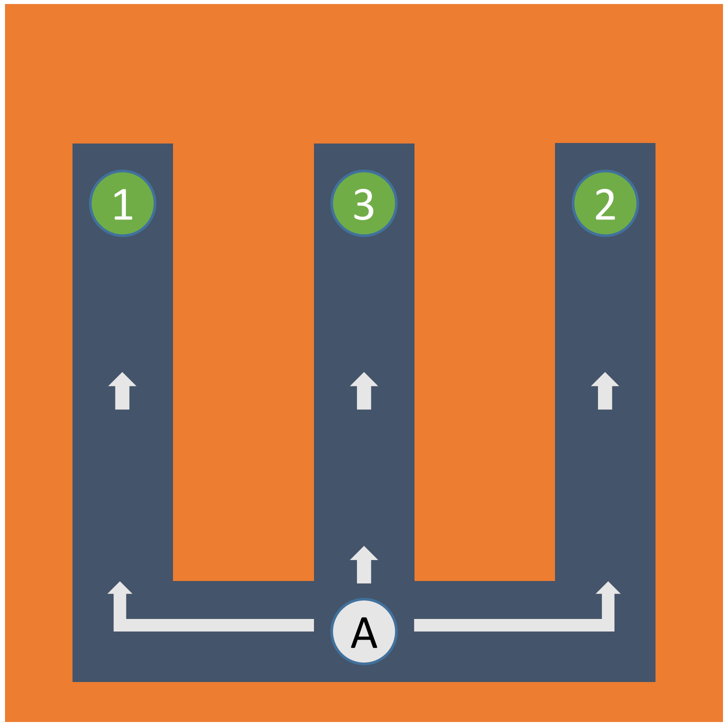
Uncertainty-based Exploration for DQN

Motivation: Exploration can depend on

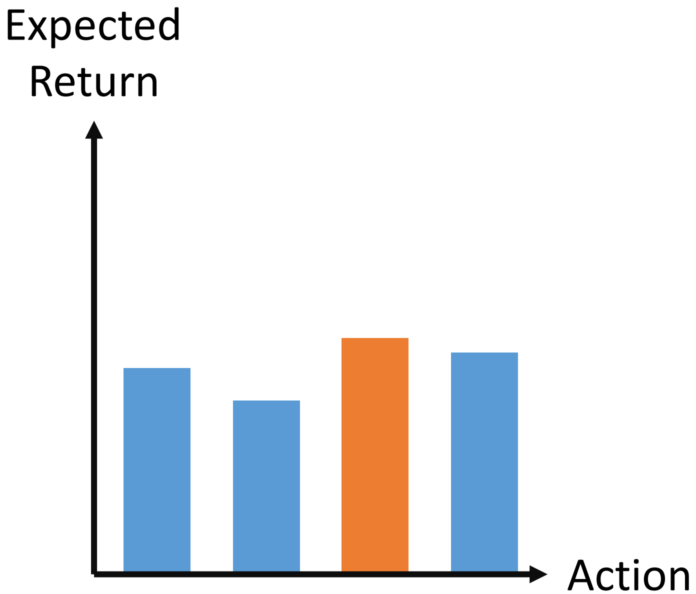
- Current state
- Learning progress
- ”Number of good actions”

→ Classic exploration strategies (ϵ- or ϵz-greedy) do not take all of these into account

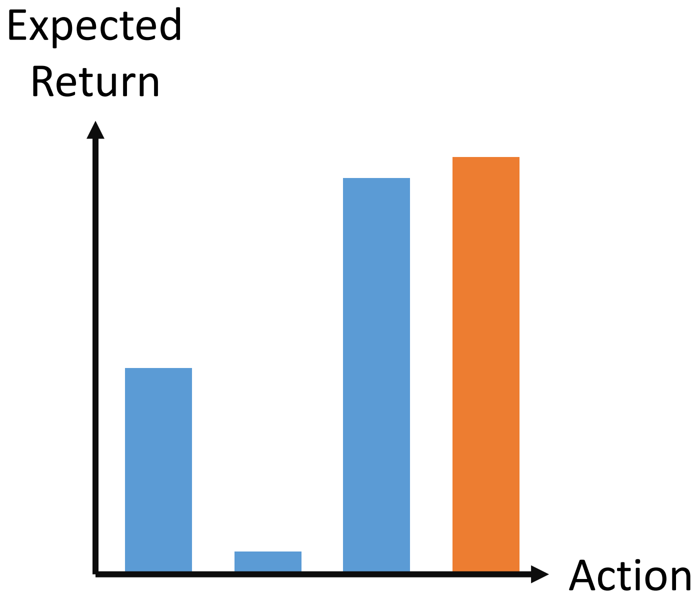


Uncertainty-based Exploration for DQN

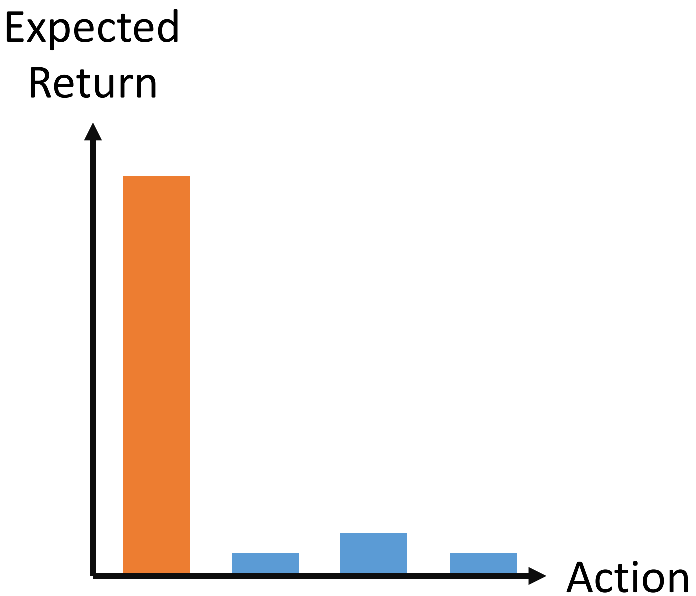
Idea: Adjust ϵ (or ϵ and z) online, based on **uncertainty of Q-values**



State s_0



State s_1



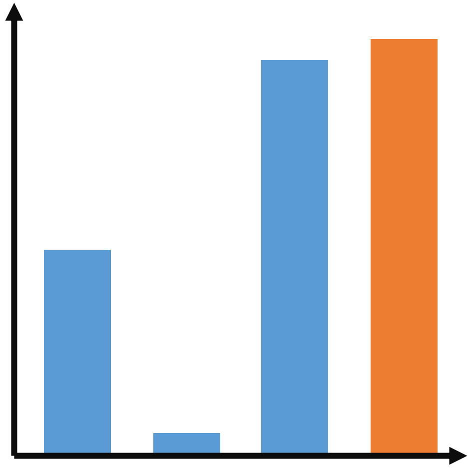
State s_2

e.g. $\boldsymbol{\epsilon = 0.8}$ $\boldsymbol{\epsilon = 0.5}$ $\boldsymbol{\epsilon = 0.1}$

Uncertainty-based Exploration for DQN

Challenge: How to compute ϵ?

Expected Return



Action

State s_1

1) $\epsilon = \frac{|\mathcal{A}_{good}|}{|\mathcal{A}|}$, where $a \in \mathcal{A}_{good}$ if $Q(s,a) > \bar{Q}(s)$

   $\bar{Q}(s) =$ mean, median, X-percentile, etc.

2) $\epsilon = f(\sigma(Q(s, \cdot)))$, where f normalizes the standard deviation to $(0, 1]$

3) Statistical measures (skewness), …

$\boldsymbol{\epsilon = 0.5}$

Uncertainty-based Exploration for DQN

Possible Advantages:

- State-based exploration
- Assuming Q-value distributions ”converge”, this approach would inherently lead to ϵ-schedules for each state

However:

- Requires reliable calculation of good ϵ values
- Could we maybe learn this in some way?
- **Problem**: What is a good target?