

APPENDICES: Revisiting Fundamentals of Experience Replay

A. Experimental details

A.1. The Dopamine Rainbow agent

Our empirical investigations in this paper are based on the Dopamine Rainbow agent (Castro et al., 2018). This is an open source implementation of the original agent (Hessel et al., 2018), but makes several simplifying design choices. The original agent augments DQN through the use of (a) a distributional learning objective, (b) multi-step returns, (c) the Adam optimizer, (d) prioritized replay, (e) double Q-learning, (f) duelling architecture, and (g) noisy networks for exploration. The Dopamine Rainbow agent uses just the first four of these adjustments, which were identified as the most important aspects of the agent in the original analysis of Hessel et al. (2018).

A.2. Atari 2600 games used

A 14 game subset was used for the grid measuring the effects of varying replay capacity and oldest policy. A 20 game subset, which is comprised of the 14 games used for the grid with 6 additional games, was used for all other experiments.

14 game subset: AIR RAID, ASTERIX, BREAKOUT, FREEWAY, GRAVITAR, JAMES BOND, MONTEZUMA’S REVENGE, MS. PACMAN, PRIVATE EYE, Q*BERT, SEAQUEST, SPACE INVADERS, VENTURE, ZAXXON.

20 game subset: The 14 games above in addition to: ASTEROIDS, BOWLING, DEMON ATTACK, PONG, WIZARD OF WOR, YARS’ REVENGE.

B. Additive and ablative studies

B.1. DQN additions

We provide game-level granularity on the performance of each supplemented DQN agent in Figure 10.

B.2. Rainbow ablations

We provide game-level granularity on the performance of each ablated Rainbow agent in Figure 11.

C. Error analysis for rainbow grid

We provide an error analysis for each of the elements in Figure 2 (reproduced here as Figure 12) by providing the 25% and 75% percentile improvements for each combination of replay capacity and oldest policy. These results are given in Figure 13.

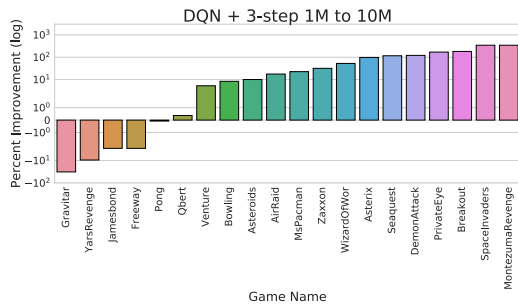
We present an alternative view of the data using a bootstrap estimation technique. Instead of fixing the seeds for both the baseline agent and our new agent at each cell, we sample, with replacement, the seeds. We carry out this procedure repeatedly and report the mean and standard deviations in Figure 14.

D. Replay buffer size

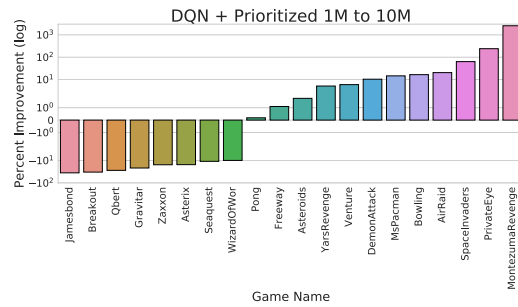
We provide a different perspective on the data from Figure 2 in Figure 15, illustrating a general relationship between replay ratio and performance improvement. We provide game-level granularity on the performance of Rainbow with varying buffer sizes in Figure 17. In Figure 16 we also gives results for varying replay buffer size and age of oldest policy for DQN, 3-step DQN, and Rainbow.

E. Batch RL learning curves

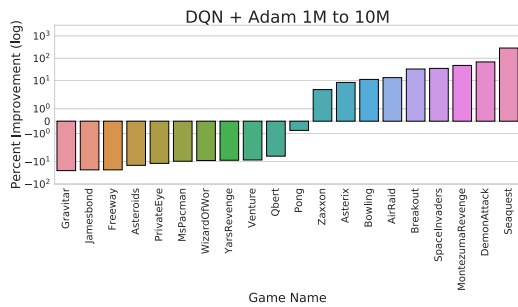
In Figures 18 and 19, we provide learning curves for the batch RL agents described in Section 4.2.



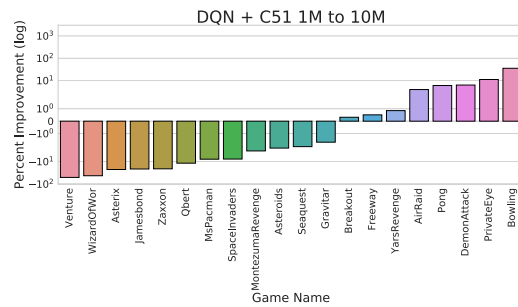
(a) DQN + 3-step provides a median performance change of +24.8%.



(b) DQN + PER provides a median performance change of +1.5%.



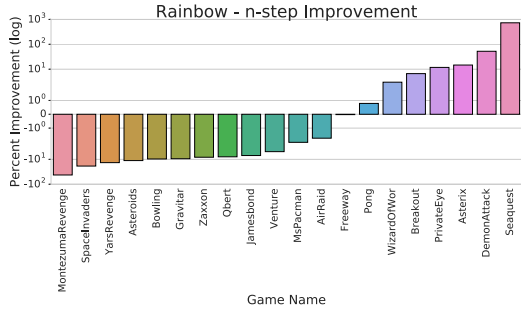
(c) DQN + Adam provides a median performance change of -0.6%.



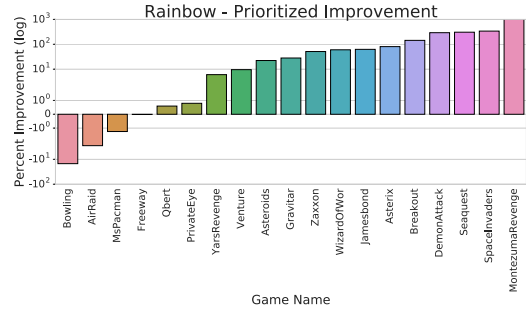
(d) DQN + C51 provides a median performance change of -3.0%.

Figure 10. Only DQN with n -step improves with increased capacity. DQN with an additional component results at a per-game level, measuring performance changes when increasing replay capacity from 1M to 10M.

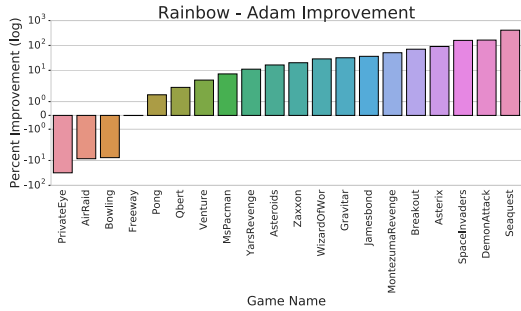
Revisiting Fundamentals of Experience Replay



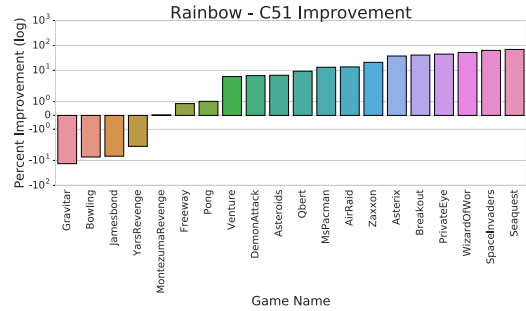
(a) The performance difference for a Rainbow agent without n-step returns when increasing the replay buffer size from 1M to 10M. We find that the resultant agent does not benefit from larger replay buffer sizes, reporting a median performance decrease of 2.3%. This implies the importance of n-step returns.



(b) The performance difference for a Rainbow agent without prioritized experience replay when increasing the replay buffer size from 1M to 10M. Even without prioritization, the algorithm still benefits +17.3%.



(c) A Rainbow agent without Adam optimizer has a median performance increase of +27.0% when the replay buffer size is increased from 1M to 10M.



(d) The performance difference for a Rainbow agent without C51 optimizer when increasing the replay buffer size from 1M to 10M. Median performance change of +26.6%.

Figure 11. Rainbow ablation results at a per-game level.

		Replay Capacity				
		100,000	316,228	1,000,000	3,162,278	10,000,000
Oldest Policy	25,000,000	-74.9	-76.6	-77.4	-72.1	-54.6
	2,500,000	-78.1	-73.9	-56.8	-16.7	28.7
	250,000	-70.0	-57.4	0.0	13.0	18.3
	25,000	-31.9	12.4	16.9	--	--

Figure 12. Performance consistently improves with increased replay capacity and generally improves with reducing the age of the oldest policy. We reproduce the median percentage improvement over the Rainbow baseline when varying the replay capacity and age of the oldest policy in Rainbow on a 14 game subset of Atari.

		Replay Capacity				
		100,000	316,228	1,000,000	3,162,278	10,000,000
Oldest Policy	25,000,000	-78.2	-77.6	-78.8	-76.6	-58.5
	2,500,000	-82.1	-74.2	-64.9	-18.8	26.8
	250,000	-74.6	-58.0	-1.0	8.8	13.2
	25,000	-37.5	5.6	14.5	--	--

		Replay Capacity				
		100,000	316,228	1,000,000	3,162,278	10,000,000
Oldest Policy	25,000,000	-74.1	-76.3	-77.3	-66.6	-53.2
	2,500,000	-77.5	-70.1	-56.5	-15.2	29.5
	250,000	-67.0	-54.8	1.1	16.2	20.3
	25,000	-26.2	18.7	18.6	--	--

Figure 13. 25% (left) and 75% (right) percentile performance improvement over the Rainbow baseline when the replay capacity and age of the oldest policy in Rainbow on a 14 game subset of Atari.

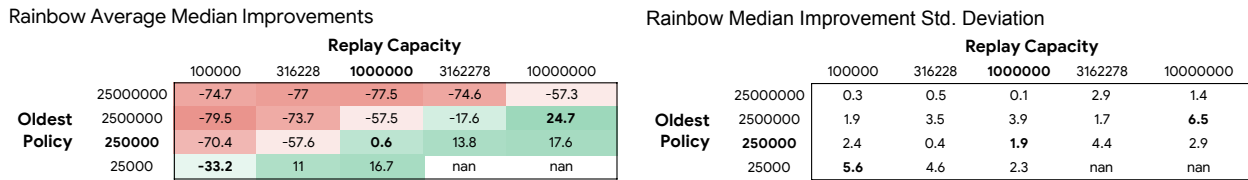


Figure 14. **Bootstrap estimate of variance for each cell.** For each cell, rather than using the same 3 seeds for the baseline and the same 3 seeds for each agent of each cell, we consider sampling seeds with replacement. The left grid shows the mean median improvement and the right grid shows the standard deviation of the median improvement.

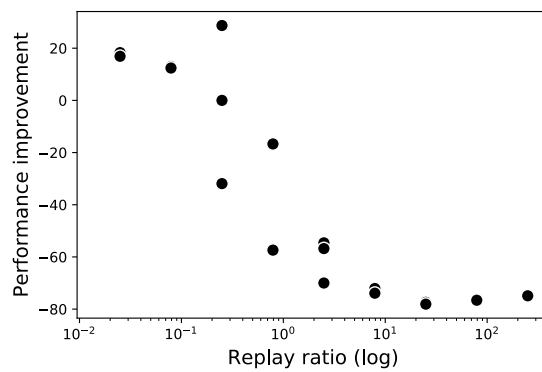
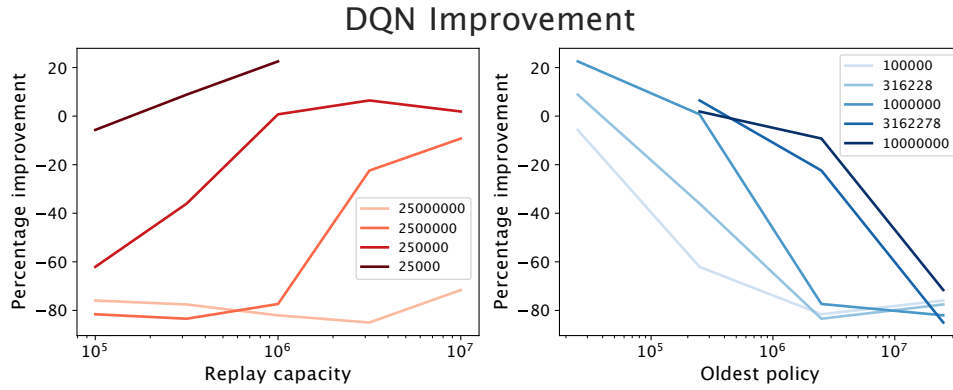
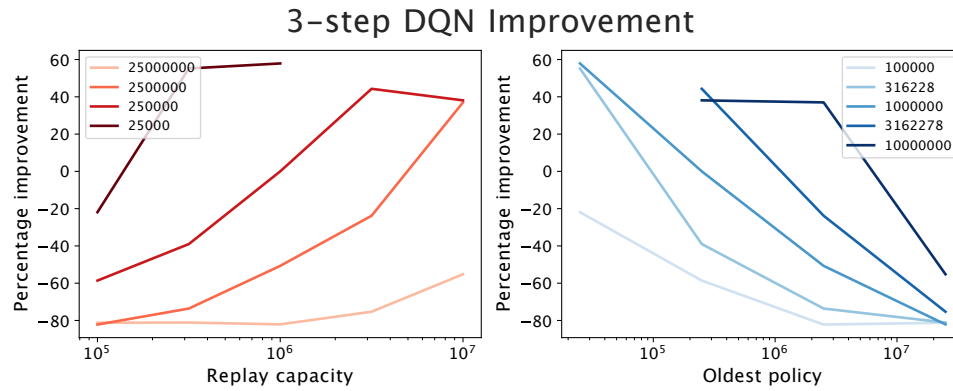


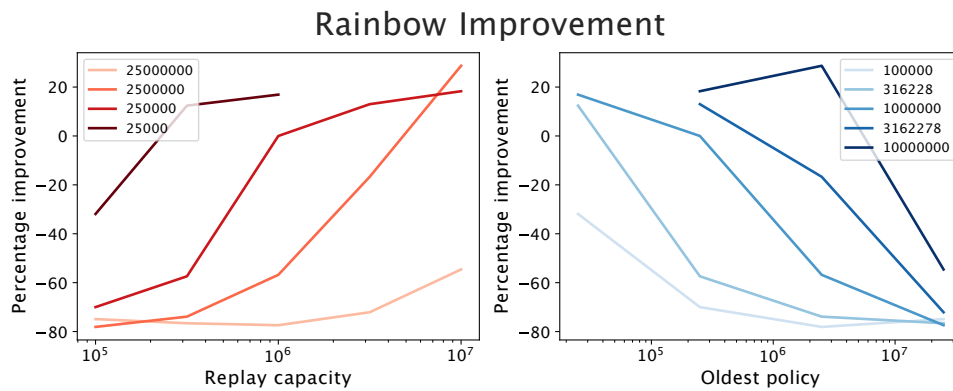
Figure 15. **Performance improvements increase as replay ratio drops.** We plot the results of Figure 2 with respect to the replay ratio, which shows the general trend.



(a) DQN does not show benefits of replay larger than IM, unless older policies are used



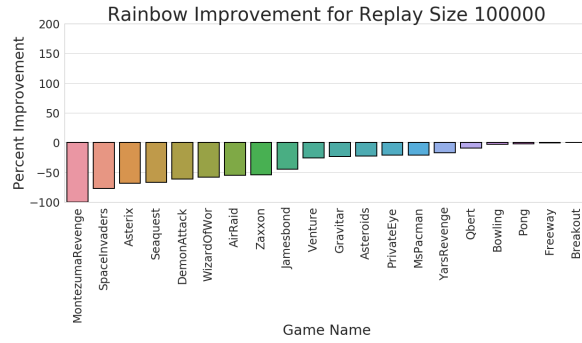
(b) 3-step DQN continues to improve with larger replay capacity



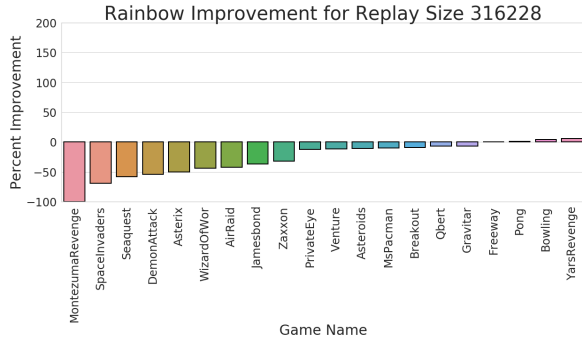
(c) Rainbow continues to improve with larger replay capacity

Figure 16. Performance improves with increased replay capacity and reduced oldest policy age.

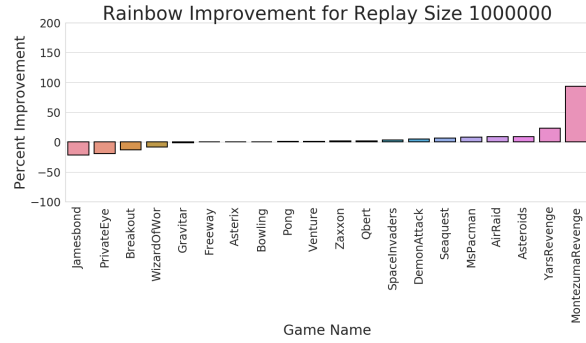
Revisiting Fundamentals of Experience Replay



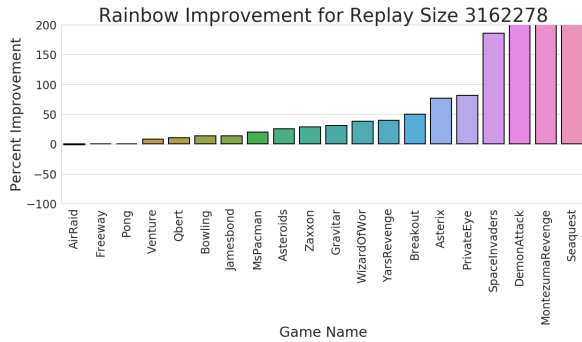
(a) 1M to 100K buffer. Median improvement: -24.6%.



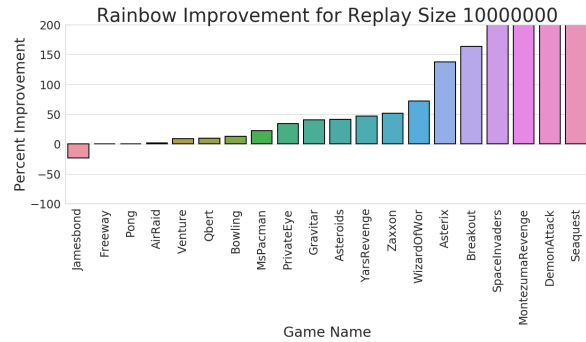
(b) 1M to 316K buffer. Median improvement: -11.8%.



(c) 1M to 1M buffer. Median improvement: 1.6%.



(d) 1M to 3M buffer. Median improvement: 30.2%.



(e) 1M to 10M buffer. Median improvement: 41.0%.

Figure 17. Rainbow replay buffer effects at a per-game level.

Revisiting Fundamentals of Experience Replay

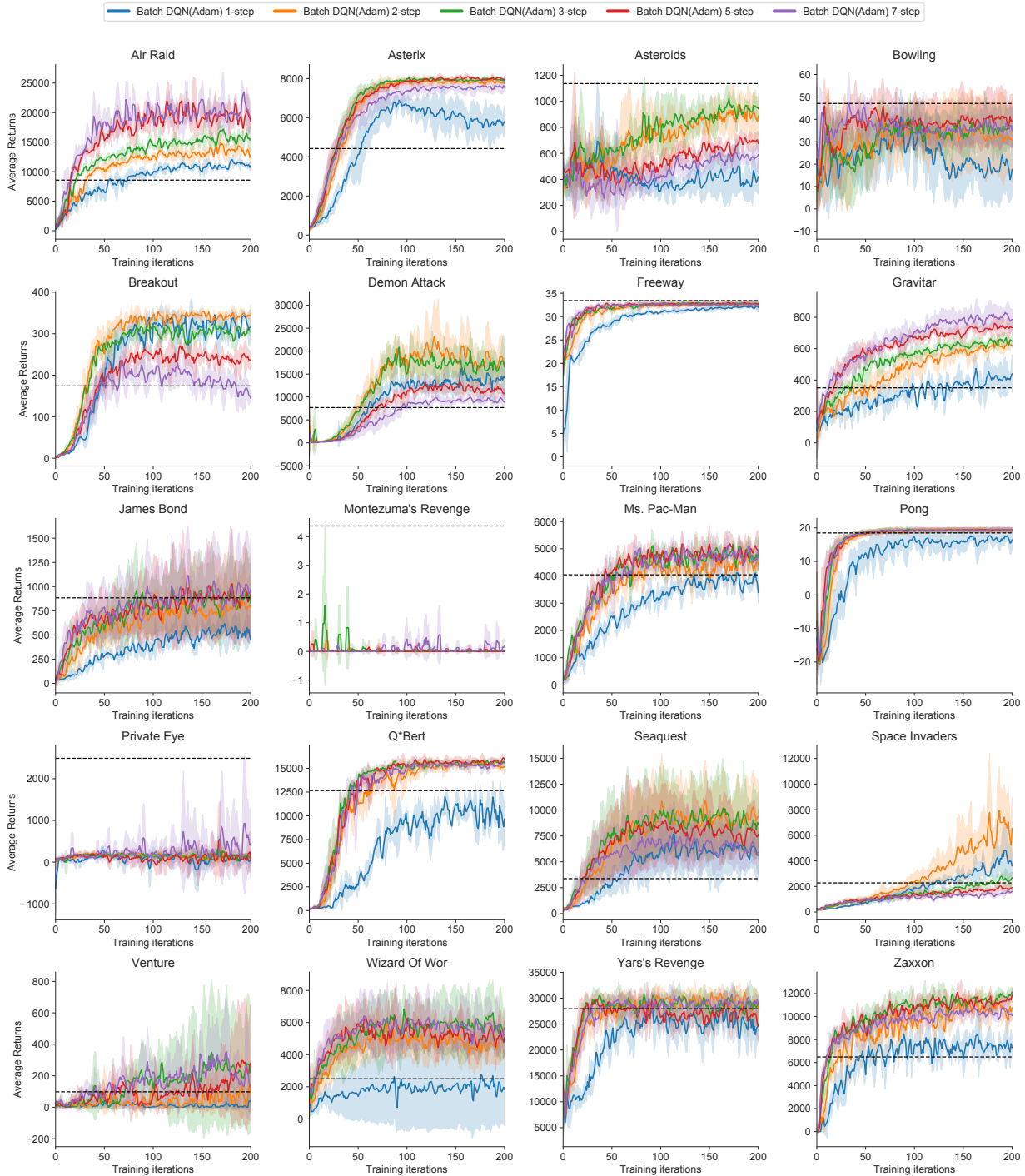


Figure 18. Average evaluation scores across 20 Atari 2600 games of batch DQN (Adam) agent with different n -step horizons trained offline using the DQN replay dataset (Agarwal et al., 2020). The horizontal line shows the evaluation performance of a fully-trained online DQN agent. The scores are averaged over 3 runs (shown as traces) and smoothed over a sliding window of 3 iterations and error bands show standard deviation.

Revisiting Fundamentals of Experience Replay

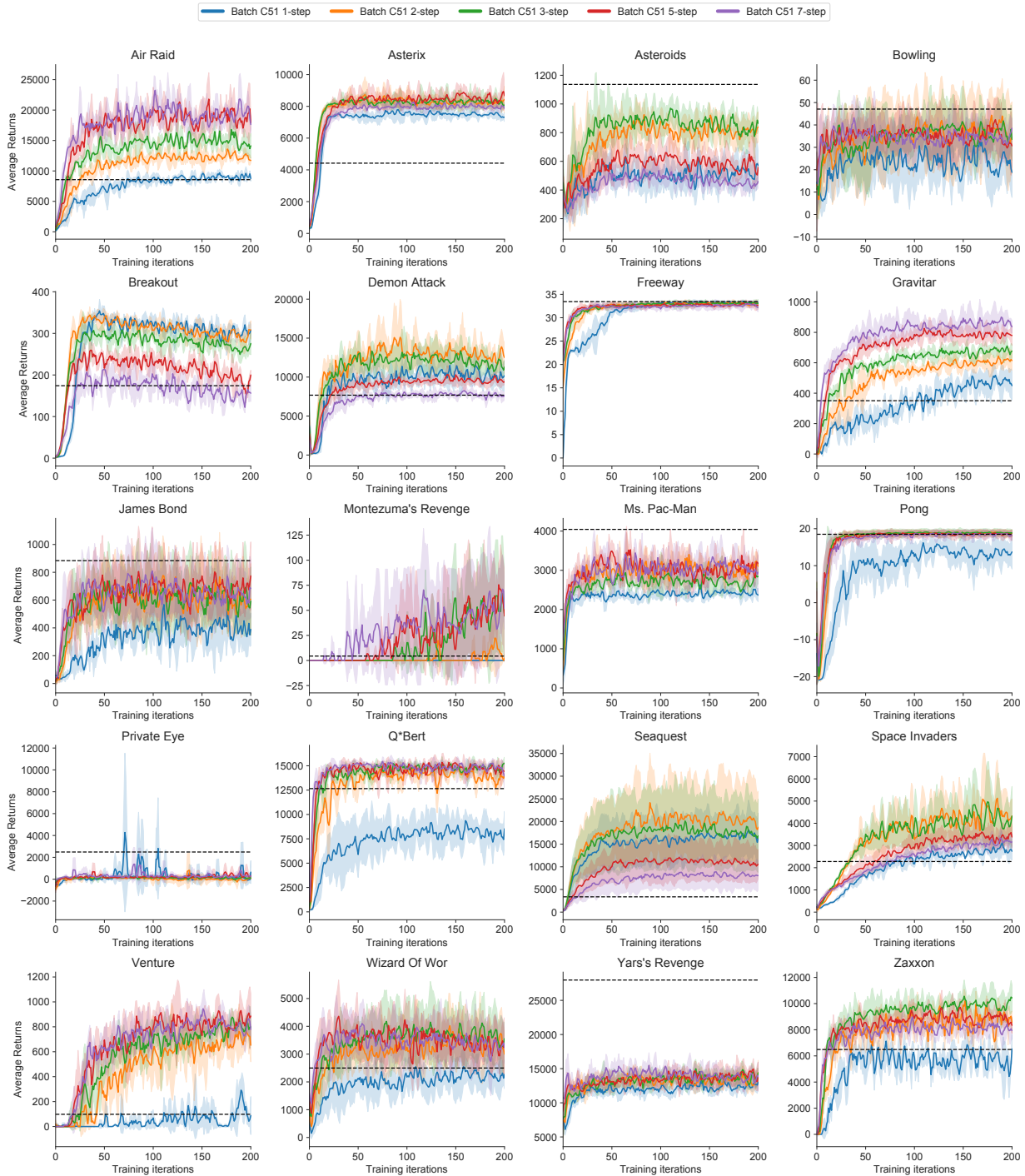


Figure 19. Average evaluation scores across 20 Atari 2600 games of a batch C51 agent with different n -step horizons trained offline using the DQN replay dataset (Agarwal et al., 2020). The horizontal line shows the evaluation performance of a fully-trained online DQN agent. The scores are averaged over 3 runs (shown as traces) and smoothed over a sliding window of 3 iterations and error bands show standard deviation.