We thank all reviewers for their valuable comments. Before we answer specific questions raised by the reviewers, we would like to first address the common concern on the broad appeal of our results outside the online learning community and its practical value. While our results are stated for two particular online learning problems (expert problem and multi-armed bandits), it is worth pointing out that 1) these are two fundamental learning problems and have numerous applications in both theory and practice; 2) our black-box approach provides a much more intuitive understanding of the problem and also gives an easy way to design algorithms with long-term memory, which we believe could be applied to other problems; 3) algorithms with long-term memory have been applied to practical applications such as TCP round-trip time estimation [4], intrusion detection system [3], and multi-agent systems [5], and we believe that our algorithms (especially the adaptive one) could potentially lead to better practical performance.

Reviewer 2:

— "the different regret bounds proved in this paper improve existing regret bounds in very specific setting":
This is admittedly true from a theoretical viewpoint. However, it is worth noting that algorithms with long-term memory indeed often exhibit superior empirical performance than those without, as shown in previous works (such as [1, 2]). Therefore, we believe that the significance of our results goes beyond the theoretical improvement of regret bounds.

— "Some discussion should be made somewhere about the optimality of the bounds."
We will add more discussion on this in the next version of our paper, as suggested by the reviewer. For the full information setting, as far as we know there is no existing lower bound. Note that, however, our upper bound (and that of [1]) essentially matches the bound of the computationally inefficient approach of running Hedge over all sequences with S switches among n experts, an approach that usually leads to the information-theoretically optimal regret bound. For the bandit setting, again there is no known lower bound. We do not believe that our bound is optimal and characterizing the optimal regret in this case is left as a future direction.

— "it would be nice to add figures that compare the rates of the existing bound and the one of Thm. 8"
We thank the reviewer for this suggestion. We will add this to the next version of our paper.

—"About the existing results for the stochastic setting (see line 47-52): ...":
The existing results refer to any results for switching regret in the stochastic environment. We are not aware of any existing results for tracking a small set of experts with stochastic losses. The problem is explicitly stated as an open problem in [6].

— Whether Corral algorithm can be used to improve the results for the sparse bandit setting:
We indeed have thought about this carefully, but in short we could not make it work. Note that there are two important differences here compared to the Corral setup: 1) we need to avoid polynomial dependence on K (except for lower-order terms) and 2) we need to have switching regret bound (instead of static regret, as in Corral) for the sub-routines.

Reviewer 3:

— "It is worth noting that none of the full-information results are *new* by themselves. ... AdaNormalHedge.TV gets similar guarantees in the stochastic setting although suboptimal in log factors"
We respectfully disagree with this comment. Our best-of-both-worlds result (or even just the result for the stochastic case) is new and resolves the open problem of Koolen and Warmuth [6]. What AdaNormalHedge.TV achieves is the typical switching regret bound, involving a term $S \ln T + S \ln K$ (for either adversarial or stochastic setting), while our results improve this term to $S \ln T + n \ln K$ (not just log factors), which is the typical and desirable improvement for this problem and is meaningful for large K (for example, the first paper on this topic by Bousquet and Warmuth [1] obtains the exact same improvement, but only in the adversarial case).

References

[1] O. Bousquet and M. K. Warmuth. Tracking a small set of experts by mixing past posteriors. *Journal of Machine Learning Research*, 3(Nov):363–396, 2002.

[2] R. B. Gramacy, M. K. Warmuth, S. A. Brandt, and I. Ari. Adaptive caching by refetching. In *Advances in Neural Information Processing Systems*, pages 1489–1496, 2003.

[3] H. T. Nguyen and K. Franke. Adaptive intrusion detection system via online machine learning. In *2012 12th International Conference on Hybrid Intelligent Systems (HIS)*, pages 271–277. IEEE, 2012.

[4] B. A. A. Nunes, K. Veenstra, W. Ballenthin, S. Lukin, and K. Obraczka. A machine learning framework for tcp round-trip time estimation. *EURASIP Journal on Wireless Communications and Networking*, 2014(1):47, 2014.

[5] T. Santarra. *Communicating Plans in Ad Hoc Multiagent Teams*. PhD thesis, UC Santa Cruz, 2019.

[6] M. K. Warmuth and W. M. Koolen. Open problem: Shifting experts on easy data. In *Conference on Learning Theory*, pages 1295–1298, 2014.