We thank the reviewers for their careful review and valuable suggestions. We will revise the paper accordingly. Below, we address some specific comments and questions.

1 Reviewer 1

Thank you for your helpful comments and suggestions, we will include them in the final version of the paper. We do have one main comment which is relevant to consider raising your score (point 1 below).

- "Significance: Medium in my opinion, perhaps not too high because of the relatively narrow focus and results...I don't think it could be improved much except by expanding the scope.":

  While this paper focuses on proving a single concrete statement, this statement bridges two well-researched areas within machine learning: *Online Learning* and *Private Learning*. Moreover, it is one of the only such papers which combines non-trivial techniques from both areas (e.g. the representation dimension and online boosting). As such, we believe it could be of interest to researchers from both areas and lead to more interaction between them. Consequently, the scope of this paper is actually wider than appears at first sight and spans two separate research communities.
- "Line 7-8: a bit confusing...": indeed the reduction is from online learning to private learning. We will fix it.
- "How Def 6 interacts with oblivious...": our proof needs weak online learners in both the oblivious and adaptive setting. We will clarify this in the definition.
- "...What is the expectation over in Def 6?..." the expectation in definition 6 is w.r.t. the randomness of both the learner and the adversary.

2 Reviewer 2

- "One comment I have would be...": We agree that Lemma 2 is counter-intuitive at a first sight. This is largely due to the fact that Pure Differential Privacy is a very strong requirement. More precisely, the notion of Differential Privacy requires the learner to be stable w.r.t. modification of one example in the sample. This enables replacing the entire sample and still getting a non-trivial correlation.

  This result is not new. As we mention in the proof overview (section 1.1), Lemma 9 is very close to a similar characterization shown in [4].
- "The most obvious improvement...to address either of the open problems...": see the last item in our reply to Reviewer 3.

3 Reviewer 3

- "Lemma 10: The big-O...": indeed, the $\log(T)$ term in lemma 10 is redundant.
- "Theorem 12: Could the $\ln(1/\epsilon)$ term in equation (1)...": indeed, the bound in Theorem 12 can be improved to $T\epsilon + \sqrt{\ln(1/\epsilon)}$. As a result, the dependence on T in Theorem 1 can be improved to $\sqrt{\ln T}$. Thanks!
- Discussing possible directions for extending our results: an earlier draft of this paper had a longer and more detailed discussion of such possible directions and difficulties, but we decided to omit it as we felt that it was a bit too long and technical. As two reviewers expressed interest, we will integrate (at least some of) it back. Essentially, the main challenges are:
  - Extension to pure Differential Privacy: we don't know of any useful extension of the structural characterization of pure Differential Privacy (Lemma 9) to the approximate case.
  - We do not know of any (online) boosting algorithm for the agnostic setting whose guarantees are useful for our setting.