

# Model-Based and Model-Free Decision-Making

Neural Modelling 2023

Georgy Antonov

Computational Neuroscience department  
Max Planck Institute for Biological Cybernetics

# Outline

- ▶ Model-based and model-free control
- ▶ Dyna
- ▶ Hippocampal replay
- ▶ Exploration
- ▶ Assignment: part 1
- ▶ Assignment: part 2
- ▶ Questions

# Model-based and model-free control

## Model-based control

- ▶ Learns a model of the environment
- ▶ Performs prospective evaluation (planning)

Pros:

- ▶ Reflective; affords behavioural flexibility

Cons:

- ▶ Expensive; slow

## Model-free control

- ▶ Learns and stores expected outcomes associated with each state-action pair

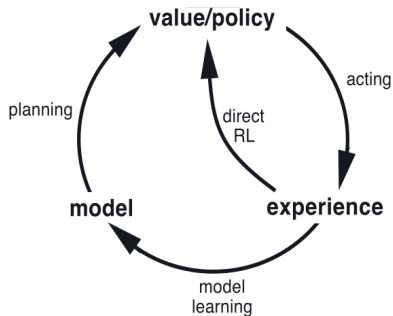
Pros:

- ▶ Reflexive; fast
- ▶ Computationally cheap

Cons:

- ▶ Stubborn; inflexible

# Dyna

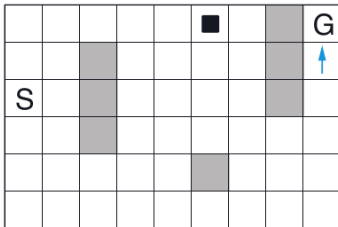


Sutton (1990)

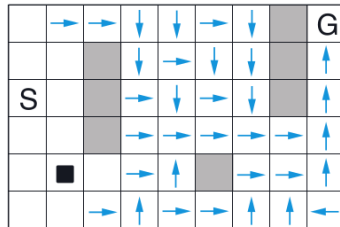
- DYNA is an integrated architecture
- Combines a *reflexive* MF policy and a *reflective* MB system
- MB system is used offline to provide additional training for MF values

# Dyna

WITHOUT PLANNING ( $n=0$ )

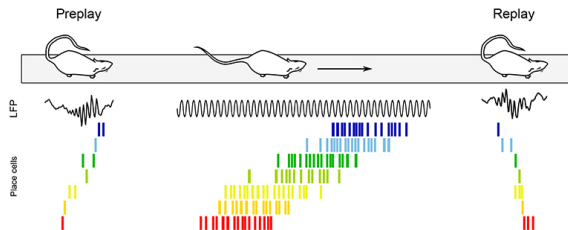


WITH PLANNING ( $n=50$ )



- Agent discovers online prediction errors (e.g., a goal)
- Model inversion (planning) to additionally train MF values

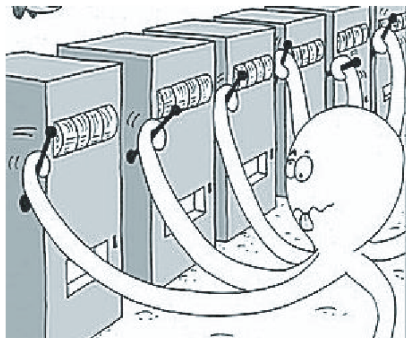
# Hippocampal replay



Drieu et al. (2019); Diba et al. (2007)

- Reinstatement of behaviourally-relevant neural activity during periods of quiet wakefulness and sleep [offline periods] (M. A. Wilson et al., 1993)
- The order of the replayed experiences is highly specific
- Forward replay seems to be predictive of the subsequent animal choices (Pfeiffer et al., 2013); reverse replay is highly sensitive to reward (Ambrose et al., 2016)

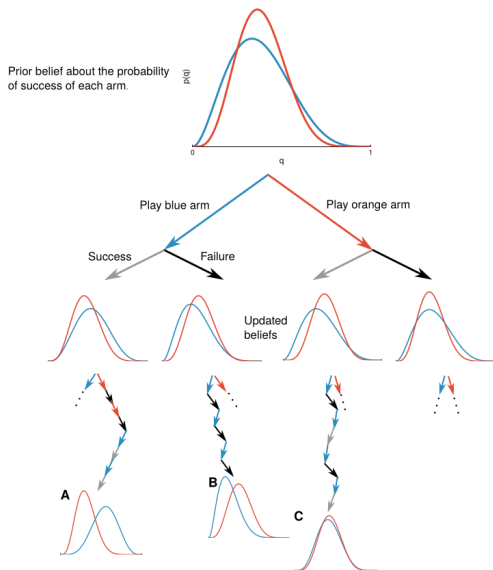
# Exploration



Source: [link](#)

- ▶ Multi-arm bandit is the classic problem for studying the exploration-exploitation tradeoff
- ▶ The objective is to maximise discounted expected reward
- ▶ Payoff probabilities are unknown
- ▶ One of the few problems for which an optimal solution is possible to compute: the Gittins index (Gittins, 1979)
- ▶ Some animals explore near-optimally (Krebs et al., 1978)

# Exploration



- Optimal exploration amounts to performing optimal control in belief space
- Belief spaces are continuous so forget about tractability in most problems more complex than simple bandits
- Good approximations exist, such as for instance BAMCP (Guez et al., 2012)



# Exploration

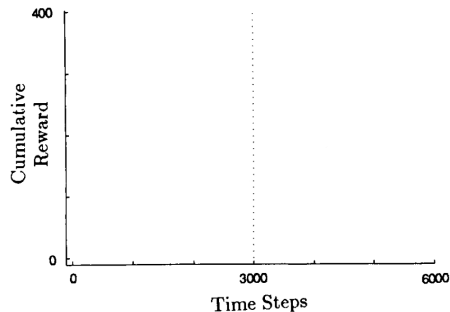
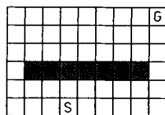
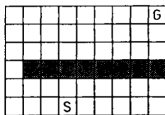
- Undirected
  - ▶  $\epsilon$ -greedy
  - ▶ Softmax (Boltzmann)
- Directed, 'optimism in the face of uncertainty'
  - ▶ upper confidence bound (Auer, 2002)

$$a = \arg \max_a \left[ Q_t(s, a) + c \sqrt{\frac{\log N(s)}{N(s, a)}} \right]$$

Exploration bonus

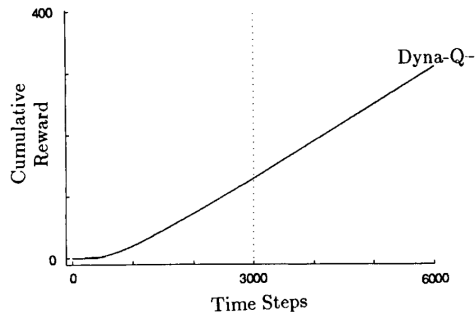
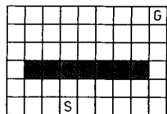
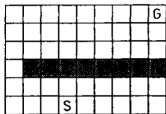
Sometimes humans' and other animals' exploration is random (undirected) (Daw, O'Doherty, et al., 2006), sometimes directed (R. C. Wilson et al., 2021)

# Exploration



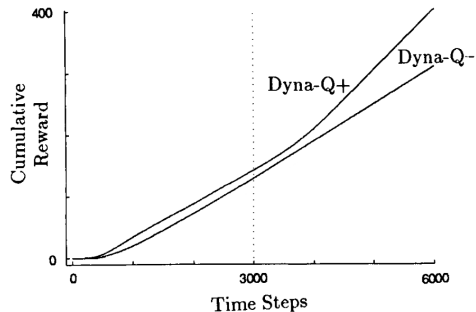
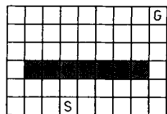
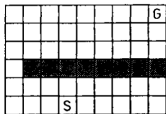
- ▶ Sutton (1990)'s changing world example
- ▶ Will a naive  $Q$ -learning agent discover the shortcut?

# Exploration



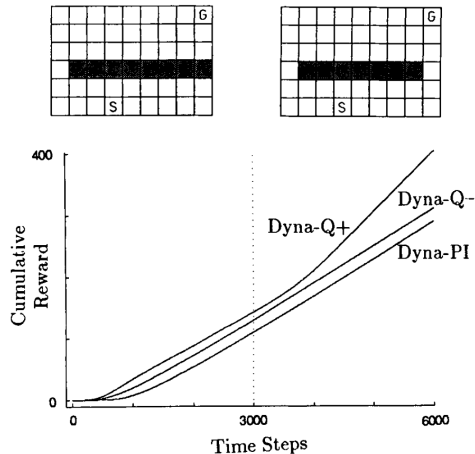
- ▶ Sutton (1990)'s changing world example
- ▶ Will a naive  $Q$ -learning agent discover the shortcut? No
- ▶ What if we encourage exploration?

# Exploration



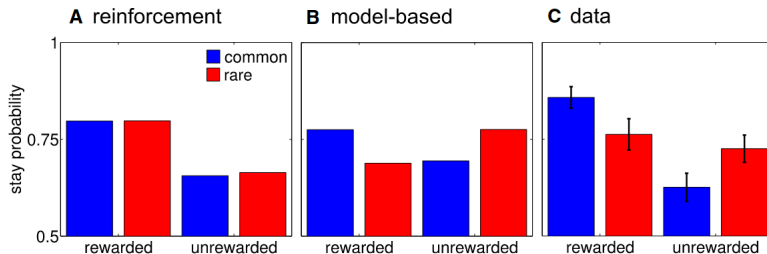
- ▶ Sutton (1990)'s changing world example
- ▶ Will a naive Q-learning agent discover the shortcut? **No**
- ▶ What if we encourage exploration? **Yes**

# Assignment: part 1



- ▶ One of the original intentions of Dyna was to improve exploration efficiency
- ▶ By incorporating an exploration bonus into the planning updates, uncertainty can propagate to distal states and therefore encourage exploration
- ▶ Your task is to reproduce this figure; focus only on Dyna-Q+ and Dyna-Q-

# Assignment: part 2



- The iconic RL task (Daw, Gershman, et al., 2011) to probe the relative contributions of MB and MF control to subjects' choices
- In this part of the assignment, your task is to reproduce the above figure

# Questions?

- ▶ You will find the assignment and all the necessary details in my github repository:  
[https://github.com/geoant1/GTC\\_Neural\\_Modelling\\_Tutorial](https://github.com/geoant1/GTC_Neural_Modelling_Tutorial)
- ▶ For part 1 the code is already written for you; the task is to fill in the missing implementation
- ▶ For part 2 you have to write most of the code yourself