Advanced Topics in AI

Taxonomy

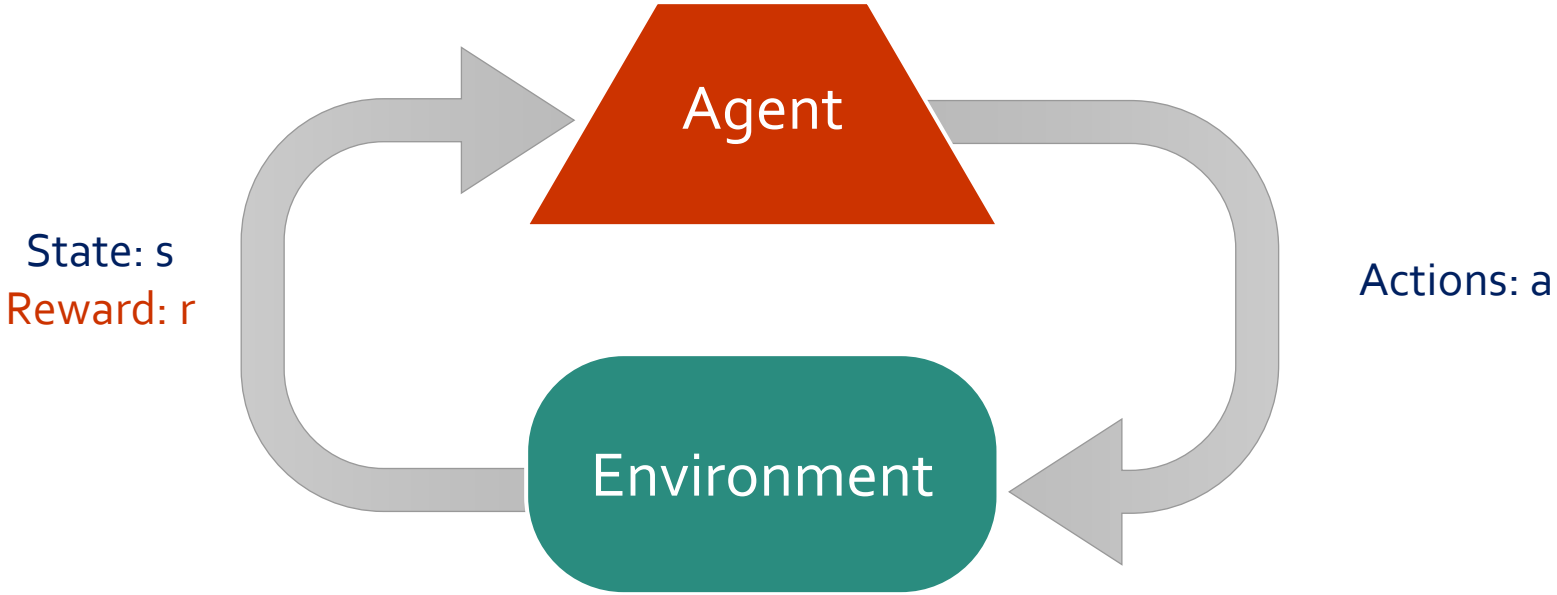
Instructor: Prof. Dr. techn. Wolfgang Nejdl

Leibniz University Hannover

Co-financed by the Connecting Europe Facility of the European Union

[These slides were created by Dan Klein and Pieter Abbeel for CS188 Intro to AI at UC Berkeley. All materials are available at http://ai.berkeley.edu.]

Reinforcement Learning



- Basic idea:
  - Receive feedback in the form of rewards
  - Agent's utility is defined by the reward function
  - Must (learn to) act so as to maximize expected rewards
  - All learning is based on observed samples of outcomes!

Co-financed by the Connecting Europe Facility of the European Union

xAIM

Reinforcement Learning

- Still assume a Markov decision process (MDP):
  - A set of states $s \in S$
  - A set of actions (per state) a $\in A$
  - A model $T(s, a, s')$
  - A reward function $R(s, a, s')$
- Still looking for a policy $\pi(s)$

Cool

Warm

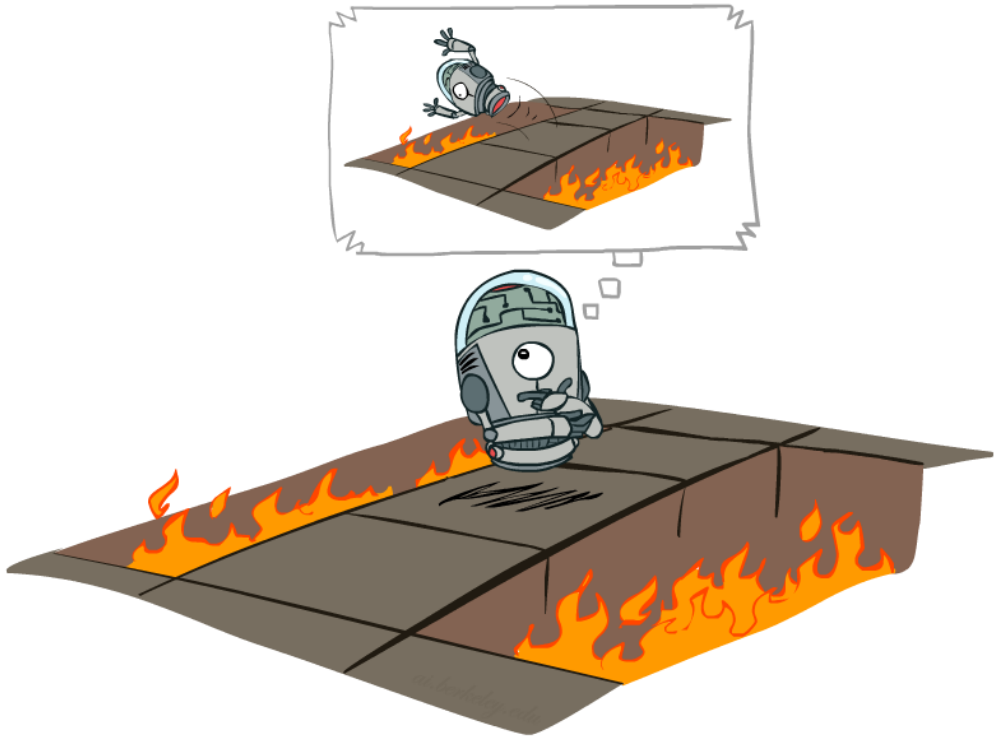
Overheated

- New twist: don't know T or R
  - I.e. we don't know which states are good or what the actions do
  - Must actually try actions and states out to learn

Co-financed by the Connecting Europe Facility of the European Union

xAIM

Offline (MDPs) vs. Online (RL)



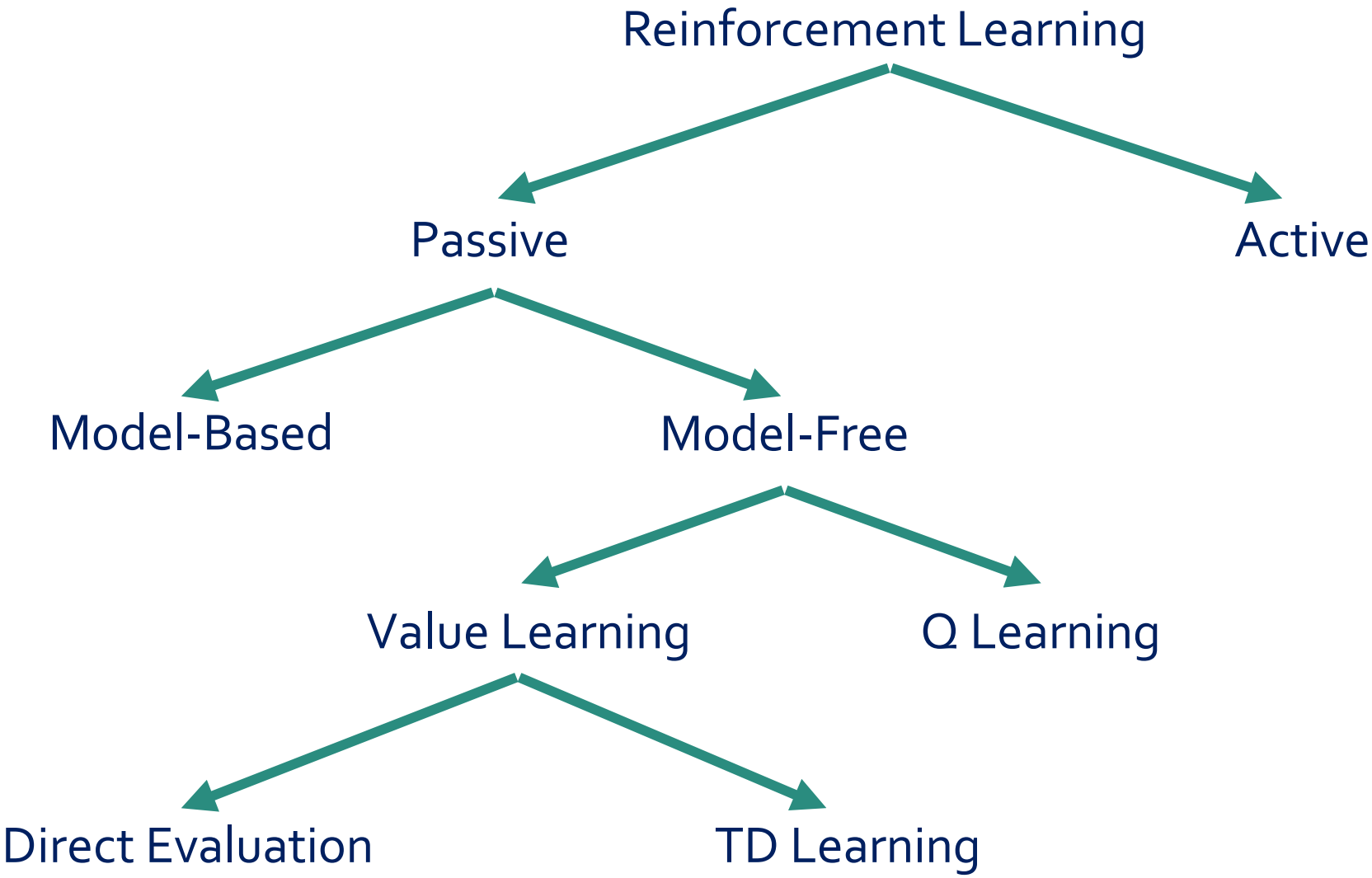
Offline Solution

Online Learning

Co-financed by the Connecting Europe Facility of the European Union

xAIM

Reinforcement Learning Taxonomy



Co-financed by the Connecting Europe Facility of the European Union

Reinforcement Learning Overview

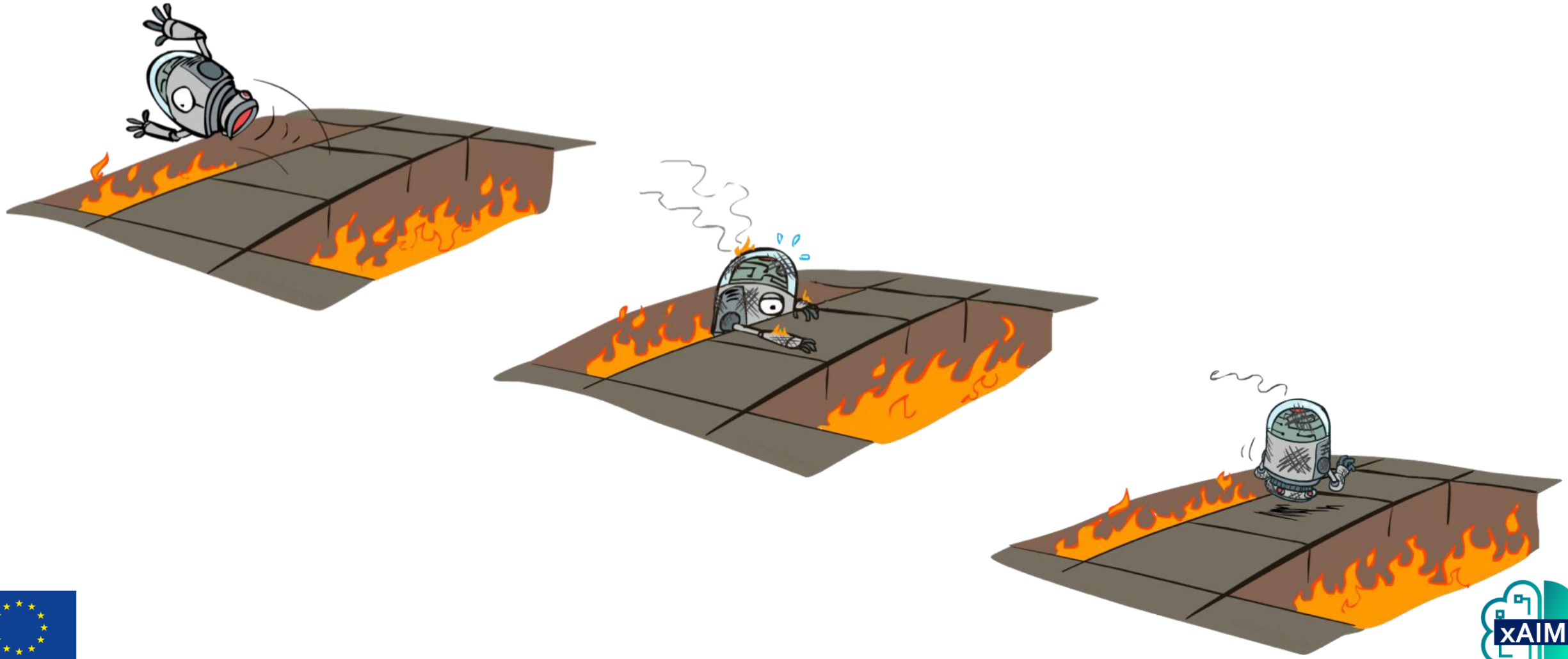
- Passive Reinforcement Learning (how to learn from experiences)
  - **Model-Based RL**: Learn MDP model from experiences, then solve with value / policy iteration
  - **Model-Free RL**: Skip learning MDP model, directly learn V or Q
    - **Value Learning**: learn values of fixed policy (Direct Evaluation or TD value learning)
    - **Q-Learning**: learn Q-values of optimal policy (Q-based version of TD learning)

- Active Reinforcement Learning (also decide how to collect experiences)
  - Challenges: how to **explore and minimize regret**

- Approximate Reinforcement Learning (how to deal with large state spaces)
  - **Approximate Q-Learning**
  - **Policy Search**

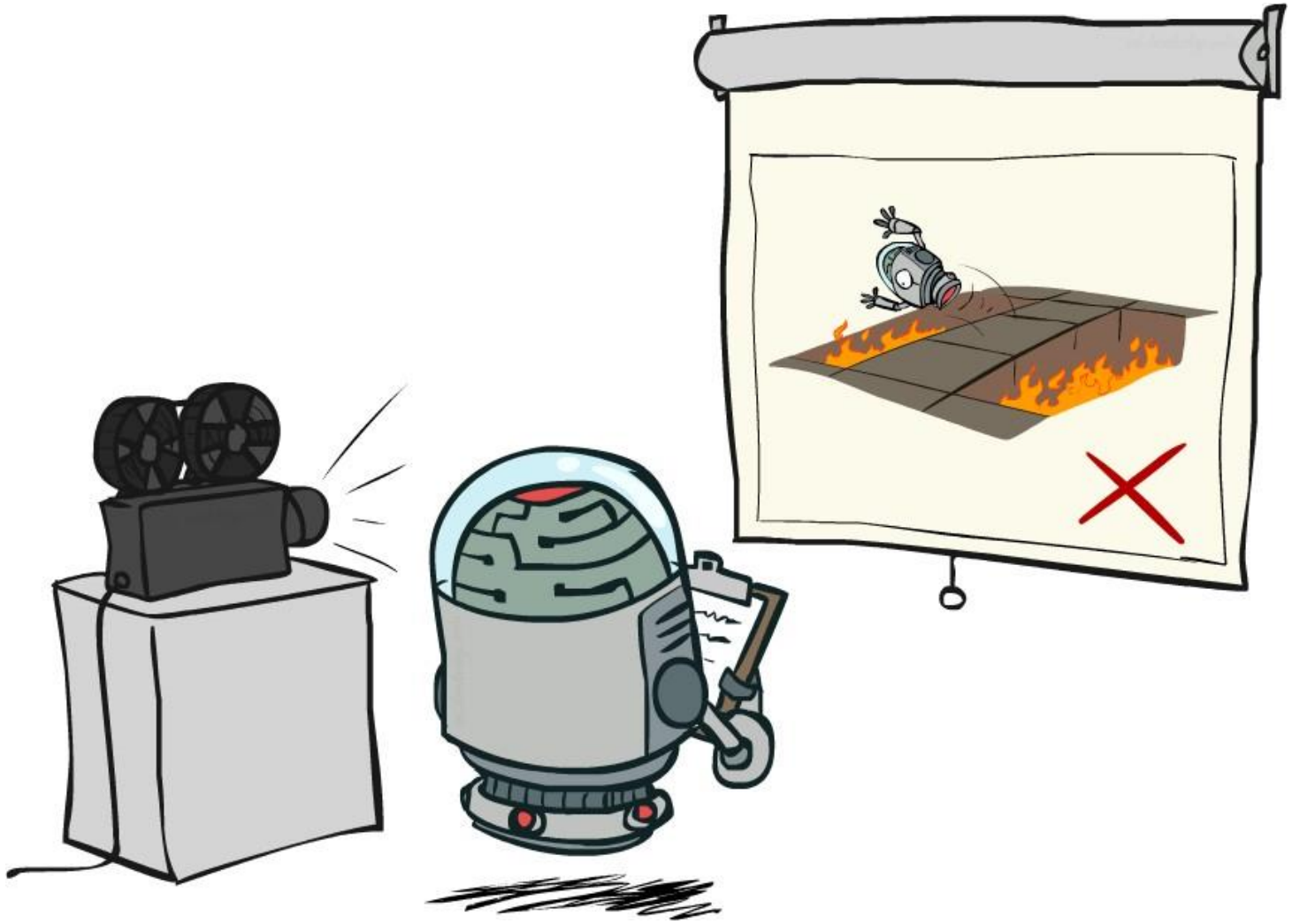
Co-financed by the Connecting Europe Facility of the European Union

Active Reinforcement Learning



Co-financed by the Connecting Europe Facility of the European Union

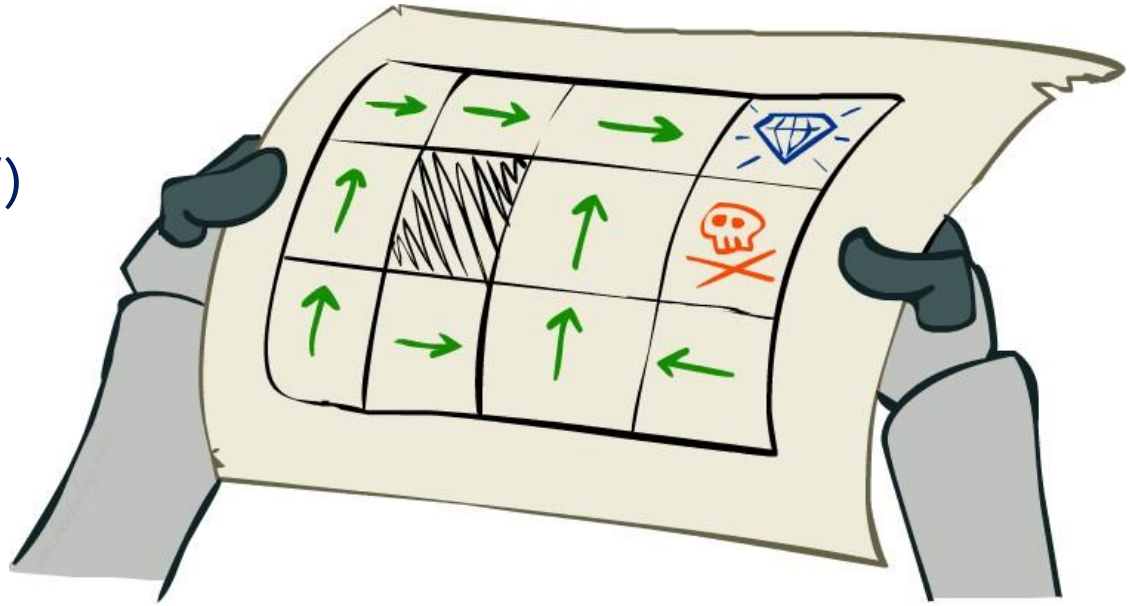
Passive Reinforcement Learning



Co-financed by the Connecting Europe Facility of the European Union

Passive Reinforcement Learning

- Simplified task: policy evaluation
  - Input: a fixed policy $\pi(s)$
  - You don't know the transitions T(s,a,s')
  - You don't know the rewards R(s,a,s')
  - Goal: learn the state values



- In this case:
  - Learner is "along for the ride"
  - No choice about what actions to take
  - Just execute the policy and learn from experience
  - This is NOT offline planning! You actually take actions in the world.

Co-financed by the Connecting Europe Facility of the European Union

Analogy: Expected Age

Goal: Compute expected age of students

Known P(A)

$$E[A] = \sum_a P(a) \cdot a \quad = 0.25 \cdot 25 + \ \ldots$$

Without $P(A)$, instead collect samples $[a_1, a_2, \ldots\ a_N]$

Unknown $P(A)$: "Model Based"

$$\hat{P}(a) = \frac{\text{num}(a)}{N}$$

$$E[A] \approx \sum_a \hat{P}(a) \cdot a$$

Why does this work? Because eventually you learn the right model.

Unknown $P(A)$: "Model Free"

$$E[A] \approx \frac{1}{N} \sum_i a_i$$

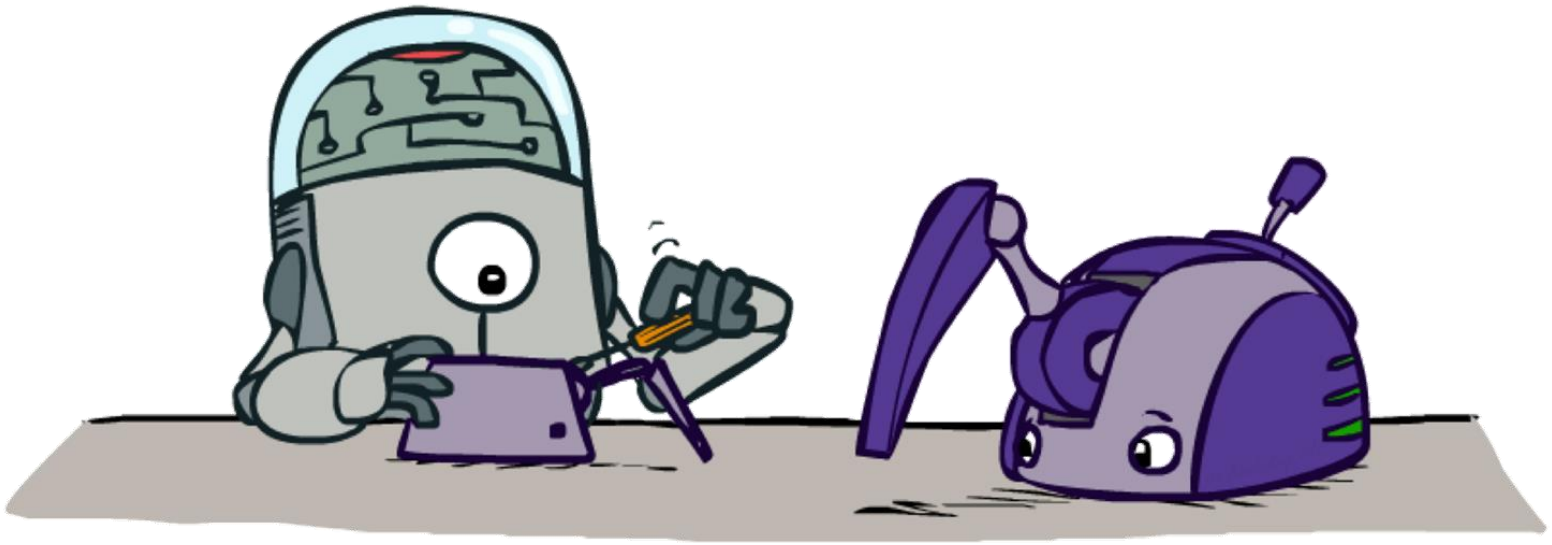
Why does this work? Because samples appear with the right frequencies.

Co-financed by the Connecting Europe Facility of the European Union

xAIM

Advanced Topics in AI

Next: Model-Based RL



Instructor: Prof. Dr. techn. Wolfgang Nejdl

Leibniz University Hannover

[These slides were created by Dan Klein and Pieter Abbeel for CS188 Intro to AI at UC Berkeley. All materials are available at http://ai.berkeley.edu.]

Co-financed by the Connecting Europe Facility of the European Union