Multi-Agent Systems

Introduction to Reinforcement Learning

Part 3B: Model-free Methods: DYNA-Q

Eric Pauwels (CWI & VU)

December 12, 2023

Reading

- Sutton & Barto: chapters 5 & 6

Outline

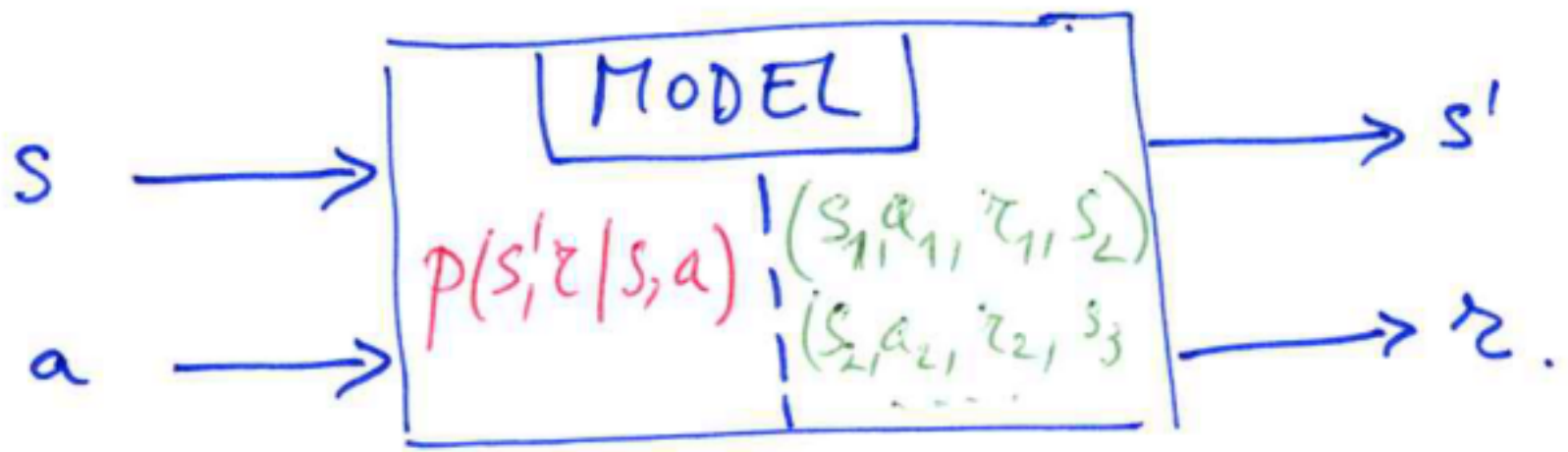
Integrating Planning and Learning

Dyna-Q; Integrating planning and learning

Model: tells agent what will happen next ...

- **Model-based**: planning
- **Model-free**: learning

Distributional vs. Sample-based Model

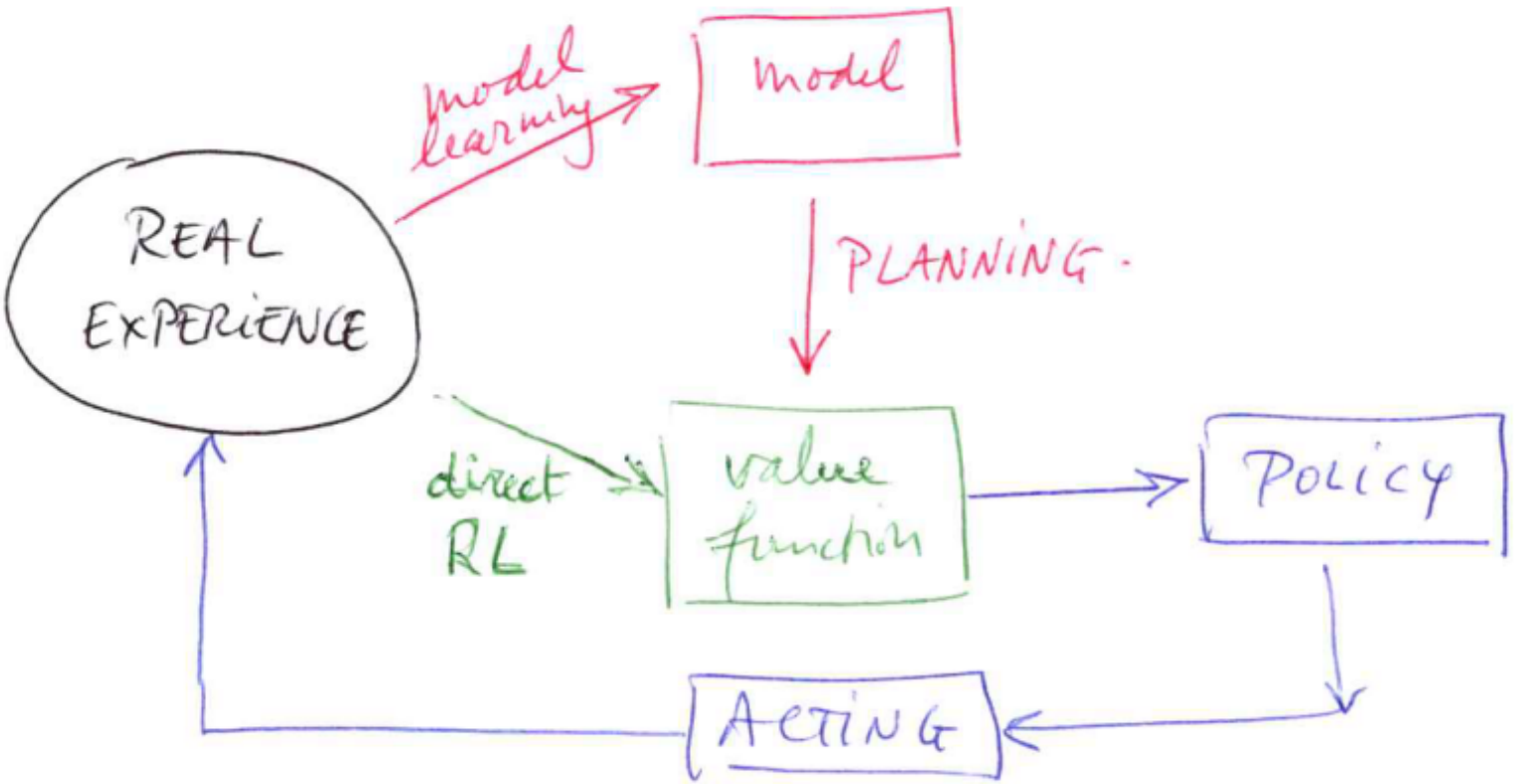


Model-based learning

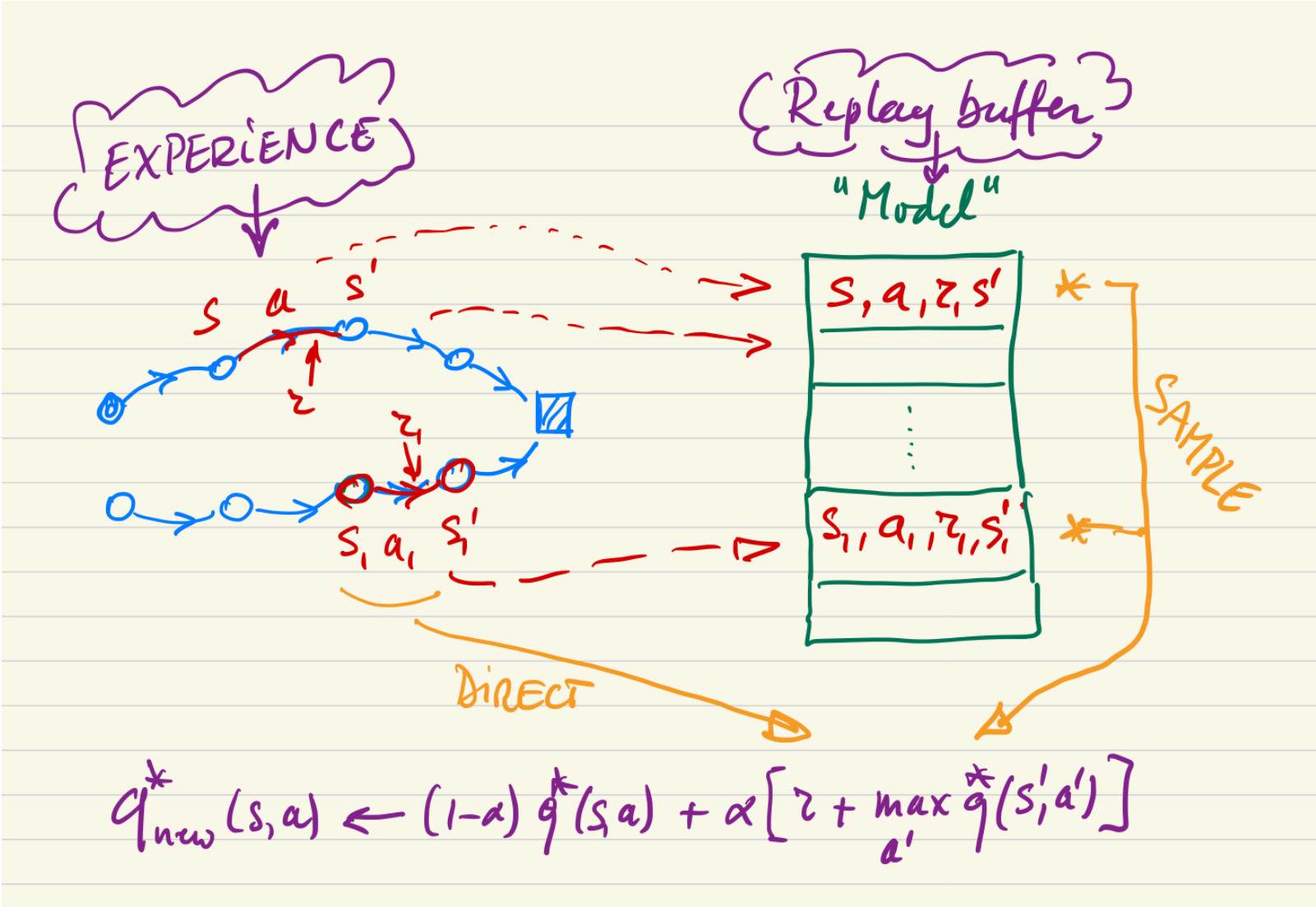
- Until now: Model = fully specified MDP!
- More generally: anything that helps the agent to plan:
- More accurate models are more effective (myths vs. science):
  - Folklore and weather saying:
    A wet and windy May fills the barn with corn and hay.
  - Meteorological models running on supercomputer

Dyna-Q; Integrating **planning** and **learning**

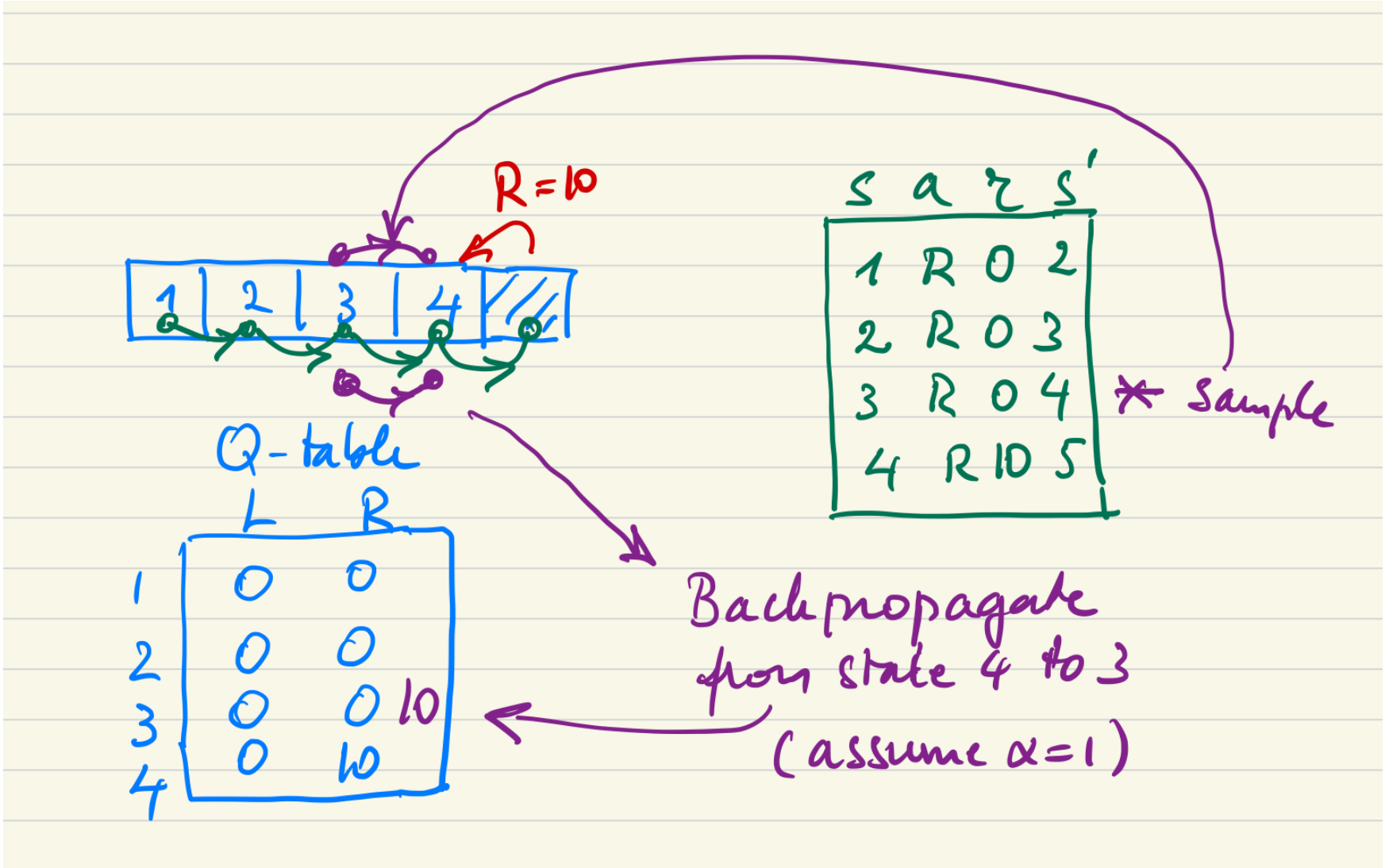
Real experience can be used in two ways:



Dyna-Q; Integrating **planning** and **learning**

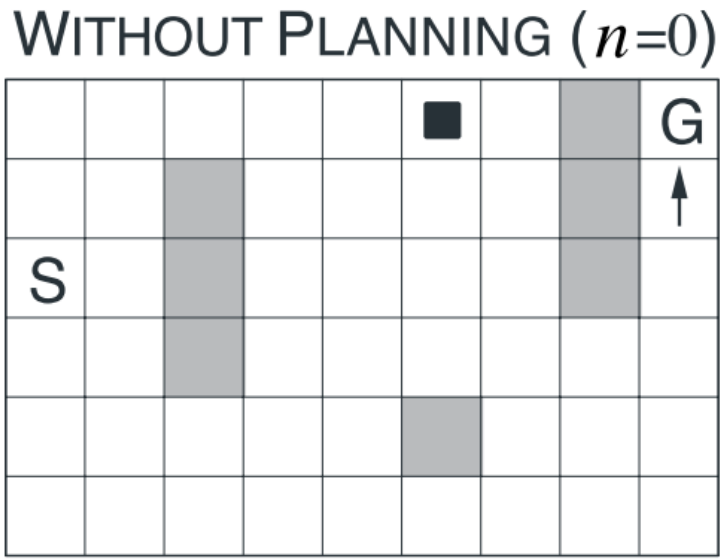


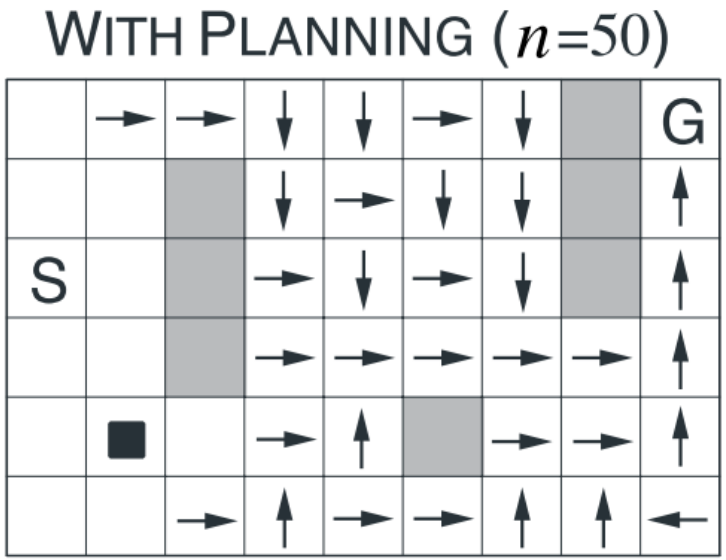
Dyna-Q; Integrating **planning** and **learning**



DYNA-Q: Maze example

Greedy policy midway through 2nd episode:





More info: Sutton and Barto, sections 8.2-8.3

DYNA-Q: Algorithm

Initialize $Q(s,a)$ and $Model(s,a)$ for all $s \in \mathcal{S}$ and $a \in \mathcal{A}(s)$

Do forever:

(a) $S \leftarrow$ current (nonterminal) state

(b) $A \leftarrow \varepsilon$-greedy(S,Q)

(c) Execute action A; observe resultant reward, R, and state, S'

(d) $Q(S,A) \leftarrow Q(S,A) + \alpha\big[R + \gamma \max_a Q(S',a) - Q(S,A)\big]$

(e) $Model(S,A) \leftarrow R, S'$ (assuming deterministic environment)

(f) Repeat n times:

$S \leftarrow$ random previously observed state

$A \leftarrow$ random action previously taken in S

$R, S' \leftarrow Model(S,A)$

$Q(S,A) \leftarrow Q(S,A) + \alpha\big[R + \gamma \max_a Q(S',a) - Q(S,A)\big]$

Prioritized Sweeping

- **Uniform sampling** from experience database: Wasteful! ;
- Idea: *work "backwards" from "goal state"*
- More generally: **backward focusing:** work backwards from any state the value of which has changed!
- ...but prioritize: Most significant changes first!