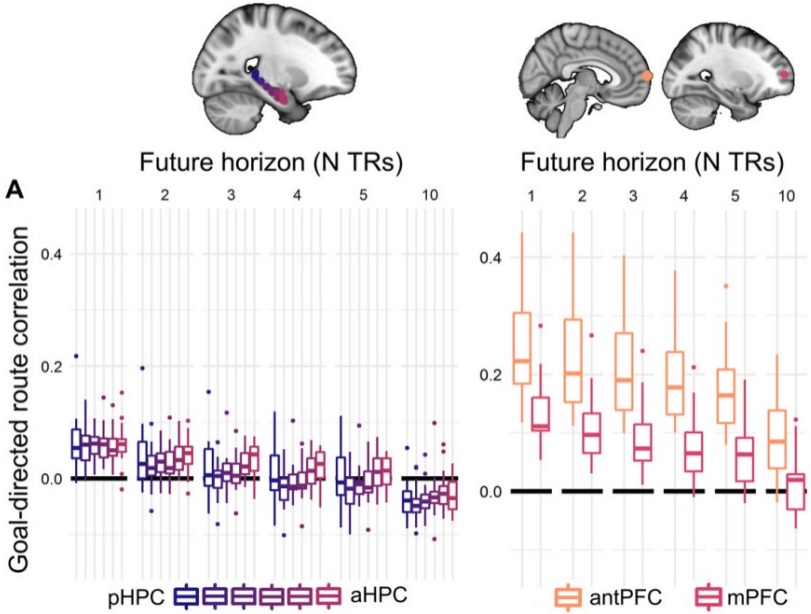
(Brunec et al., 2018)

Experiment conditions

Experiment conditions



Results


A
Future horizon (N TRs)
Future horizon (N TRs)
1
2
3
4
5
10
Goal-directed route correlation
0.4
0.2
0.0
pHPC
aHPC
antPFC
mPFC

Terms of RL and SR used here

TR

- “steps” (spatial and/or temporal)
- time resolution of fMRI scans
- 1 TR = 1 fMRI scan of whole brain volume
  → 0.5 Hz → 2 s → ~ 25 m on average
- TR / fMRI pattern of the TR corresponds to state function

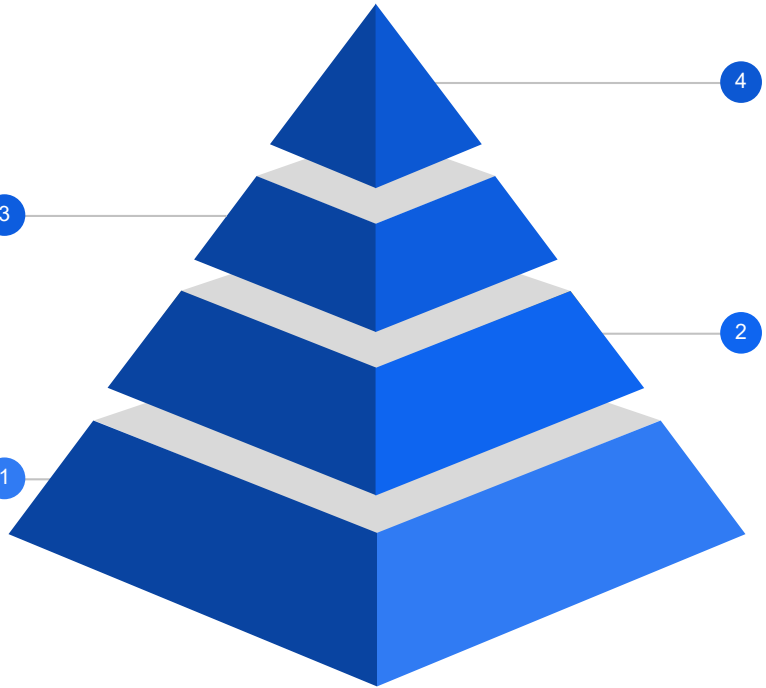
γ

- discount factor, weighs states to value function
  → with higher γ, value of farther in the future state representations decreases less
  → more future representations contribute significantly to the prediction

What corresponds to the **reward function** of the SR model?

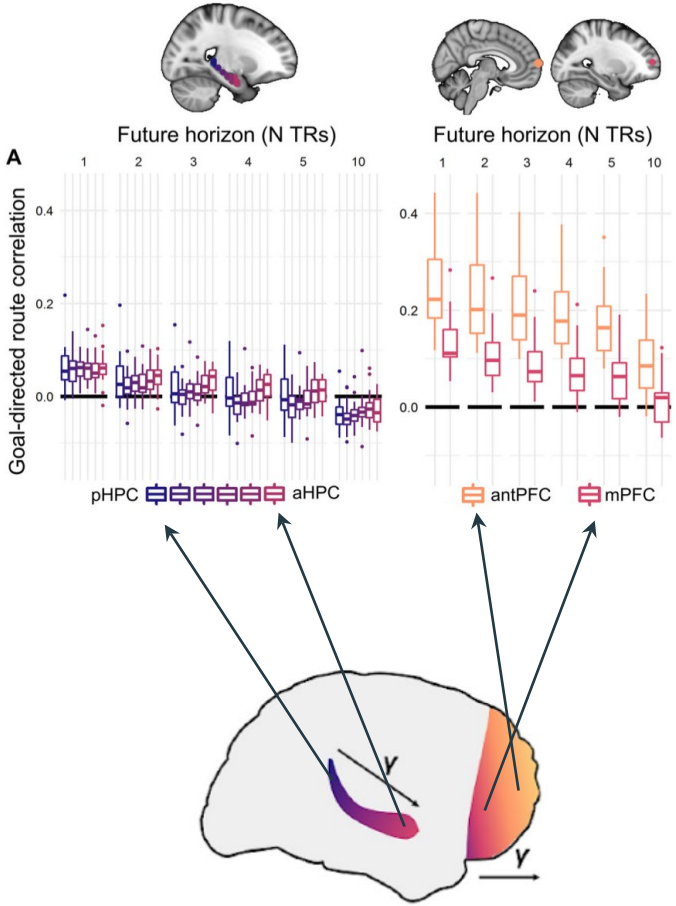
Reaching of the goal and goal directed movement?
Or high evaluation of similarity (=correlation) of current representation to weighted future representations?
Main analysis tool: **Testing for correlation (=similarities) between fMRI voxel patterns at different temporal/spatial points (TRs)**

Progressional analysis:



4. Search light analysis (accounting for brain area volume)
3. Accounting for path distance
2. Multi-scale SR - Correlating TR with γ-weighted sum of future TRs
1. Correlating TR with average of n future TRs.

Similarity between TR and mean of future TRs



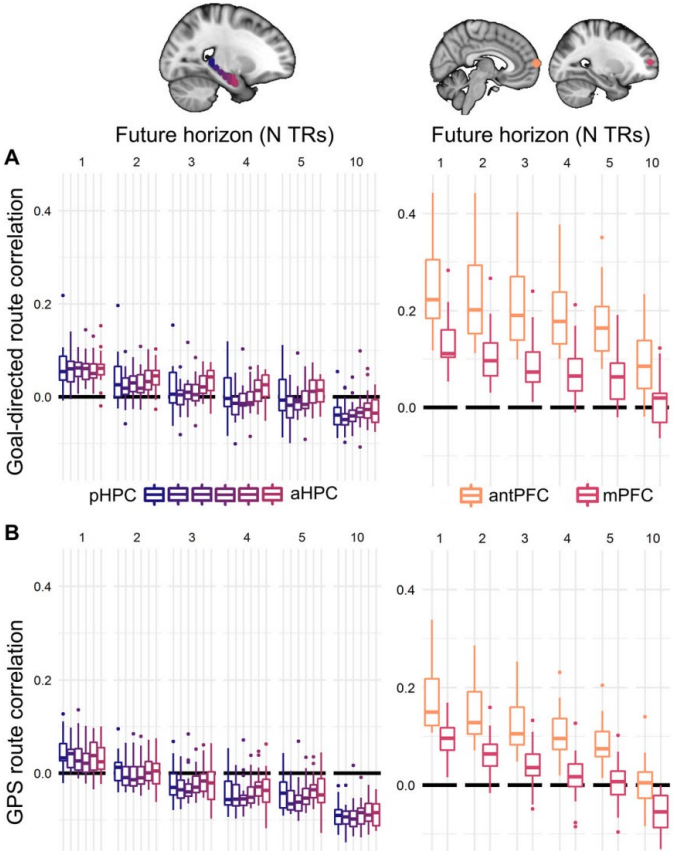
Similarity between TR and mean of future TRs

First analysis:

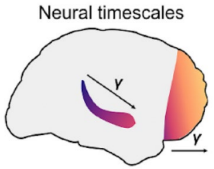
- Linear mixed effects models (dependent variable = similarity; axial segment, number of TRs and hemisphere as fixed effects, participants as random effects)
- Representational similarity between TR (state ~ fMRI voxel pattern) and **unweighted** mean of n future TR

→ larger scale representation have higher similarity in fMRI voxel pattern to future representations and slower changes among states

here: unweighted
→ gives us a basic idea, how similar representations at the TRs are to their “neighboring” TRs

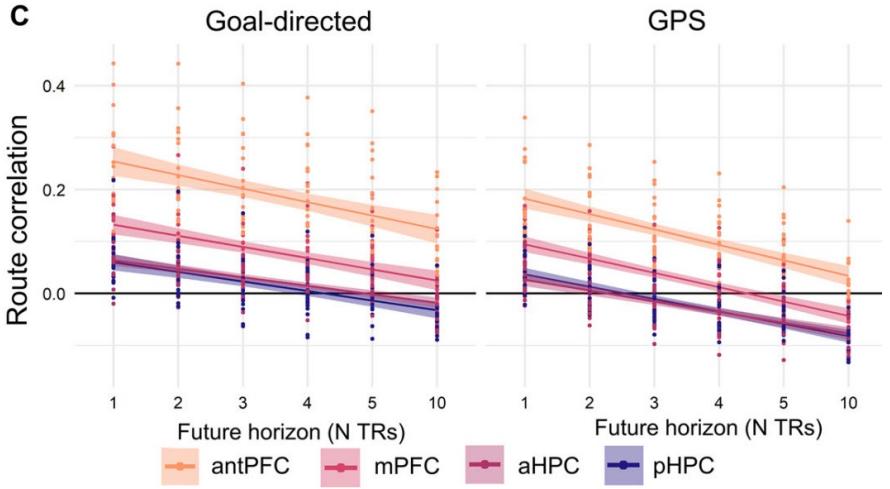


Similarity between TR and mean of future TRs



Comparing representational similarities of ROIs and conditions (goal-directed vs. GPS based) against 0
→ 0 equals no significant correlation
→ only the representations of the anterior prefrontal cortex still correlate to thinking 10 steps (TRs) ahead

→ not very well pronounced difference in the hippocampal gradient, better multiscale representation in following analyses



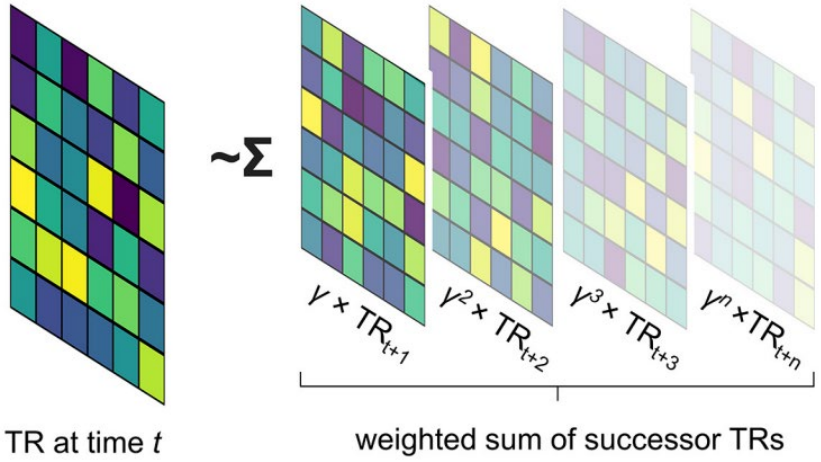
	pHPC	aHPC	mPFC	antPFC
Goal directed [TR]	1	4	5	10
GPS based [TR]	2	1	3	5

Second approach: Multi-scale SR

M = SR matrix
T = Transition matrix
s_t = current TR/state
s_{t1}= future TRs/states
t = number of steps

$$M(s_1, s_g) = \sum_{t=t1}^{g} \gamma^t T(s_{t1}, s_t)$$

$$Rep(s_i) = \sum_{i=t}^{n} \gamma^t Rep(s_t)$$



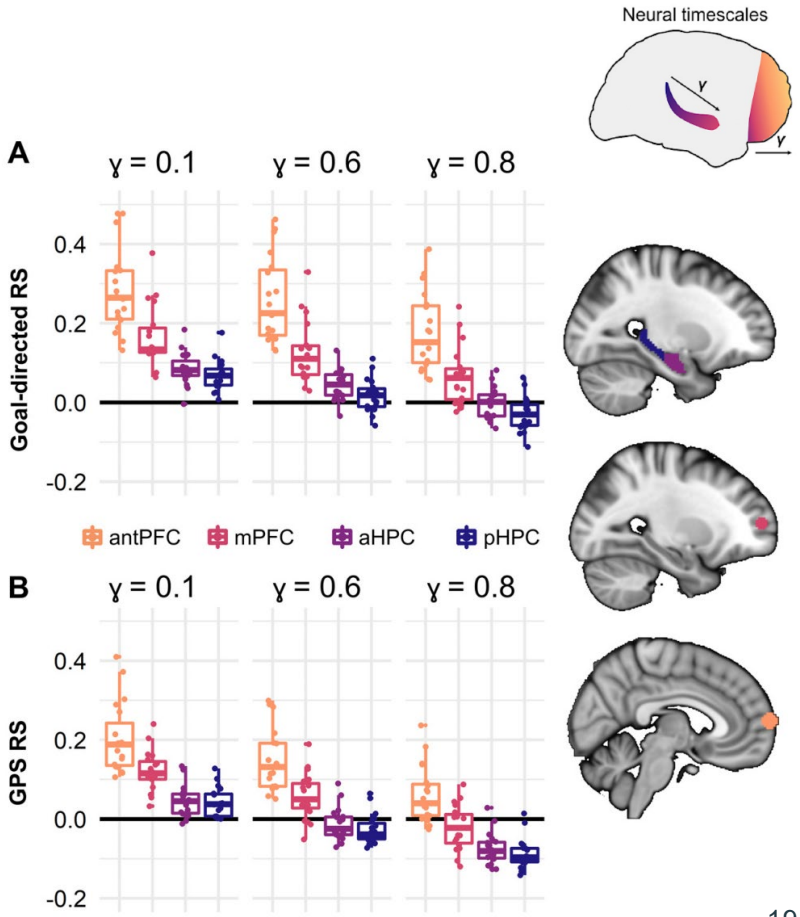
Predictive similarity between TR and discounted weight of future TRs

Second approach

Multiscale SR

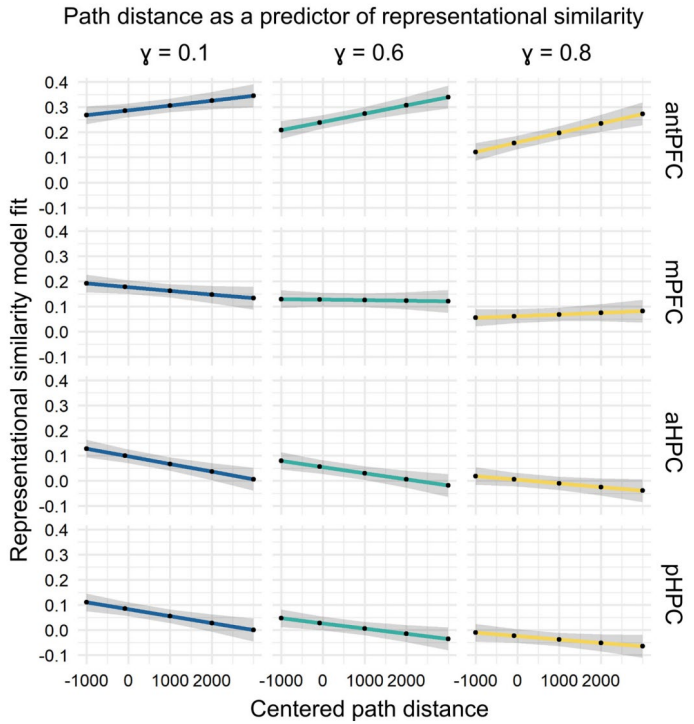
(Momennejad and Howard, 2018)

- mixed effects model
- Representational similarity between TR (state ~ fMRI voxel pattern) and sum of discount-weighted future TRs
- here: 4 ROIs at 3 different scales
- gradient in hippocampal representations is more defined
- same trends than in previous analysis
- note: mostly no overlapping of most anterior HPC and medial PFC



Accounting for path distance

- path distance included as factor in mixed effects model
  → calculated by summed change in coordinates between neighboring TRs
- normalized the number of TRs per route to account for time to goal as factor
- linear mixed effects model with ROIs
  → similarity not only depending on gamma but additionally dependent on path distance
- with euclidian distance as predictor



BREAK!!

Searchlight analysis

- identifies hierarchies of representations within PFC
- ROIs: spherical searchlights, 6mm radius
- 4 γ values: 0.1, 0.6, 0.8, 0.9
- Fig. 6A: significant representations of future states
- Fig. 6B: correlation threshold > 0,06

→ the farther in the future the states, the more they are preferred by the most anterior parts of PFC

→ in GPS-directed navigation, steps are not represented as far into the future

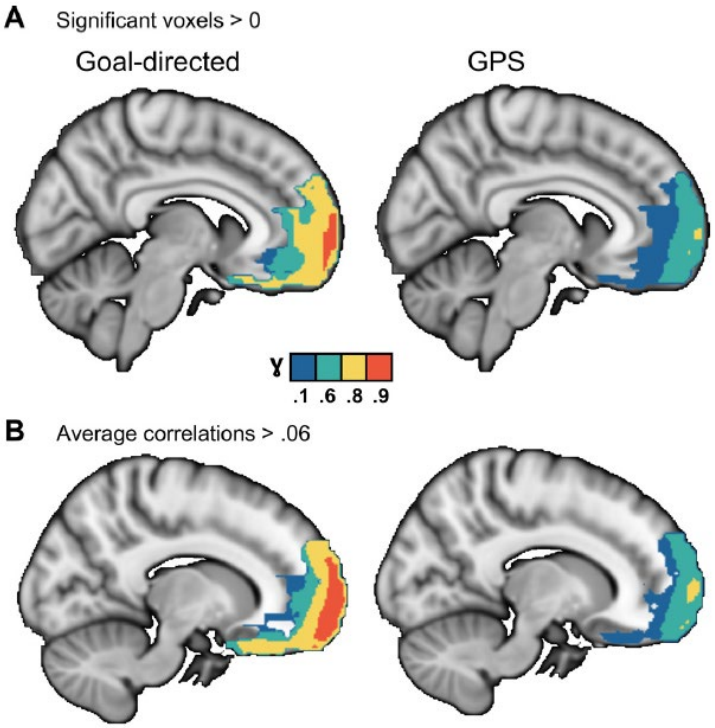


Fig. 6 A, B

Searchlight analysis

- per TR, the subjects travelled ca. 25m
- apply weights to each successor TRs
- γ = 0,1 - 25m
- γ = 0,6 - 175m
- γ = 0,8 - 375m
- γ = 0,9 - 800m

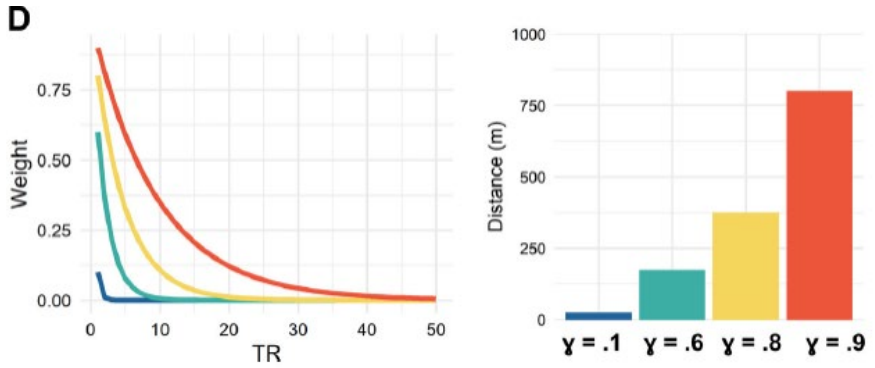


Fig. 6D

Prefrontal hierarchy

- similarity of voxels within posterior-anterior slices

  → more predictive representation in anteriormost part

  → more predictive representation of the nearer future (γ = 0.1)

  → more predictive representation in goal-directed condition

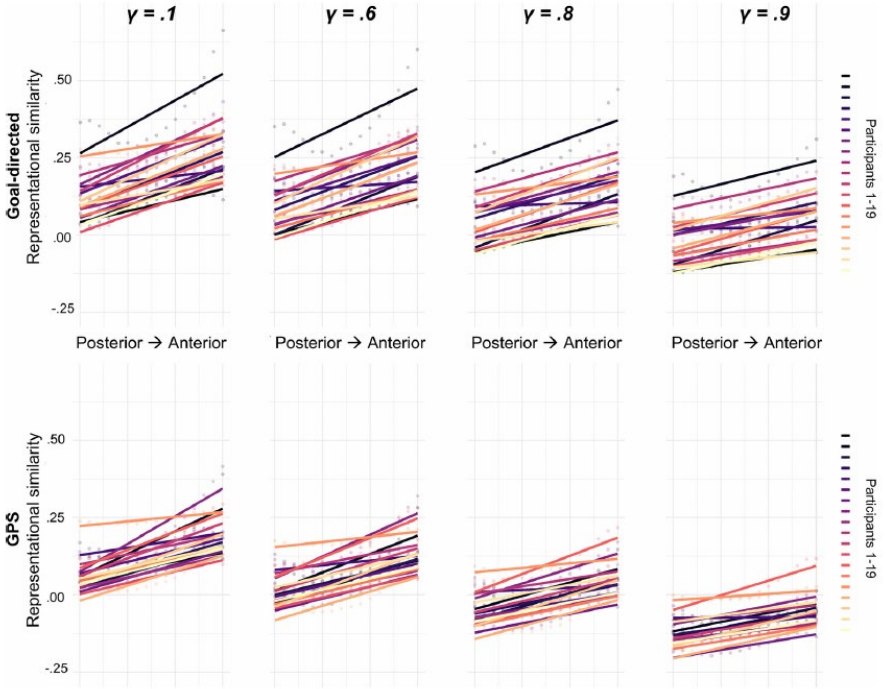


Fig. 7

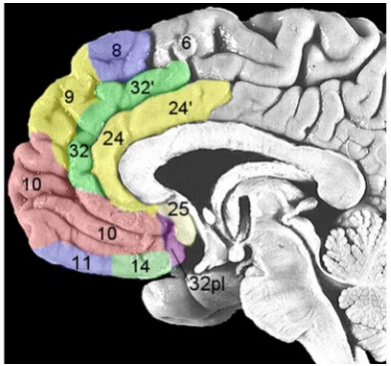
Prefrontal hierarchy - Brodmann areas

- voxel activation in different Brodmann areas

  → anterior areas are in general more activated

  → anterior areas represent higher γ (states further in future)

  → goal-directed navigation includes planning further in the future than GPS-directed navigation



https://www.researchgate.net/profile/Giovanni-Mirabella/publication/268793611/figure/fig1/AS:271472207855625@1441735426305/Areas-composing-the-prefrontal-cortex-PFC-according-to-the-parcellation-of-Petrides-and.png

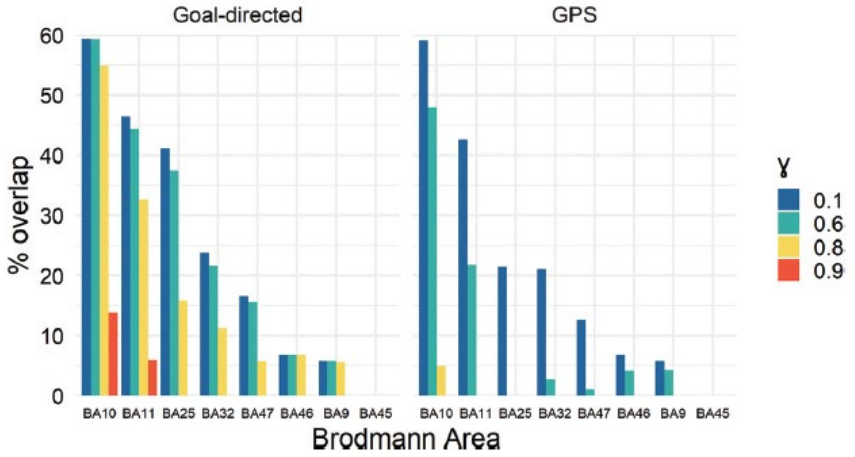


Fig. 8

Controlling for distance

- distances of the routes significantly differ between the two conditions (3.5 vs. 2.5 km) (see Fig. 2E)
- manual selection of pairs of routes for both conditions for each participant (see Fig. 9A)

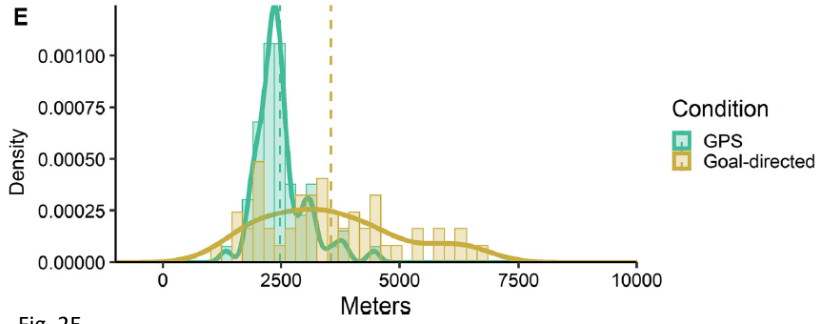


Fig. 2E

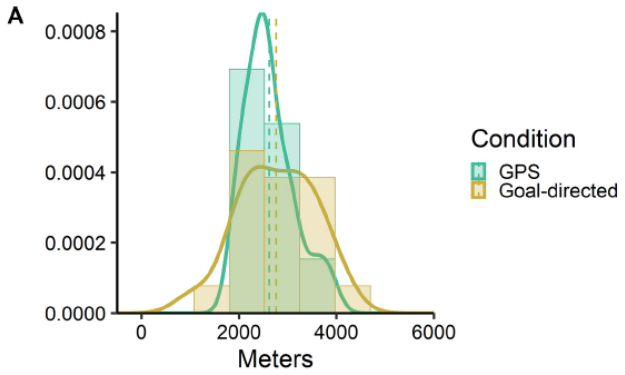


Fig. 9A

Controlling for distance

- few clusters differed significantly between the two conditions
- set of clusters may differ between the two conditions in rostrocaudal PFC
- smaller horizons: only orbitofrontal clusters differed
- larger horizons: more dorsal, ventral & caudal PFC clusters differed

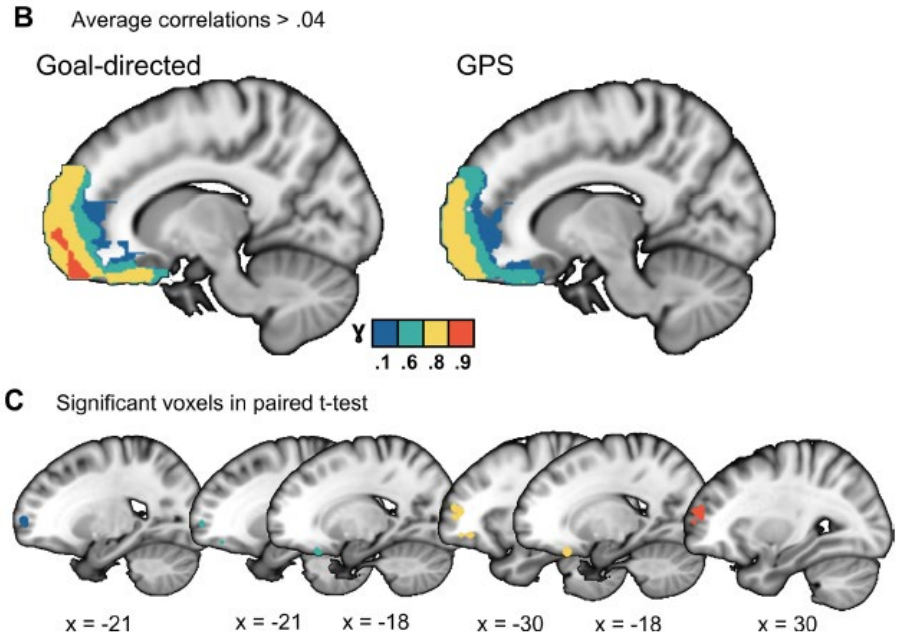


Fig. 9 B-C

Nonspatial relevance

- representational hierarchy along prefrontal & hippocampal gradients: relational knowledge, category generalization, reward predictions, schema learning
- anterior PFC: encoding & retrieval of memory task-sets and goals
  → lesions: impair multitasking & prospective memory
- more anterior PFC = higher levels of integration & abstraction

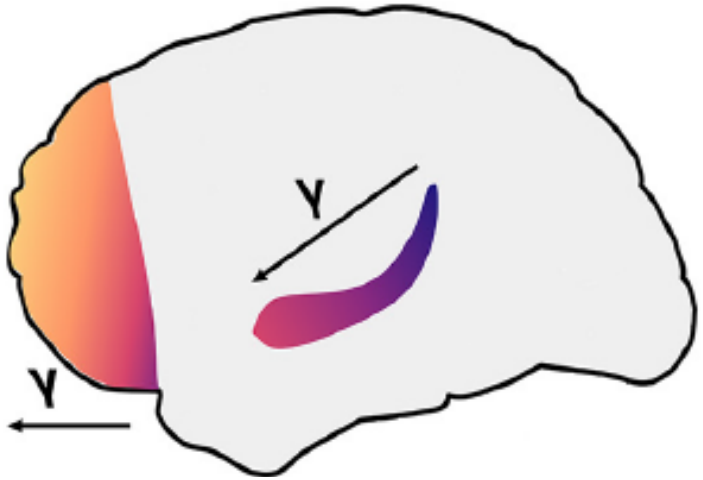
Drawbacks of the study

- no comparison between pre- & posttraining data
- GPS routes were shorter & included more turns than goal-directed navigation routes
- a priori selection of ROIs
- selection of routes doesn’t include multiple past & future trajectories for each state

Future studies:

- investigate compressed representation & abstraction
- investigate temporal hierarchy of navigation

Conclusion



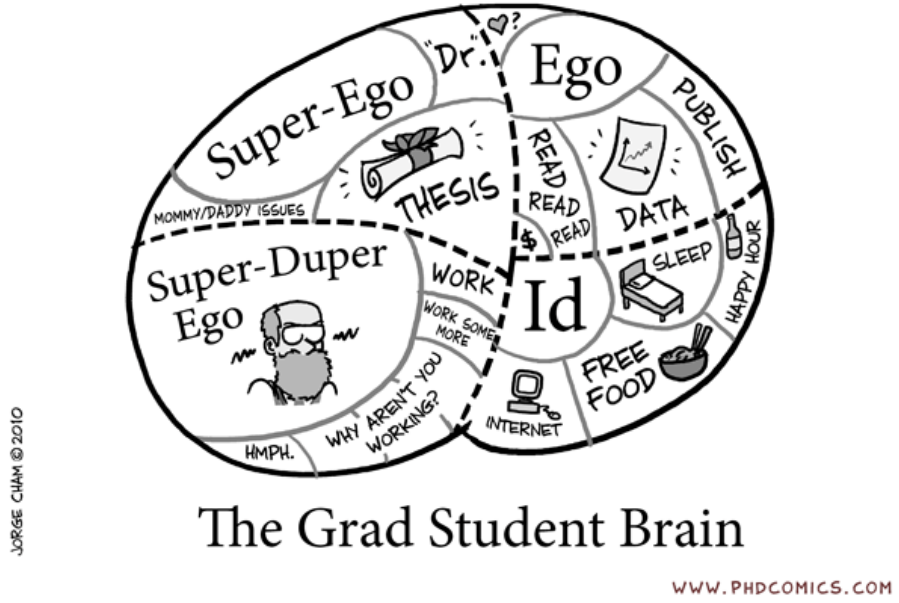
large γ → small γ

(ventromedial) PFC → anterior hippocampus → posterior hippocampus

largest → smallest predictive scales

https://thumbs.dreamstime.com/b/level-indicator-progress-bar-element-set-royalty-free-vector-illustration-81801033.jpg

Thank you for your attention!



https://www.stressmarq.com/wp-content/uploads/2014/11/comic-brain.gif

Sources/Literature:

- Brunec, I. K., & Momennejad, I. (2022). Predictive Representations in Hippocampal and Prefrontal Hierarchies. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *42*(2), 299–312. https://doi.org/10.1523/JNEUROSCI.1327-21.2021
- Abstract Memory Representations in the Ventromedial Prefrontal Cortex and Hippocampus Support Concept Generalization. Caitlin R. Bowman, Dagmar Zeithamova. *Journal of Neuroscience* 7 March 2018, 38 (10) 2605-2614; DOI: 10.1523/JNEUROSCI.2811-17.2018
- Momennejad, Ida & Howard, Marc. (2018). Predicting the Future with Multi-scale Successor Representations. 10.1101/449470.
- Brunec, I.K., Bellana, B., Ozubko, J.D., Man, V., Robin, J., Liu, Z., Grady, C.L., Rosenbaum, R.S., Winocur, G., Barense, M.D., & Moscovitch, M. (2018). Multiple Scales of Representation along the Hippocampal Anteroposterior Axis in Humans. *Current Biology, 28*, 2129-2135.e6.
- Phillip (2022). RL in the brain (+ behaviour) - slides from Cognitive Maps Seminar 9.11.2022
- Momennejad I, Russek EM, Cheong JH, Botvinick MM, Daw ND, Gershman SJ. The successor representation in human reinforcement learning. Nat Hum Behav. 2017 Sep;1(9):680-692. doi: 10.1038/s41562-017-0180-8

Discussion questions

- **Although the study includes the GPS navigation task and routes in areas where the subjects were unfamiliar, they analyzed that overall the difficulty ratings of the tasks were similar and "all navigated routes were perceived to be similarly undemanding". With this in mind, do you think we could expect similar results in more difficult tasks where shifting and replanning would be more required? How adaptive we can expect predictive representations to be?**

- **In the experiment, subjects moved through a virtual version of Toronto. Do you think it will be possible one day to conduct true field studies (i.e. walking through real Toronto and not a simulated version) and still be able to measure valuable neurophysiological data?**

- **How do you think that the findings of the study (i.e. multiscale predictive representations) could be applied in a practical way? For example, do you think those representations could be integrated in a navigational AI system and if so how?**

Discussion questions

- **Is figure 1a supposed to attest to the goal-directed and predictive 'behavior' of place cells? Is this a representation of 3 different place cells (that have different distances between each other) firing, or just the rate at which one would reach the next one in the distances to the targets? Or is it showing something else completely?**
- **In your opinion, would these experiments also work in another, unfamiliar setting i.e. not the hometown of the study participants? I am also curious about the influence of Toronto as the city of choice here… do you think the distinct marks (skycrapers, CN tower, etc.) and the grid-like system facilitate hierarchical planning? How can we imagine this in "flat" cities/regions in Europe?**
- **How much do you think real-life extra features like smells or particular sounds (which can be strongly linked to memory) could influence navigation and predictive representation versus the virtual environment presented here, no matter how spatially accurate and familiar?**

Discussion questions

- **What does the effect of hemisphere presented by the authors (that the similarity values are higher for the right compared to left hippocampus) tell us about a potential lateralization of predictive representations? Could this finding indicate that the two hippocampi (in one person) differ in their function and when yes how?**
- **In the study they excluded trials in which the subjects did not reach their goal (got lost). It would be interesting to compare the presence and extent of predictivity gradients (in HPC and PFC) in successful vs failed trials. In failed trials, would we expect lower predictivity for future states in both HPC and PFC? One could imagine that short-term predictions are un-altered (in both HPC and PFC), while long-term predictions are reduced (in PFC) due to the long-term plan constantly changing. Does this make sense?**
- **The authors suggest that similar PFC-HC networks are used for modalities other than space. Can you think of any other modalities in which one could test this assertion? To put it another way, what would different levels of horizon be for say, social cognition?**

Discussion questions

- **What do you think if the other relational knowledge such as category and semantic also show similar gradients of predictive representation along the posterior and anterior axis in PFC and hippocampus?**

- **Would you suggest that *hierarchical* coding of successor representation states implies utilizing graph representations as well?**
- **What is the direction of information flow in the hippocampus and PFC? I suppose it is posterior to anterior (or can it be both, on different frequencies? I found an illustration indicating this). I am asking because I initially thought that it would make more sense to 'descend' from the coarse to fine representations, but with e.g. Fig. 1B and overall it seems to be the other way around.**
- **I have a more validity question, would the virtual nature of the task be successfuly comparable to actual locomotion in a real setting? In other words, would doing this experiment in a virtual environment capture the complexity of a real setting?**

Discussion questions

- **Why do you think do the authors not even mention the entorhinal cortex? We previously learned that entorhinal cortex has multiple functions for cognitive maps, like distance measurement. Why do those functions seem to be completely irrelevant for the task in this paper.**

- **If hierarchical representations exist, what shall we expect at cellular level (e.g., for place cells in the hippocampus from anterior to posterior)?**

- **How can RL agents learn hierarchical state representations similar to the human hippocampus and plan to achieve a long-run goal?**

Discussion questions

- **(ref Fig. 6) more anterior regions of the prefrontal cortex seem to be representing the coarser grained, longer range and larger predictive horizons, but mainly in the goal-directed experimental condition as opposed to the GPS condition. Could this have something to do with not needing to look ahead as much with the GPS instructing you about the new environment, as opposed to needing to rely on deeper look-aheads when you are exploring environments without external instruction?**

- **An interesting experiment would be to see what would happen when the planned route suddenly becomes impossible by an surprising obstacle. How long would the original predictions still last (as it is possible that the obstacle "disappears" again and the original plan can still be followed.**