We thank the reviewers for their time, comments and feedback. The reviewers unanimously appreciated the significance of our theoretical and empirical contributions. As a result, this rebuttal will mainly focus on minor comments and questions.

(R1, R4) – Analysis on LAP vs. PAL. A common question was with respect to the understanding the performance difference between LAP and PAL. Our theoretical analysis shows that prioritization reduces the variance of the gradient. Given LAP and PAL have the same expected gradient, our hypothesis is that LAP offers performance gains over PAL when the variance reduction matters. As suggested by R1, in an environment like MuJoCo with dense rewards, this may have limited impact. However, in Atari which has sparse rewards, and therefore, higher variance in the gradient, prioritization samples the sparse rewards more frequently and propagates the signal faster. In the camera-ready version we will attempt to validate this hypothesis in some simple toy domains, where the properties of the environment as well as the gradient of the loss function can be more readily analyzed.

(R2) – The authors also claim that the suggested methods are faster in execution, which is a useful side-effect if verified. Run time improvements appear in two ways. When using PAL (over LAP or PER), we save the cost of building and sampling from the priority structure. For LAP/PER, we don't find a meaningful run time difference, however we did find our implementation of PER outperformed existing, commonly used implementations. This result can be found in Appendix B. Notably the run time TD3 + LAP is less than using vanilla SAC.

(R2) – How was the number of time steps for the experiment decided upon? The choice of a horizon of 3 million is a common choice in the literature. SAC [1, 2] and ERE [3] use 3 million in most environments, and OAC (from NeurIPS 2019) [4] uses 2.5 million. We fixed 3 million across all environments as we felt this was the most transparent choice.

(R2) – Various terms are used without previous introduction. [...] Thank you for pointing these out. We will add definitions in the camera-ready version.

(R2) – Another ambiguous statement is the summation when computing p(i) just before Equation (6). The summation is indeed over the batch. We will clarify this.

(R2) – Another clarification is whether $\delta(j)$ in the denominator should also be an absolute value. The absolute value is correct and comes from the prioritization scheme being the absolute value of the TD error.

(R3) – Reproducibility. We include several pages of these details in Appendix D along with code submitted in the supplementary. Please check these out. We believe it will satisfy your concerns. We absolutely believe in being as transparent and reproducible as possible.

(R3) – The sentence leading with 'surprisingly' put me off a bit in the abstract without further information that comes later in the paper [...] We considered this result surprising as PER has some intuitive benefits regarding signal propagation, but we agree the language could be clarified here. Thank you.

(R4) – Broader Impact. We will expand our discussion in the broader impact. Thank you for your comments.

References

[1] Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In *International Conference on Machine Learning*, volume 80, pages 1861–1870. PMLR, 2018.

[2] Tuomas Haarnoja, Aurick Zhou, Kristian Hartikainen, George Tucker, Sehoon Ha, Jie Tan, Vikash Kumar, Henry Zhu, Abhishek Gupta, Pieter Abbeel, et al. Soft actor-critic algorithms and applications. *arXiv preprint arXiv:1812.05905*, 2018.

[3] Che Wang, Yanqiu Wu, Quan Vuong, and Keith Ross. Towards simplicity in deep reinforcement learning: Streamlined off-policy learning. *arXiv preprint arXiv:1910.02208*, 2019.

[4] Kamil Ciosek, Quan Vuong, Robert Loftin, and Katja Hofmann. Better exploration with optimistic actor critic. In *Advances in Neural Information Processing Systems*, pages 1787–1798, 2019.