Nature 2018
A. Banino, Caswell Barry et.al
[redacted] 201711060109

VECTOR-BASED NAVIGATION USING GRID-LIKE REPRESENTATIONS IN ARTIFICIAL AGENTS

——神经导航 GRID NAV

- Nature, 557:429-433, 2018

- Andrea Banino, Caswell Barry, Benigno Uria, Charles Blundell, Timothy Lillicrap, Piotr Mirowski, Alexander Pritzel, Martin J. Chadwick, Thomas Degris, Joseph Modayil, Greg Wayne, Hubert Soyer, Fabio Viola, Brian Zhang, Ross Goroshin, Neil Rabinowitz, Razvan Pascanu, Charlie Beattie, Stig Petersen, Amir Sadik, Stephen Gaffney, Helen King, Koray Kavukcuoglu, Demis Hassabis, Raia Hadsell & Dharshan Kumaran

- DeepMind & University College London

Motivations

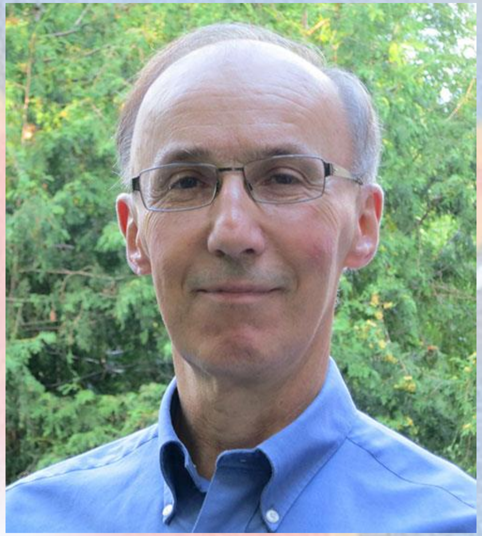
- Train a deep network to perform self-localisation (path integration) and show that the grid cells emerge as a consequence of this objective
- Show that grid cells are an effective basis for vector based navigation (i.e. computing vectors within euclidean framework)

Basic background①

A cognitive map

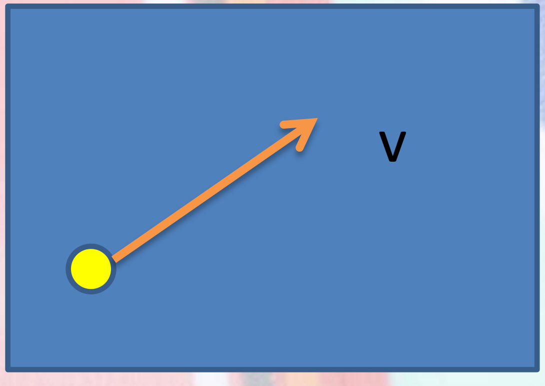
A cognitive map

Hippocampus as a cognitive map



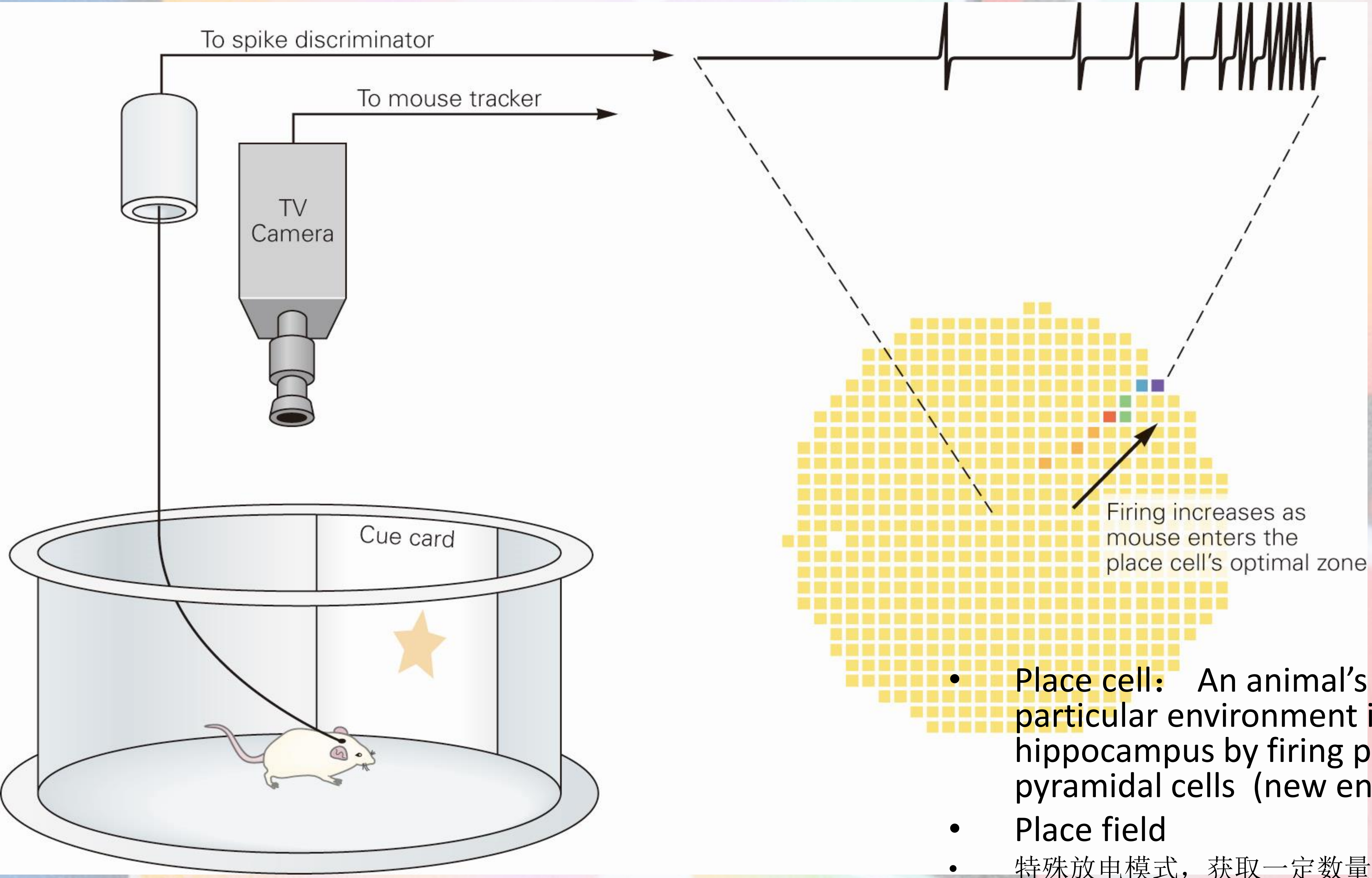
- A spatial map
  - (1971)
  - Brain Research
  - 4347c. Google s.

2014年诺贝尔生理学或医学奖获奖者：N. O’Keefe
1971年论文合作学生著名学者：Jonathan Dostrovsky



Path Integration(积分)

[1]图片来自http://en.wikipedia.org/wiki, http://can-acn.org/jonathan-dostrovsky-honorary-member



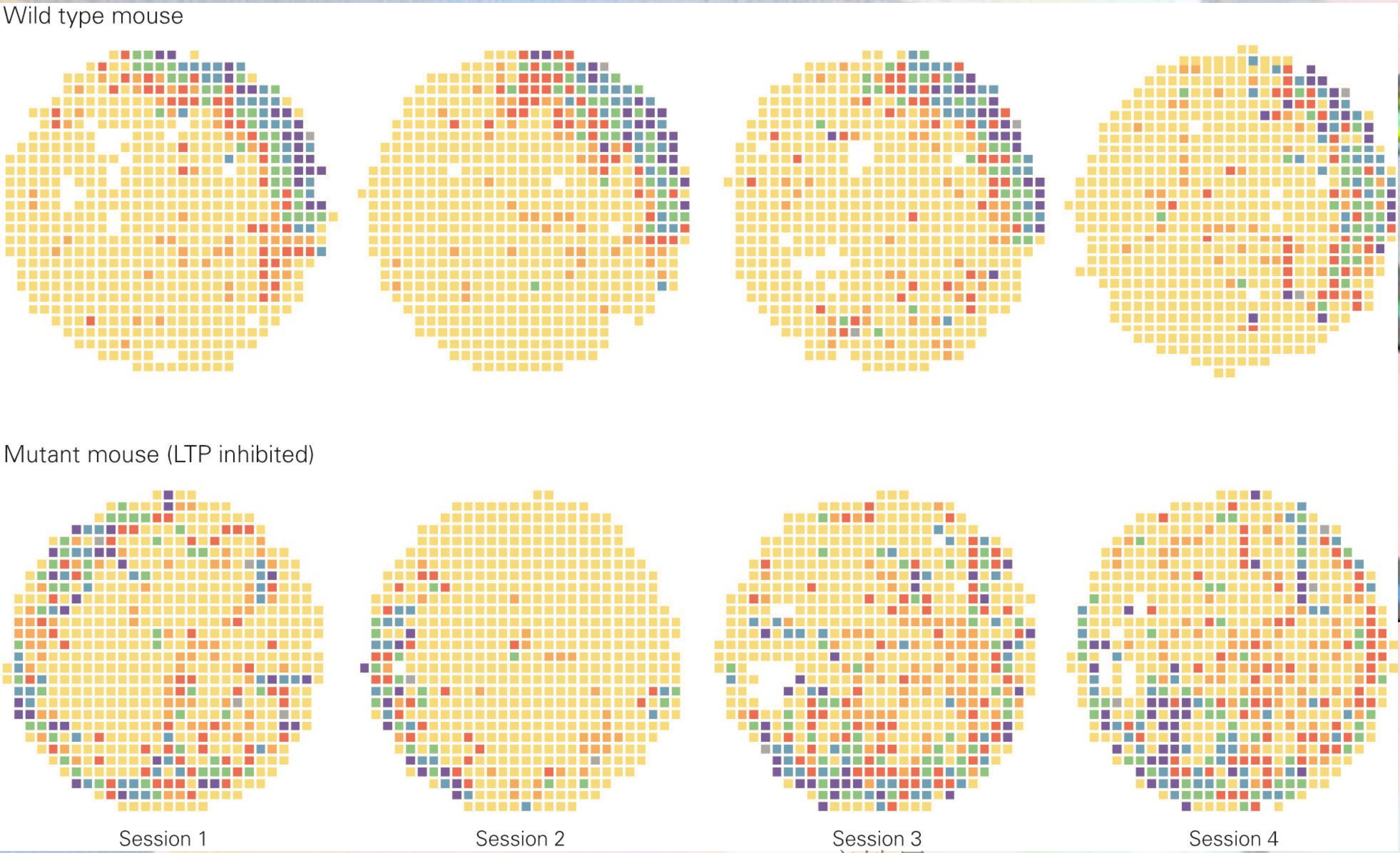
当鼠处于活动空间的某一特定区域时，海马中的一些锥体细胞强烈放电，最大放电频率达到几十赫兹，而在其他区域很少或者几乎没有发放。这类发放活动强烈依赖于动物位置的细胞，称为位置细胞，对应的局部空间区域被称作位置细胞的位置野。

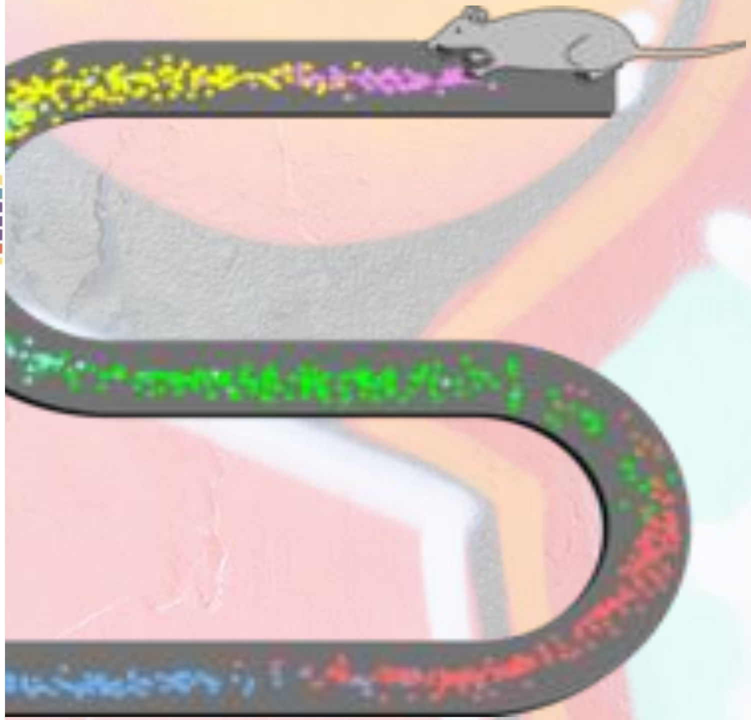
Firing rate 在中心点达到最大，进入与离开相应增强或减弱。

- Place cell： An animal's familiarity with a particular environment is represented in the hippocampus by firing pattern of populations of pyramidal cells (new env. Minutes->week|mon.)
- Place field
- 特殊放电模式，获取一定数量的电信号活动能够预测动物在环境中的位置。

[2] The firing patterns of pyramidal cells in the hippocampus create an internal representation of the animal's location in its surroundings. (principles of neural science 5th edition)

Place cell & Place field





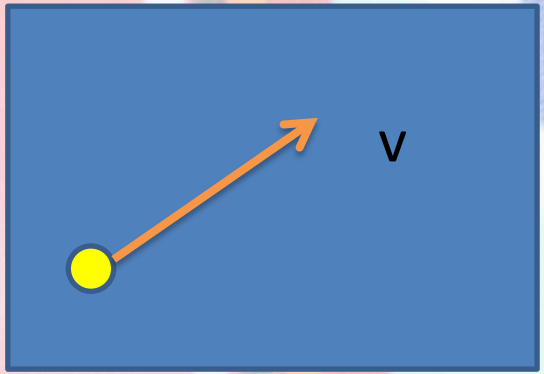
CA1,CA3

A cognitive map

Lynn Nadel

- **The hippocampus as a cognitive map**
  - **1978, 10134 on google scholar**
- The hippocampus contains an invariant representation of space that does not depend on mood or desire.
- Place cell fire: current location

Ranck 1984
头朝向细胞
（head direction cell）
方向信息信号,
45 degree

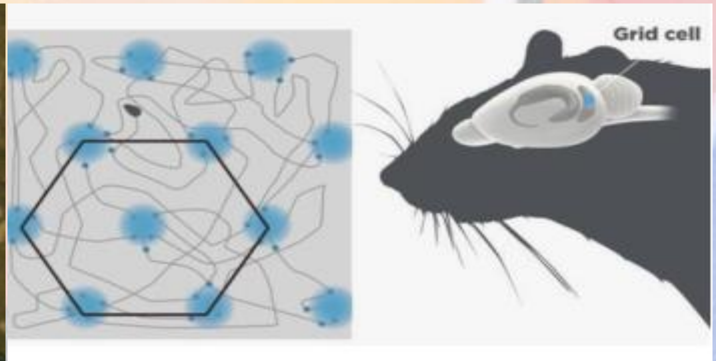


Path Integration(积分)

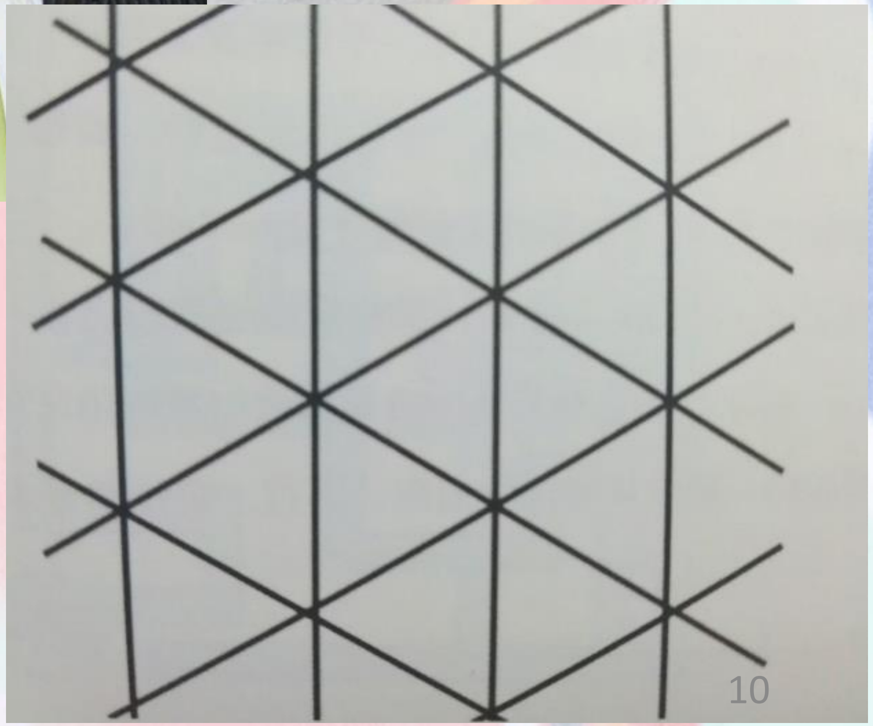
Grid cell

- Entorhinal cortex
- 与place cell 类似的对位置敏感，更高的秩序性
- 2005 Nature 2220
- [Microstructure of a spatial map in the entorhinal cortex]

Moser.

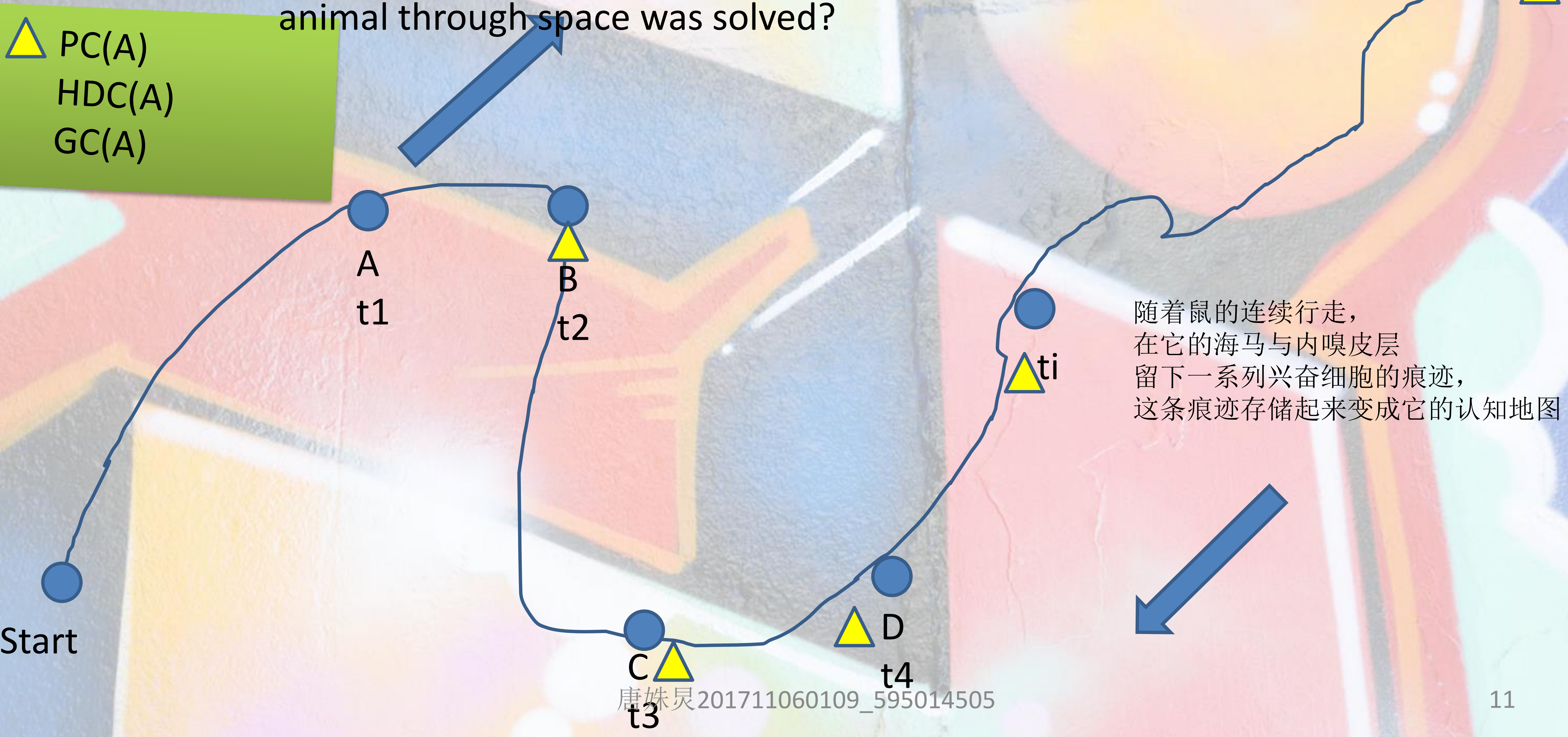


2014年诺贝尔生理学或医学奖获奖者



O’Keefe and the Mosers were awarded the Nobel Prize in physiology or medicine

The role of the hippocampus in guiding an animal through space was solved?



随着鼠的连续行走，
在它的海马与内嗅皮层
留下一系列兴奋细胞的痕迹，
这条痕迹存储起来变成它的认知地图

Get a grid-like network②

Supervised self-localisation

Supervised learning: architecture

- Goal

  learn a grid-like network , then test whether it can achieve self-localisation.
- Method

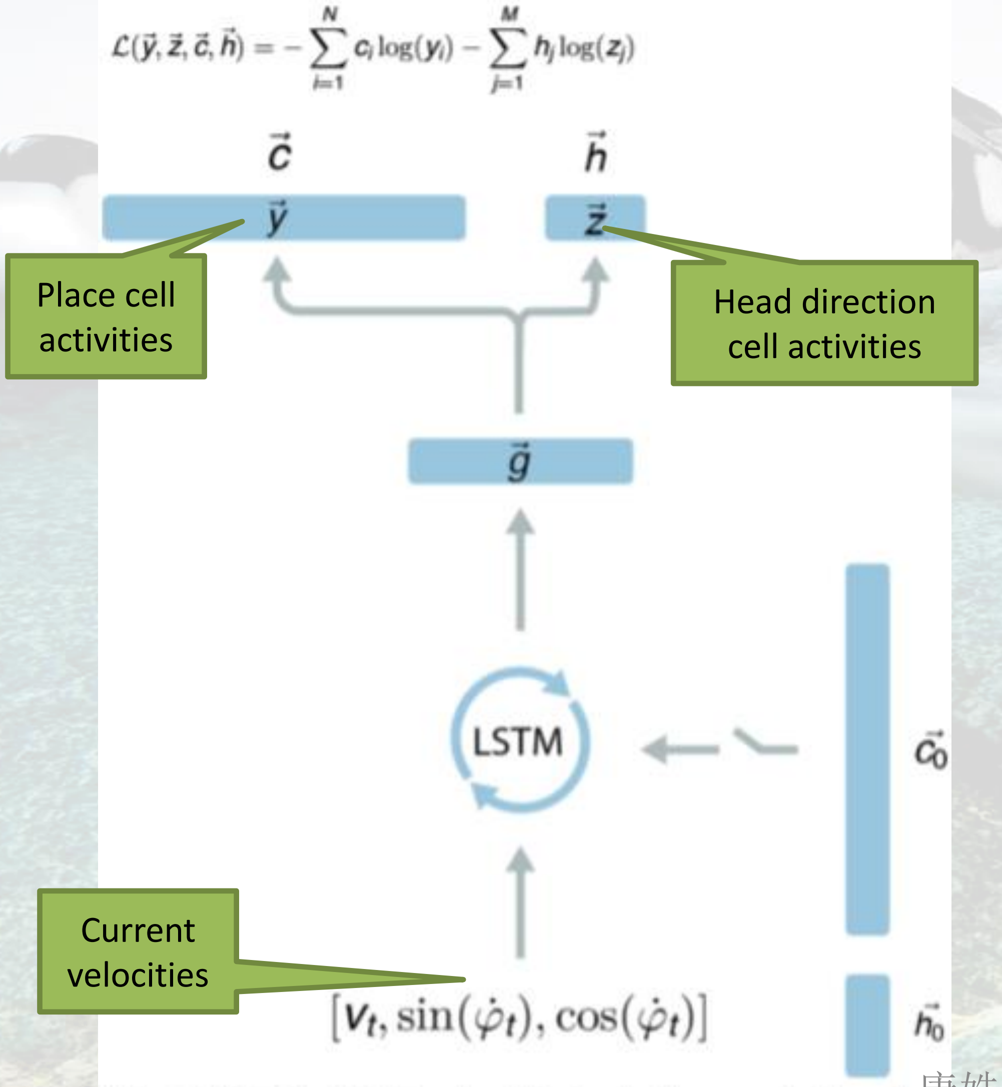
  supervised learning + lstm

  input: velocity(移动速度、回转速度)

  output: prediction(place cell, head direction cell)

Self-localisation task

$$c_i = \frac{e^{-\frac{\|\vec{x}-\vec{\mu}_i^{(c)}\|_2^2}{2(\sigma^{(c)})^2}}}{\sum_{j=1}^{N} e^{-\frac{\|\vec{x}-\vec{\mu}_j^{(c)}\|_2^2}{2(\sigma^{(c)})^2}}}$$



Extended Data Fig. 1 | Network architecture in the supervised learning experiment. The recurrent layer of the grid cell network is an LSTM with

- Architecture(left)
- Platfrom: deepmindla

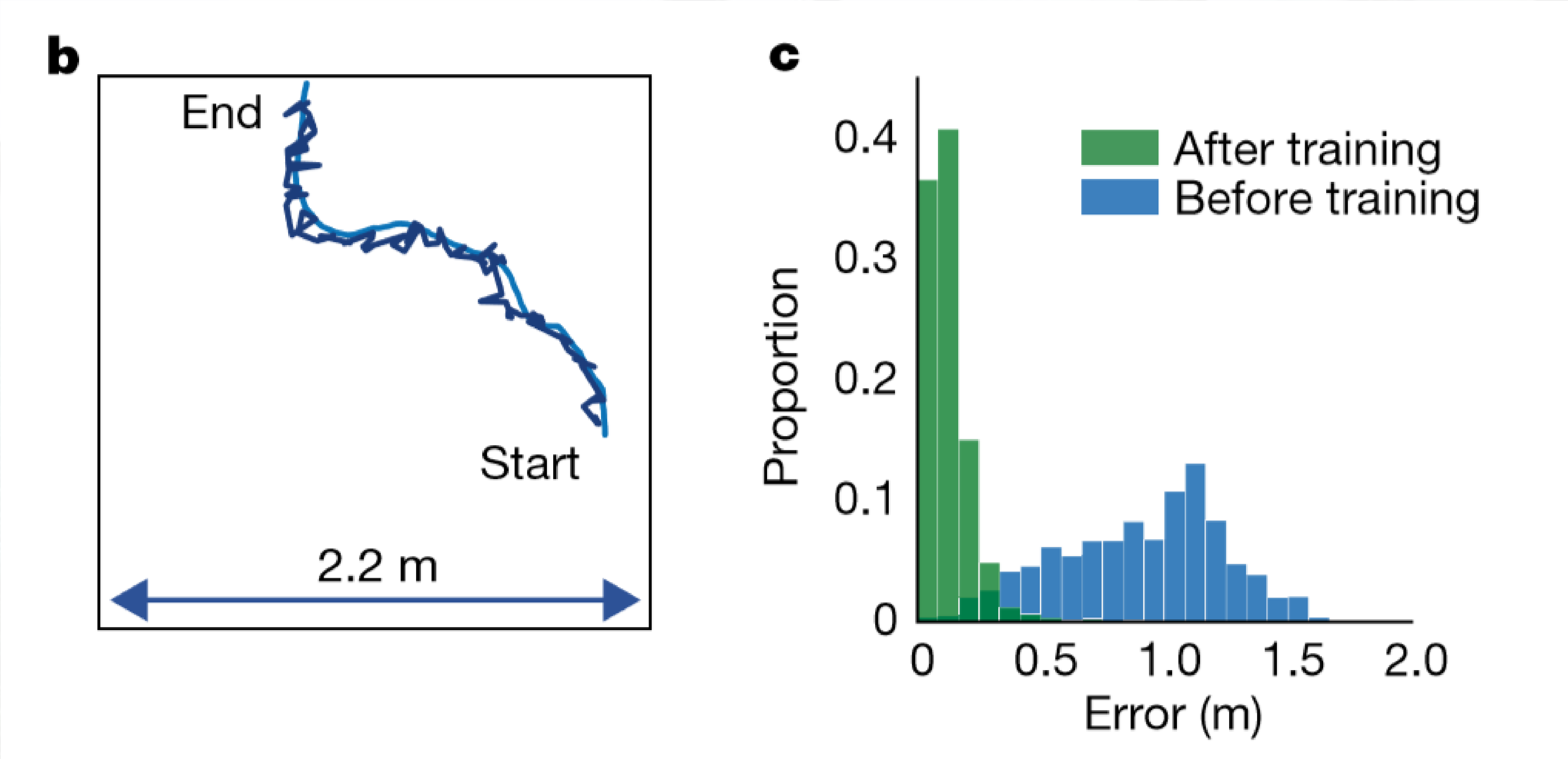
$$h_i = \frac{e^{k^{(h)}\cos(\varphi-\mu_i^{(h)})}}{\sum_{j=1}^{M} e^{k^{(h)}\cos(\varphi-\mu_j^{(h)})}}$$

- Ground truth
  - Place cell ci [gaussian]
  - Head direction cell hi[von mises]
- dropout(linear layer

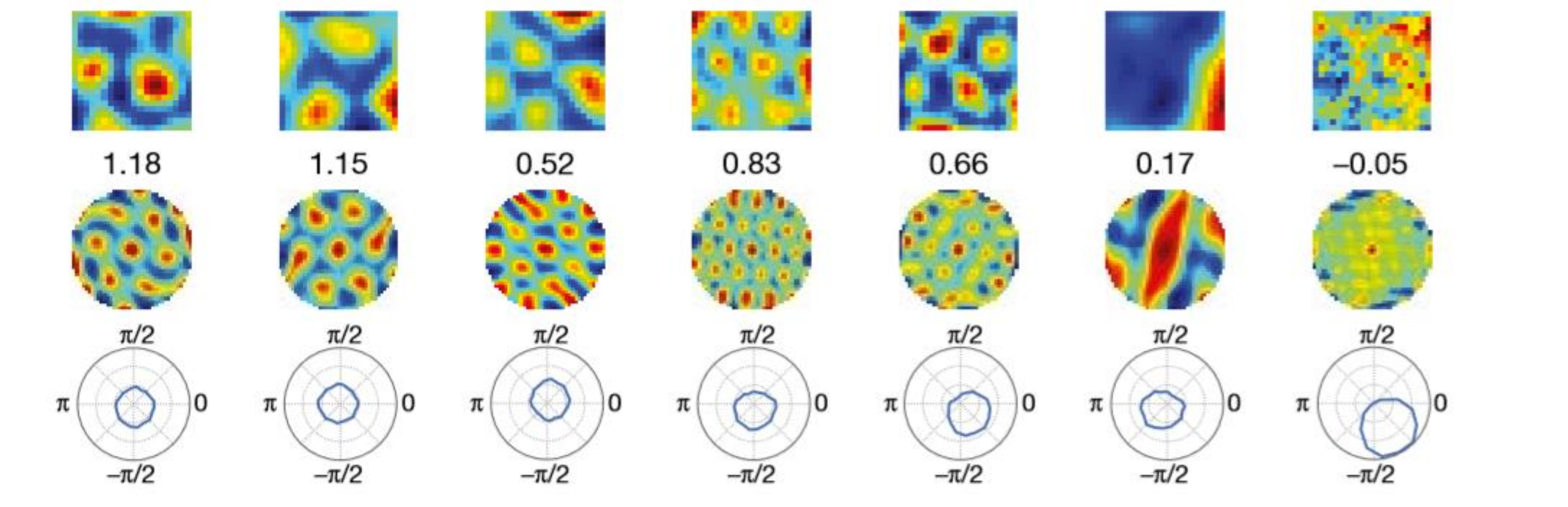
$$\vec{l}_0 = W^{cp}\vec{c}_0 + W^{(cd)}\vec{h}_0$$
$$\vec{m}_0 = W^{hp}\vec{c}_0 + W^{(hd)}\vec{h}_0$$

  - 50%

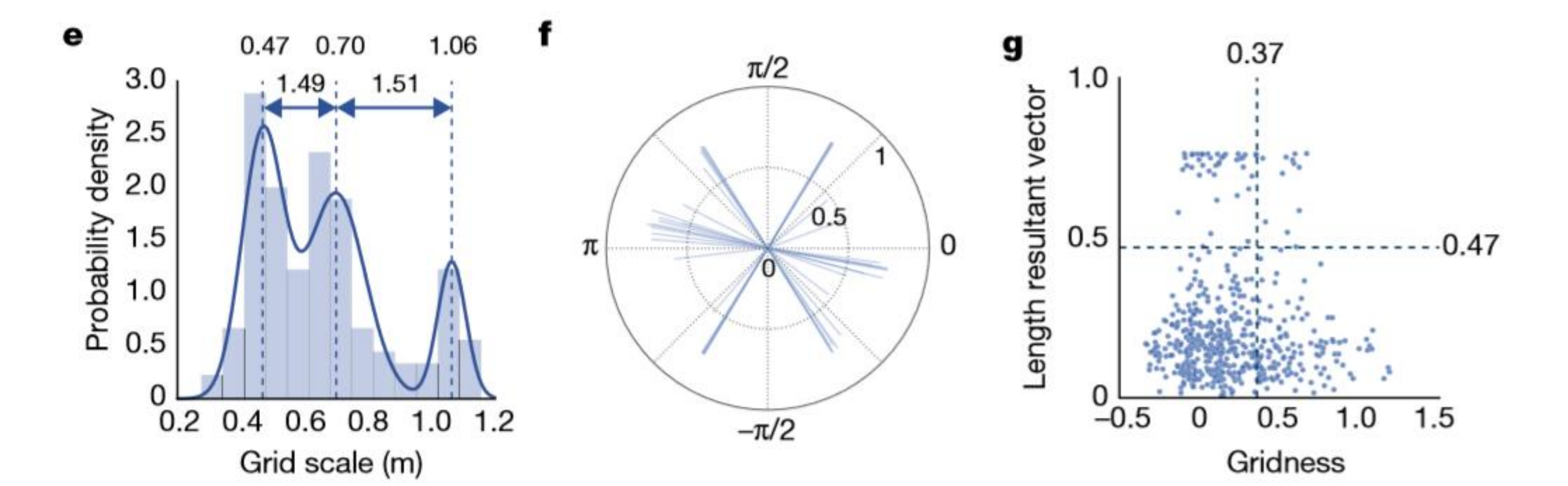
Path integration task



Linear layer activations



Linear layer activations: properties of units



Stemmier et al 2015
Wei et al.2015

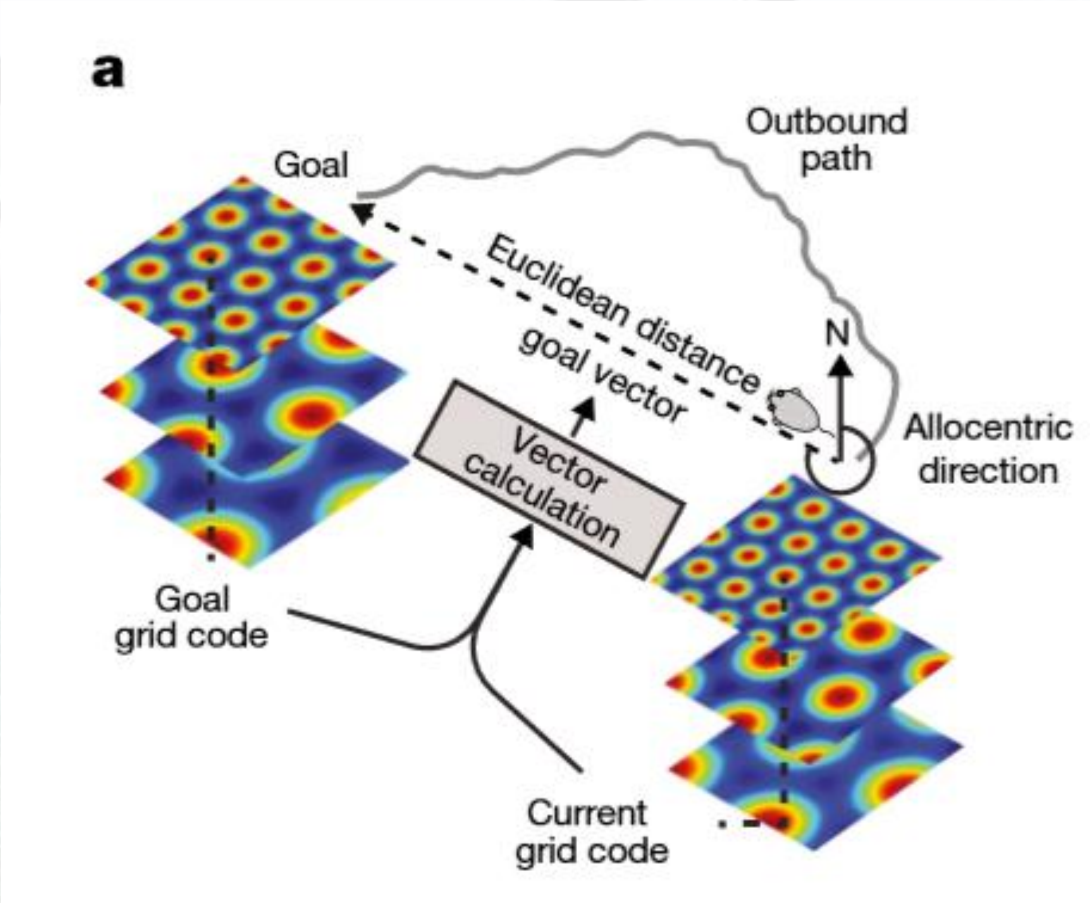
Linear layer activations: whole population

- Grid-like units: 25.2%
- HD-like units:10%
- Border-like units: 8.7%
- Unclassified units: 56.1%

Navigation ① plus RL ③

Vector-based navigation

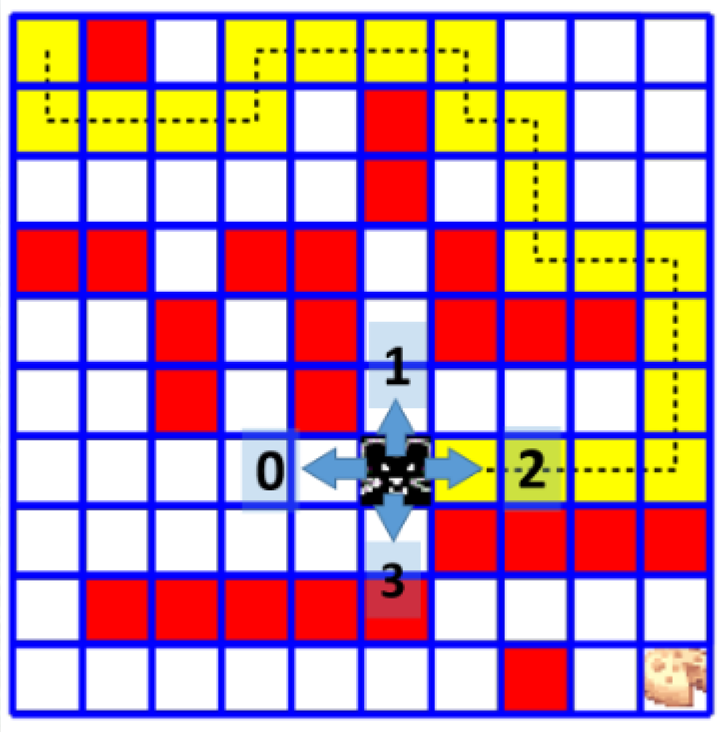
Vector-based navigation



Traditional Machine Learning is supervised pattern recognition

"CAT"

- We want agents that can act on their own, not just recognize patterns
- “GO LEFT?”

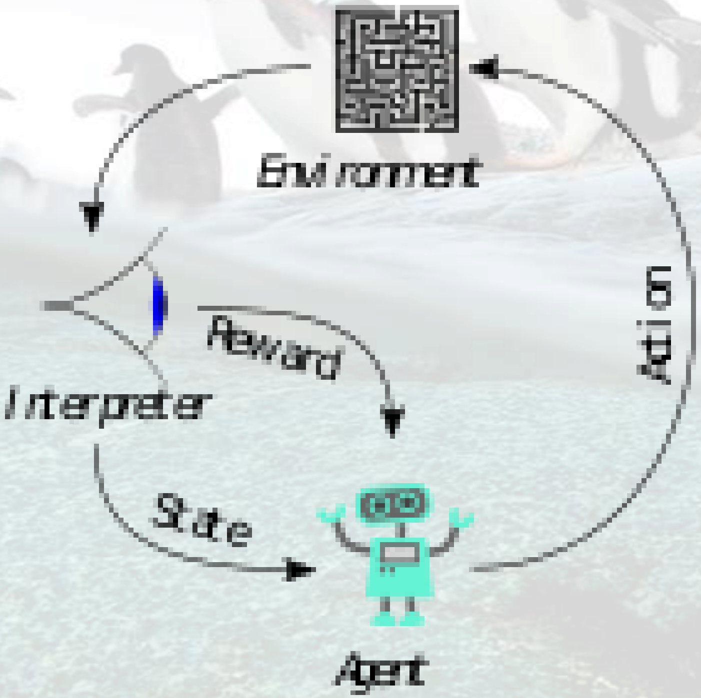


Sequential decision making processes

Decision Making in Markov Decision Processes

- **Agent** (The thing doing the acting)
- **Environment** (What the agent interacts in)
- **States** $\boldsymbol{s}$ (The state of the environment)
- **Actions** $\boldsymbol{a}$ (What the agent can do in the environment)
- **Transitions** $\boldsymbol{P}(\boldsymbol{s}' \mid \boldsymbol{s}, \boldsymbol{a})$ (How actions change the state)
- **Rewards** $\boldsymbol{R}(\boldsymbol{r} \mid \boldsymbol{s}, \boldsymbol{a})$ (How good/bad an action in a state is)
- **Policy** $\boldsymbol{\pi}(\boldsymbol{a} \mid \boldsymbol{s})$ (How the agent picks actions)

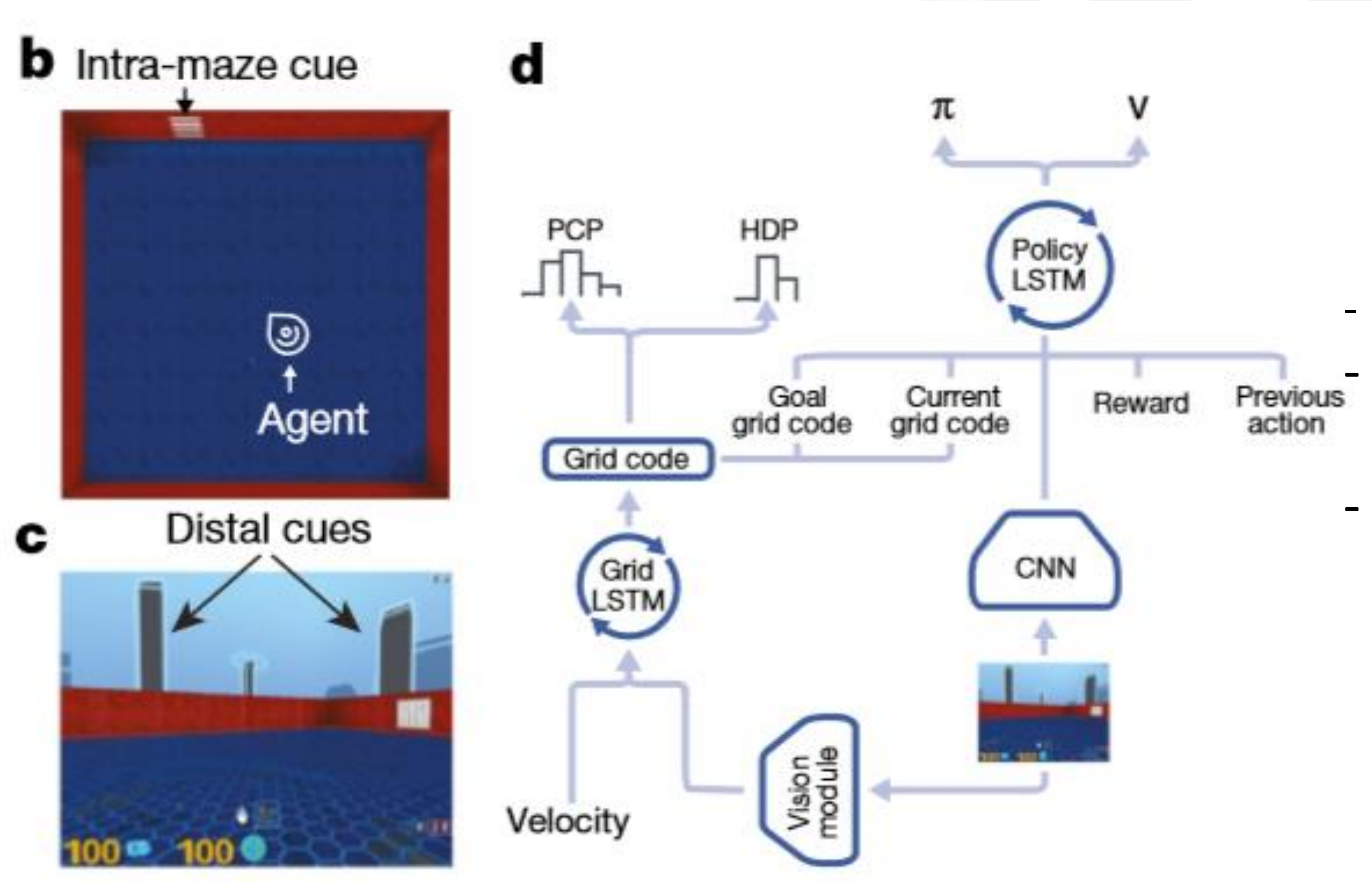
Deep reinforcement learning



$$G_t = \mathbb{E}[\sum_{j=1}^{\infty} \gamma^{j-1} R_{t+j}]$$

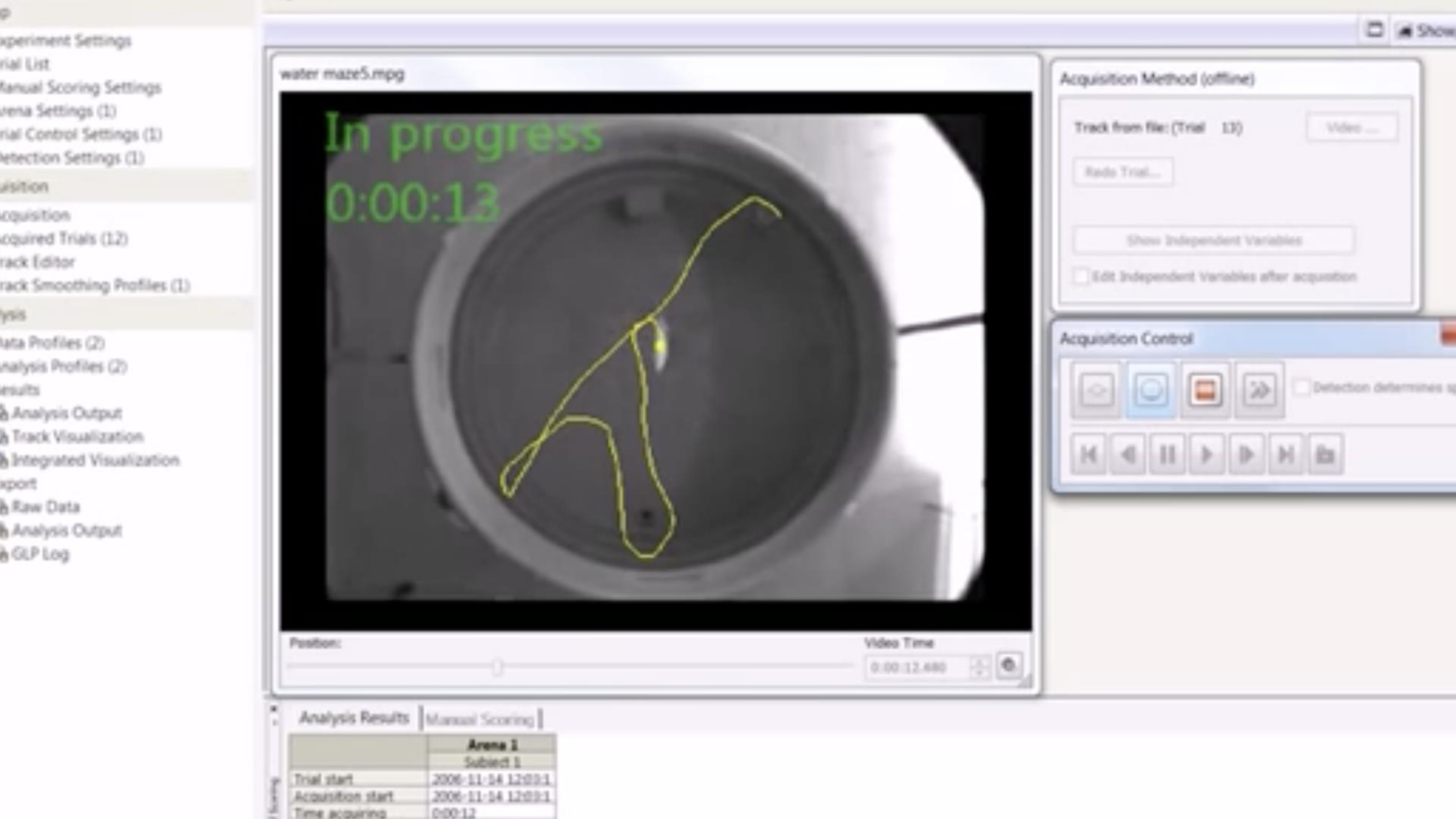
Goal : maximize expected cumulative discounted future reward

Grid cell agent: architecture



Goal: 学习到的grid-like网络能够用于导航实际环境，及grid的作用（code）

- Goal grid code + current grid code
- 基于二者可以算出goal vector的假说验证【Fietel&Brokings 08】
- Morris watermaze task的完成。


xperiment Settings
rial List
Manual Scoring Settings
rena Settings (1)
rial Control Settings (1)
etection Settings (1)
uisition
cquisition
cquired Trials (12)
rack Editor
rack Smoothing Profiles (1)
ysis
ata Profiles (2)
nalysis Profiles (2)
esults
Analysis Output
Track Visualization
Integrated Visualization
xport
Raw Data
Analysis Output
GLP Log
water maze5.mpg
In progress
0:00:13
Position:
Video Time
Acquisition Method (offline)
Track from file: (Trial 13)
Show Independent Variables
Edit Independent Variables after acquisition
Acquisition Control
Analysis Results
Manual Scoring
Arena 1
Subject 1
Trial start
Acquisition start
Time acquiring

Asynchronous Advantage Actor-Critic (A3C)

$$Q^{\pi_\theta}(s_t^n, a_t^n) - V^{\pi_\theta}(s_t^n)$$

$$V^{\pi_\theta}(s_t^n)$$

baseline

$$\nabla \bar{R}_\theta \approx \frac{1}{N}\sum_{n=1}^{N}\sum_{t=1}^{T_n}\boxed{\left(\underbrace{\sum_{t'=t}^{T_n}\gamma^{t'-t} r_{t'}^n}_{G_t^n} - b\right)}\nabla \log p_\theta(a_t^n|s_t^n)$$

G_t^n : obtained via interaction

$$E[G_t^n] = Q^{\pi_\theta}(s_t^n, a_t^n)$$

Advantage Actor-Critic

$$Q^{\pi}(s_t^n, a_t^n) - V^{\pi}(s_t^n)$$

$$r_t^n + V^{\pi}(s_{t+1}^n) - V^{\pi}(s_t^n)$$

Estimate two networks? We can only estimate one.

Only estimate state value
A little bit variance

$$Q^{\pi}(s_t^n, a_t^n) = E[r_t^n + V^{\pi}(s_{t+1}^n)]$$

$$Q^{\pi}(s_t^n, a_t^n) = r_t^n + V^{\pi}(s_{t+1}^n)$$

Advantage Actor-Critic

π interacts with the environment

TD or MC

Learning $V^{\pi}(s)$

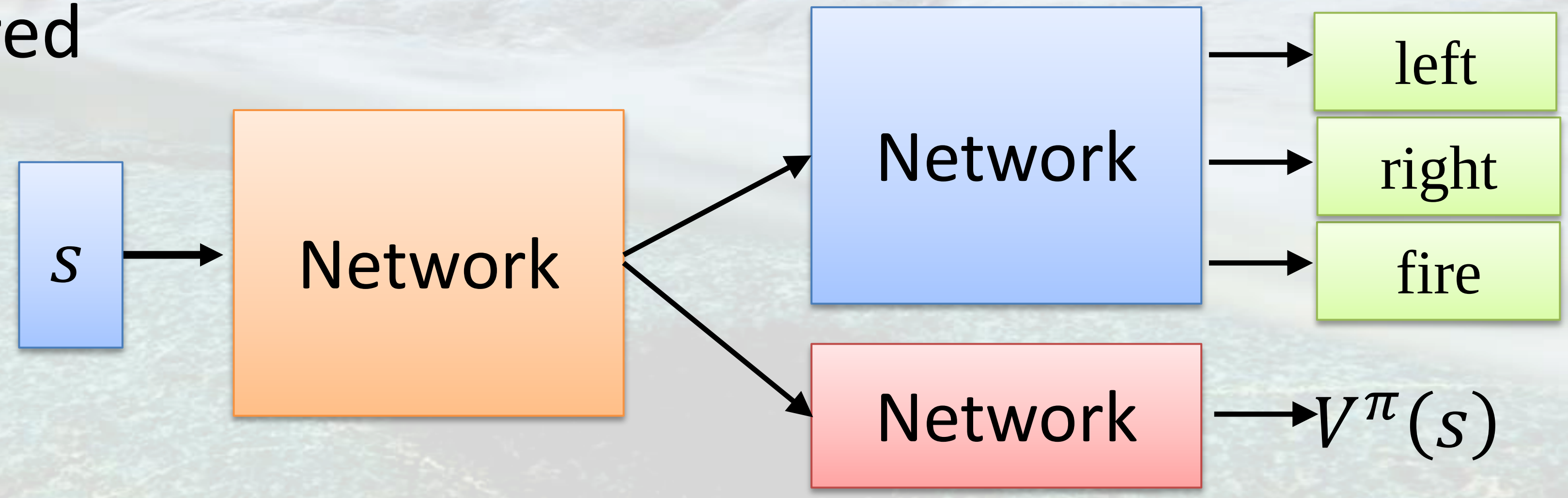
Update actor from $\pi \rightarrow \pi'$ based on $V^{\pi}(s)$

$\pi = \pi'$

$$\nabla \bar{R}_{\theta} \approx \frac{1}{N} \sum_{n=1}^{N} \sum_{t=1}^{T_n} \left(r_t^n + V^{\pi}(s_{t+1}^n) - V^{\pi}(s_t^n) \right) \nabla log p_{\theta}(a_t^n | s_t^n)$$

Advantage Actor-Critic

- Tips
  - The parameters of actor $\pi(s)$ and critic $V^\pi(s)$ can be shared

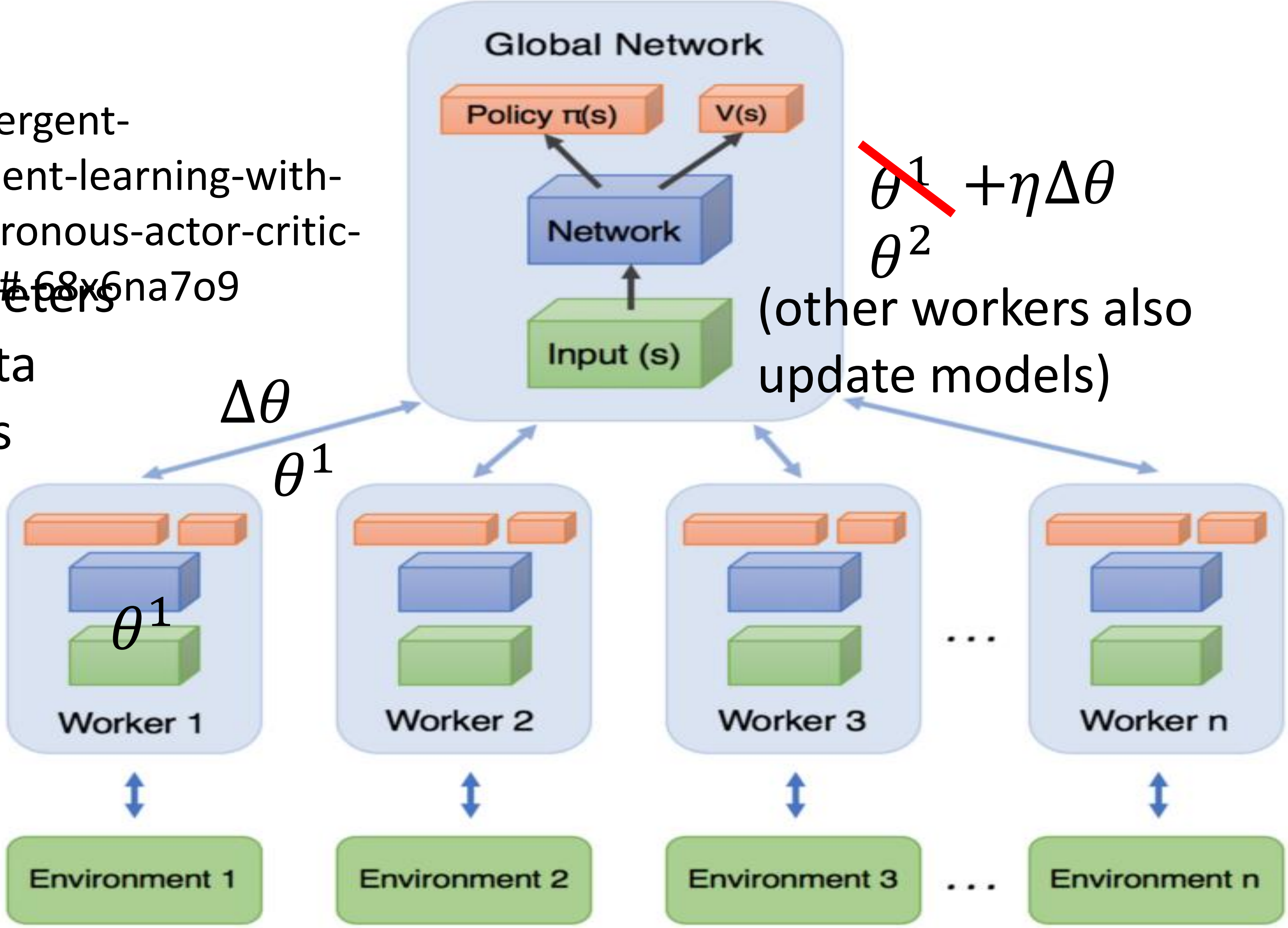


  - Use output entropy as regularization for $\pi(s)$
    - Larger entropy is preferred → exploration

Asynchronous

Source of image: https://medium.com/emergent-future/simple-reinforcement-learning-with-tensorflow-part-8-asynchronous-actor-critic-agents-a3c-c88f72a5e9f2#.68x6na7o9

1. Copy global parameters
2. Sampling some data
3. Compute gradients
4. Update global models



$\cancel{\theta^1} + \eta\Delta\theta$

θ^2

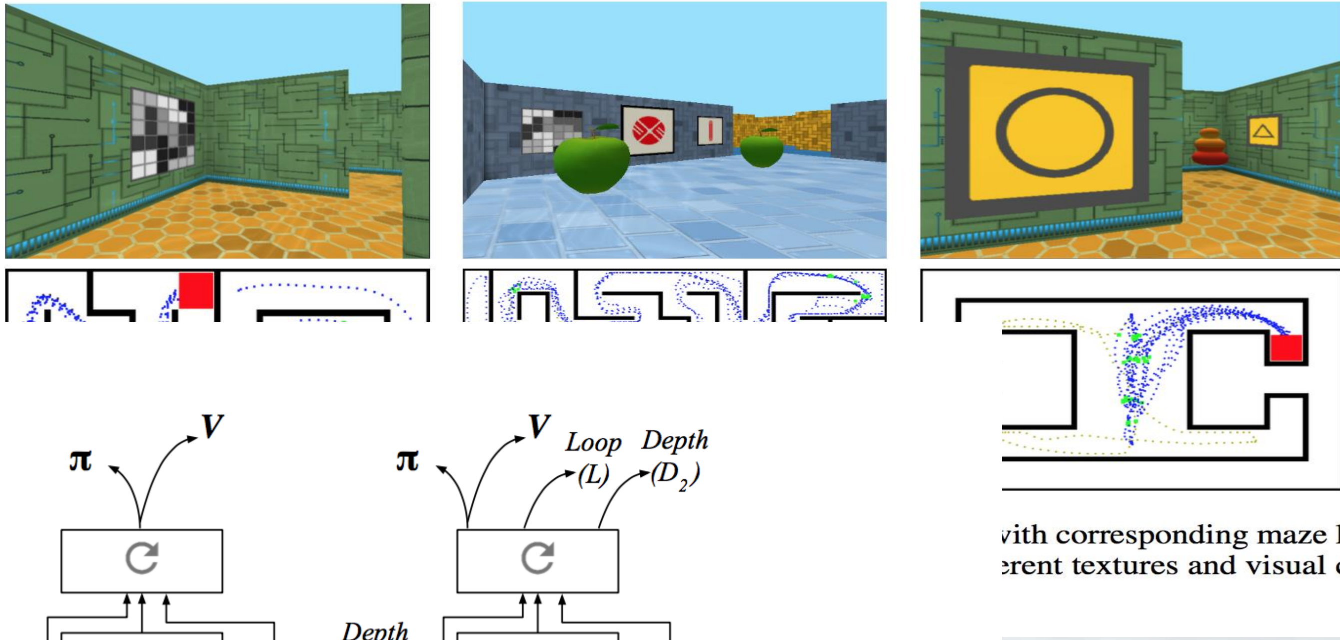
(other workers also update models)

$\Delta\theta$

θ^1

θ^1

$\Delta\theta$



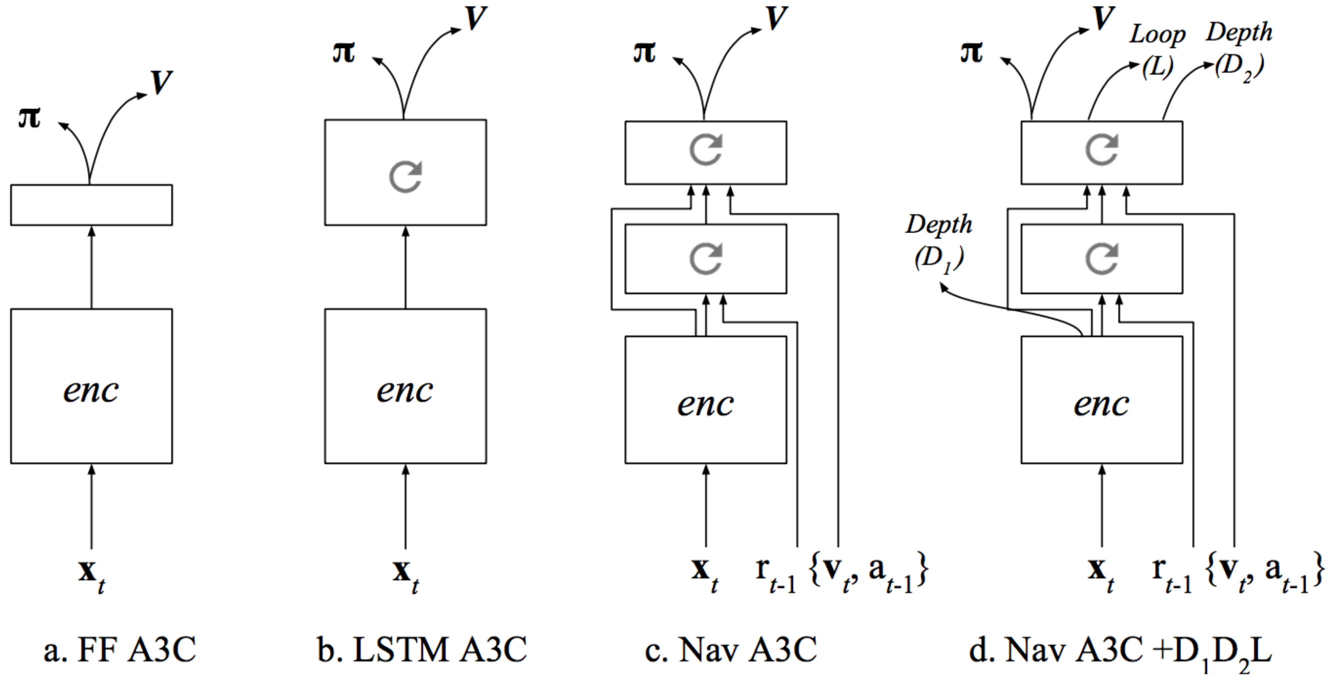
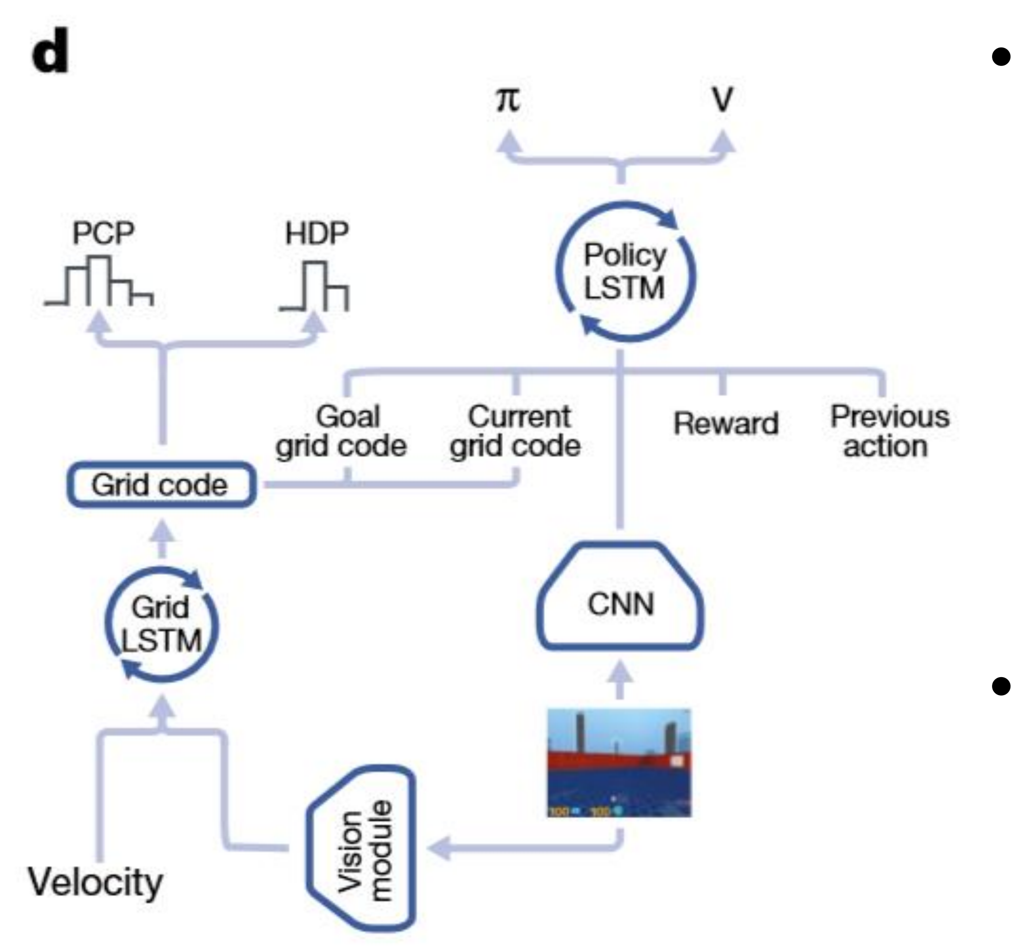


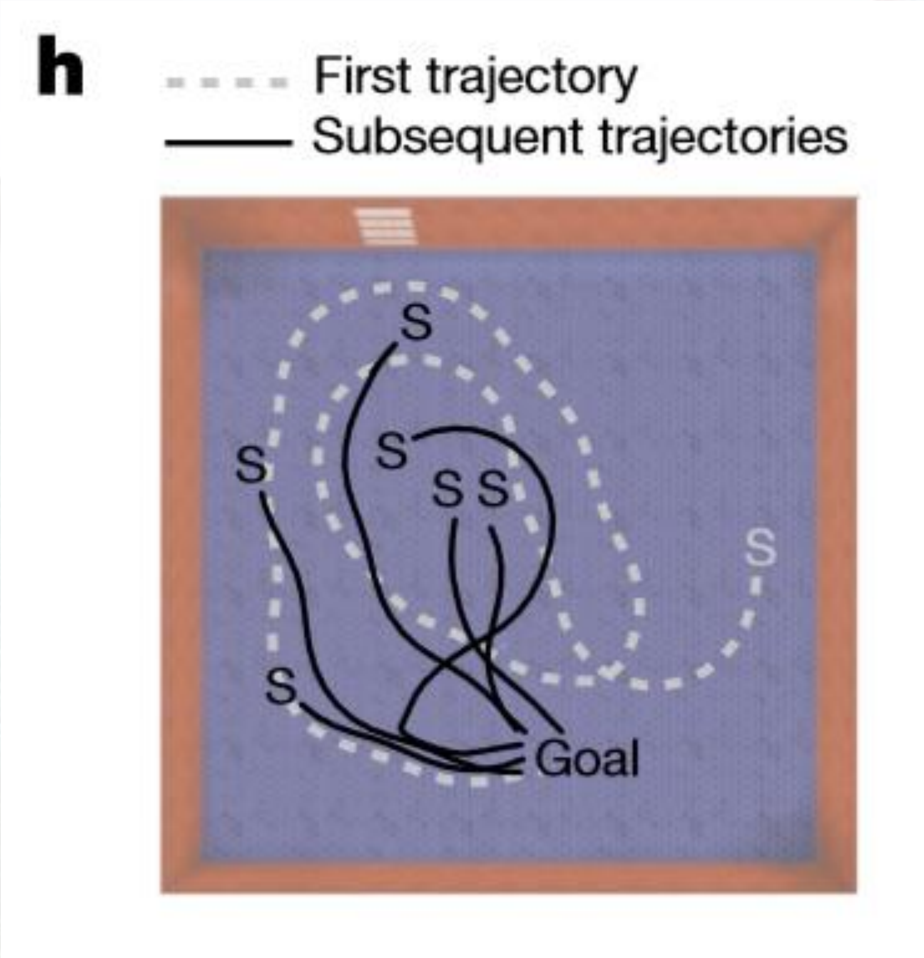
Figure 2: Different architectures: (a) is a convolutional encoder followed by a feedforward layer and policy (π) and value function outputs; (b) has an LSTM layer; (c) uses additional inputs (agent-relative velocity, reward, and action), as well as a stacked LSTM; and (d) has additional outputs to predict depth and loop closures.

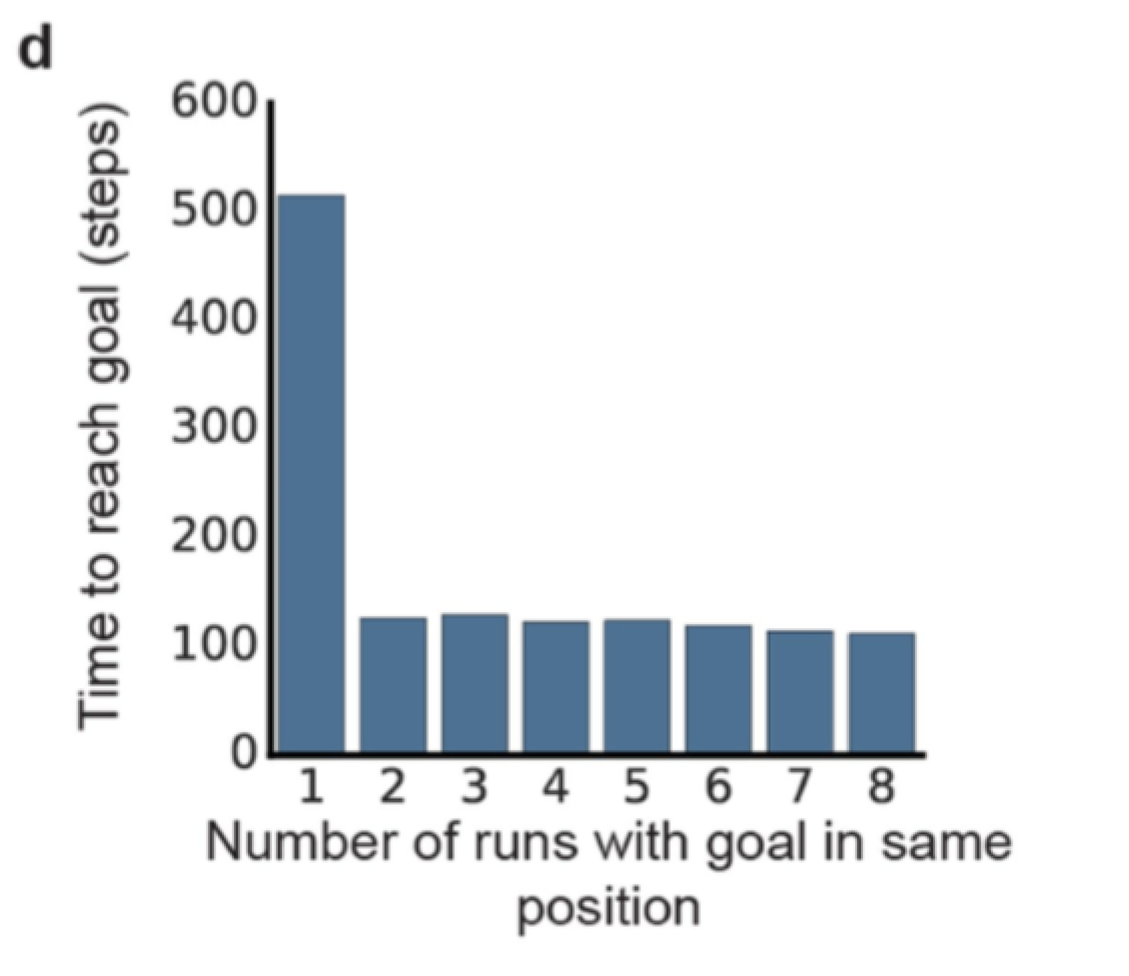
Grid cell agent: architecture



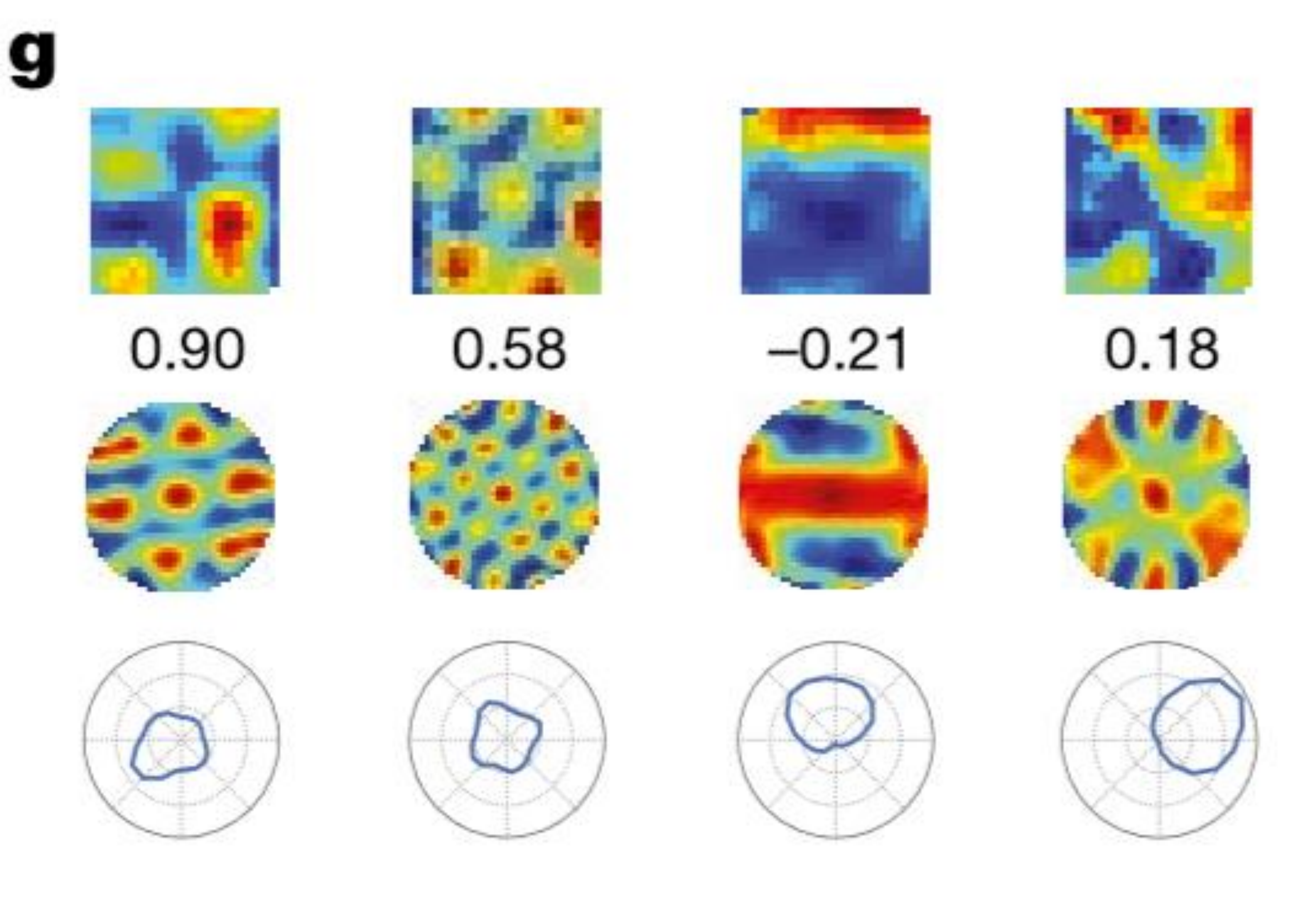
- Grid cell module
  - Input:
    - velocity(with noise)
    - Vision module(5% input)
  - Output:
  - Goal code(last time arrived)
  - Current code
- Vision module
  - Input: pixel
  - Output: place&head cell activities
  - Network designed according to [barry+ 07]

Morris water maze

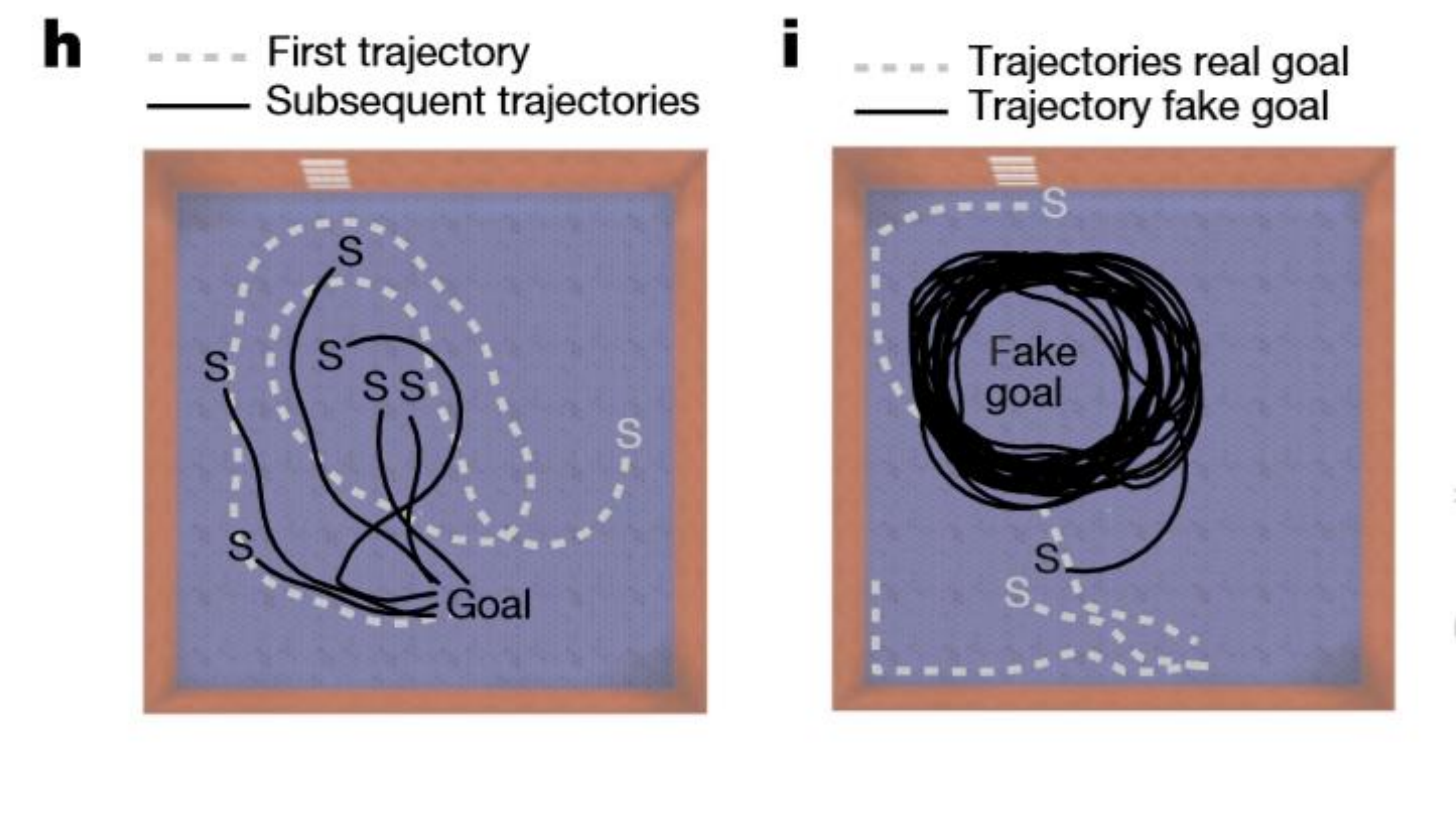




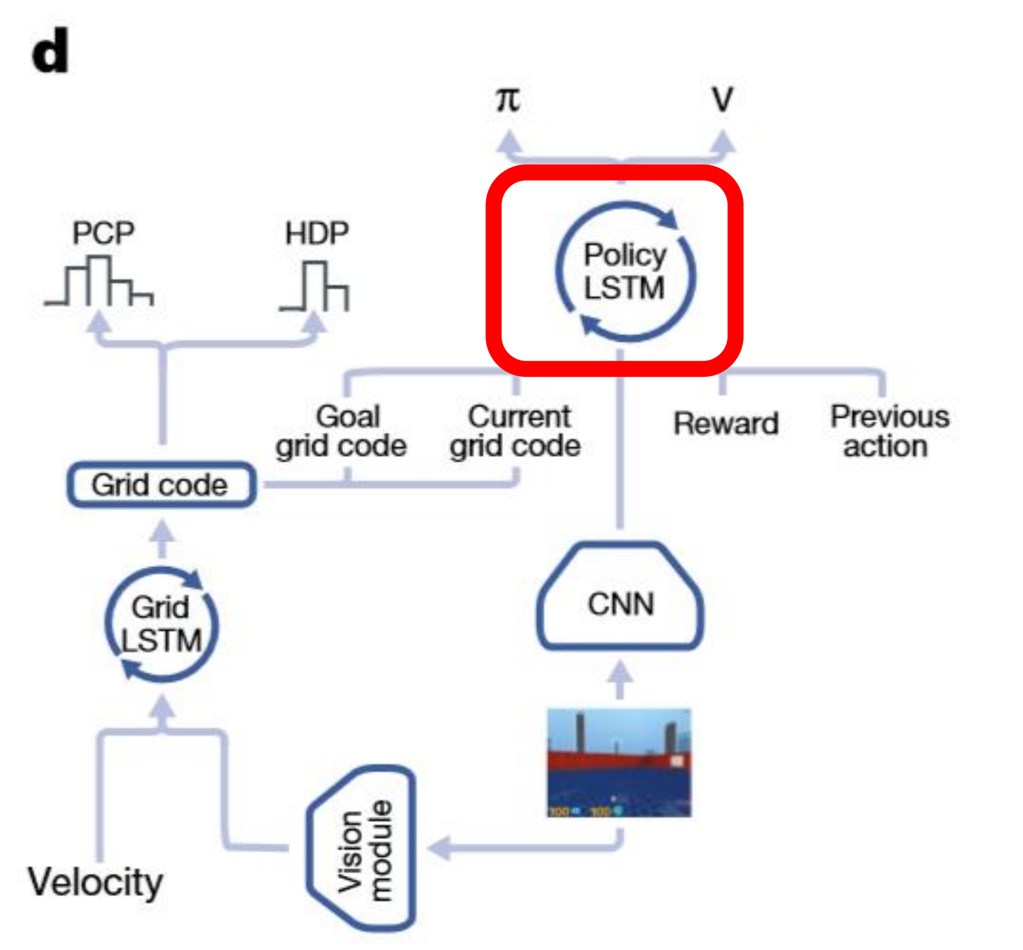
Morris water maze: grid-like units

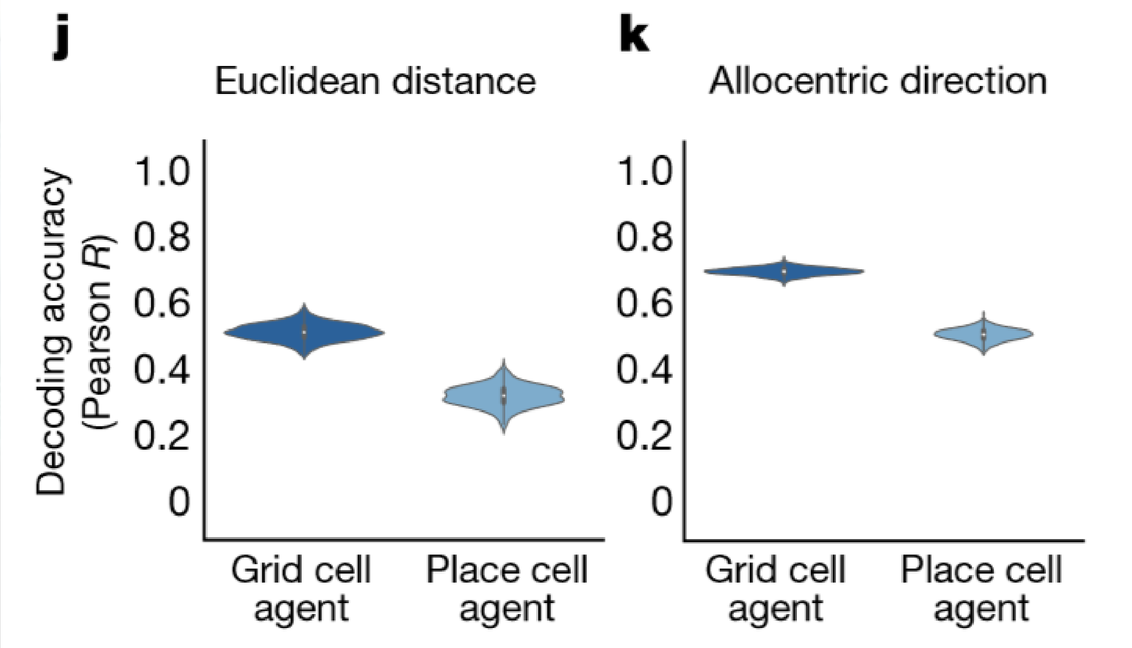


Morris water maze: “mind control”

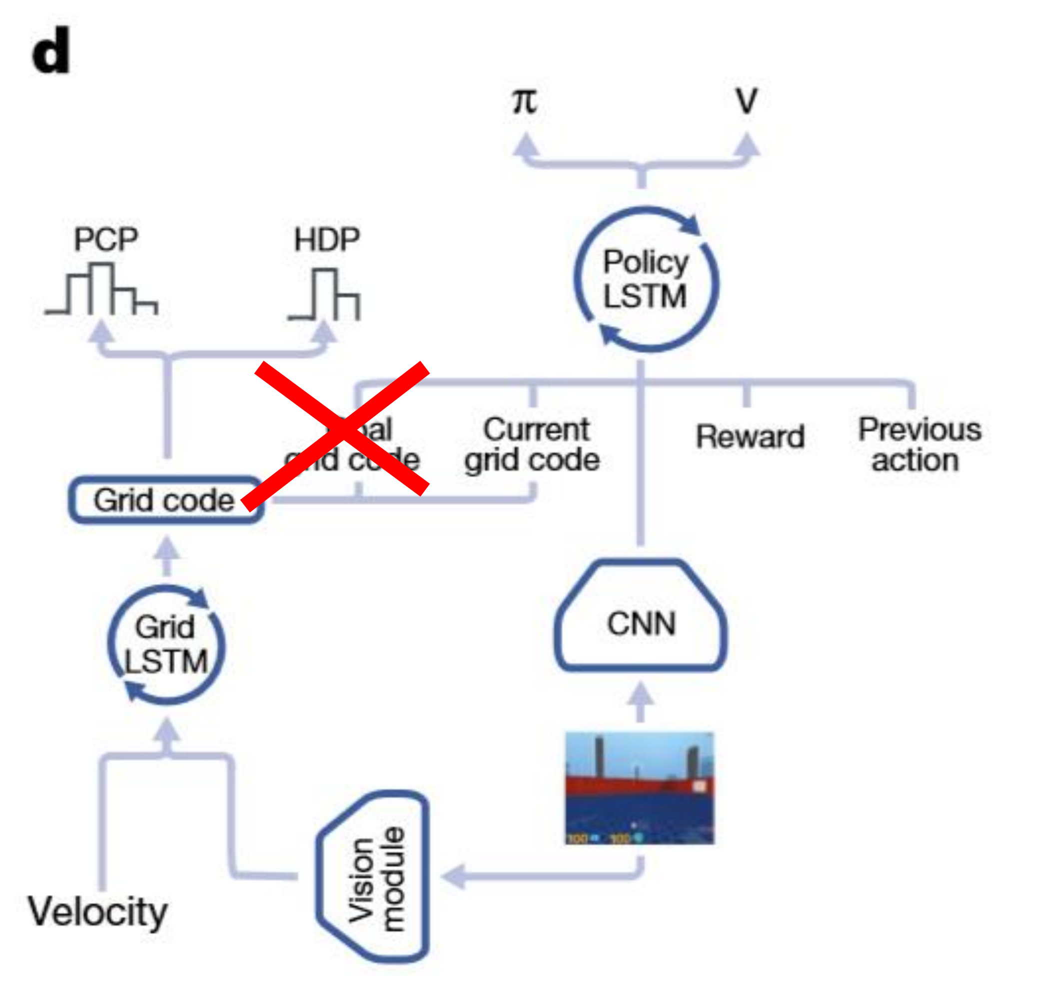


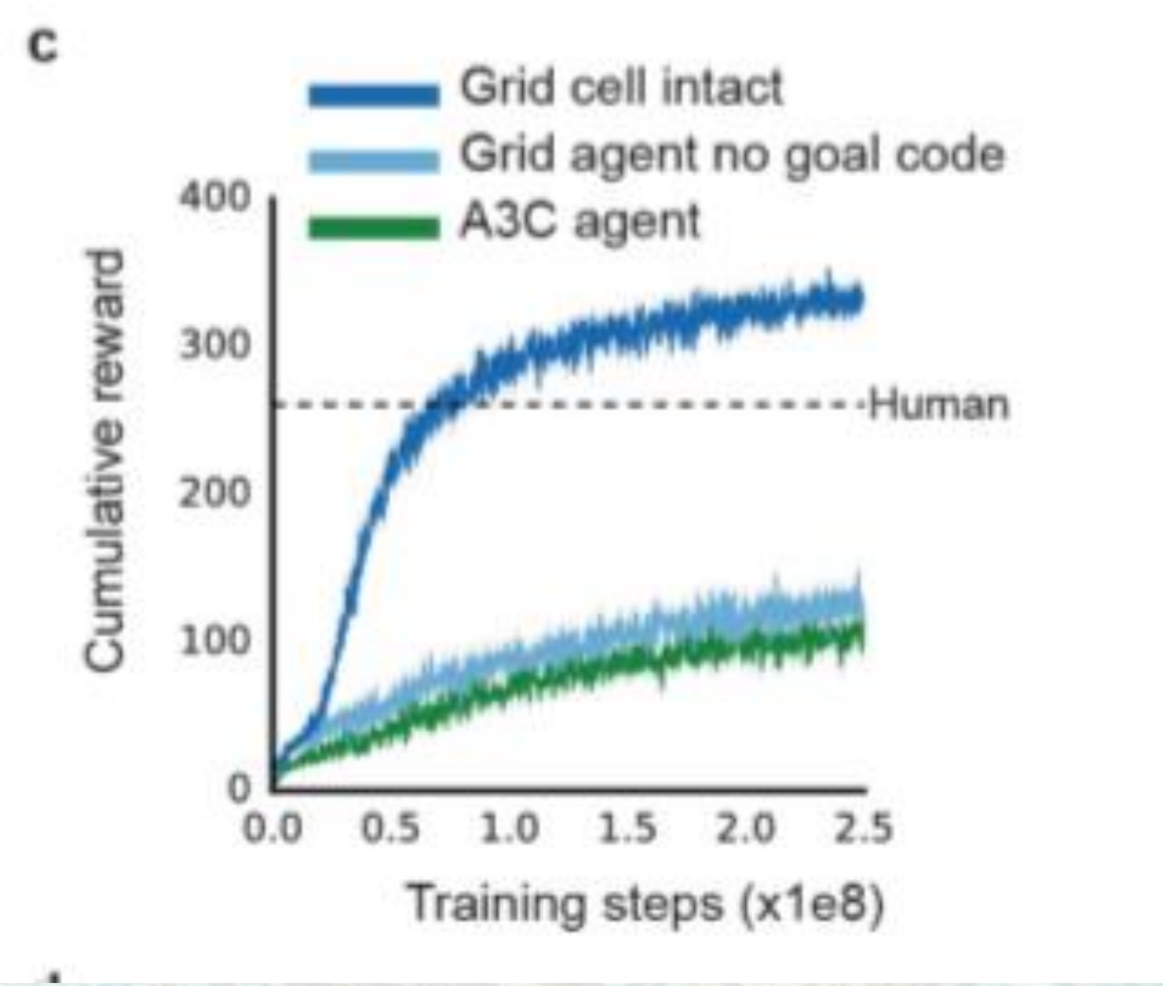
Morris water maze: multivariate decoding



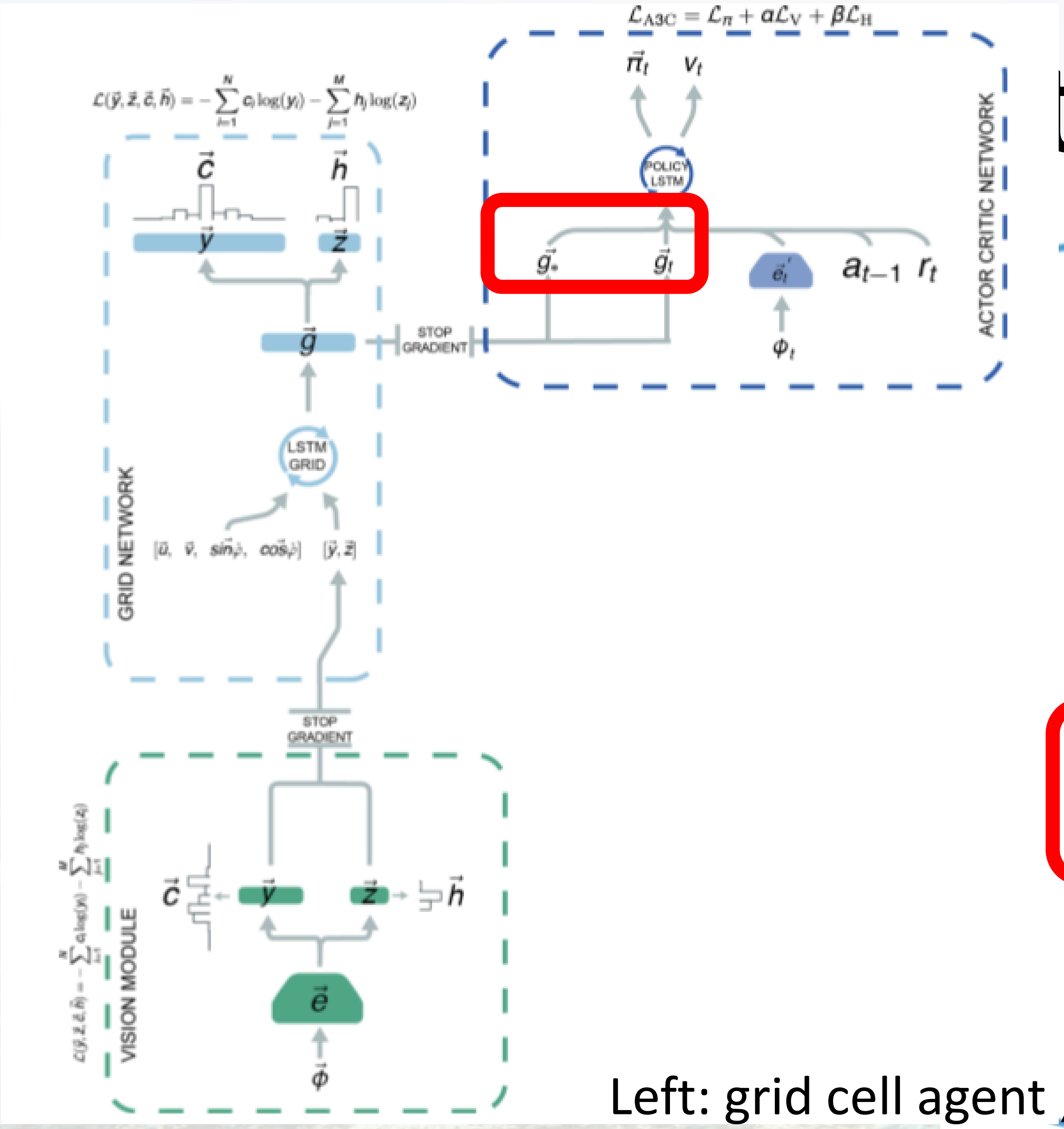


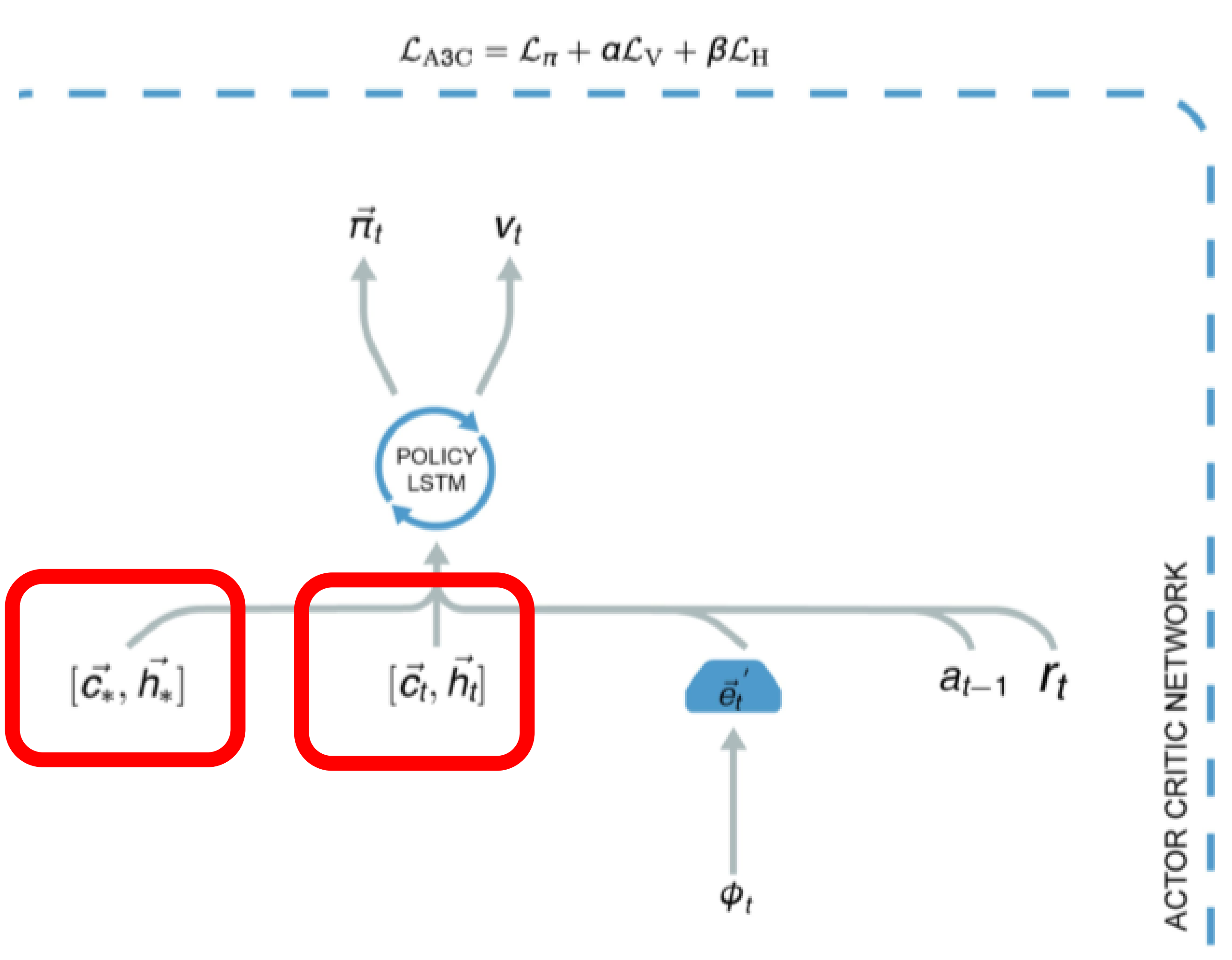
Morris water maze: goal-code ablation





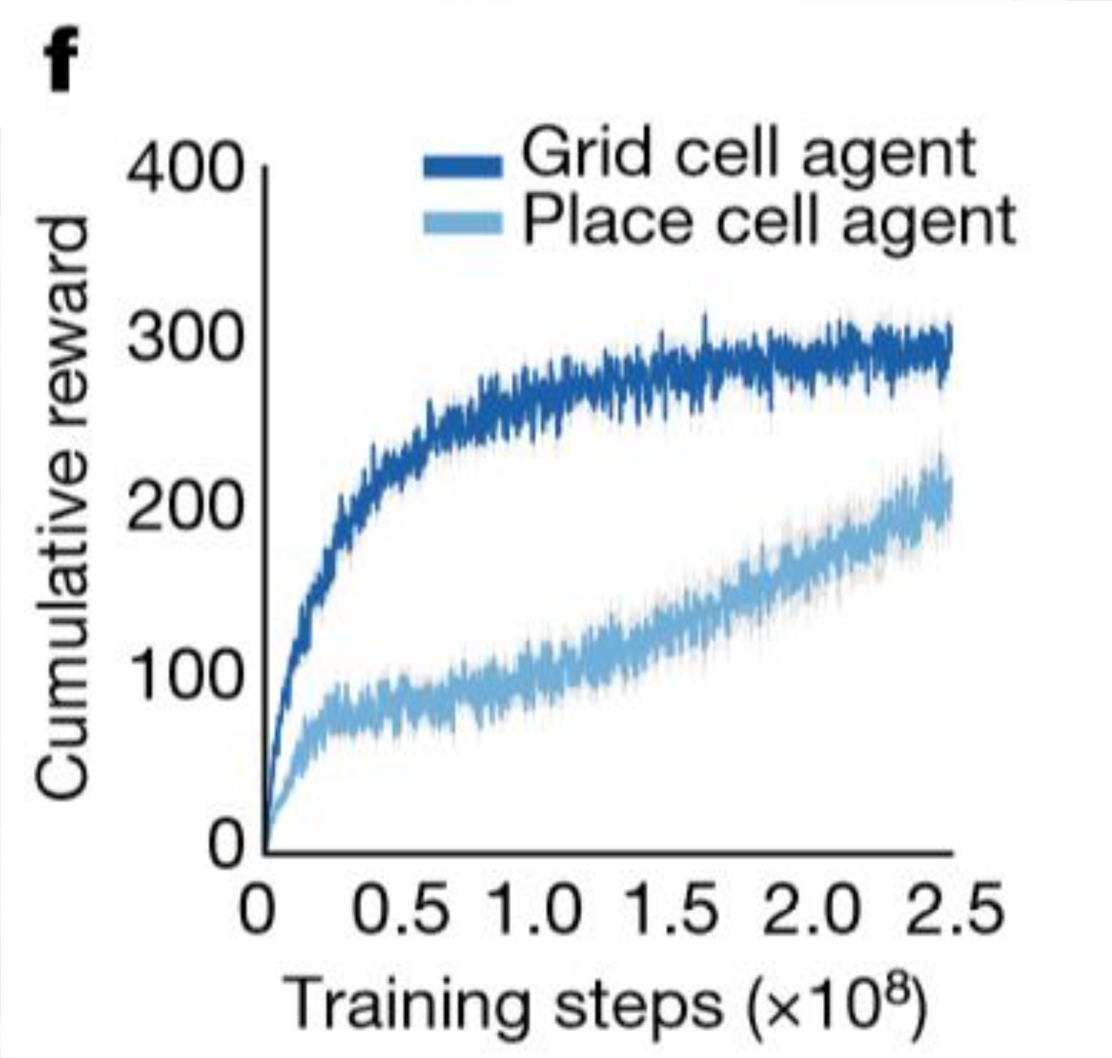
th place cell agent





Left: grid cell agent ; right : place cell agent

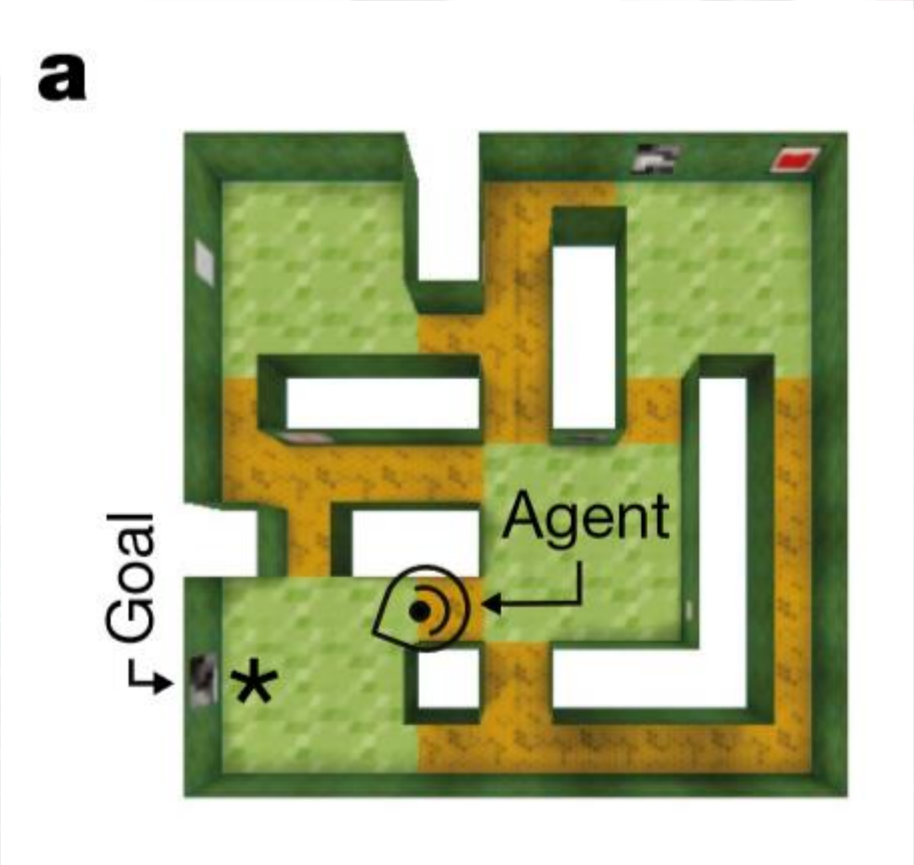
Comparison with place cell agent

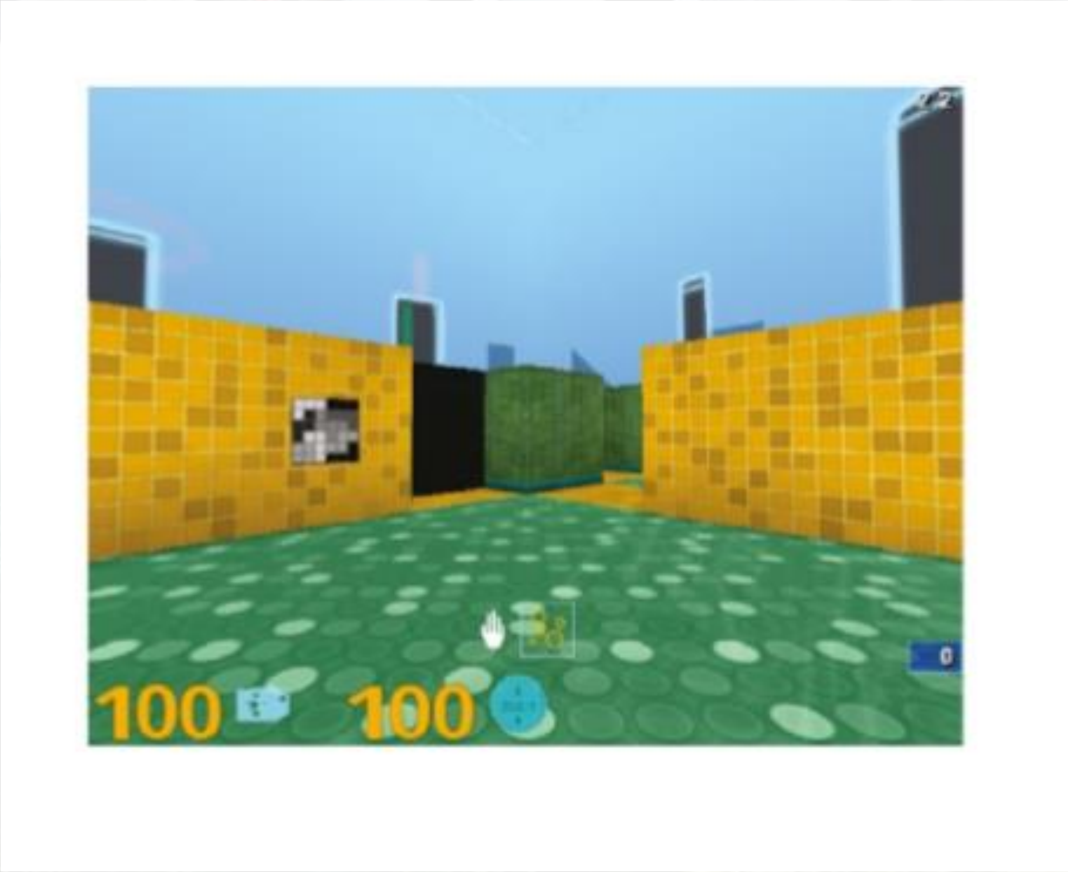


Navigation in complex ④

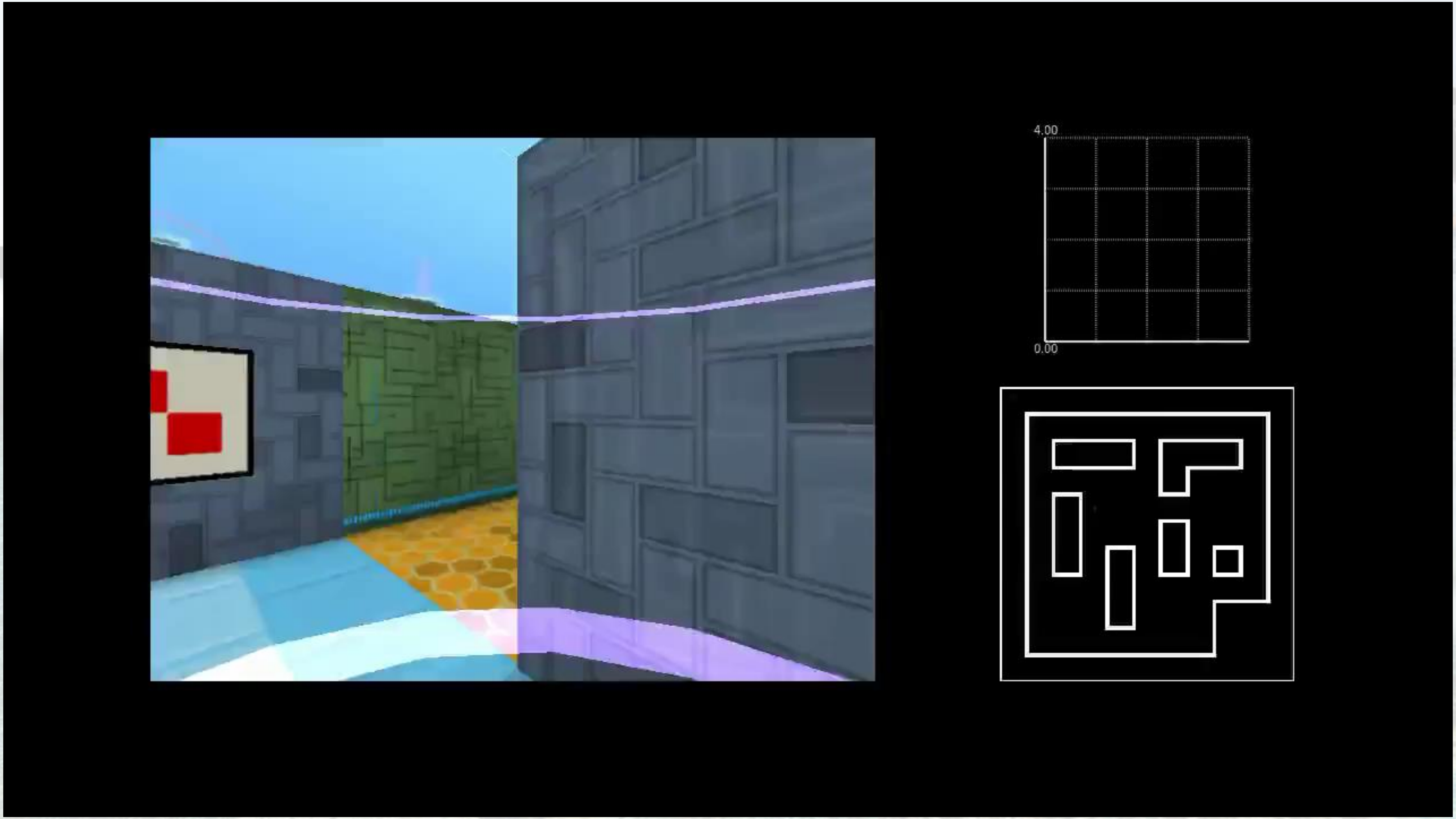
Navigation in complex env.

Complex mazes: stochastic doors



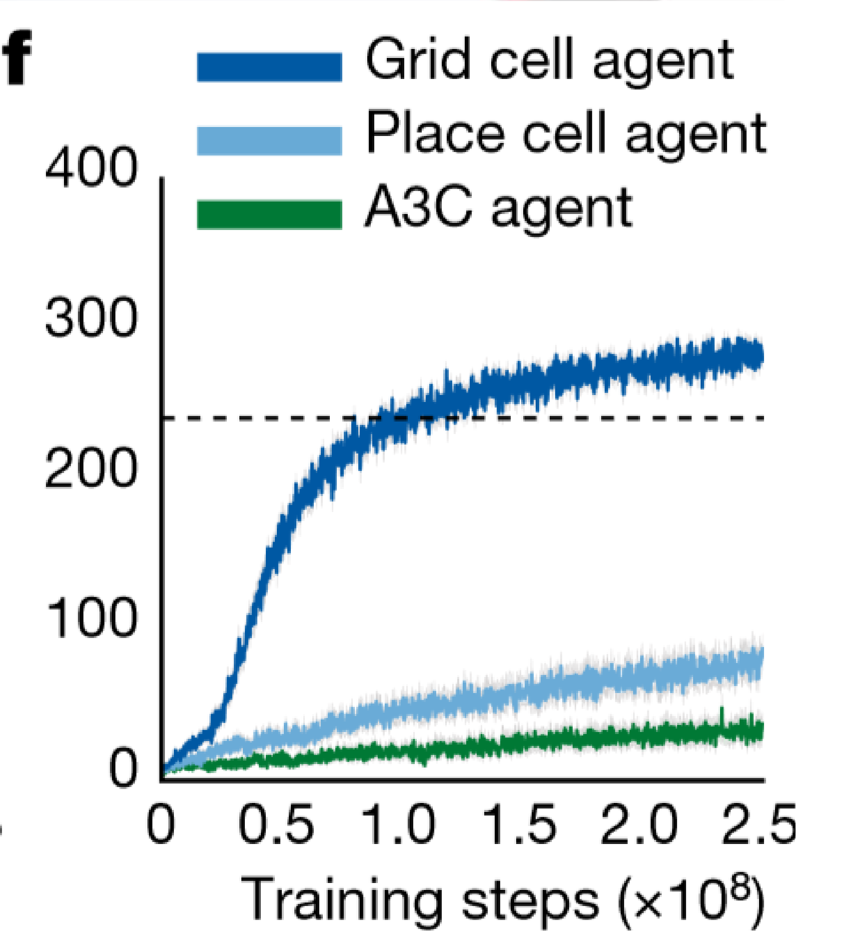


Deepmindlab: new maze configurations(colors , goal location, wall location)

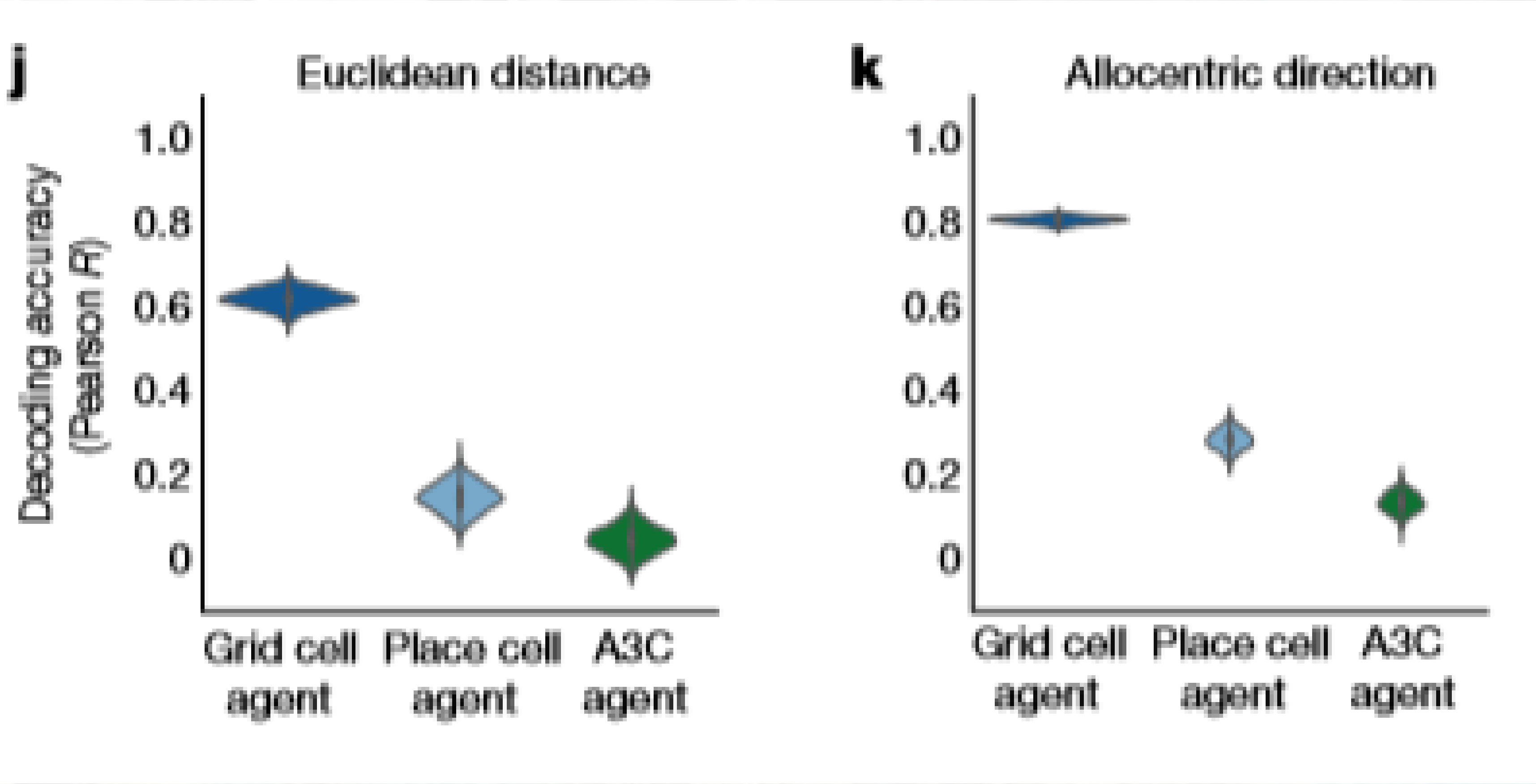

4.00
0.00

Complex mazes: analysis

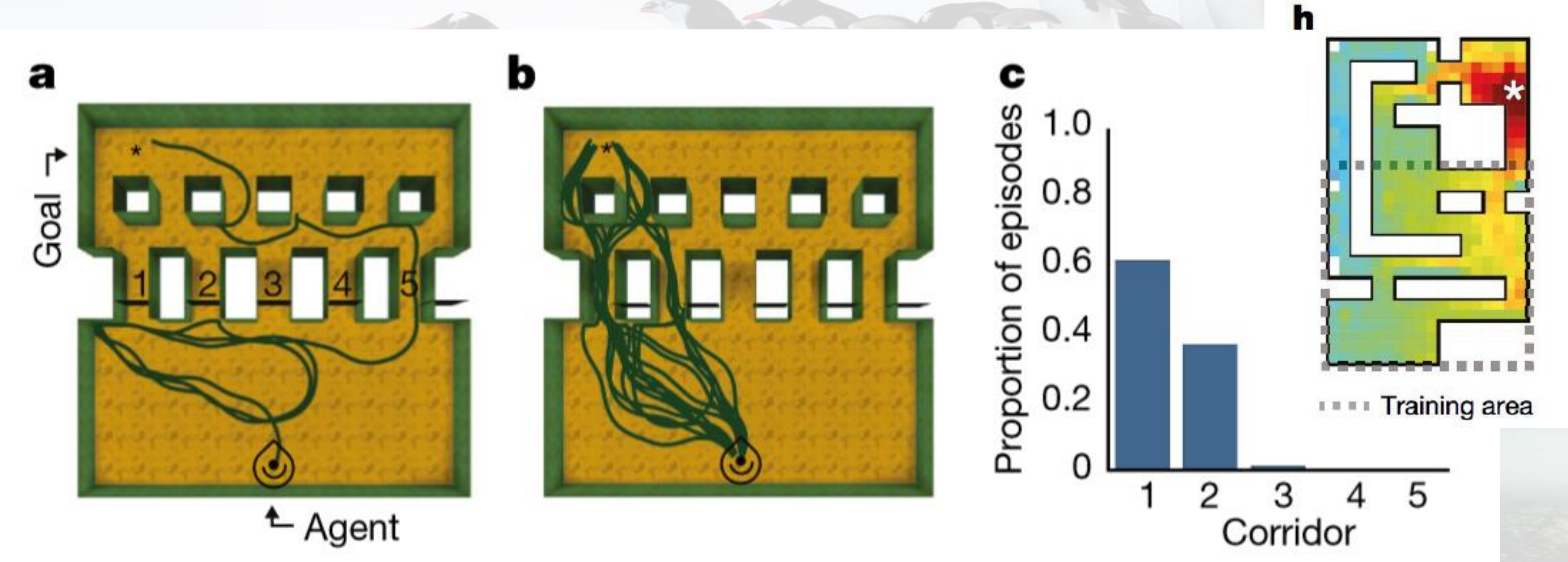
performance



Multivariate decoding



Short-cut: linearised Tolman maze



Summary part

Conclusions

Conclusion

- Designed Grid like network can perform self-localisation (path integration) and show that the grid cells emerge as a consequence of this objective
- Show that grid cells are an effective basis for vector based navigation (i.e. computing vectors within euclidean framework)

Conclusion

- 传统的SLAM技术需要构建正确、完美的环境地图，goal的位置和特征也需要外部提供，策略常需hand-coded.
- 此方法
  – 自我定位（sl rnn）得到grid cell功能
  – +RL 获得goal vector的同时可执行复杂的 policy、端到端学习、generalization能力强，且显示出探索的有效性。
- 其它研究
  认知、学习
  - Place cell & grid cell的合作组成的定位系统参与更多功能
  - Place cell如果作为deep successor representation形式加入可能构成更自动化的系统[nature neuroscience2017.11]

Review by E.moser

- It is tempting to look at this study in unidirectional fashion: using cues from biology and behavior, the authors were able to generate an artificial agent that could navigate through the environment. But the study also works in the reverse direction, and there are several new insights we neuroscientists can glean from the artificial agent. In particular, the fact that the percentages of grid cells and other functional cell types and the modular organization of grid cells simply "fall out" of the learning process suggests that perhaps we should not search endlessly for an anatomical grounding for these functional cell types. Instead, we might consider that the alternative explanation that the population is, to a large extent, a blank slate whose development is guided by experience.

  The paper highlights how the fields of neuroscience and artificial intelligence can be mutually reinforcing. In this case, the artificial agent demonstrated the power of the grid code for facilitating spatial navigation. In the future, though, such artificial agents might actually create completely new predictions for us to look for in nature.

https://f1000.com/prime/733198068?key=nvlnlWetE8dlZTv

Navigation +

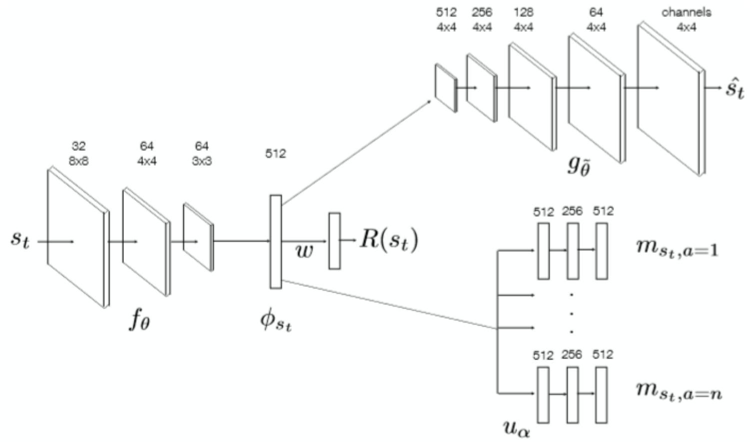


Figure 1: **Model Architecture:** DSR consists of: (1) feature branch f_θ (CNN) which takes in raw images and computes the features ϕ_{s_t}, (2) successor branch u_α which computes the SR $m_{s_t,a}$ for each possible action $a \in \mathcal{A}$, (3) a deep convolutional decoder which produces the input reconstruction $\hat{s}_t$ and (4) a linear regressor to predict instantaneous rewards at s_t. The Q-value function can be estimated by taking the inner-product of the SR with reward weights: $Q^\pi(s,a) \approx m_{sa} \cdot \mathbf{w}$.

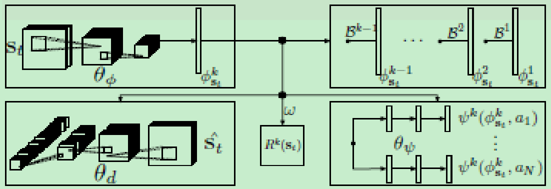
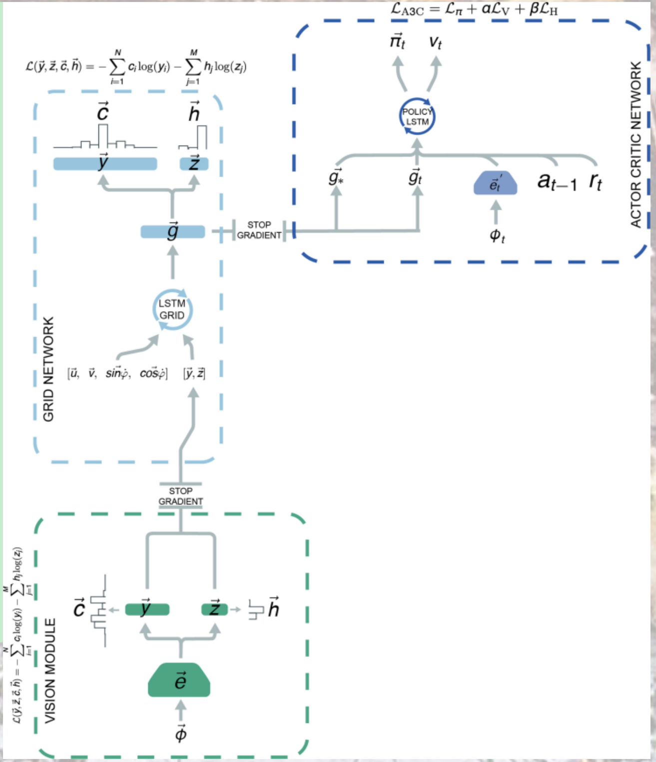


Fig. 2: Visualization of the model architecture: θ_ϕ parameterizes a convolutional network for extracting features $\phi^k_{\mathbf{s}_t}$ (k is the current target task) from $\mathbf{s}_t$ (contains three convolutional layers, with the first layer consisting of 32 8×8 filters with stride 4, the second of 64 4×4 filters with stride 2 and the 3rd of 64 3×3 filters with stride 1, each followed by a rectifying nonlinearity; the last layer is followed by one fully-connected layer with 512 units); θ_d reconstructs $\mathbf{s}_t$ back from $\phi^k_{\mathbf{s}_t}$ (contains five deconvolutional layers, with feature sizes $\{512, 256, 128, 64, 4\}$ and increasing spatial dimensionality in factors of 2); ω regresses the immediate reward $R^k(\mathbf{s}_t)$ out of the state representation $\phi^k_{\mathbf{s}_t}$; θ_ψ computes the successor features $\psi^k(\phi^k_{\mathbf{s}_t}, a_n; \theta_{\psi^k})$ for each $a_n \in \mathcal{A}$ (contains two fully-connected layers); $\mathcal{B}^i$ maps the features of the current task k back to those of the old tasks.



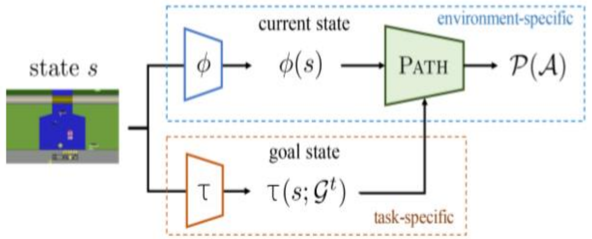


Figure 2: Proposed *universal agent*, which consists of three parts: a ϕ function mapping raw observation to feature space, a PATH function as an environment actor, and a τ function for future state planning.

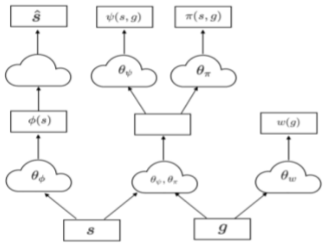


Figure 1: Model Architecture

……

201711060109
唐姝灵

We are beginning to get a clearer view of how this key brain area makes us who we are.

References

- [O'Keefe&Dostrovsky 71] O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial
- map: Preliminary evidence from unit activity in the freely-moving rat. Brain research.
- - [Hafting+ 05] Hafting, T., Fyhn, M., Molden, S., Moser, M. B., & Moser, E. I. (2005). Microstructure
- of a spatial map in the entorhinal cortex. Nature, 436(7052), 801.
- - [Gothard+ 96] Gothard, K. M., Skaggs, W. E., & McNaughton, B. L. (1996). Dynamics of mismatch
- correction in the hippocampal ensemble code for space: interaction between path integration and
- environmental cues. Journal of Neuroscience, 16(24), 8027-8040.
- - [Fiete&Brokings 08] Fiete, I. R., Burak, Y., & Brookings, T. (2008). What grid cells convey about
- rat location. Journal of Neuroscience, 28(27), 6858-6871.
- - [Barry+ 07] Barry, C., Hayman, R., Burgess, N. & Jefery, K. J. Experience-dependent rescaling of
- entorhinal grids. Nat. Neurosci. 10, 682–684 (2007).
- - [Cueva&Wei 18] Cueva, C. J., & Wei, X. X. (2018). Emergence of grid-like representations by
- training recurrent neural networks to perform spatial localization. arXiv preprint
- arXiv:1803.07770.
- [Raudies&Hasselmo 12] Raudies, F., & Hasselmo, M. E. (2012). Modeling boundary vector cell
- firing given optic flow as a cue. PLoS computational biology, 8(6), e1002553.
- Volodymyr Mnih, Adrià Puigdomènech Badia, Mehdi Mirza, Alex Graves, Timothy P. Lillicrap, Tim Harley, David Silver, Koray Kavukcuoglu, “Asynchronous Methods for Deep Reinforcement Learning”, ICML, 2016