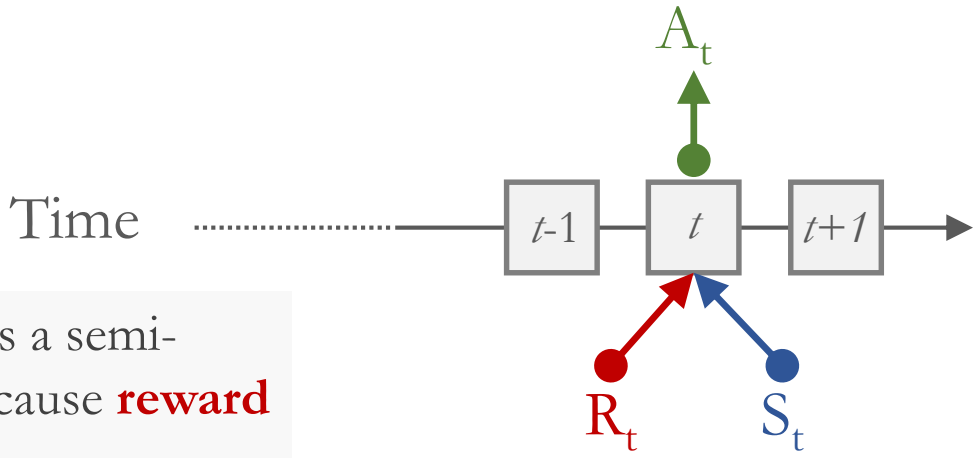
Reinforcement learning

Introduction to reinforcement learning and deep reinforcement learning

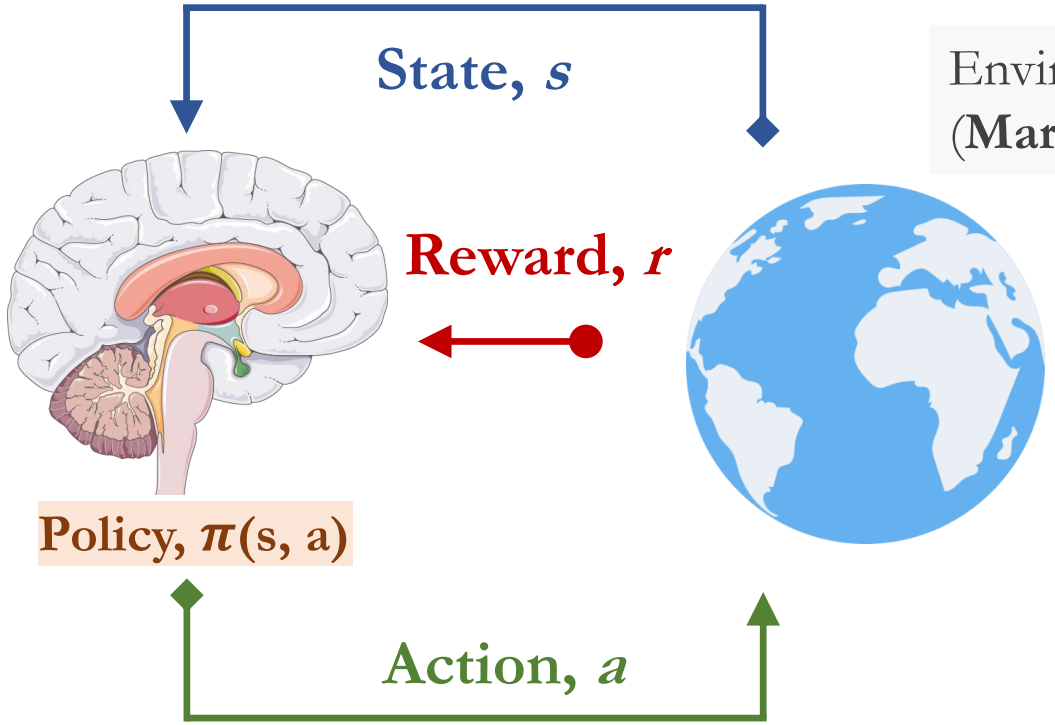
Markov Decision Process (MDP)

Reinforcement learning

Reinforcement learning is a framework for learning how to interact with the environment from experience.



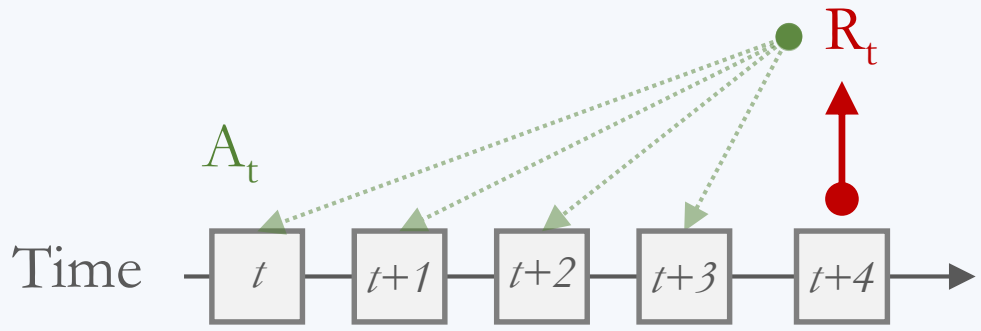
Most of the time, RL is a semi-supervised learning because **reward** is time-delayed



Environment is modelled as probabilistic (**Markov Decision Process, MDP**)

Exploration | Exploitation

Credit Assignment Problem



Source: https://www.youtube.com/watch?v=0MNVhXEX9to

Key concepts

Model: predict what the environment will do next.

$$p(s', r|s, a) = P(S_t = s', R_t = r|S_{t-1} = s, A_{t-1} = a)]$$

Value function: prediction of expected rewards.

$$\vartheta_\pi(s) = \mathbb{E}[R_t + \gamma R_{t+1} + \gamma^2 R_{t+2} + \cdots | S_t = s]$$

Discount rate

The value of a state **s** given a policy **π** is my expectation of how much reward I will get in the future if I start in that state and enact that policy.

Policy: how the agent pick its actions.

Deterministic	$a = \pi(s)$
Stochastic	$a \sim \pi(a|s)$

Policy learning | Value learning

Q-learning

$$Q^\pi(s, a) = \text{quality of state/action pair}$$

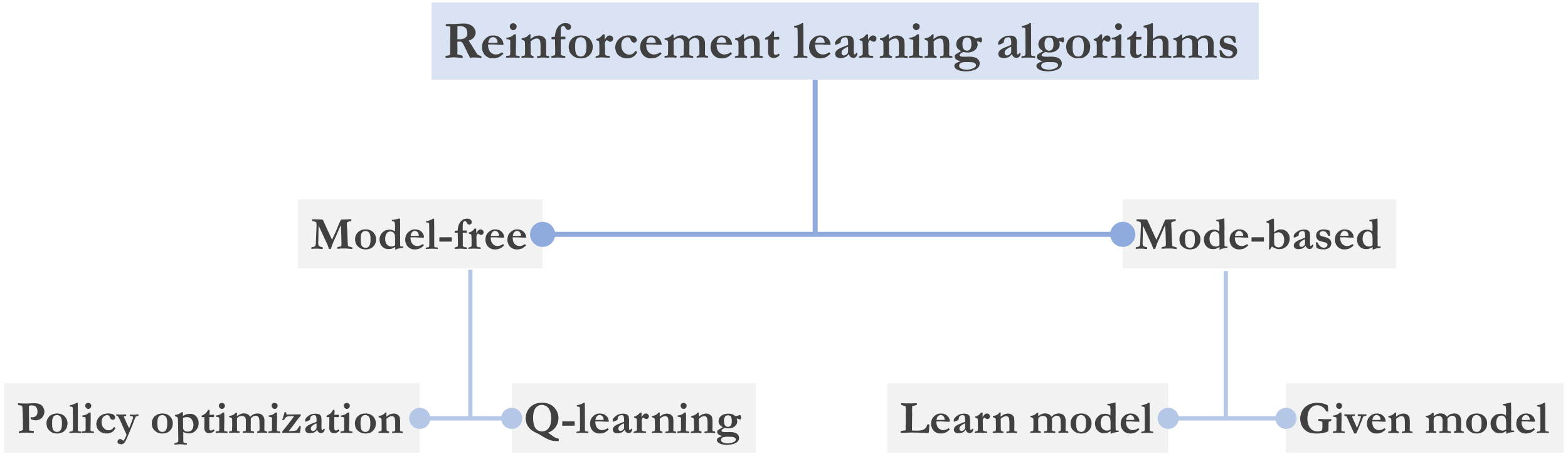
$$Q(s, a) = Q^{old}(s_t, a_t) + \alpha(r_t + max_a Q(S_{t+1}, a) - Q^{old}(s_t, a_t))$$

Given a state **s** and an action **a**, and assuming that I will do the best thing I can in the future, what is the quality of being in that state and taking that action.

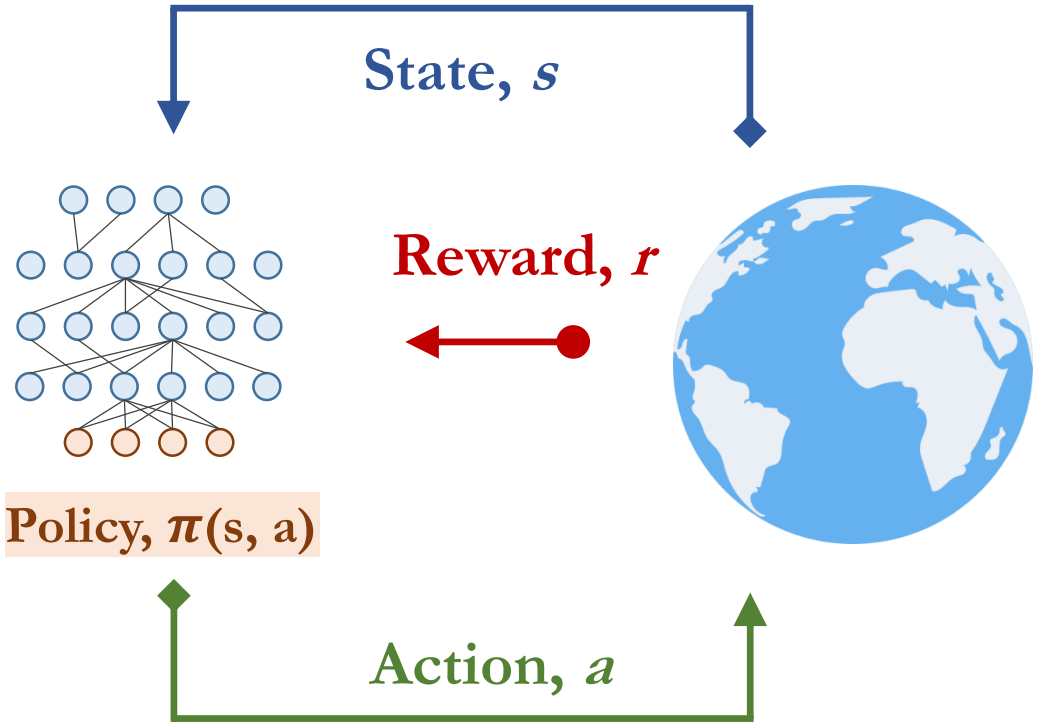
Source: https://www.youtube.com/watch?v=K67RJH3V7Yw&list=PLMsTLcO6ettgmyLVrcPvFLYi2Rs-R4JOE&index=4

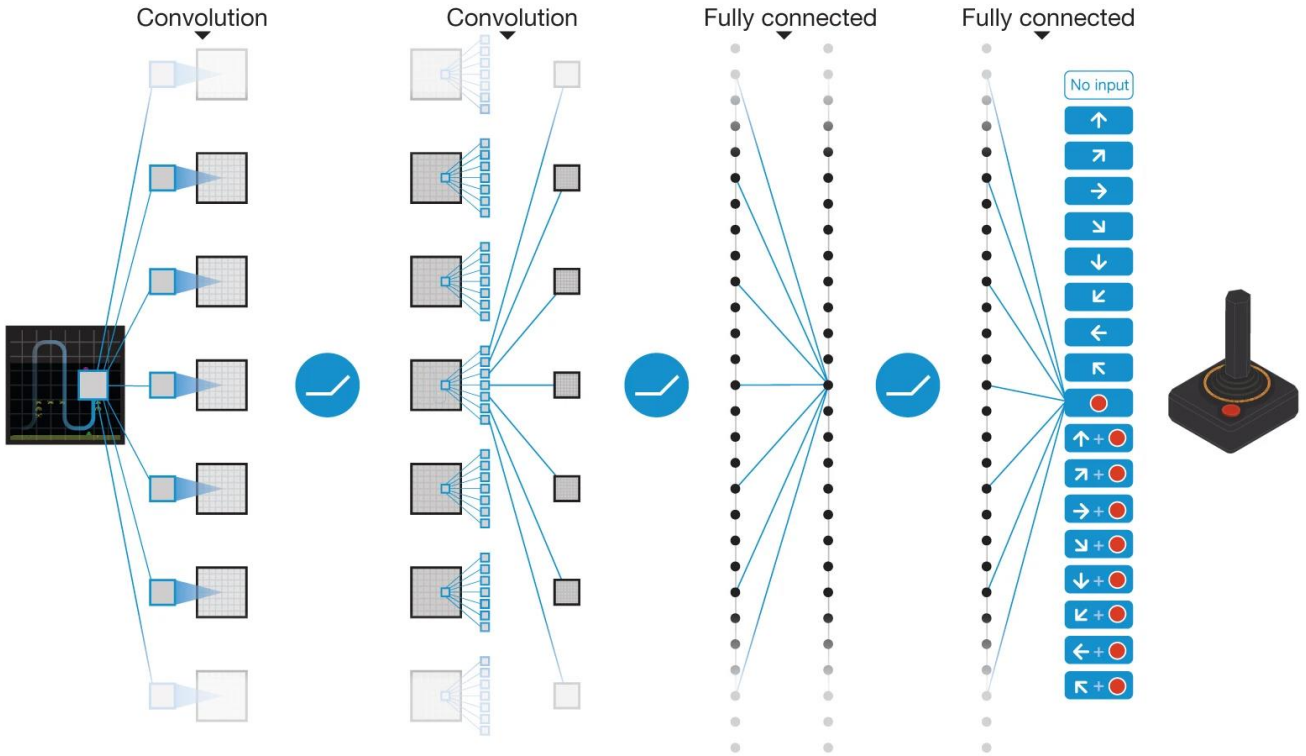
RL Algorithms

Hindsight Experience Replay
Save all behaviors and code reward for different goal.
https://www.youtube.com/watch?v=0Ey02HT_1Ho



Deep reinforcement learning





Mnih et al. (2015)

Source: https://www.youtube.com/watch?v=IUiKAD6cuTA

Examples

Hide and seek

https://www.youtube.com/watch?v=Lu56xVlZ40M

AlphaGo Move 37

https://www.youtube.com/watch?v=JNrXgpSEEIE

Flexible muscle-based locomotion for bipedal creatures

https://vimeo.com/79098420

Cart-Pole

https://www.youtube.com/watch?v=XiigTGKZfks

Atari video games

https://www.youtube.com/watch?v=TmPfTpjtdgg&t=43s

Markov Decision Processes

• Nicolas Legrand • Postdoctoral fellow • Embodied Computation Group