Dear reviewers and area chairs,

Thank you for the careful reading of our manuscript and for the helpful comments and suggestions. We will address all of the smaller suggestions for improving the next revision of the manuscript. Below, we respond to some of the more significant points the reviewers raised.

Reviewer 1. *. . . some details of the proofs are deferred to a full version . . .*

We apologize for not correctly uploading the full version of the paper as supplementary material as we had intended. The full version will be included with the final submission. At the bottom of this response, please find a proof of the unsubstantiated Proposition 18 (excerpted from the public full version).

One question I had is whether the work of Feldman and Xiao 2015 isn't also relevant around lines 52-58 . . .

Feldman and Xiao's result is indeed highly relevant to our discussion about private sample complexity vs. mistake bound. We will explain this in the next revision.

Reviewer 2. *The result, which the author describes as "barrier to barrier" is very theoretical and I'm not sure if NeurIPS is the best venue?*

The "barrier to a barrier" interpretation of our result places it in the context of an explicit question raised by prior work. Beyond this context, however, our result addresses one of the most basic questions about learning in two fundamental and well-studied models. We believe that this puts it in scope for a broad and inclusive NeurIPS community.

There is an existing huge separation in terms of sample complexity, albeit for non-efficient algorithms. I'm wondering whether that example can be padded to make the algorithms "polynomial" time?

The challenge is that known sample complexity separations hold for classes that *do* have efficient algorithms. One-dimensional thresholds over a domain of size d can be efficiently, privately PAC learned using $\approx \log^* d$ samples and efficiently online learned using $\log d$ samples (via binary search). Blum's class does as the reviewer suggests, using cryptography to amplify the hardness of online learning thresholds while keeping PAC learning easy. Since private learning implies non-private learning, any way to obtain our result would imply Blum's result that efficient non-private learning $\not\Longrightarrow$ efficient online learning; to our knowledge, Blum's class is the simplest one that achieves this.

Reviewer 3. *There is a slight drawback, which is that the authors have omitted to submit the supplementary material.*

Please see our response to Reviewer 1 and the material at the bottom of this page.

Perhaps breaking it into three parts (Gonen et al and what is uniform pure private learning, impossibility of efficient uniform pure-private learning, and relaxations that allow the reduction to be useful) would help.

This is a great suggestion for improving the readability of this section and will be incorporated in the next revision.

Proof of Proposition 18. Let $t > 0$. We will show that $\mathbb{E}[|h|] \geq t$. Let n be the number of samples used by L. Let $\mathcal{H}_t$ be the set of all functions $h : \{0,1\}^* \to \{0,1\}$ with description length $|h| \leq 2e^n t$. Lemma 1 below shows that there exists a concept $c \in \mathcal{C}$ and a pair x, y such that $c(x) = 1$ and $c(y) = 0$ but $h(x) = 0$ or $h(y) = 1$ for every $h \in \mathcal{H}_t$.

Consider the distribution $\mathcal{D}$ that is uniform over $(x, 1)$ and $(y, 0)$. Accuracy of the learner requires that $\Pr_{S' \sim \mathcal{D}^n}[L(S') \notin \mathcal{H}_t] \geq 1/2$. Since any sample S' can be obtained from S by changing at most n elements of S, pure differential privacy implies that $\Pr[L(S) \notin \mathcal{H}_t] \geq e^{-n}/2$. Hence $\mathbb{E}_{h \leftarrow L(S)}[|h|] \geq 2e^n t \cdot e^{-n}/2 \geq t$ as we wanted to show. □

Let $\mathcal{S} = \{S_1, \ldots, S_n\}$ be a collection of subsets of $\{0,1\}^*$. We say that $\mathcal{S}$ *generates* another set $T \subseteq \{0,1\}^*$ if for every pair $x, y \in \{0,1\}^*$ with $x \in T$ and $y \notin T$, there exists $i \in [n]$ such that $x \in S_i$ and $y \notin S_i$.

Lemma 1. *A collection $\mathcal{S} = \{S_1, \ldots, S_n\}$ generates at most 2^{2^n} distinct sets $T \subseteq \{0,1\}^*$.*

Proof. By doubling the size of $\mathcal{S}$ we may assume it is closed under complement, i.e., $S \in \mathcal{S}$ iff $\overline{S} \in \mathcal{S}$. Let us say that a set $R \subseteq \{0,1\}^*$ is *pairwise separated* by $\mathcal{S}$ if for every pair $x, y \in R$, there exists $i \in [n]$ such that $x \in S_i$ and $y \notin S_i$. Let r denote the maximum size of a set that is pairwise separated by $\mathcal{S}$; by induction, $r \leq 2^{n-1}$. We will show that if T is generated by $\mathcal{S}$, then determining the membership of each element of R in T completely determines the set T. Therefore, there are at most $2^r \leq 2^{2^{n-1}}$ possible choices for T.

To see this, suppose for the sake of contradiction that there are two sets T_1, T_2 that are generated by $\mathcal{S}$ for which $T_1 \cap R = T_2 \cap R := I$. Let z be an element on which T_1, T_2 disagree; say $z \in T_1$ but $z \notin T_2$. We derive our contradiction by showing that $R \cup \{z\}$ is pairwise separated by $\mathcal{S}$, contradicting the maximality of R. To do so, all we need to show is that for every $y \in R$, there exists S_i such that $z \in S_i$ and $y \notin S_i$, and that there exists S_j such that $z \notin S_j$ and $y \in S_j$. If $y \in I$, we can take S_i to be the set such that $z \notin \overline{S_i}$ and $y \in \overline{S_i}$ as guaranteed by the fact that $\mathcal{S}$ generates T_2. If $y \notin I$, we can take S_i to be the set such that $z \in S_i$ and $y \notin S_i$ as guaranteed by the fact that $\mathcal{S}$ generates T_1. A similar argument can be used to construct S_j. □