# 7 Appendix

## 7.1 Complete Results from Section 4.1

We first show individual learning curves for Rainbow (Adam optimizer) with and without resetting and different values of $K$.

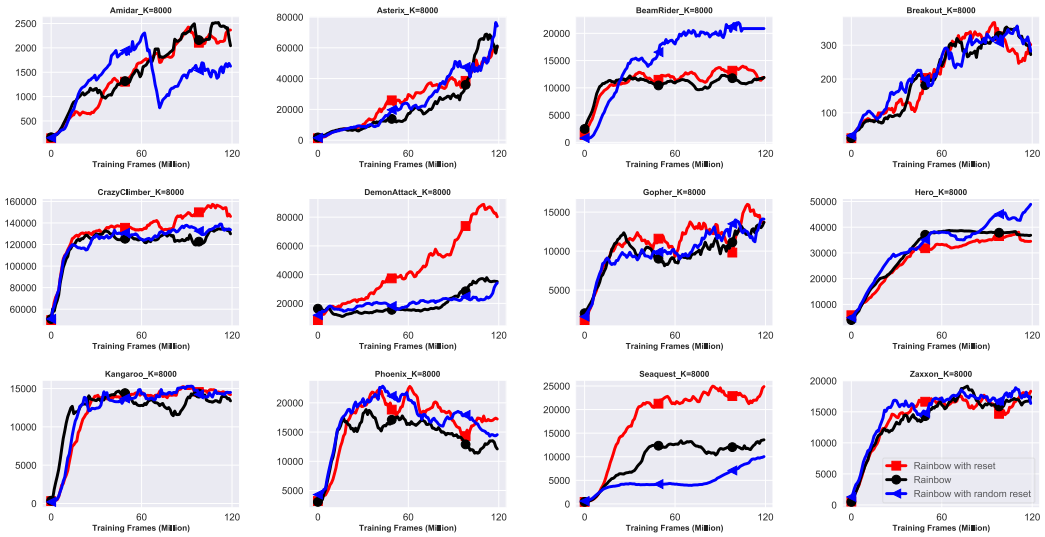


Figure 8: Performance of Rainbow with and without resetting the Adam optimizer and with a fixed value of $K = 8000$ on 12 randomly-chosen Atari games. See section 4 for a description of the random resetting case.

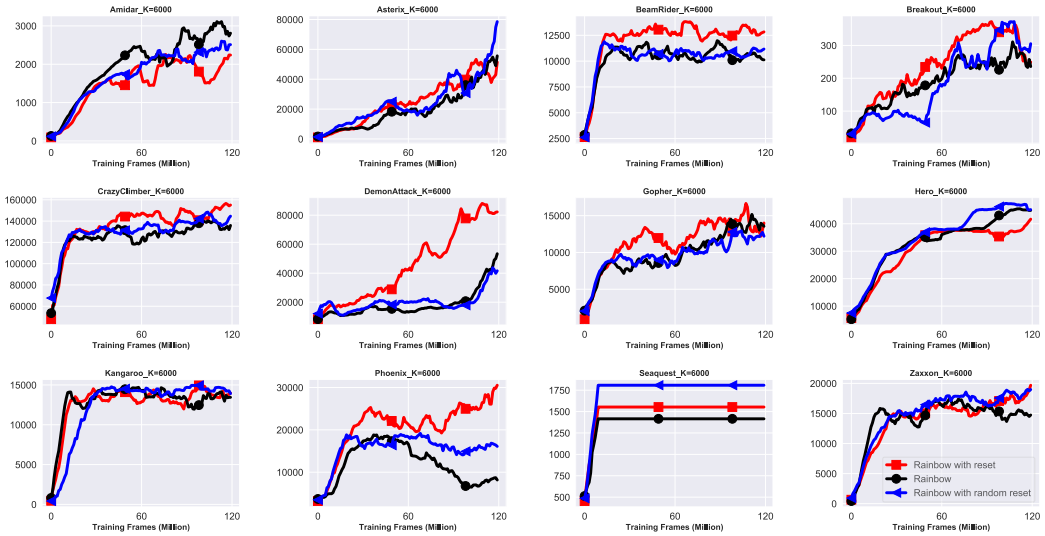


Figure 9: $K = 6000$.

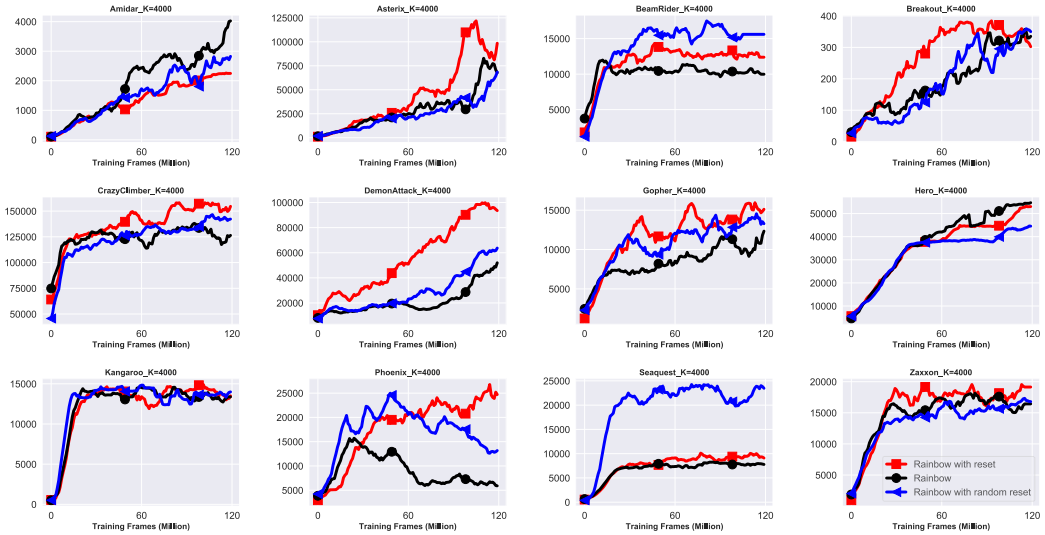


Figure 10: $K = 4000$.

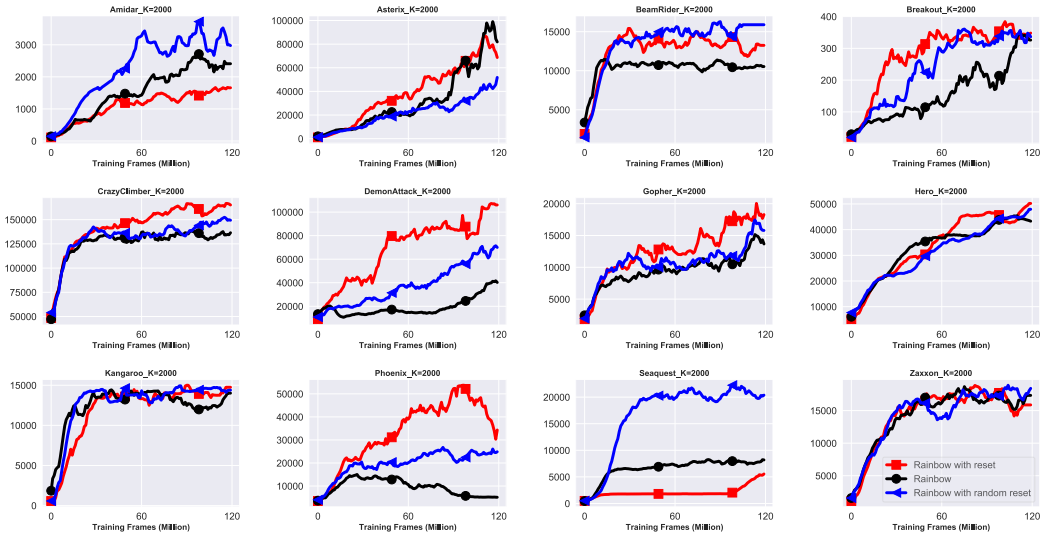


Figure 11: $K = 2000$.

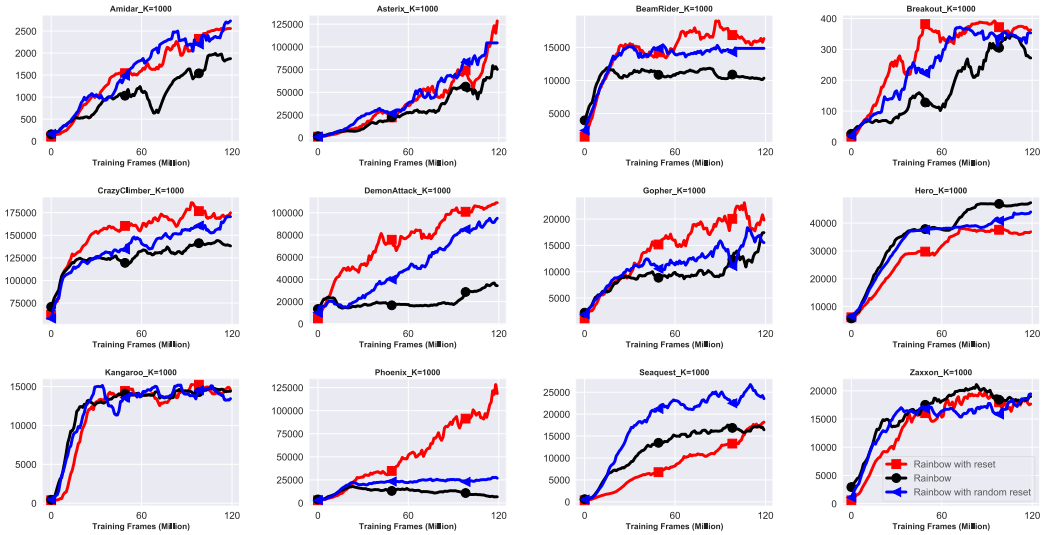


Figure 12: $K = 1000$.

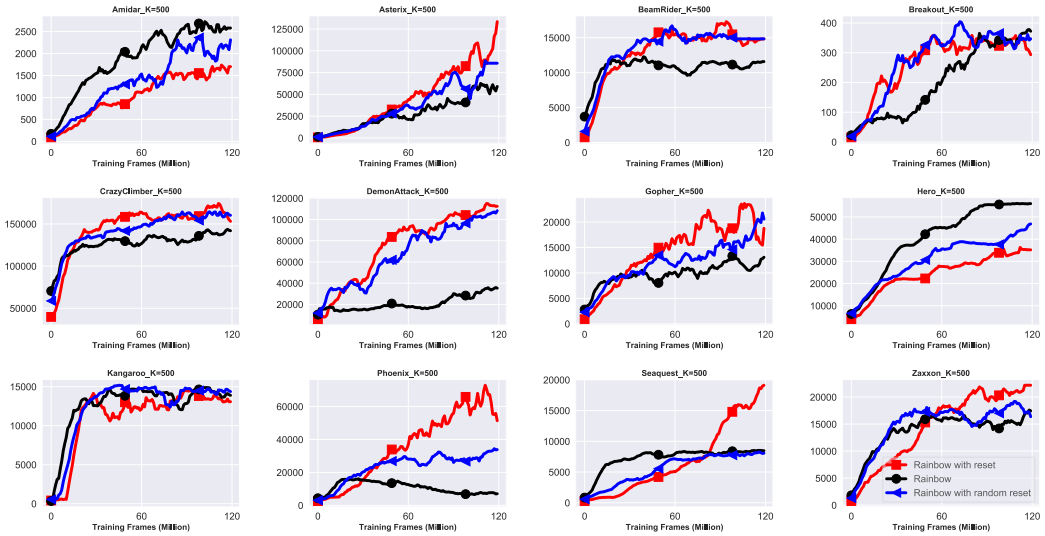


Figure 13: $K = 500$.

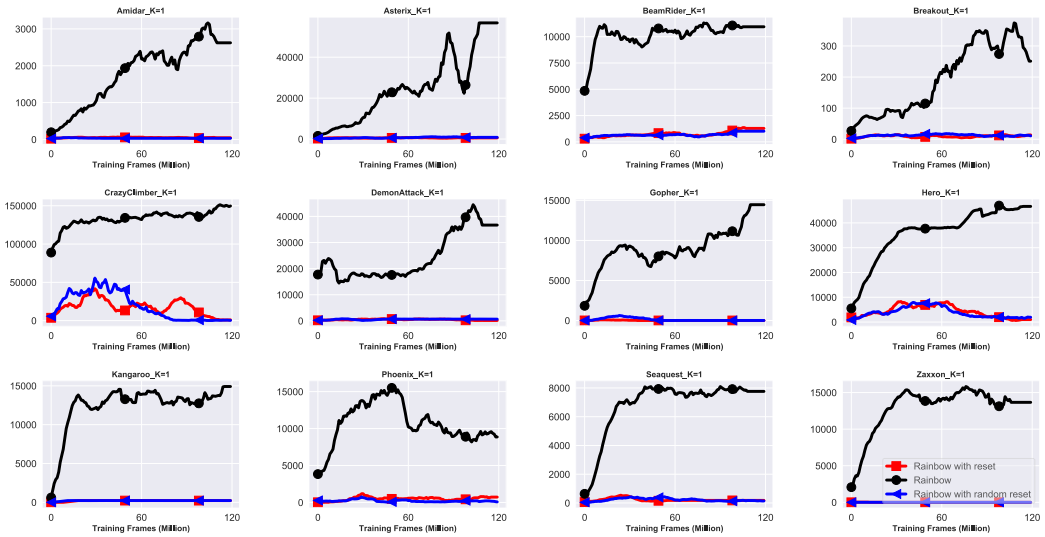


Figure 14: $K = 1$.

We now take the human-normalized median and mean of the results on 12 games and present them for each value of $K$.

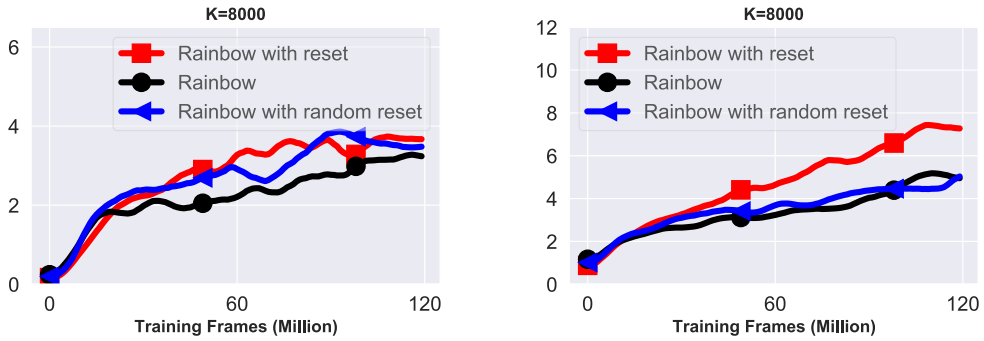


Figure 15: Human normalized median (left) and mean (right) for $K = 800$.

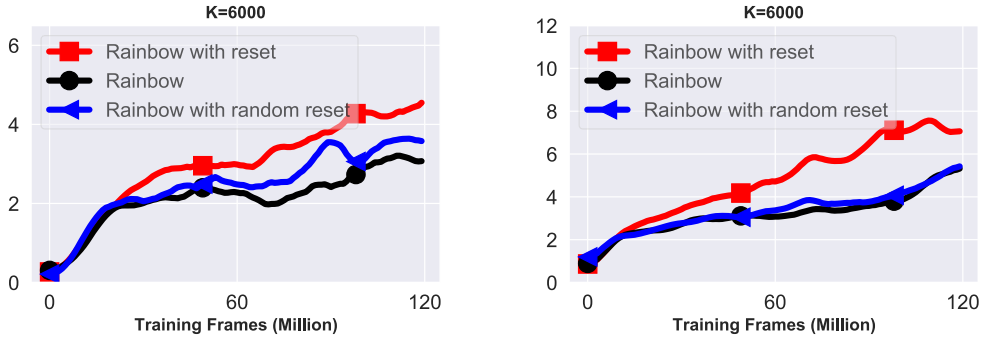


Figure 16: $K = 6000$.

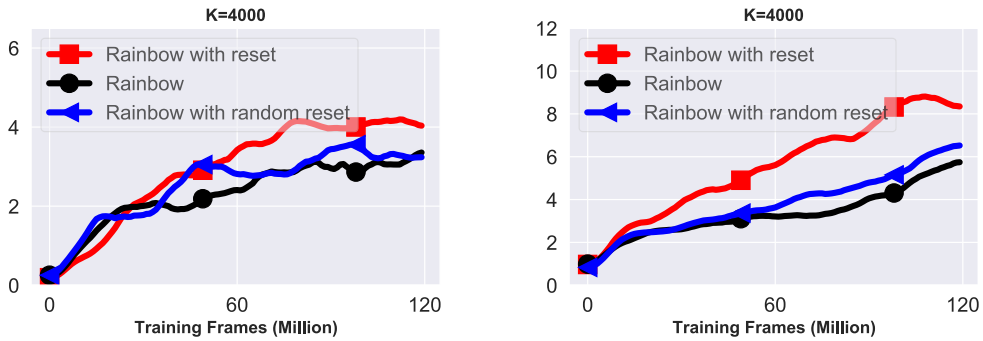


Figure 17: $K = 4000$.

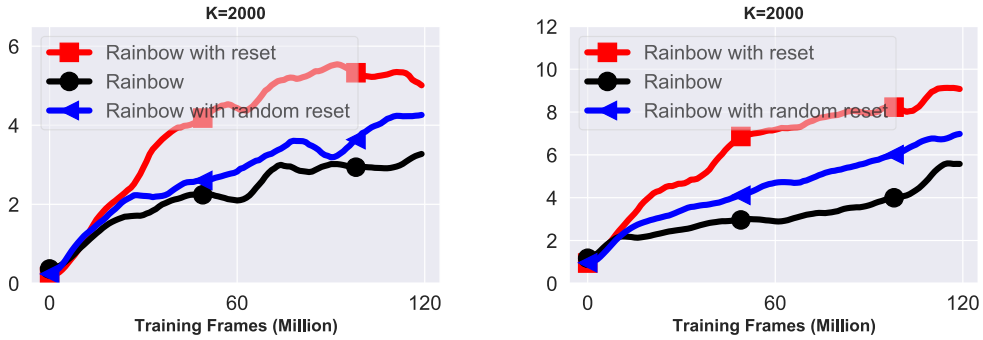


Figure 18: $K = 2000$.

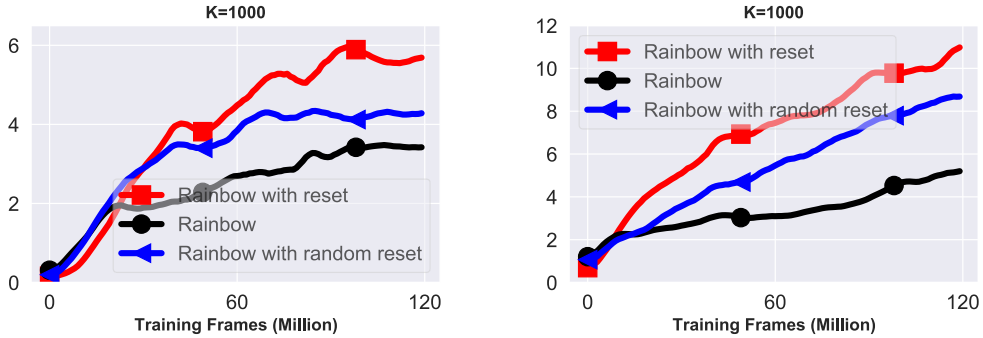


Figure 19: $K = 1000$.

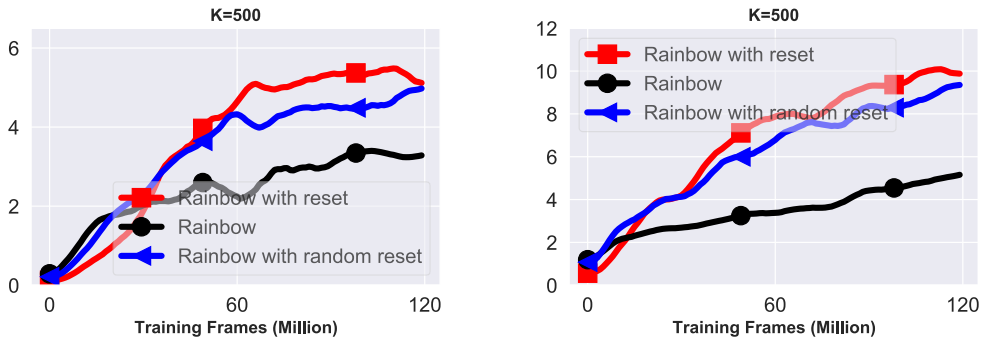


Figure 20: $K = 500$.

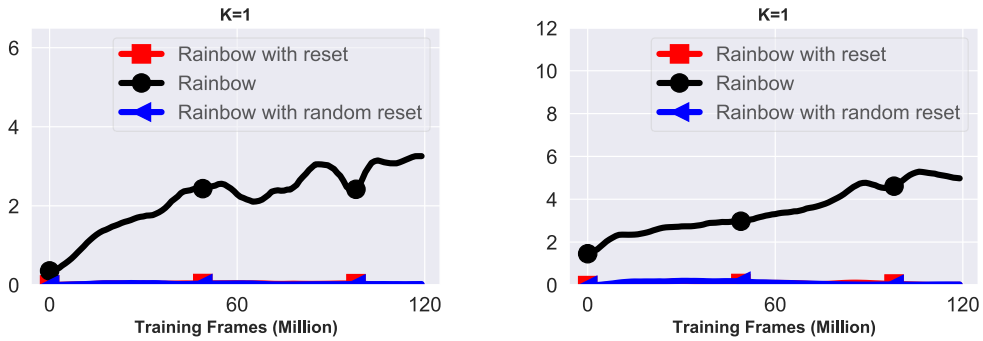


Figure 21: $K = 1$.

We also show area under the curve.

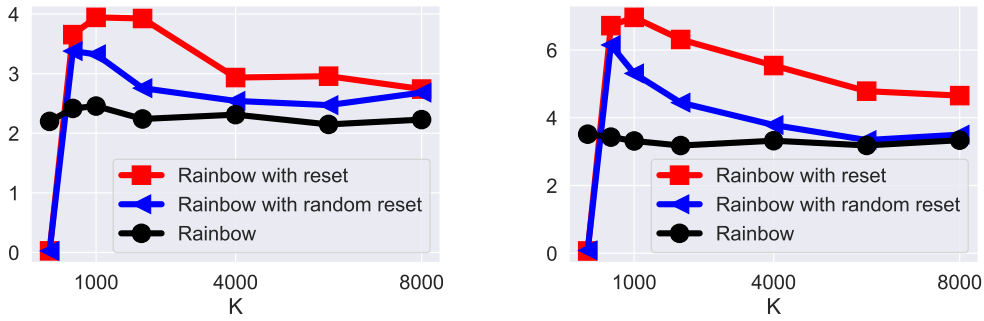


Figure 22: Area under the curve for median (left) and mean (right) of human-normalized performance.

## 7.2 Complete Results from Section 4.2

We now show complete results from section 4.2 starting with Rainbow RMSProp.

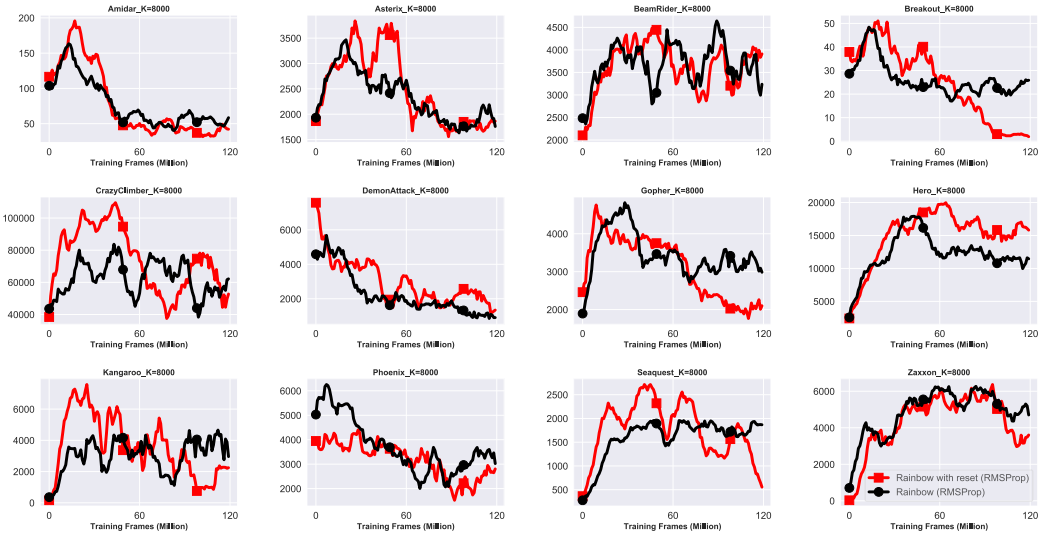


Figure 23: Performance of Rainbow with and without resetting the RMSProp optimizer and with a fixed value of $K = 8000$ on 12 randomly-chosen Atari games.

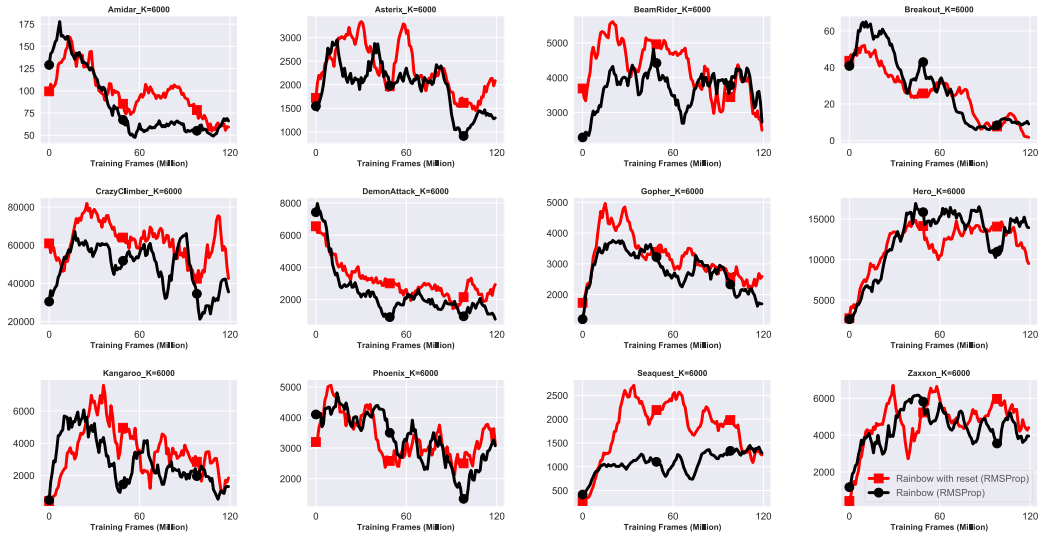


Figure 24: $K = 6000$.

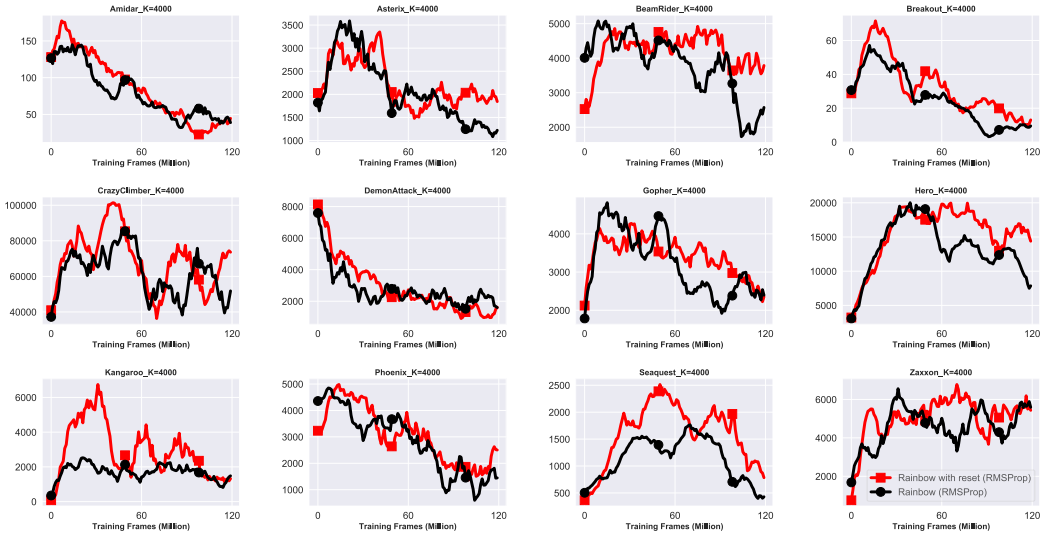


Figure 25: $K = 4000$.

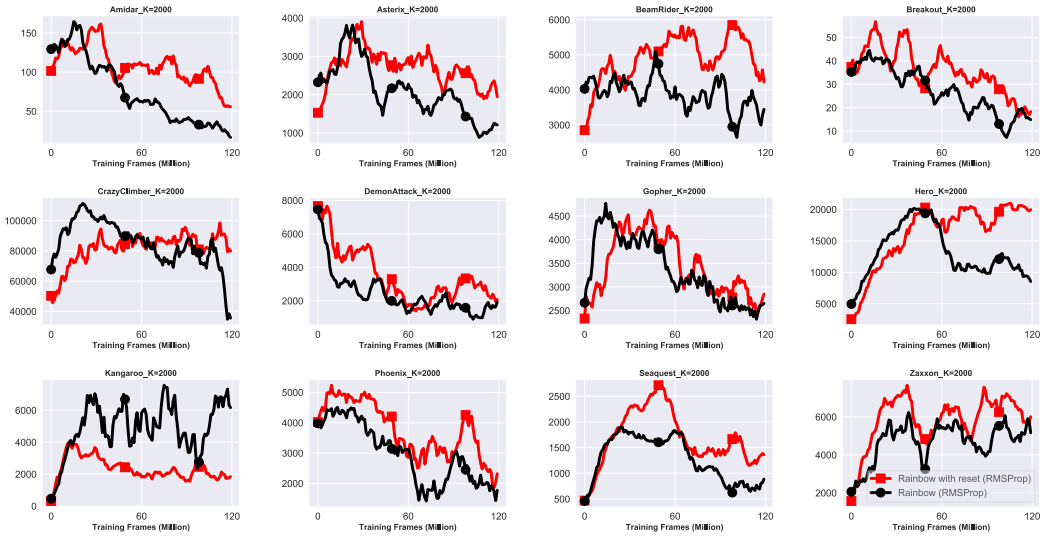


Figure 26: $K = 2000$.

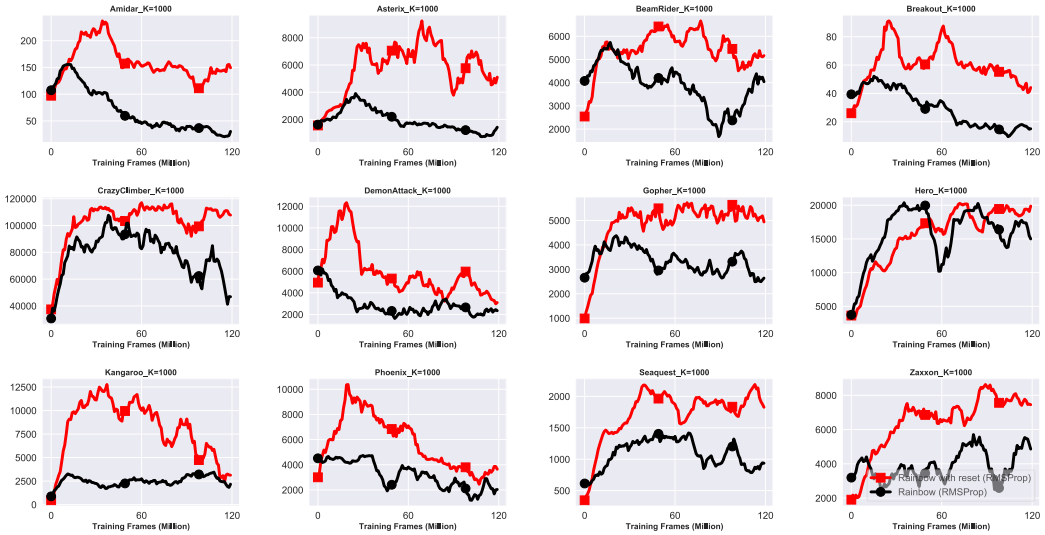


Figure 27: $K = 1000$.

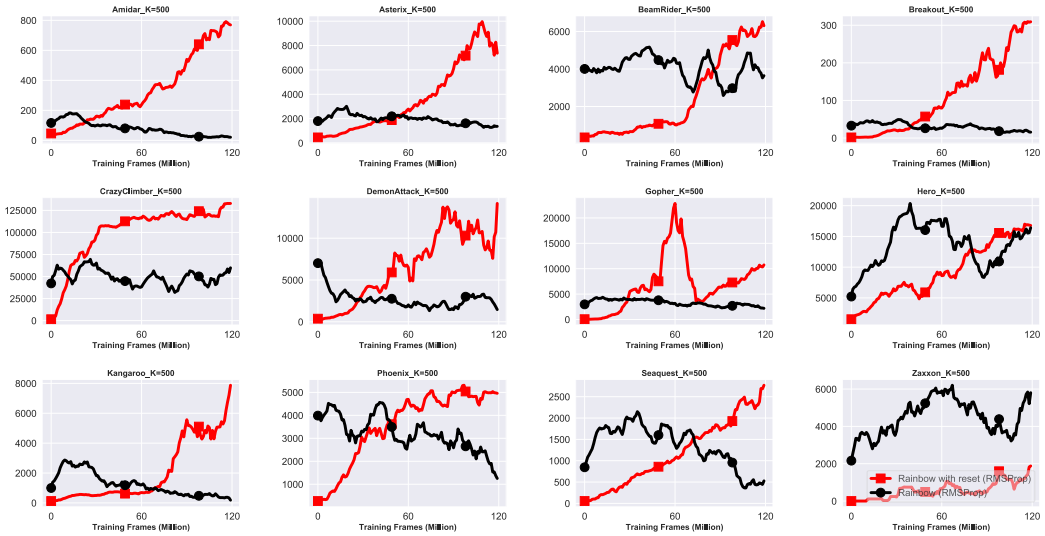


Figure 28: $K = 500$.

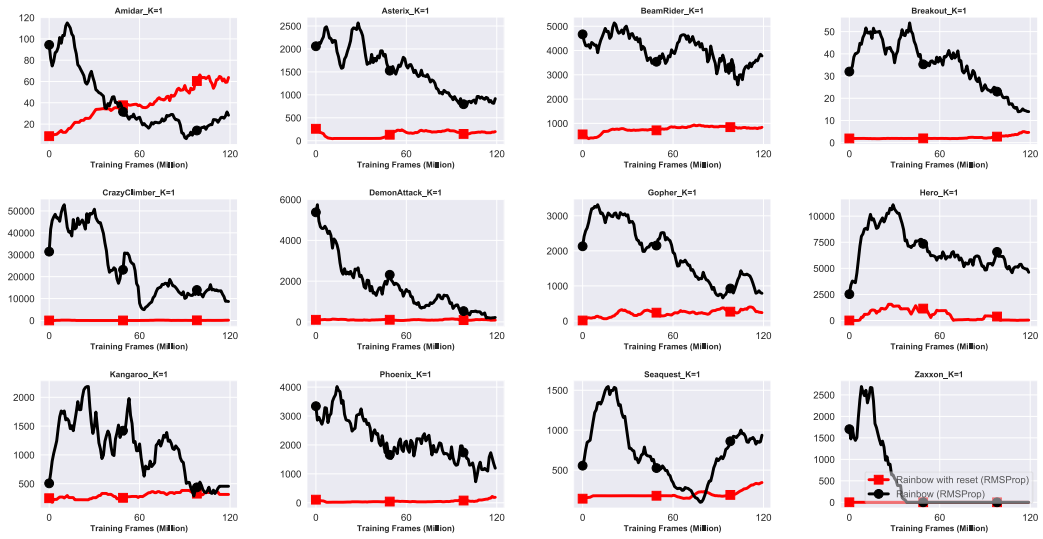


Figure 29: $K = 1$.

We now take the human-normalized median on 12 games and present them for each value of $K$.

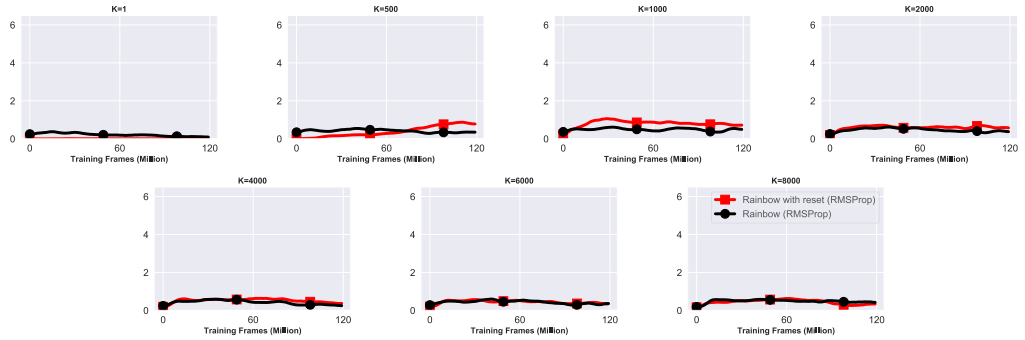


Figure 30: A comparison between Rainbow with and without resetting RMSProp on the 12 Atari games for different values of $K$.

Overall we can see that RMSProp results in poor performance, but resetting can somewhat improve the performance.

We now move to the Rainbow Pro agent.

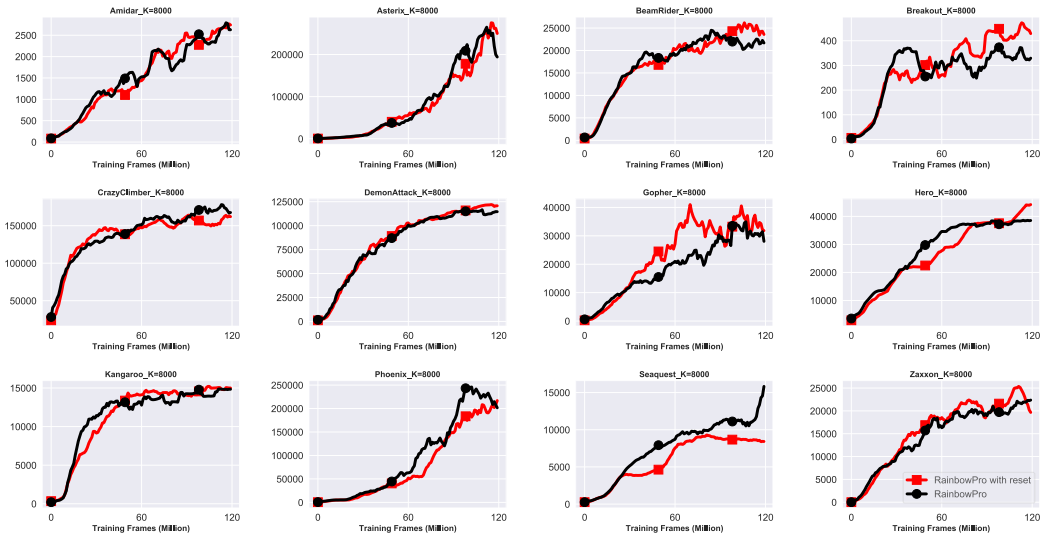


Figure 31: Performance of Rainbow Pro with and without resetting the Adam optimizer and with a fixed value of $K = 8000$ on 12 randomly-chosen Atari games.

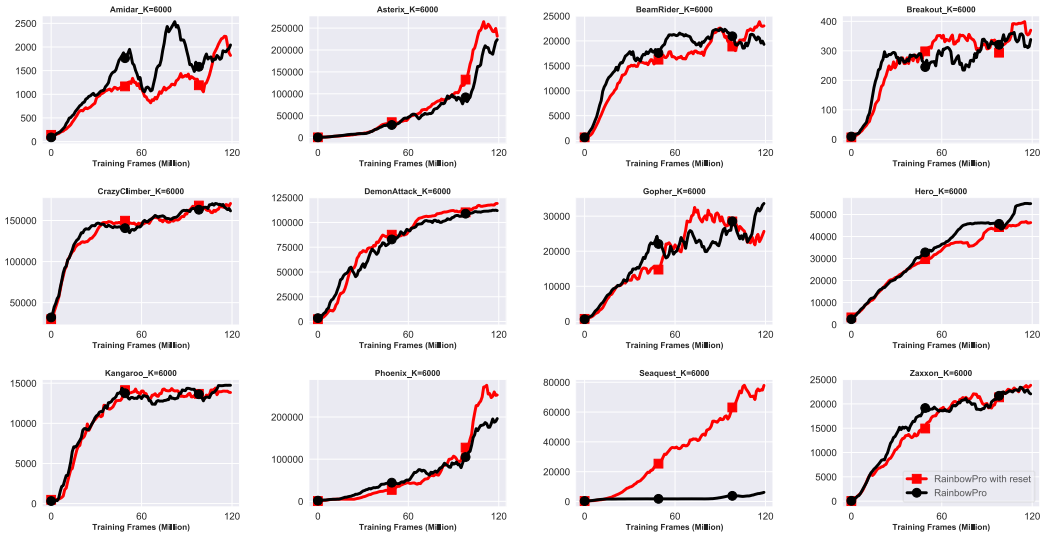


Figure 32: $K = 6000$.

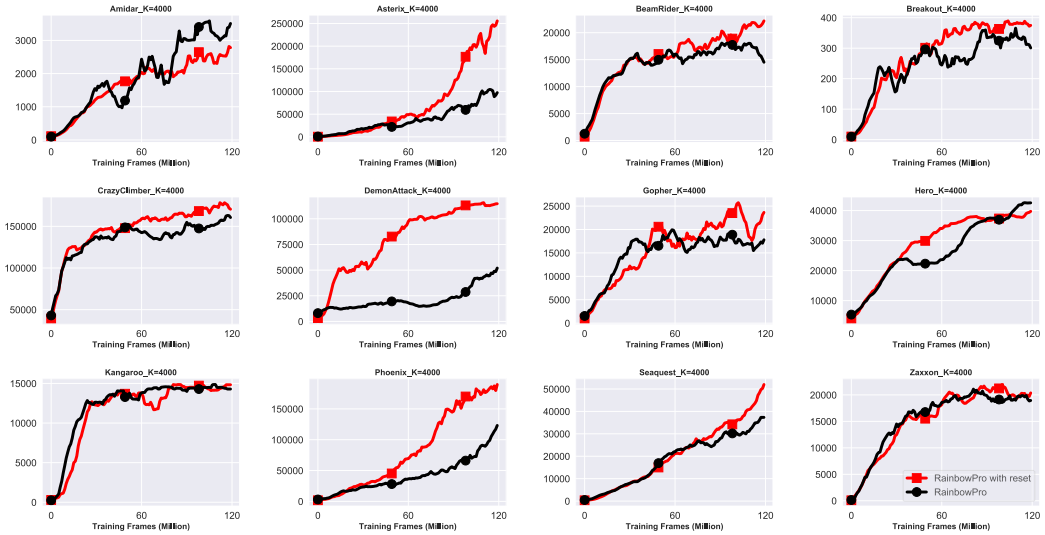


Figure 33: $K = 4000$.

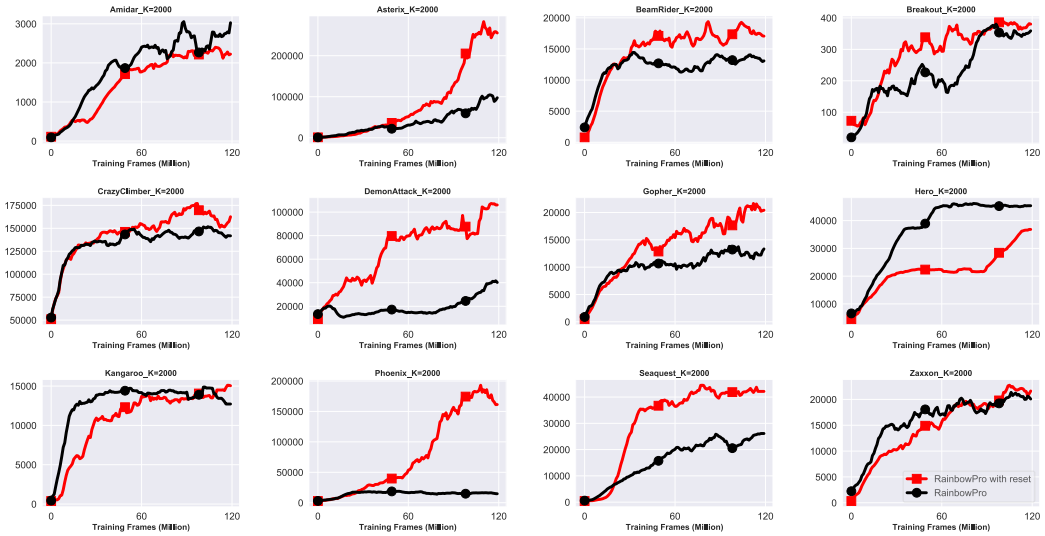


Figure 34: $K = 2000$.

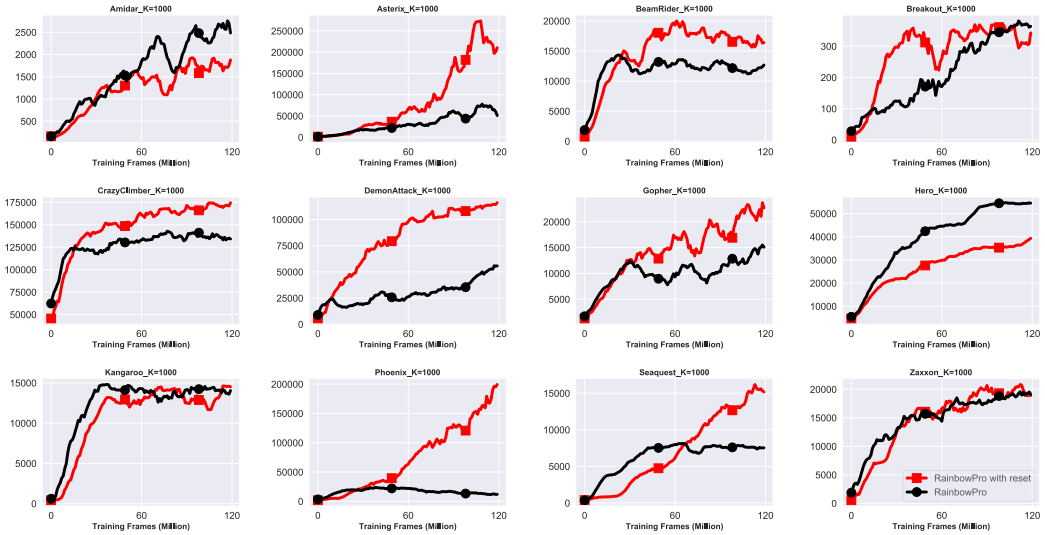


Figure 35: $K = 1000$.

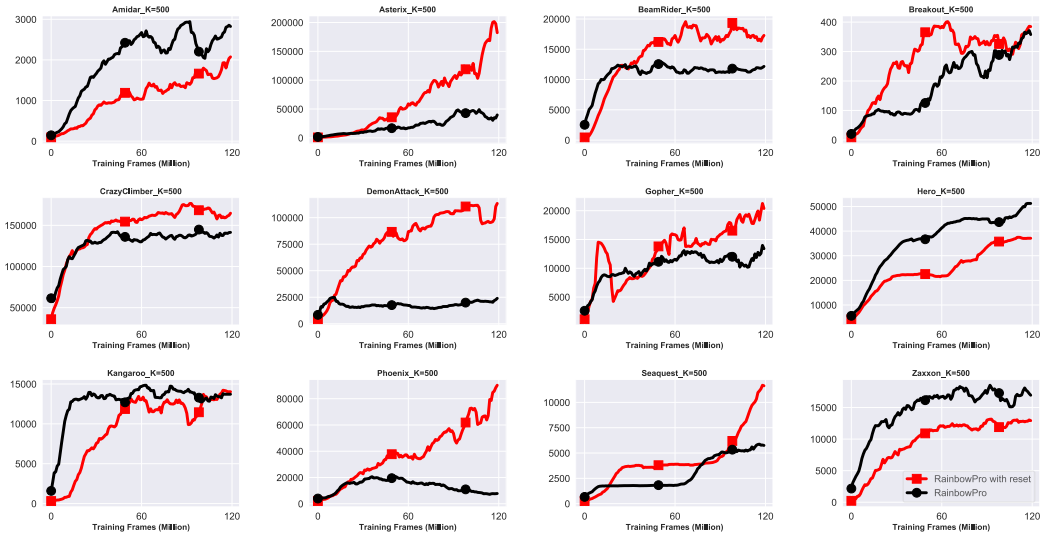


Figure 36: $K = 500$.

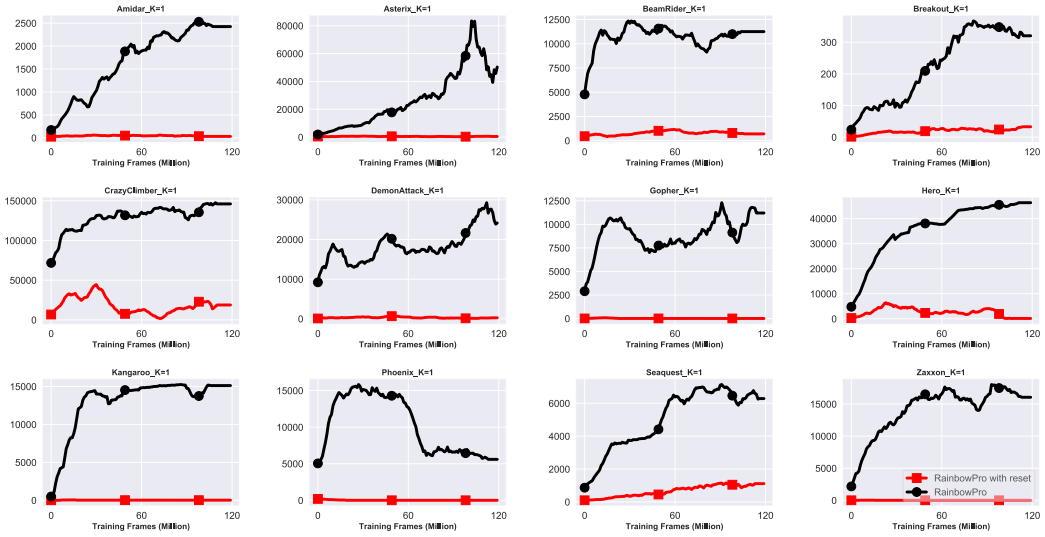


Figure 37: $K = 1$.

We now take the human-normalized median on 12 games and present them for each value of $K$.

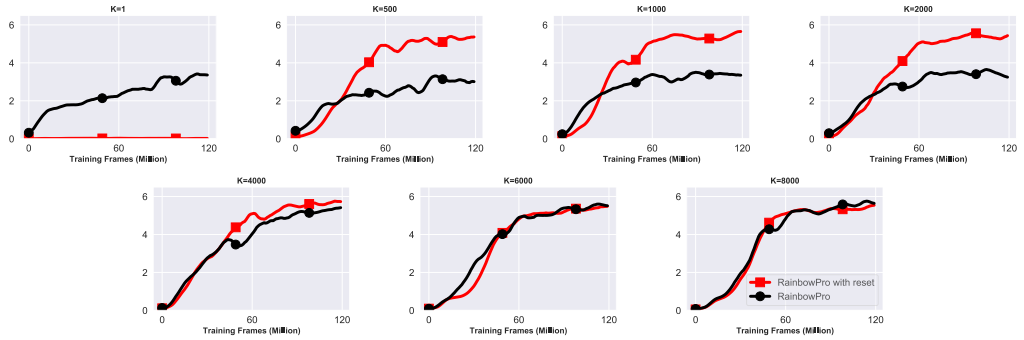


Figure 38: A comparison between Rainbow Pro with and without resetting Adam on the 12 Atari games for different values of $K$.

Overall we can see that resetting makes RainbowPro less sensitive to the $K$ hyper-parameter.

In the last result of this section, we look at Rainbow with the Rectified Adam optimizer.

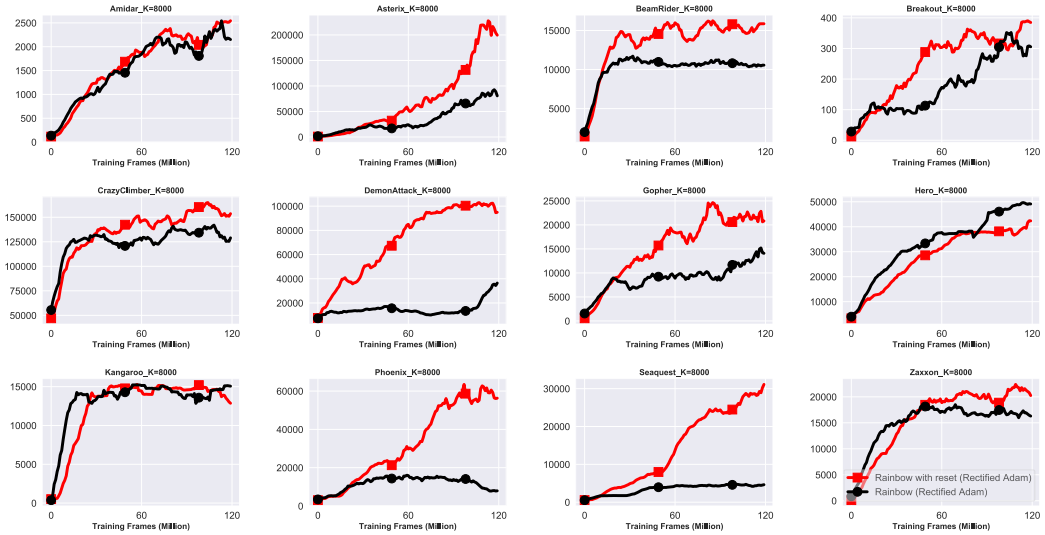


Figure 39: Performance of Rainbow with and without resetting the Rectified Adam optimizer and with a fixed value of $K = 8000$ on 12 randomly-chosen Atari games.

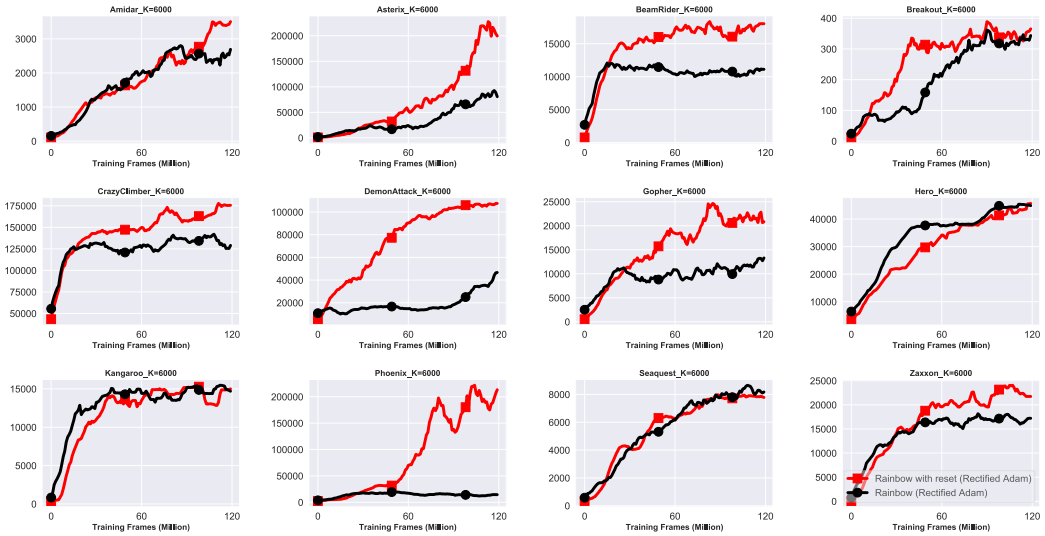


Figure 40: $K = 6000$.

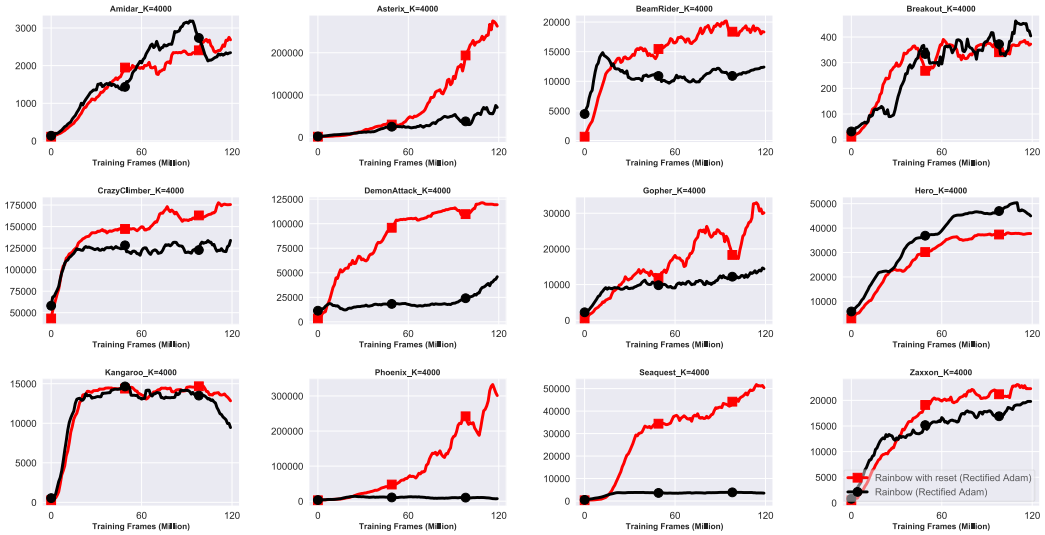


Figure 41: $K = 4000$.

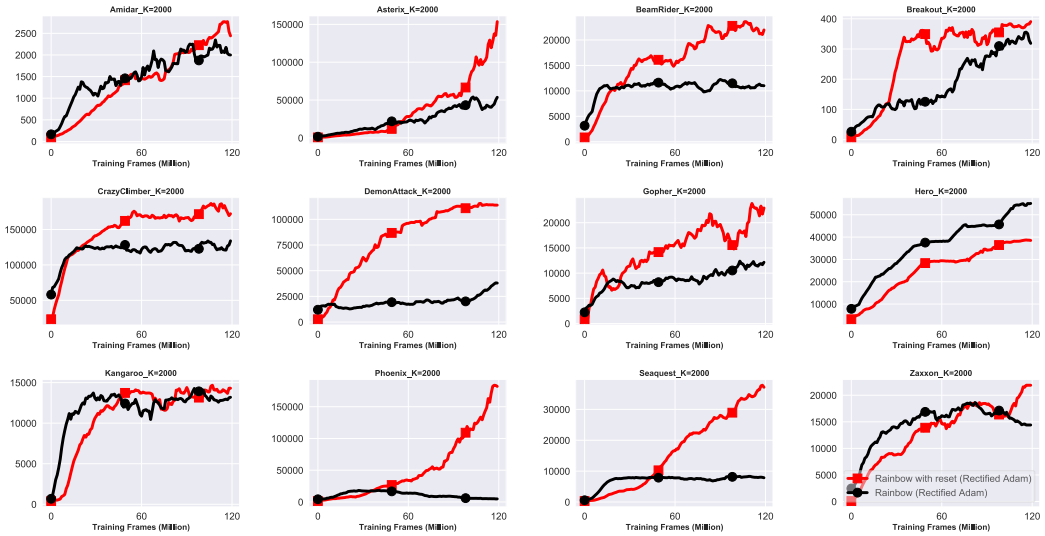


Figure 42: $K = 2000$.

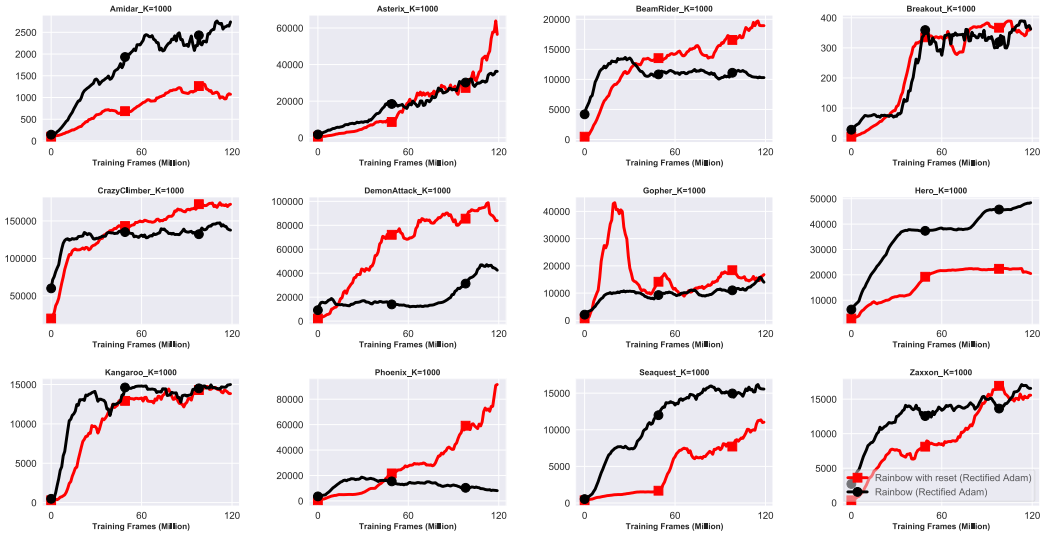


Figure 43: $K = 1000$.

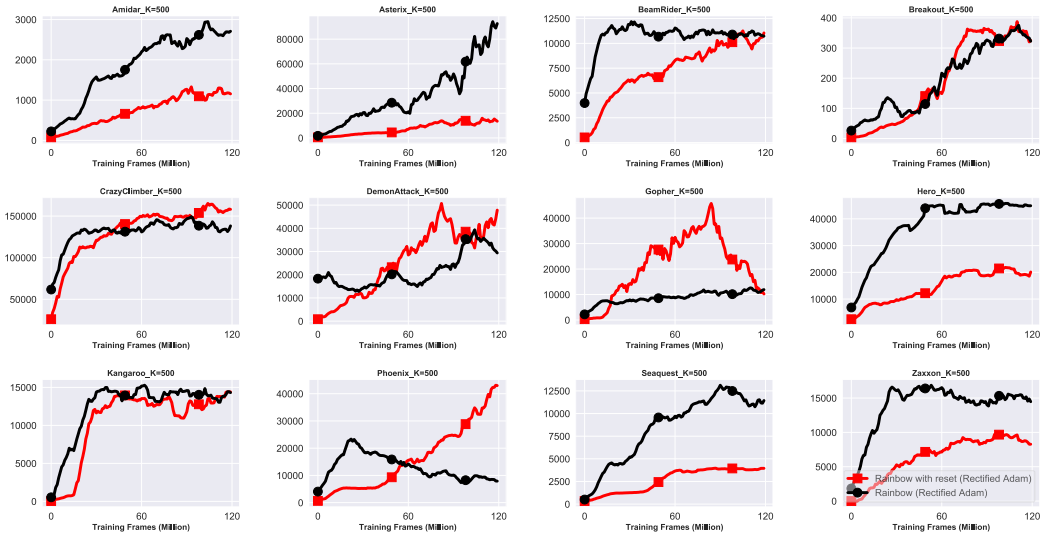


Figure 44: $K = 500$.

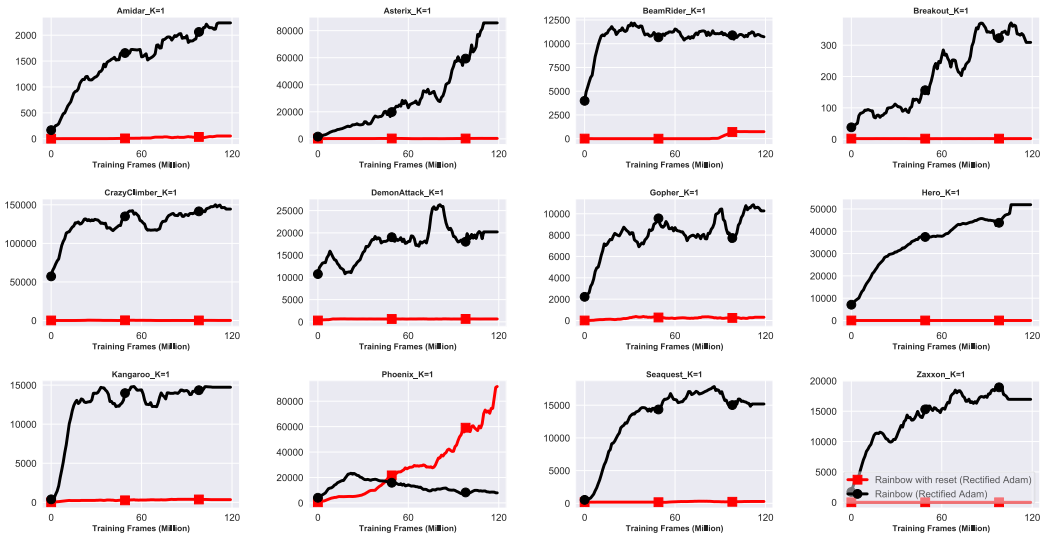


Figure 45: $K = 1$.

We now take the human-normalized median on 12 games and present them for each value of $K$.

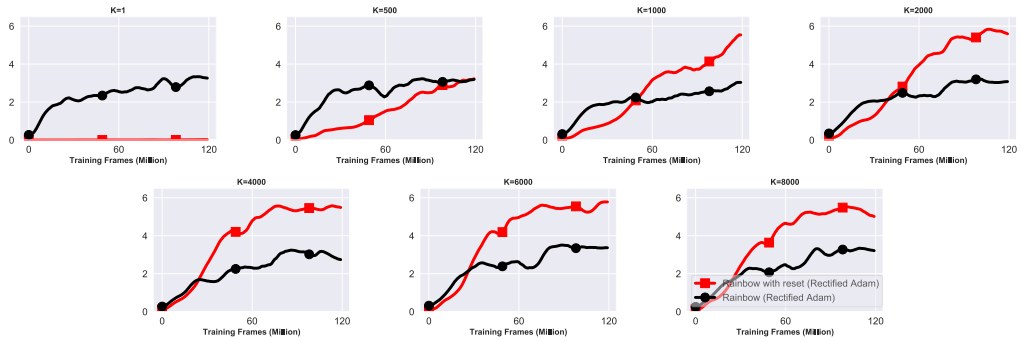


Figure 46: A comparison between Rainbow with and without resetting Rectified Adam on the 12 Atari games for different values of $K$.

Overall we can see that resetting improves Rainbow with Recitified Adam.

# 8 Complete Results From Section 4.3

We now show full learning curves for all 55 Atari games and over 10 random seeds. We benchmark three agents: the default Rainbow agent from the Dopamine (no reset), Rainbow with resetting the Adam optimizer, and Rainbow with resetting the rectified Adam optimizer.

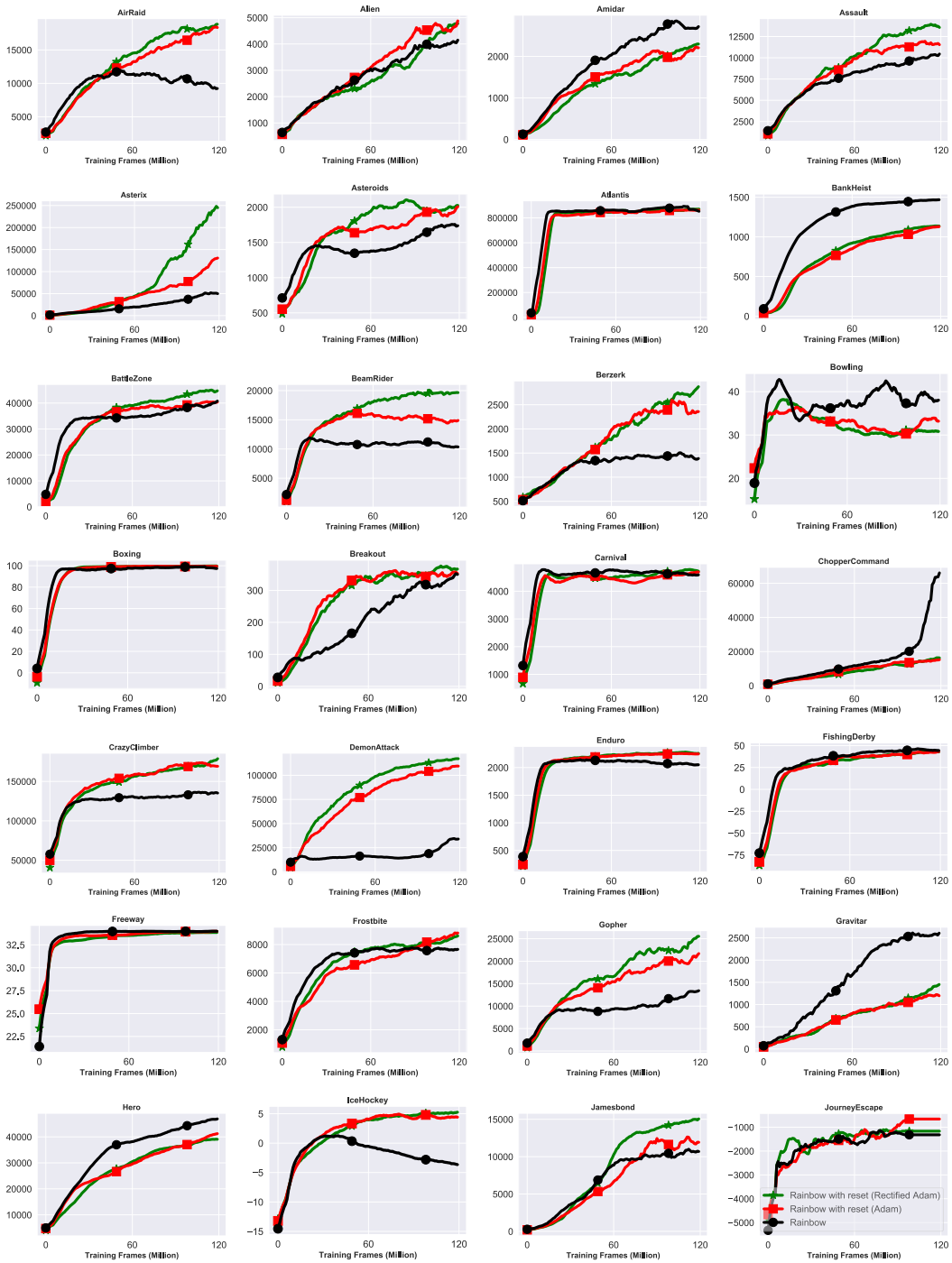


Figure 47: Full learning curves (Part I) averaged over 10 seeds.

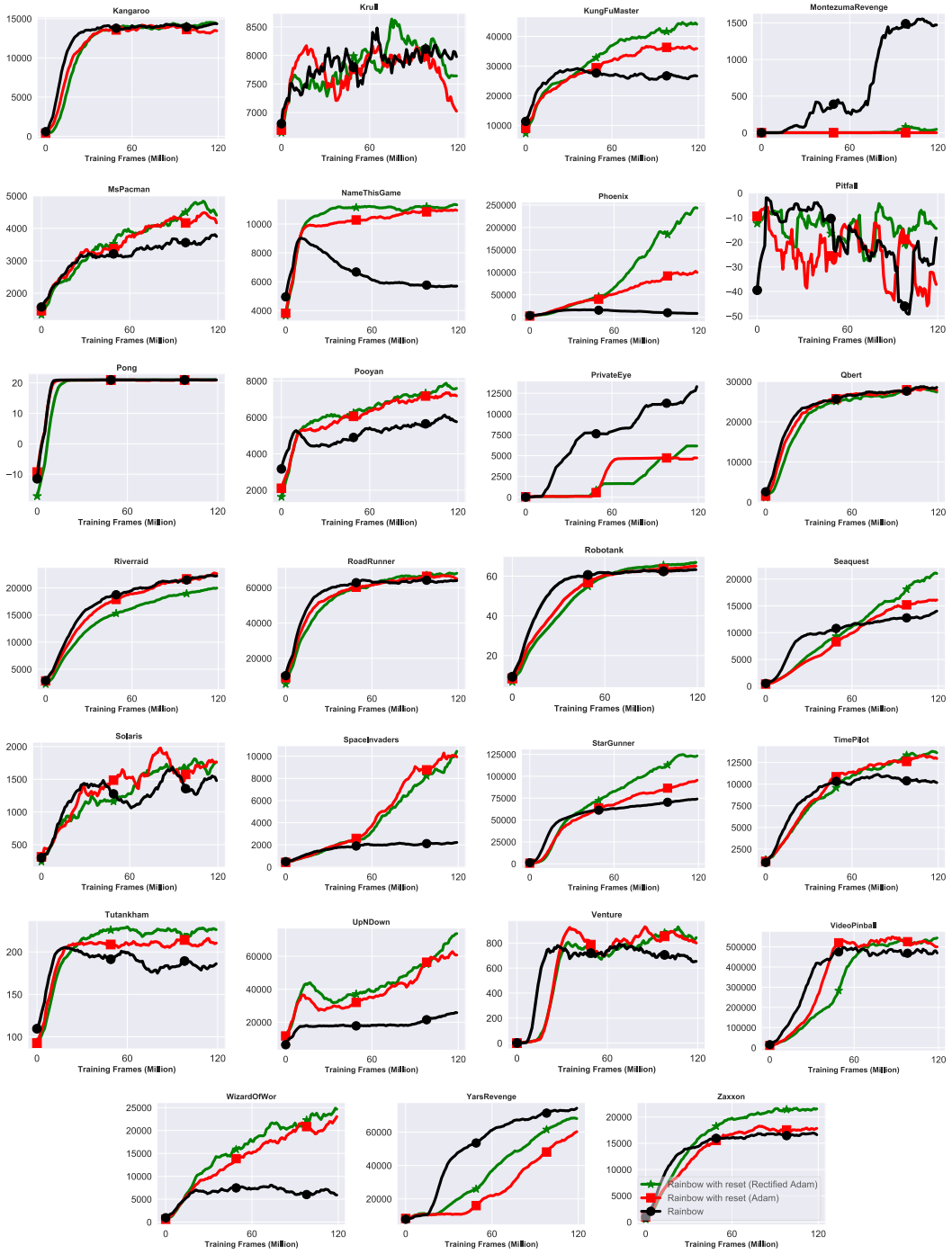


Figure 48: Full learning curves (Part II) averaged over 10 seeds..

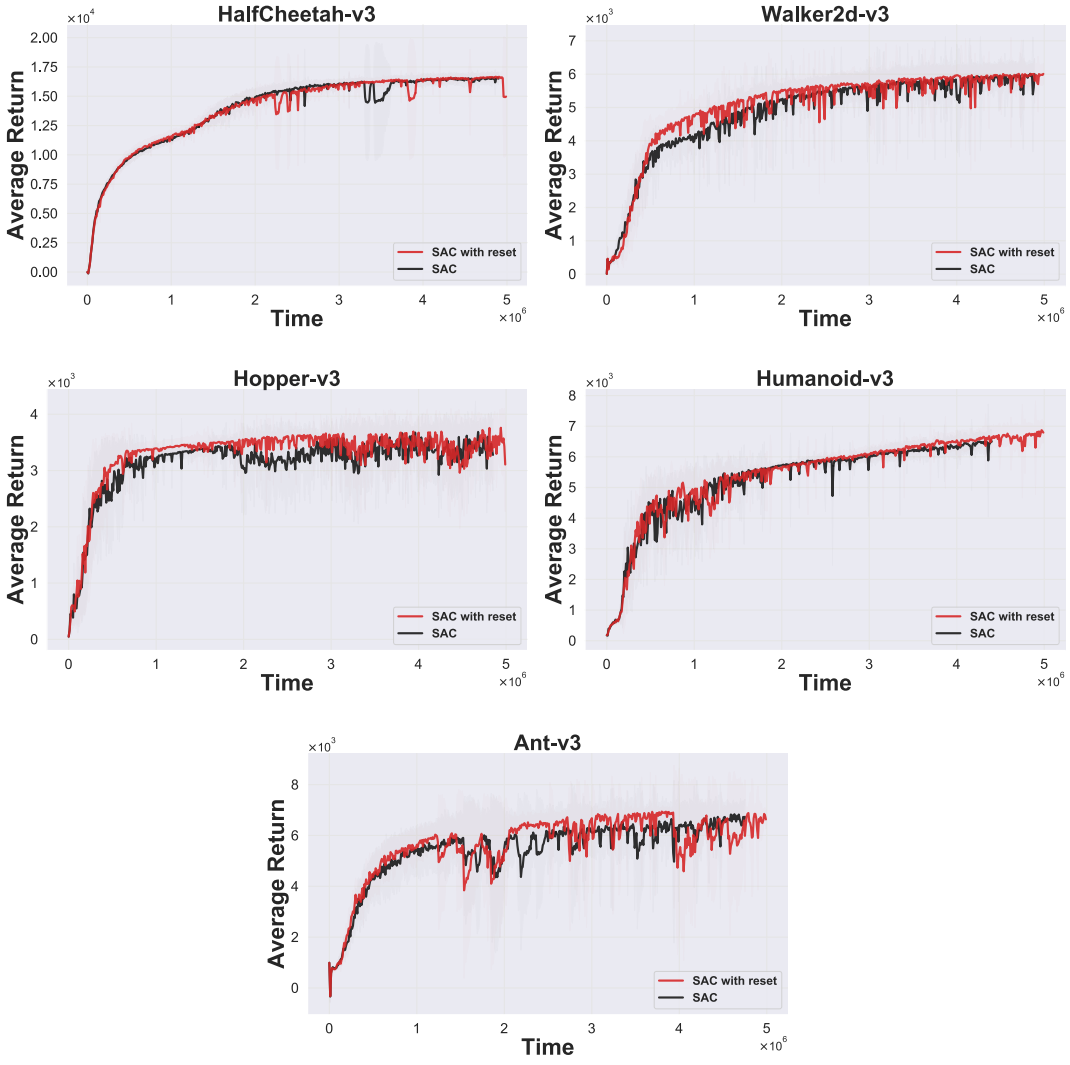


Figure 49: A comparison between Soft Actor-Critic (SAC) with and without resetting Adam on the standard MuJoCo tasks. In this study, both the actor and critic optimizers are reset every 5000 steps. The results are averaged over 10 different seeds.

## 8.1 Complete Results From Section 4.4

We finally present results on continuous control task with soft actor critic (SAC) and the Adam optimizer, where we reset the optimizers every 5000 steps. Note that, in contrsat to Atari and Rainbow, the target parameter $\theta$ is updated using the Polyak strategy, so it is less clear when to reset the optimizer. Thus we chose the simple strategy of resetting the optimizer every 5000 steps. We leave further exploration of resetting with Polyak updates to future work.