# The Primacy Bias In Deep Reinforcement Learning

Evgenii Nikishin* Max Schwarzer* Pierluca D'Oro* Pierre-Luc Bacon Aaron Courville

## Primacy in Human Learning

- "Steve is impulsive, critical, and smart"
- "Steve is smart, critical, and impulsive"

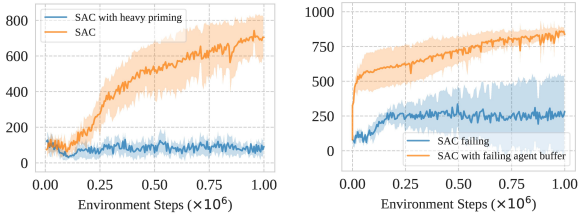
Same information, different perception

## The Primacy Bias in Deep RL

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

## Controlled Experiments

- Severe overfitting might be unrecoverable
- Data from an overfitted agent are sufficient




How can an agent properly learn from the data it collected?

## Addressing the Primacy Bias

**Periodically reset the agent's last layers while preserving its replay buffer**

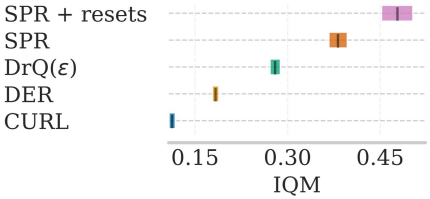
## Contributions

- Analyze a special form of overfitting
- A simple remedy: resets
- Large performance gains
- Unlock novel data-efficient training regimes

## Results

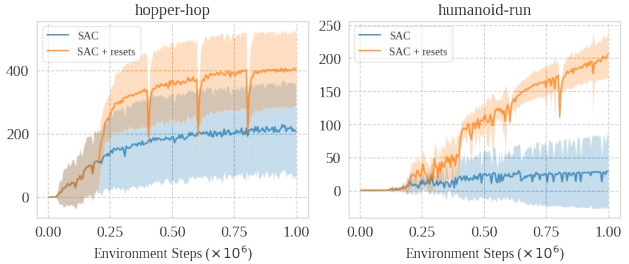
- Benchmarks
  - DMC-dense (SAC)
  - DMC-pixels (DrQ)
  - Atari 100k (SPR)
- Resets are a special form of regularization for the primacy bias

| Method | IQM |
|---|---|
| SAC | 501 (389, 609) |
| SAC + resets | **656** (549, 753) |
| SAC + dropout | 219 (160, 285) |
| SAC + L2 | 412 (299, 524) |
| DrQ | 569 (475, 662) |
| DrQ + resets | **762** (704, 815) |
| DrQ + dropout | 492 (414, 567) |
| DrQ + L2 | 463 (362, 566) |

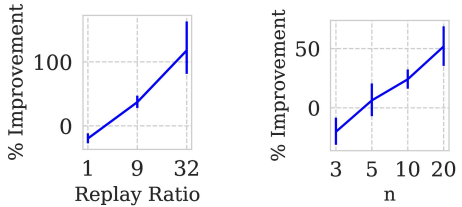



## Analysis of Learning with Resets

The agent is able to recover incredibly quickly




Resets are more beneficial with high RR and *n*




### Ablations

- Reset depth depends on representation complexity
- Resets fix TD collapses and divergence
- Resetting the optimizer does not matter
- Agents are robust to the number of resets