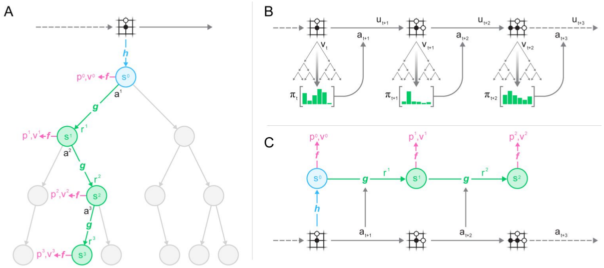
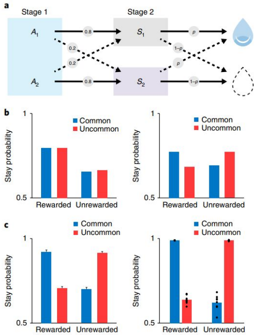
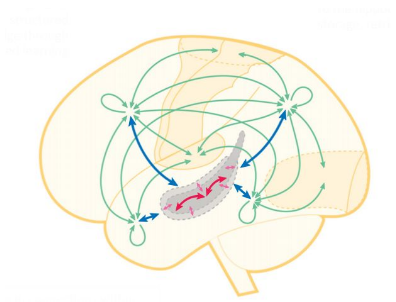
Model-Based RL

The State of the Art; the Blurred Edges of MBRL; MBRL in general intelligences





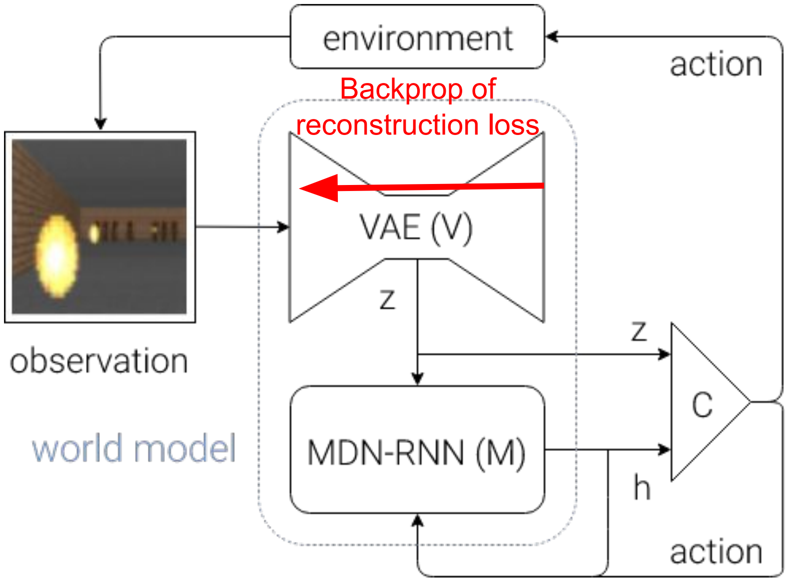


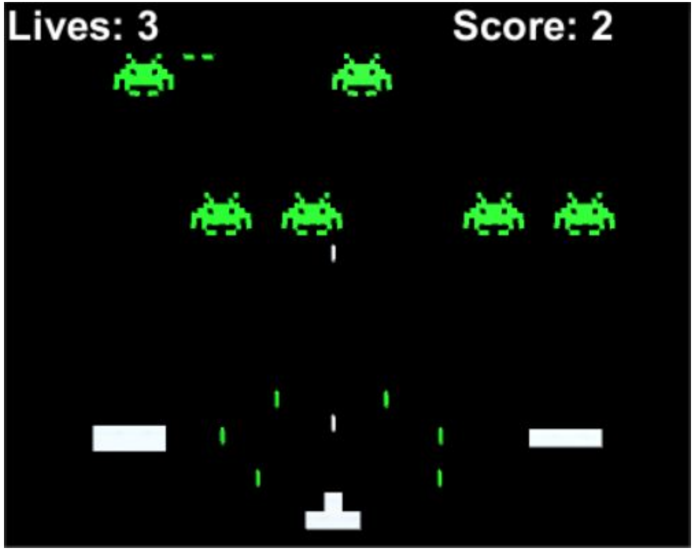
Lee Sharkey
Neural Systems and Computation MSc.
Institute of Neuroinformatics
UZH & ETHZ
leedsharkey@gmail.com

University of Zurich UZH | ETH zürich

State of the Art

What makes a good model?



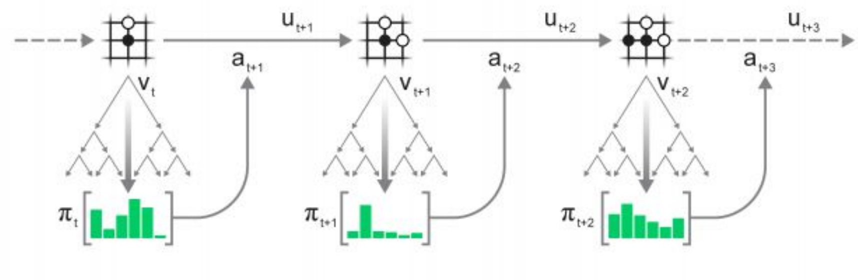


References:
Ha and Schmidhuber (2018)

State of the Art

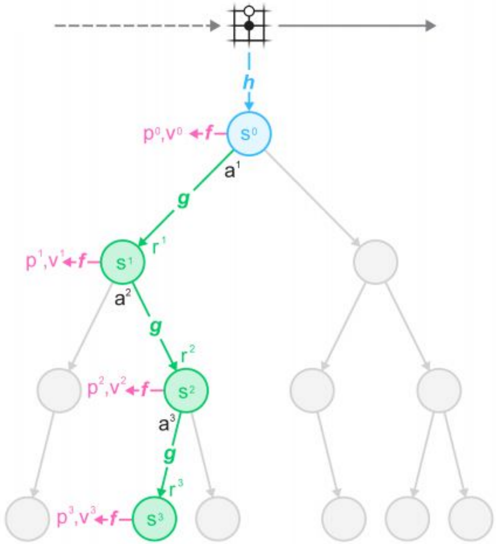
MuZero: Model-Based RL that actually works

$$\nu_t, \pi_t = MCTS(s_t^0, \mu_\theta)$$
$$a_t \sim \pi_t$$



Environment timestep $\cdot_t$

Model timestep $\cdot^k$

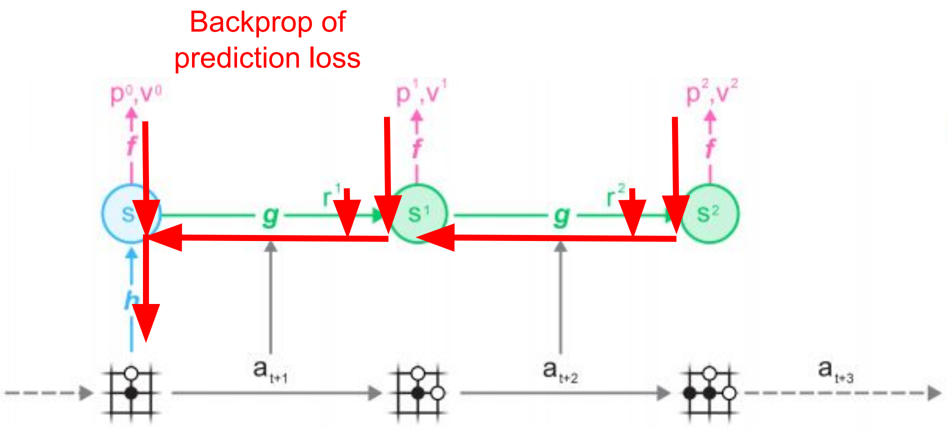


$$\begin{aligned} \text{Model} \quad & \mu_\theta = (g_\theta, f_\theta, h_\theta) \\ \text{Representation function} \quad & s^0 = h_\theta(o_1, \cdots, o_t) \\ \text{Dynamics function} \quad & r^k, s^k = g_\theta(s^{k-1}, a^k) \\ \text{Prediction function} \quad & \boldsymbol{p}^k, v^k = f_\theta(s^k) \end{aligned}$$

References:
Schritweisser et al. (2019)

State of the Art

MuZero: Model-Based RL that actually works



Learning Rule

Reward from environment

Value and policy produced by MCTS

$$\mathbf{p}_t^k, v_t^k, r_t^k = \mu_\theta(o_1, ..., o_t, a_{t+1}, ..., a_{t+k})$$

$$z_t = \begin{cases} u_T & \text{for games} \\ u_{t+1} + \gamma u_{t+2} + ... + \gamma^{n-1} u_{t+n} + \gamma^n v_{t+n} & \text{for general MDPs} \end{cases}$$

$$l_t(\theta) = \sum_{k=0}^{K} l^r(u_{t+k}, r_t^k) + l^v(z_{t+k}, v_t^k) + l^p(\pi_{t+k}, p_t^k) + c||\theta||^2$$

From model

Losses

$$l^r(u, r) = \begin{cases} 0 & \text{for games} \\ \phi(u)^T \log \mathbf{r} & \text{for general MDPs} \end{cases}$$

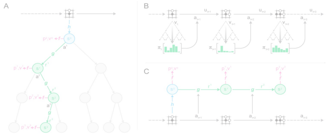
$$l^v(z, q) = \begin{cases} (z - q)^2 & \text{for games} \\ \phi(z)^T \log \mathbf{q} & \text{for general MDPs} \end{cases}$$

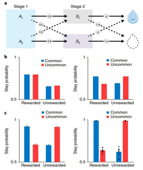
$$l^p(\pi, p) = \boldsymbol{\pi}^T \log \mathbf{p}$$

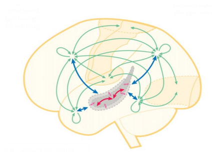
References:
Schritweisser et al. (2019)

Interlude

The State of the Art; the Blurred Edges of MBRL; MBRL in general intelligences







…[T]he actual contents of minds are tremendously, irredeemably complex; we should stop trying to find simple ways to think about the contents of minds, such as simple ways to think about space, objects, multiple agents, or symmetries. All these are part of the arbitrary, intrinsically-complex, outside world. They are not what should be built in, as their complexity is endless; instead **we should build in only the meta-methods that can find and capture this arbitrary complexity**. Essential to these methods is that they can find good approximations, but the search for them should be by our methods, not by us. We want AI agents that can discover like we can, not which contain what we have discovered. Building in our discoveries only makes it harder to see how the discovering process can be done

Richard Sutton (The Bitter Lesson; March 13, 2019)

The Blurred Edges of Model-Based RL

Why no consensus definition of MBRL?

Some properties of MBRL:

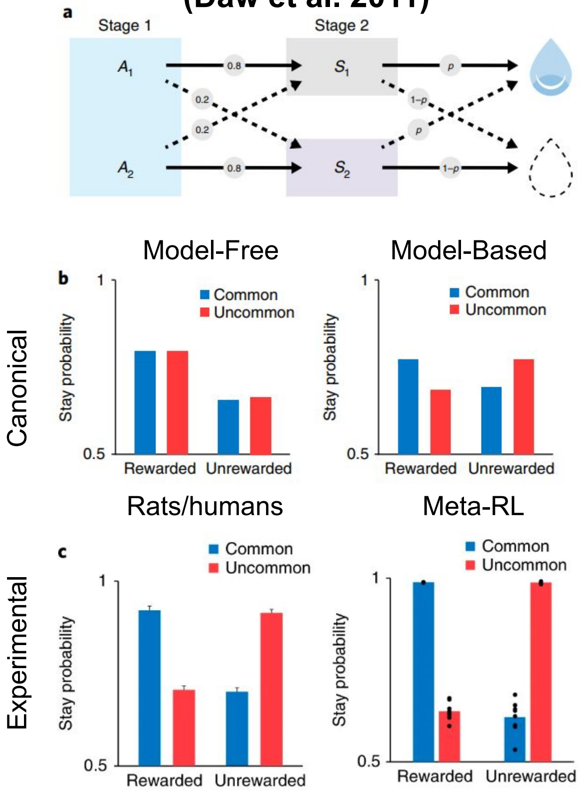
1) Using representations of task structure to select actions and predict value

2) Stricter property: Performing **explicit planning** by unrolling a forward model.

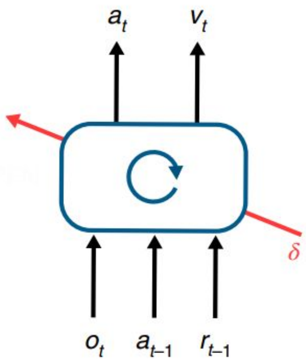
The Blurred Edges of Model-Based RL

MBRL as 1) using representations of task structure to select actions and predict value

Two-step task (Daw et al. 2011)



Meta-RL (Wang et al. 2016)



Meta-RL = Normal RL but with

1. RNN
2. trained on task distribution
3. $\{o_t, a_{t-1}, r_{t-1}\}$ as input

References:
Wang et al. (2018);
Daw et al. (2011);
Wang et al. (2016)

The Blurred Edges of Model-Based RL

Why no consensus definition of MBRL?

Some properties of MBRL:

1) Using representations of task structure to select actions and predict value

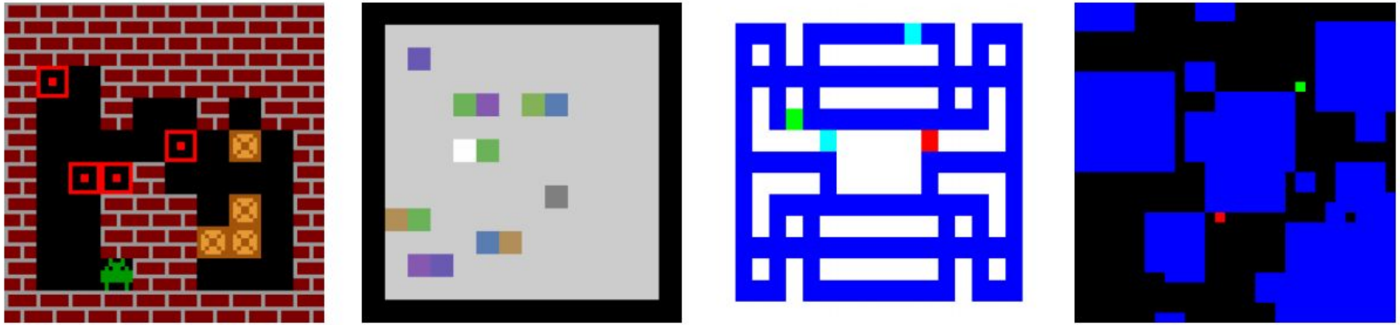
2) Stricter property: Performing **explicit planning** by unrolling a forward model.

The Blurred Edges of Model Based RL

Model-free planning

No special inductive bias toward planning, just

$$s_t = g_\theta(s_{t-1}, i_t) = \underbrace{f_\theta(f_\theta(\dots f_\theta(s_{t-1}, i_t), \dots, i_t), i_t)}_{N \text{ times}}$$



(a) Sokoban (b) Boxworld (c) MiniPacman (d) Gridworld

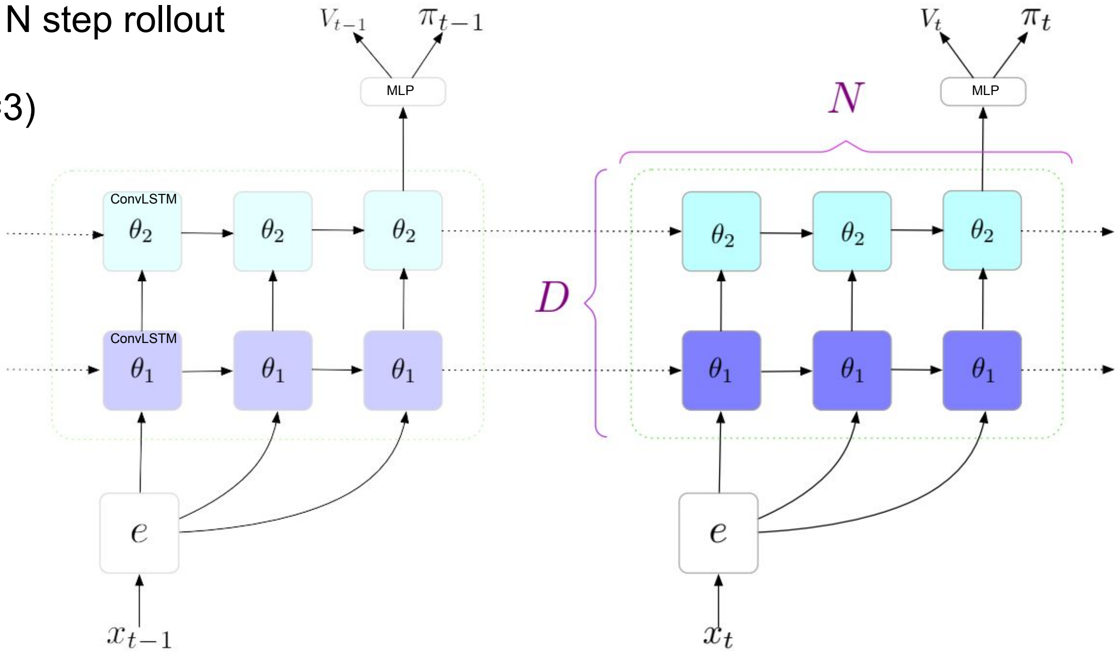
References:
Guez et al. (2018);
Tamar et al. (2016)

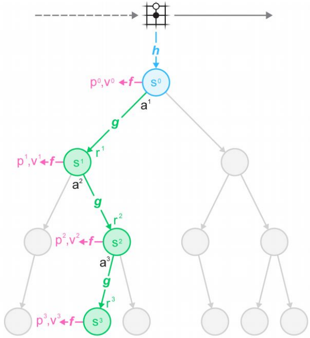
The Blurred Edges of Model Based RL

Model-free planning

Deep Repeated ConvLSTM
with depth 2 and N step rollout

i.e. DRC(D=2,N=3)





References:
Guez et al. (2018);

The Blurred Edges of Model Based RL

Model-free planning

Planner should be able to:

1. Generalize with ease to different situations
2. Learn from little experience
3. Make good use of additional thinking time

Gridworld levels

Model	% solved at $1e6$ steps	% solved at $1e7$ steps
DRC(3, 3)	30	**99**
VIN	**80**	97
CNN	3	90

Sokoban levels

Model	% solved at $2e7$ steps	% solved at $1e9$ steps
DRC(3, 3)	**80**	**99**
ResNet	14	96
CNN	25	92
I2A (unroll=15)	21	83
1D LSTM(3,3)	5	74
ATreeC	1	57
VIN	12	56

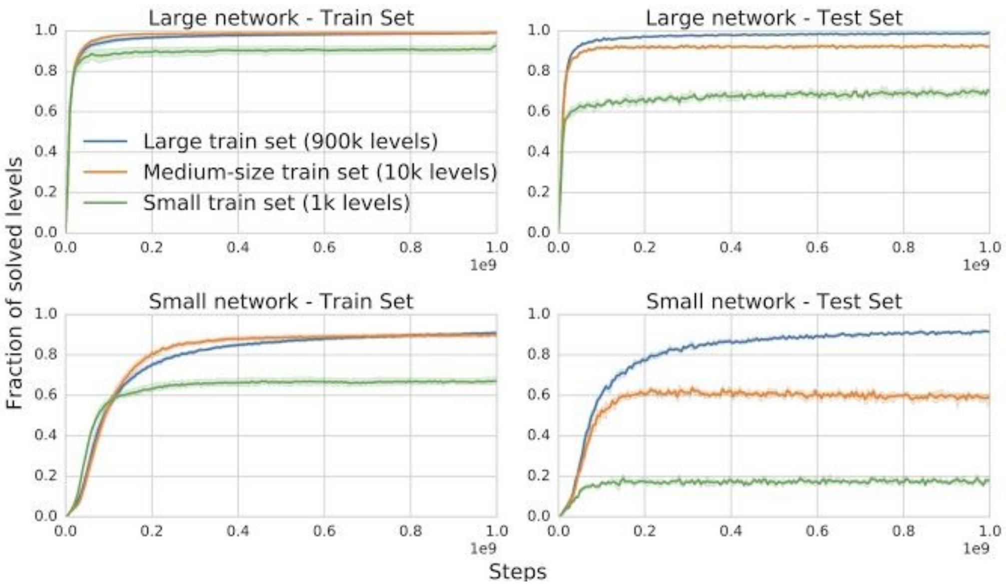
References:
Guez et al. (2018);

The Blurred Edges of Model Based RL

Model-free planning

Planner should be able to:

1. Generalize with ease to different situations
2. Learn from little experience
3. Make good use of additional thinking time



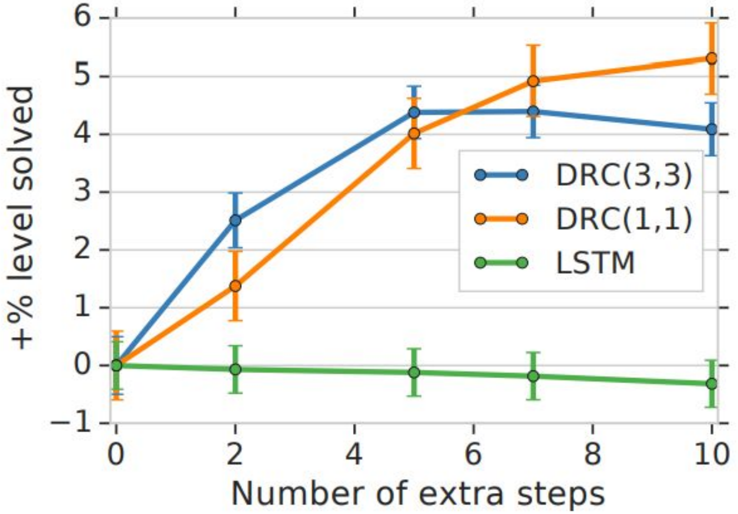
References:
Guez et al. (2018);

The Blurred Edges of Model Based RL

Model-free planning

Planner should be able to:

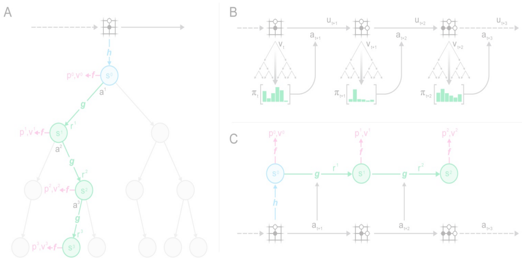
1. Generalize with ease to different situations
2. Learn from little experience
3. Make good use of additional thinking time

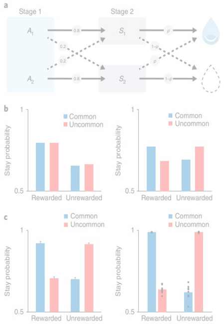


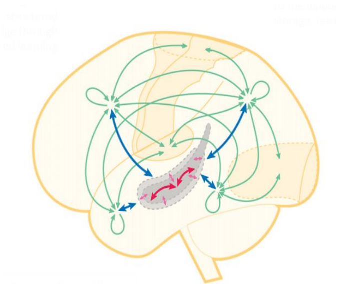
References:
Guez et al. (2018);

Interlude II

The State of the Art; the Blurred Edges of MBRL; MBRL in general intelligences





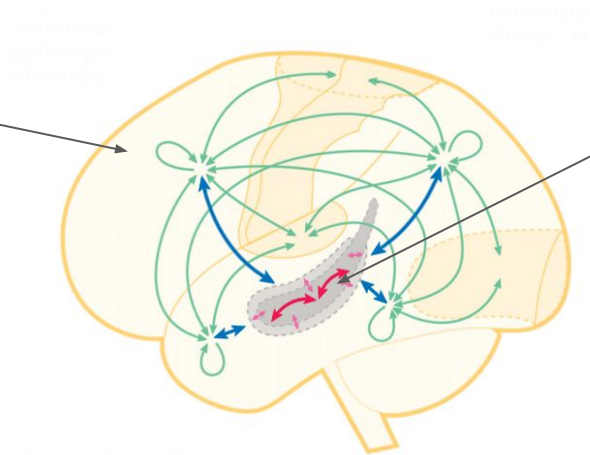


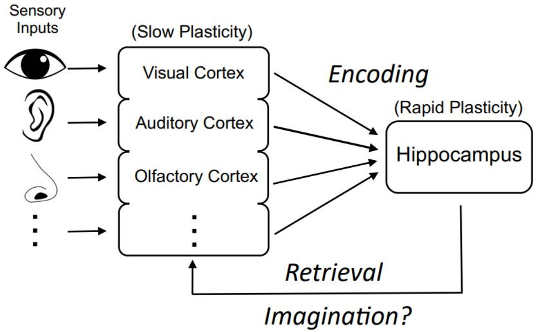
Model-Based RL in General Intelligences

Brief Intro to ‘Complementary Learning Systems’

Cortex

- Parametric model
- Slow, unsupervised learning
- Generalised features
- Many, many properties shared with deep networks (representational geometry, dynamics)





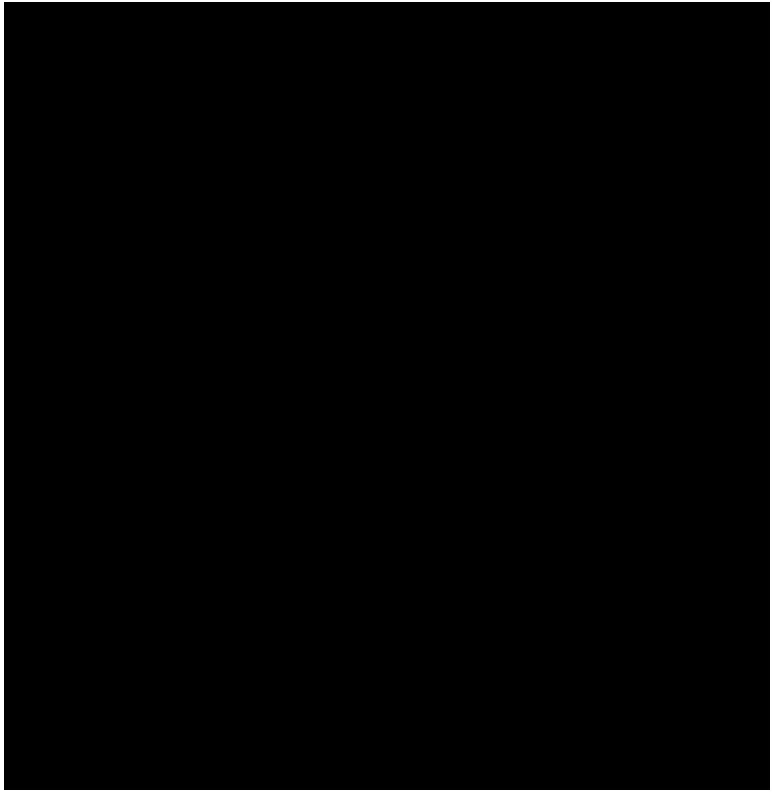
Hippocampus

- Non-parametric memory buffer
- Instantaneous learning
- Specific instances

References:
Kumaran, Hassabis, McClelland (2017);
Loren Frank (presentation fig.) (2019);
Hassabis et al. (2007);

Model-Based RL in General Intelligences

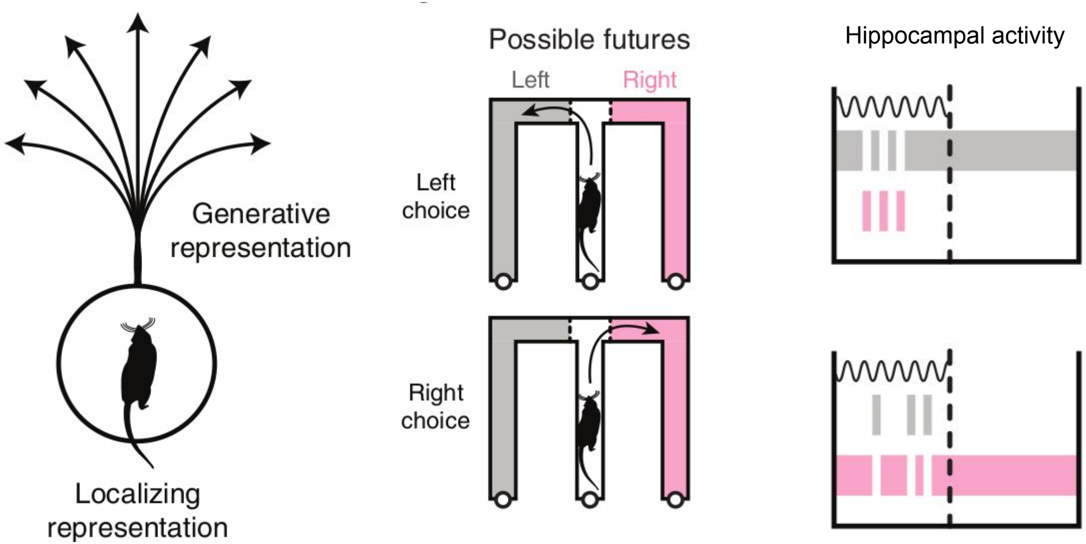
Replay and model-based planning in humans and animals



References:
Pfeiffer and Foster (2013)

Model-Based RL in General Intelligences

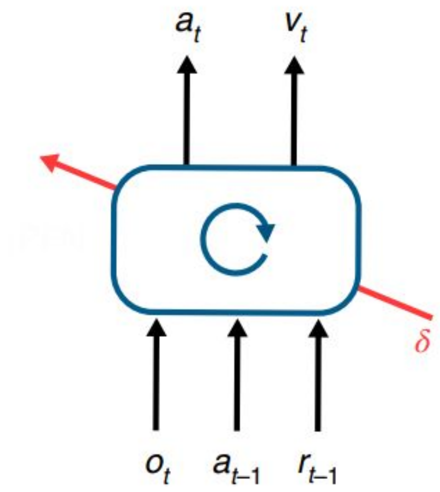
Replay and model-based planning in humans and animals

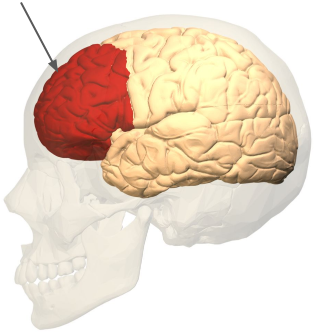


References:
Kay et al. (2020)

Model-Based RL in General Intelligences

Replay and model-based planning in humans and animals

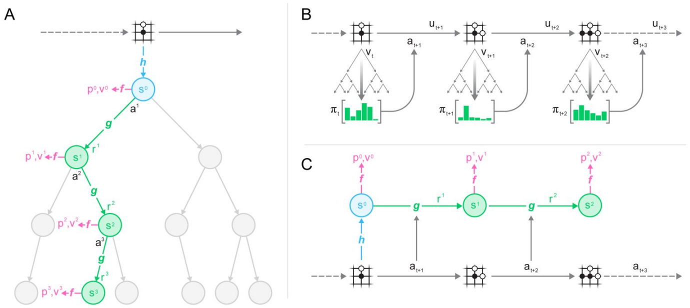


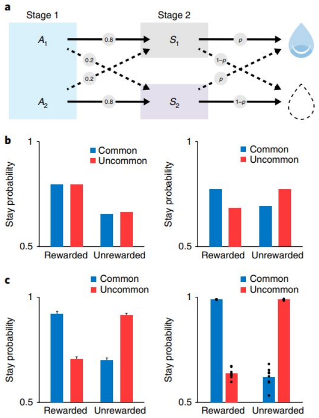


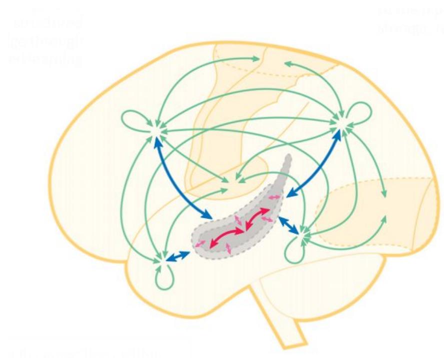
References:
Wang et al. (2018)

Thanks!

Questions?







Lee Sharkey
leedsharkey@gmail.com