The Primacy Bias in Deep RL

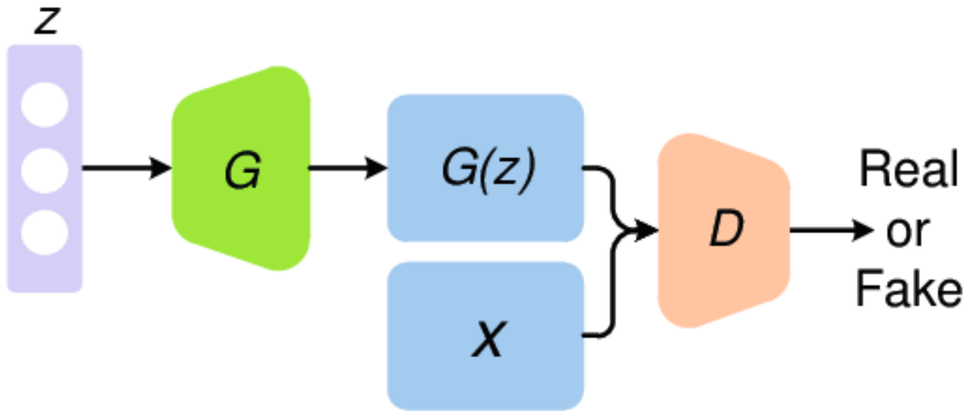
with Max Schwarzer*, Pierluca D'Oro*, Pierre-Luc Bacon, Aaron Courville




May 20, 2022

Science of Deep Learning

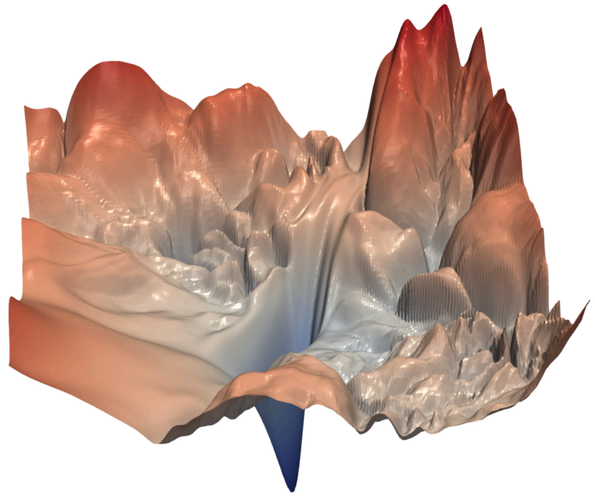
Benchmarks



Algorithms

$$\mathbb{E}_{h\sim Q}[R[h]] \leq \mathbb{E}_{h\sim Q}[\hat{R}_S[h]] + \sqrt{\frac{D(Q||P) + \log(\frac{n}{\delta})}{2(n-1)}}$$

Theory



Understanding

The first impression in human learning

«Steve is impulsive, critical, and smart.»

vs

«Steve is smart, critical, and impulsive.»

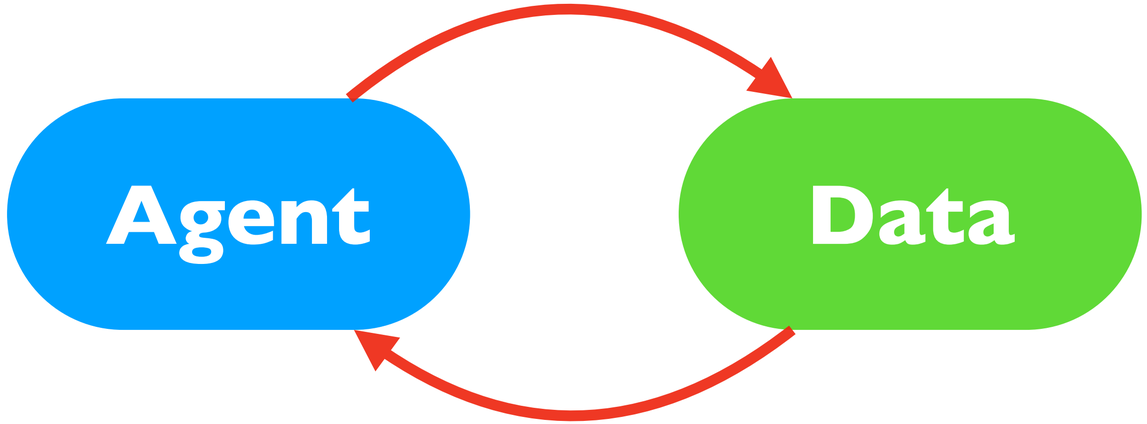
First experiences can have large effects on future behavior

The Primacy Bias in Deep RL

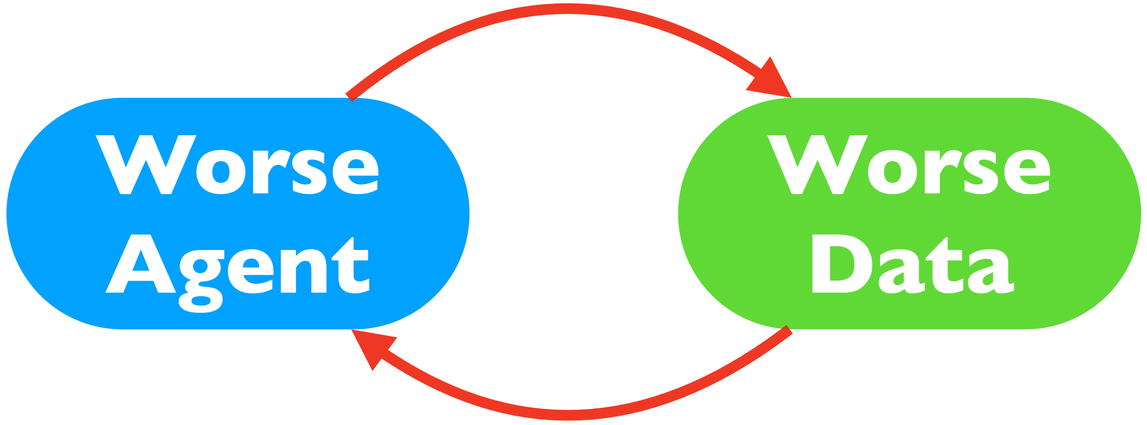
A tendency to overfit initial experiences that damages the rest of the learning process

Role of first experiences in deep RL

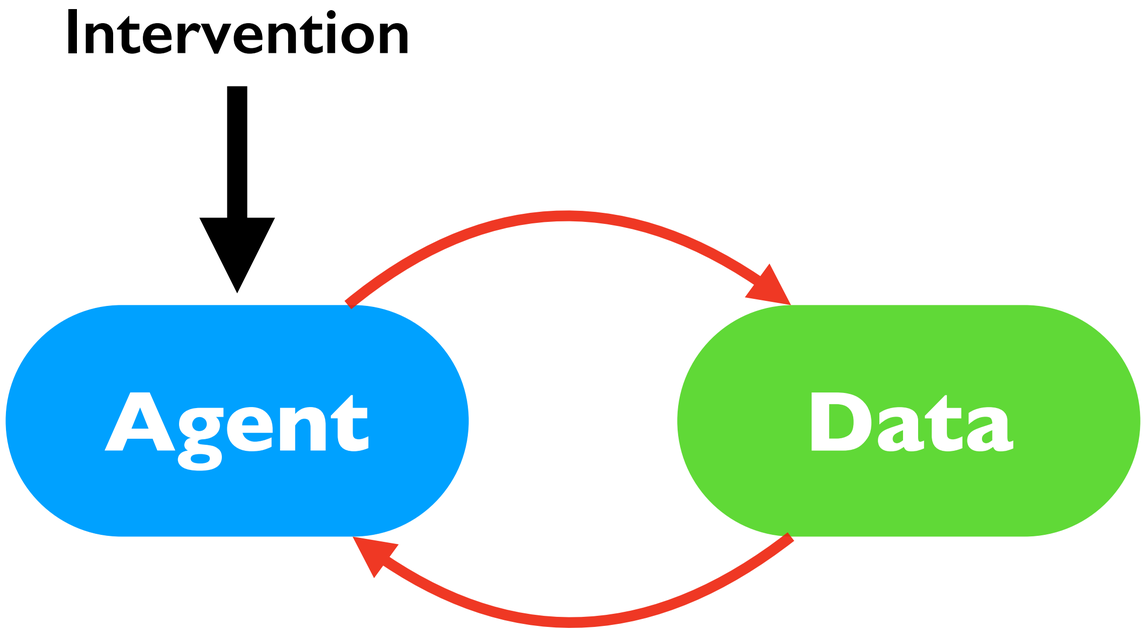
Sequential decision making:

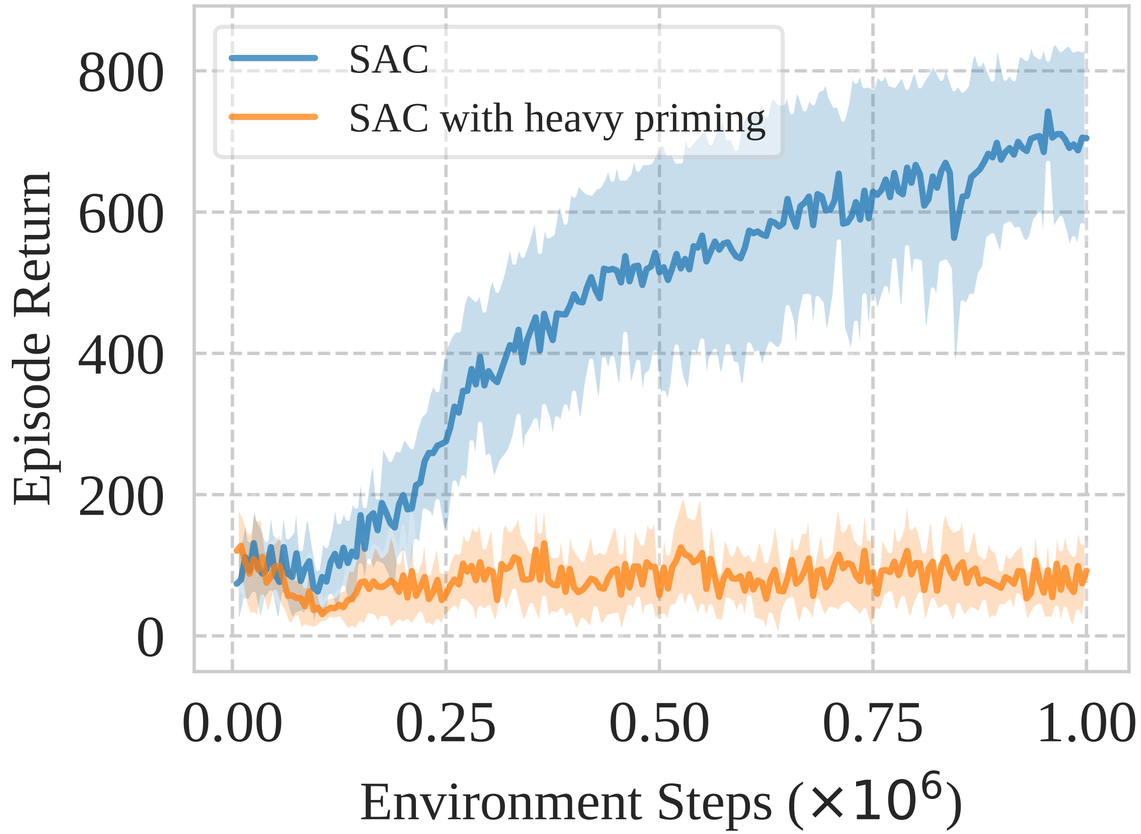


With the primacy bias:

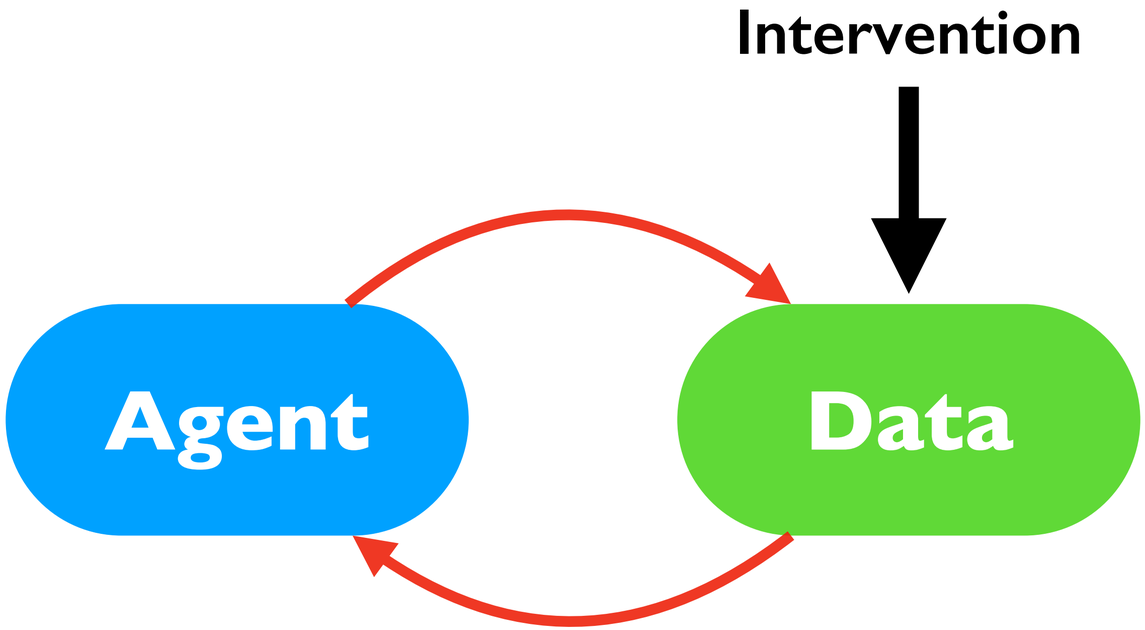


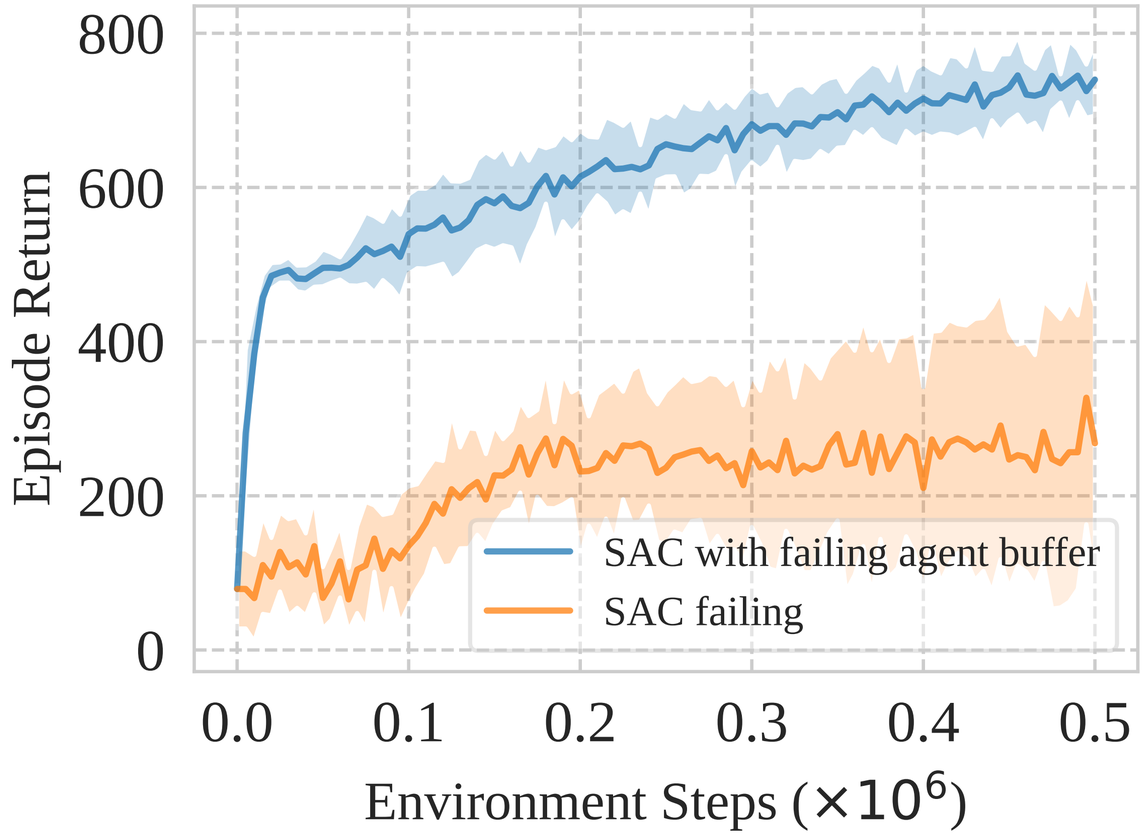
Overfitted agent does not recover





Agent starting with bad data recovers





The primacy bias and its consequences

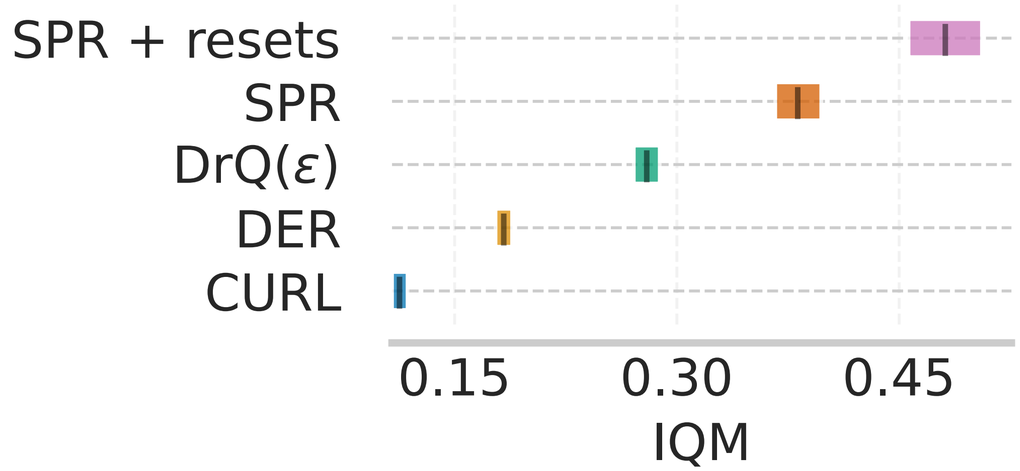
- Overfitting to early data can damage unrecoverably
- Not about data but about failure to learn
- Vicious circle of decreasing performance
- …
- Solution?

Have You Tried Resetting It?

Given an agent's network, periodically reinitialize the parameters of the last few layers while preserving the replay buffer

Simply works!

SPR



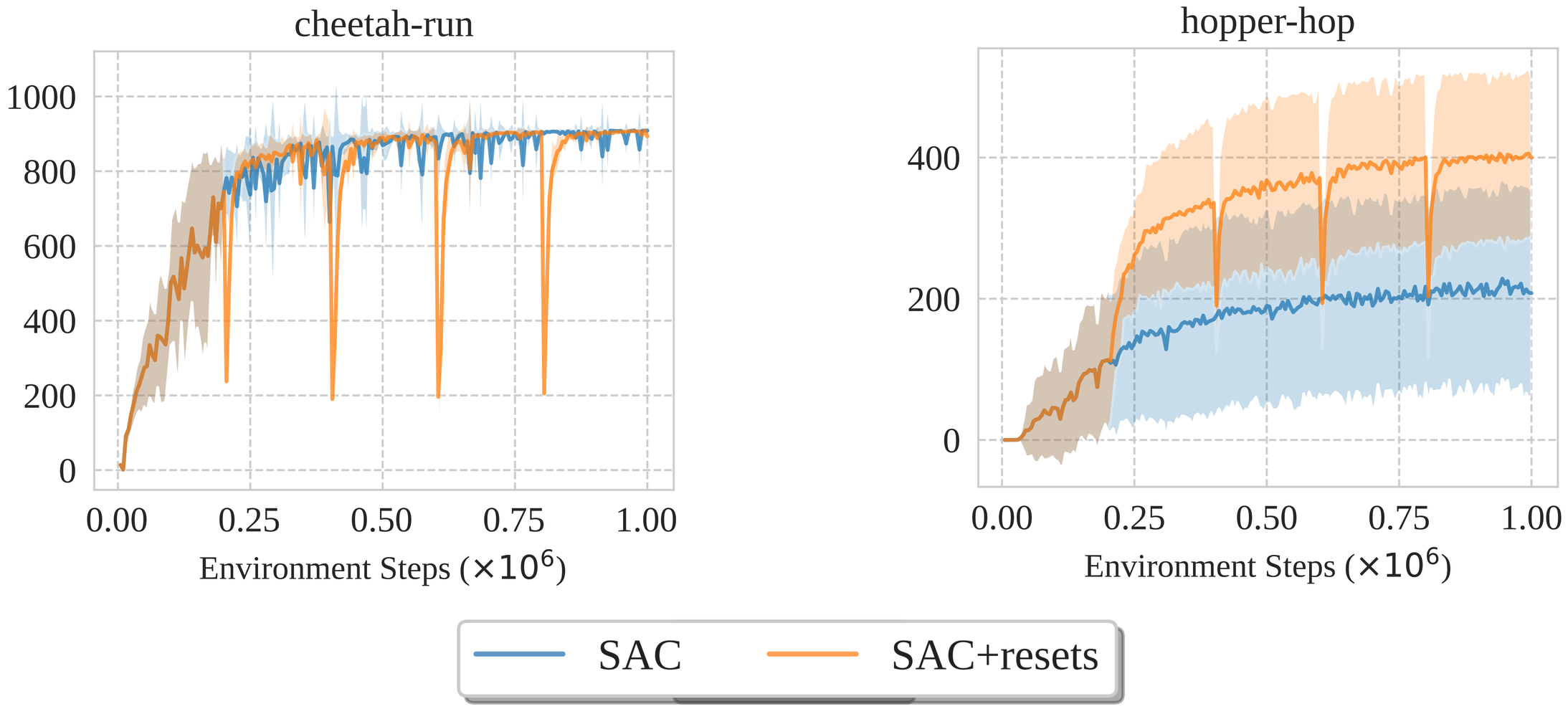
DrQ

Resets	IQM
Yes	**680** (625, 731)
No	521 (470, 600)

SAC

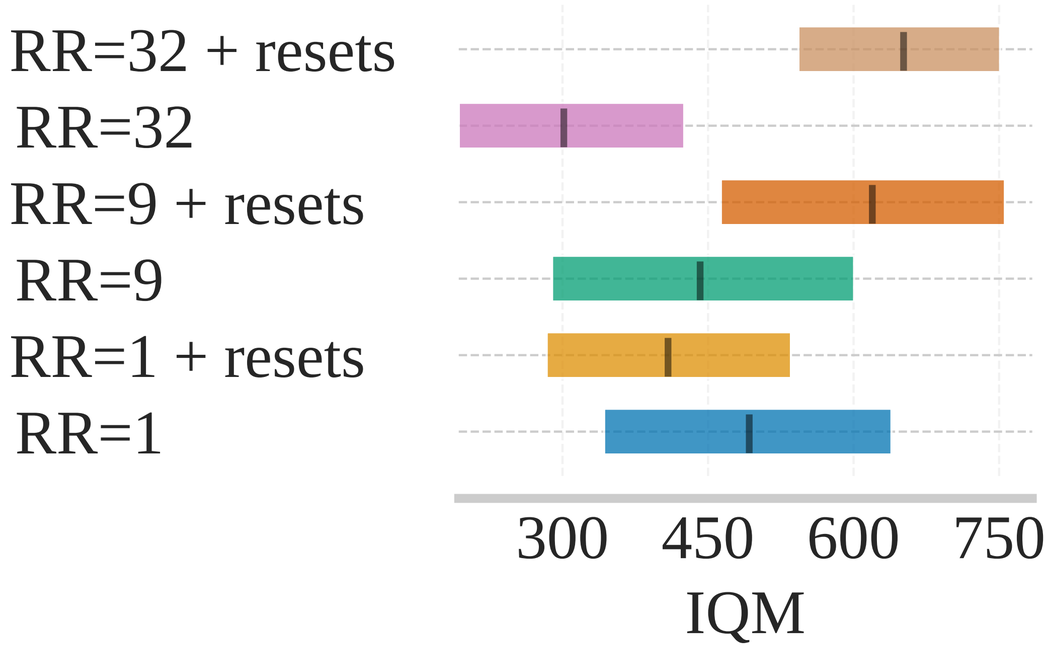
Resets	IQM
Yes	**616** (538, 681)
No	475 (407, 563)

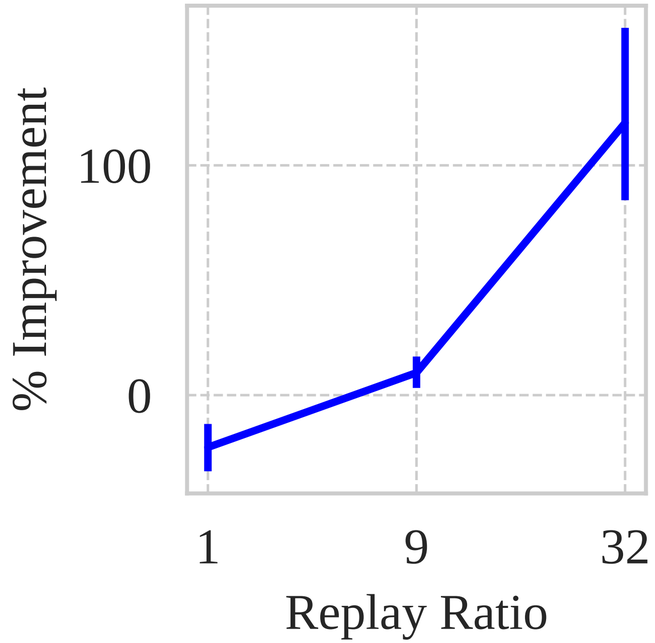
How reset training looks like



Resets allow more aggressive training

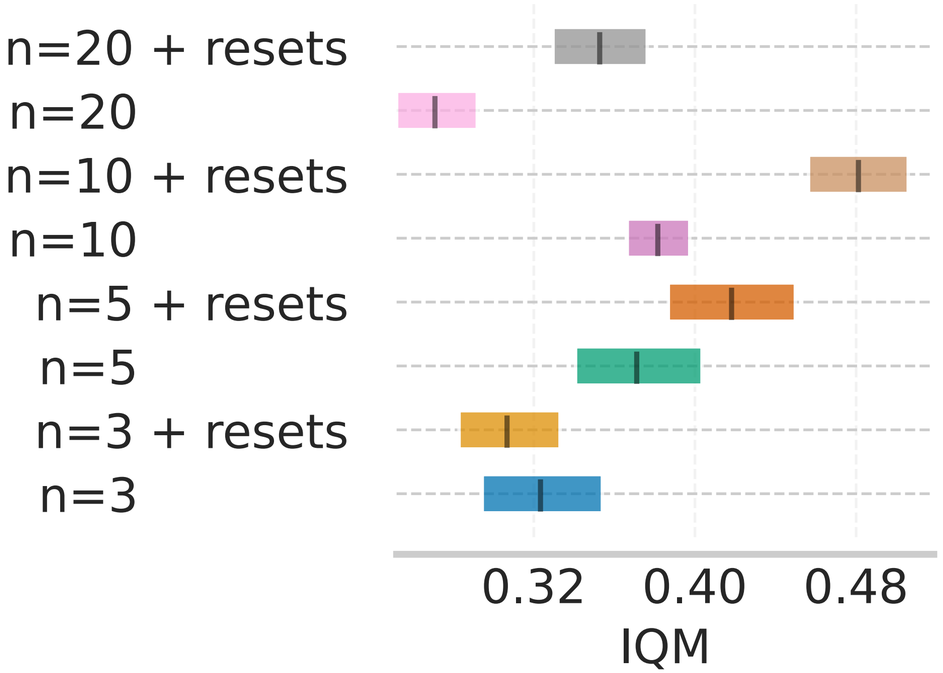
```
next_state, reward, done, info = env.step(action)
replay_buffer.insert(state, action, next_state, reward, done)
for _ in range(replay_ratio):
    batch = replay_buffer.sample(batch_size)
    agent.update(batch)
```

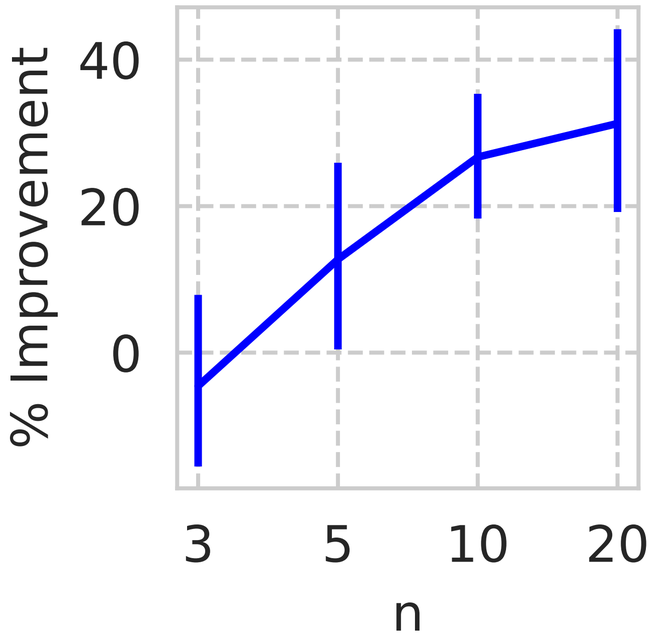




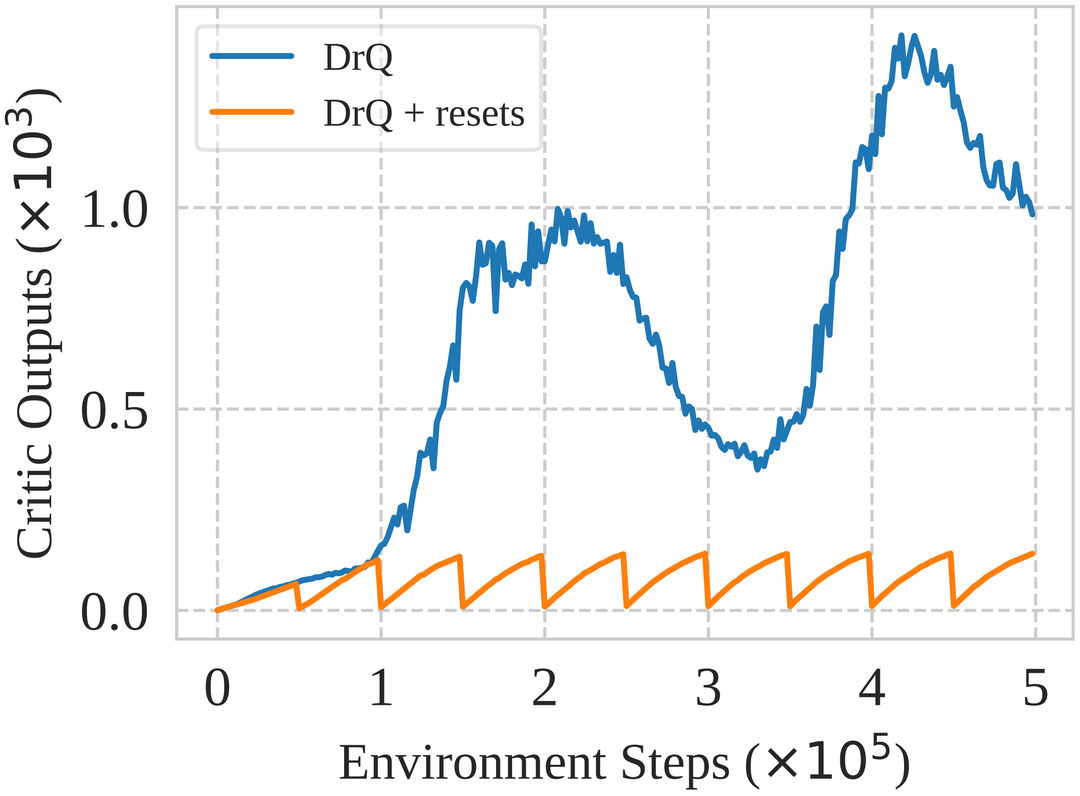
Resets prevent overfitting to noisy targets

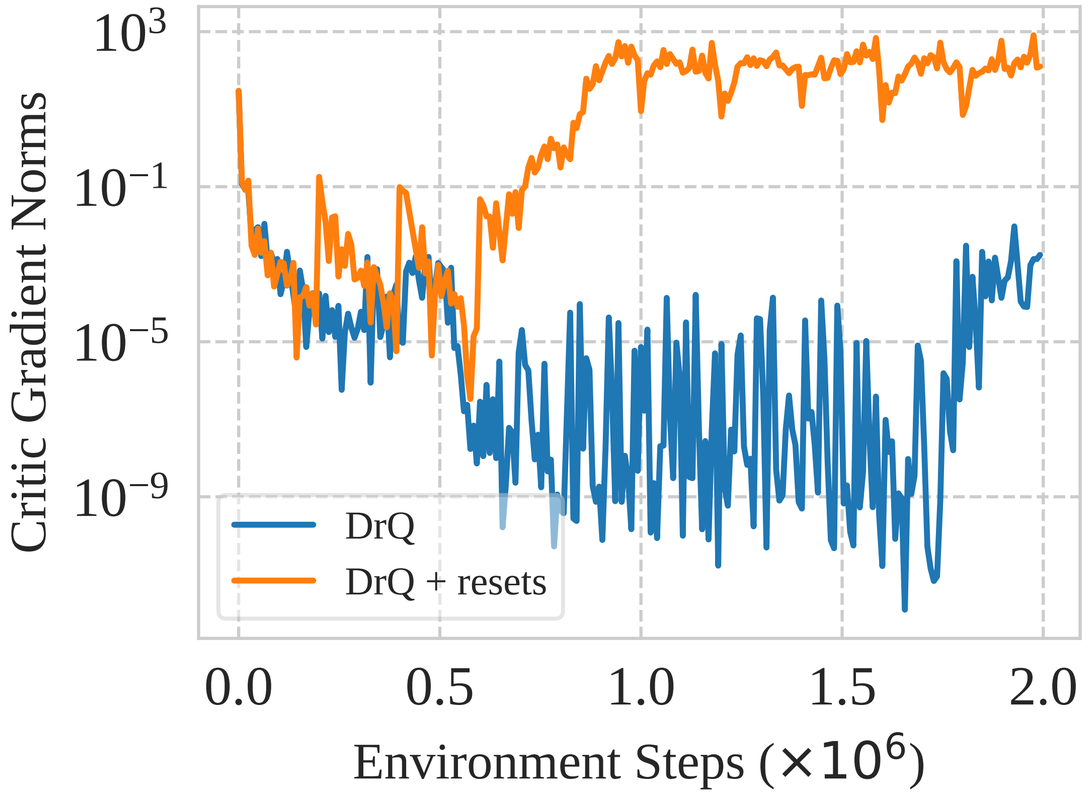
$$\mathbb{E}_\pi \left[r(s_t, a_t) + \gamma r(s_{t+1}, a_{t+1}) + \cdots + \gamma^n Q_\pi(s_{t+n}, a_{t+n}) \right]$$



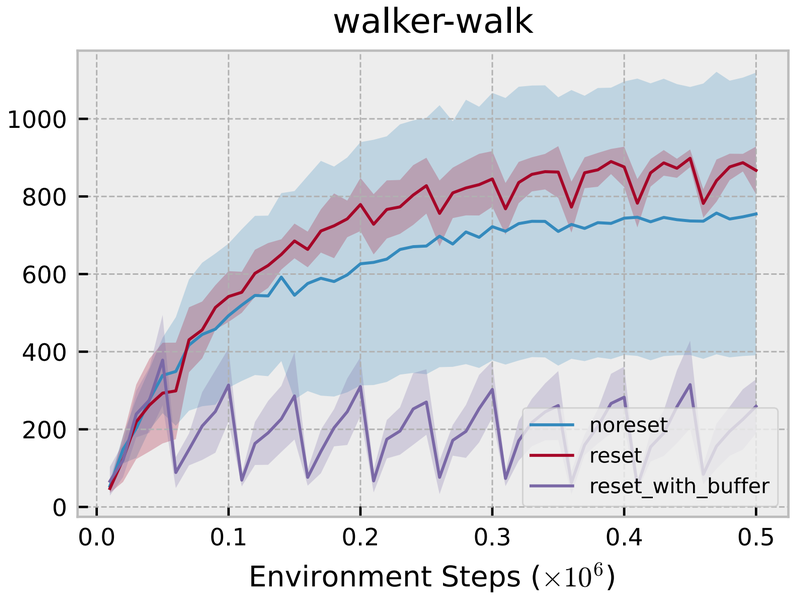


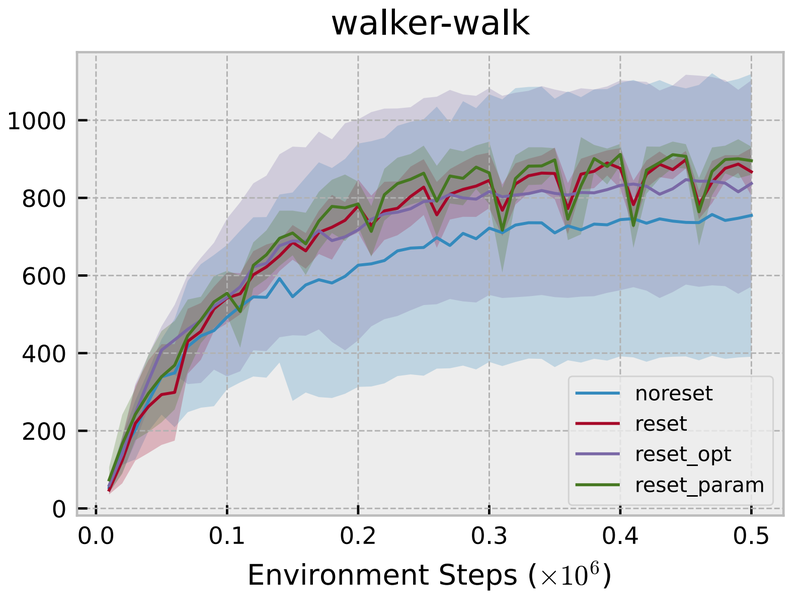
Resets avoid TD failure modes

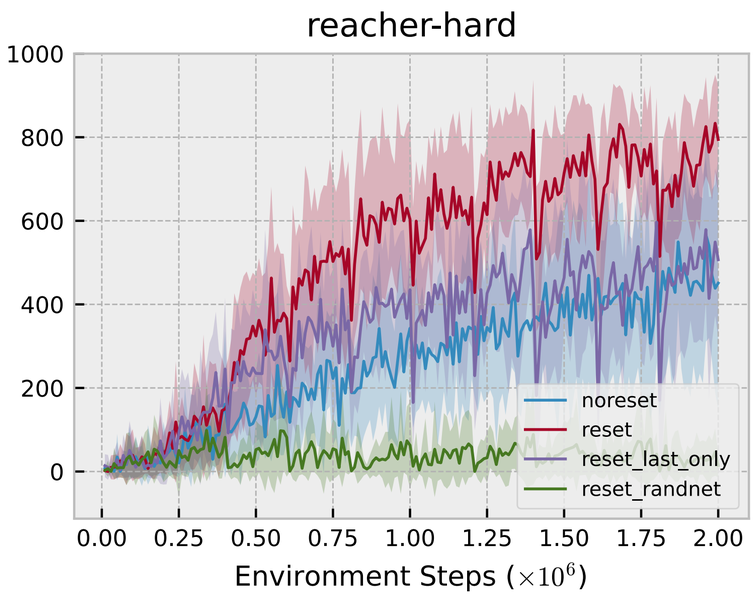




What to reset?



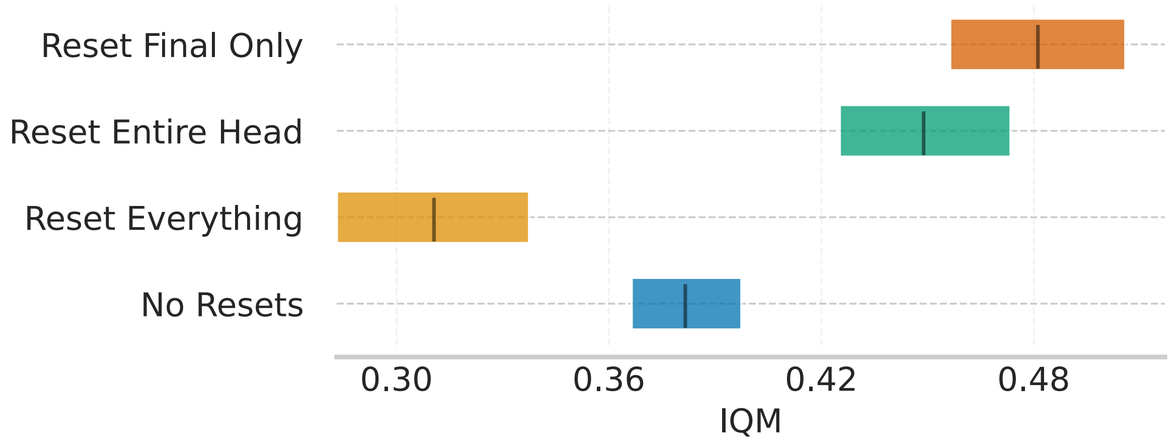




Buffer: very important to keep

Adam statistics: not important

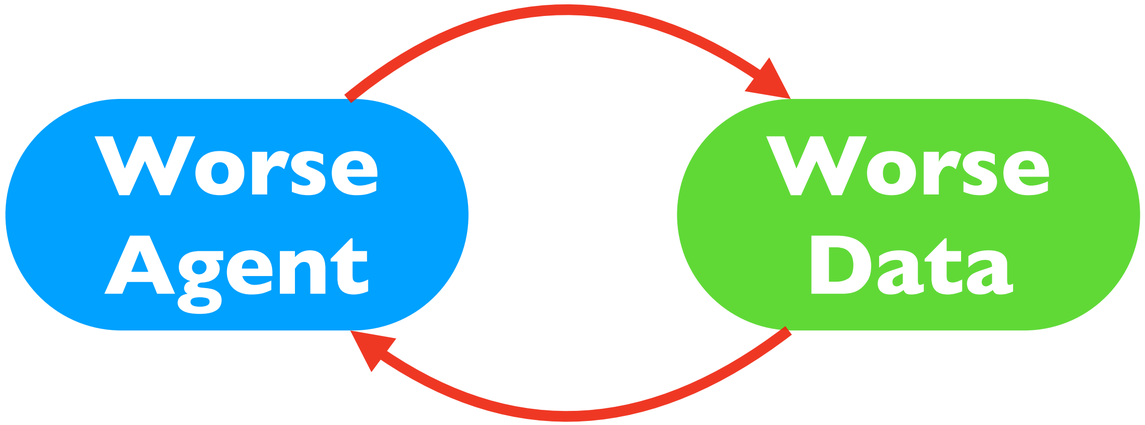
Number of layers: depends



Effects of resets

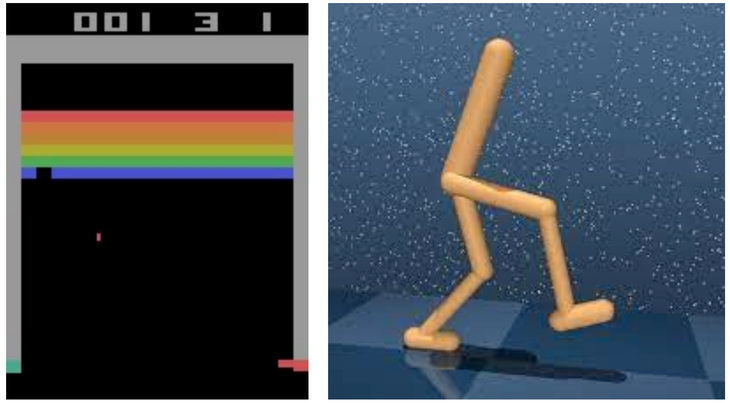
- Alleviate the primacy bias
- Allow more optimization steps per data point
- Prevent overfitting to noisy targets
- Avoid TD failure modes

Conclusion



Understanding

Algorithms



Benchmarks

?

Theory