The Primacy Bias in Deep RL

Evgenii Nikishin

Pierluca D'Oro

Max Schwarzer

Pierre-Luc Bacon

Aaron Courville

Mila

The first impression in human learning

«Steve is impulsive, critical, and smart.»

VS

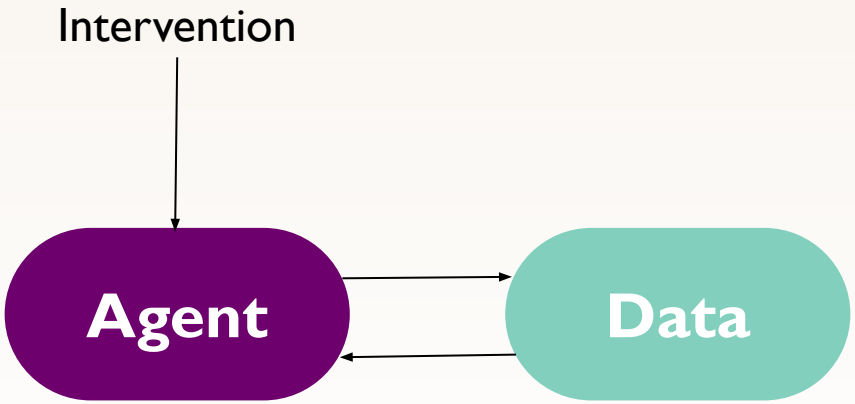
«Steve is smart, critical, and impulsive.»

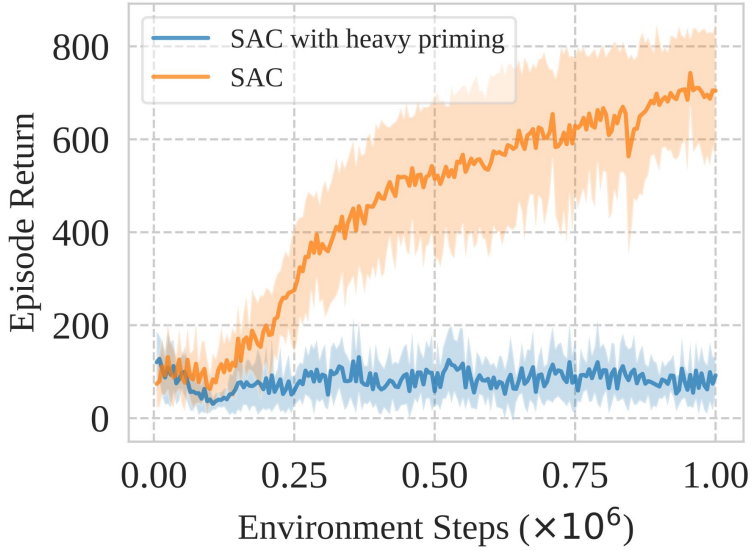
First experiences can have large effects on future behavior

The Primacy Bias in Deep RL

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

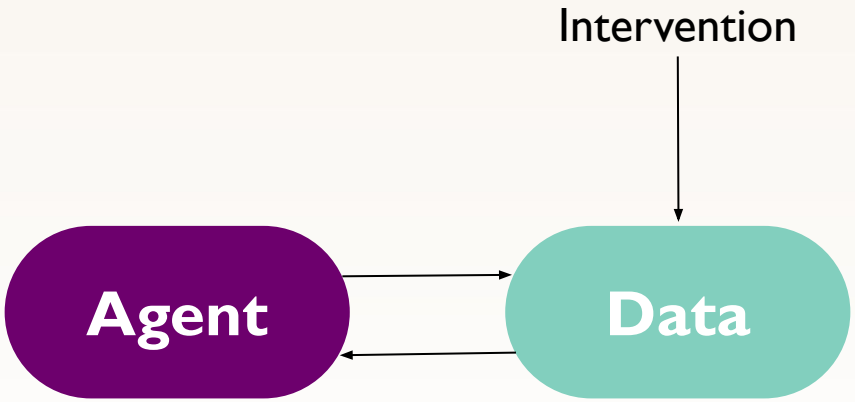
Overfitted agents do not recover

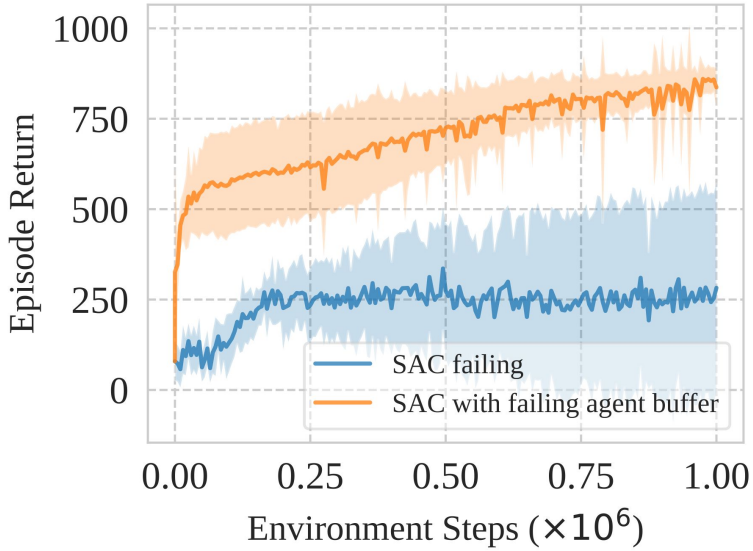




Mila

Experiences of overfitted agents suffice



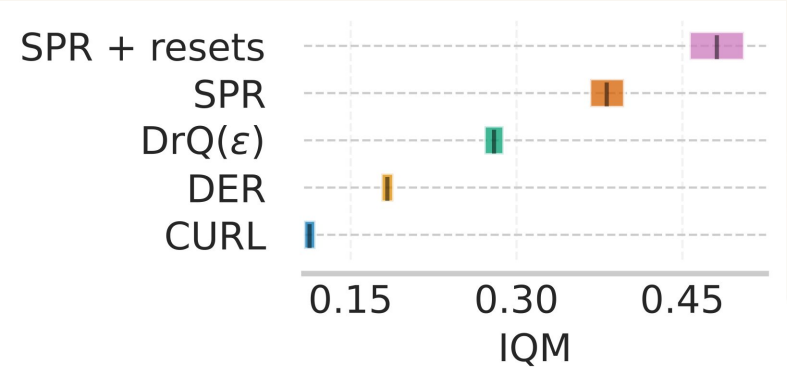


Mila

Have you tried resetting it?

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

Simply works!

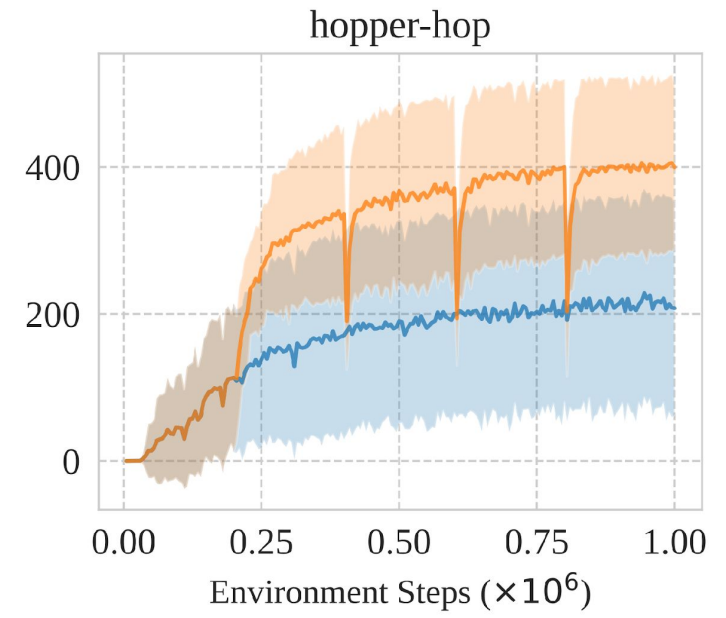


Resets	IQM
Yes	**762** (704, 815)
No	569 (475, 662)

Resets	IQM
Yes	**656** (549, 753)
No	501 (389, 609)

Mila

How reset training looks like



SAC SAC+resets

Mila

Summary

- A special form of overfitting in deep RL
- A simple way to address it
- Ablations:
  What and how to reset
  Qualitative effects of resets