We are grateful to the reviewers for their detailed comments, for their judgment of the problem setting as "interesting and well-motivated" (Reviewer 3), and for highlighting the novelty and intricacy of the fugal game (Reviewers 1-4). In this rebuttal, we address general concerns shared across reviewers rather than responding to individual comments.

Q1: The proof techniques for the lower bound and mini-batching upper bound in dimension $n > 1$.

A1: As Reviewer 3 noted "there are a variety of results over the years", so we naturally built on and clearly acknowledged prior work. Nonetheless, it was non-trivial to adapt prior techniques to the switching-constrained setting. First, concerning the adversary's strategy (the "orthogonal trick", introduced by Abernethy et al.) for the lower bound: to prove our result, we had to impose a certain switching pattern upon the adversary which was not obvious *a priori*, since the adversary is free to play any functions they wish from round to round. Without constraining the adversary to follow the player's switching pattern, the orthogonal trick cannot be applied. In addition, we found that this adversary's strategy can be adapted for $n = 2$ - thereby avoiding the need for special treatment as in $n = 1$ - via a geometric fact about the intersection of two closed half-spaces. Abernethy et al.'s original work only covered $n > 2$.

As the reviewers noted, the mini-batching algorithm of the upper bound was an existing innovation central to Arora et al. 2012, and we were careful to clearly cite this paper; we thank Reviewer 3 for bringing Dekel et al. 2011 to our attention, and will add this reference. However, part of our appreciation of the result is the counter-intuitiveness that switching at evenly sized intervals is optimal (up to a constant). This is surprising because the algorithm of Jaghargh et al. used a Poisson process to decide when to switch actions and thus had unevenly sized blocks. Note that we also included more technically involved results in the Appendix, including dimension-independent (Proposition 30) and more precise regret (Proposition 34) upper bounds, which went beyond the immediate scope of mini-batching to attain tighter bounds.

More broadly, it is not uncommon in online optimization for minimax bounds in one-dimension to be technically more demanding than for higher-dimensions, and our work is no different. However, we appreciate the elegance and coherence with past work of our bounds for $n > 1$ and do not find their value diminished as a result; quite the contrary.

Q2: One dimensional lower bound and the broader applicability of fugal game.

A2: We should highlight that for $n = 1$ we provide an alternative, shorter, and also completely new lower in Proposition 7. However, the full machinery of the fugal game was necessary to achieve a bound asymptotically tight in T (see Proposition 8). Beyond our setting, the fugal game could be a valuable tool in proving lower bounds for any of the many other natural switching-constrained continuous settings, including the convex bandit optimization and Gaussian process bandit optimization, online submodular, and submodular bandit settings.

Q3: The intuition behind the lack of phase transition in the continuous, as opposed to discrete, setting

A3: We particularly thank Reviewer 2 for their excellent observation, which noted that in the discrete but not the continuous setting, the player modifies their probability distribution each round. This observation is at the heart of the difference in phase transition between the discrete and continuous settings. In our setting, such a randomization is futile because *the adversary is stronger than in the discrete setting, and can choose the loss function f_t based even on the player's t^{th} action.* (Intuitively, randomizing over a discrete set is similar to picking deterministically from its convex hull.) By contrast, in prediction from experts, the adversary must be oblivious to avoid linear regret: they may choose f_t based only on $x_1, \ldots, x_{t-1}$ and knowledge of the player's randomized strategy. With a strong adversary, any number of extra allotted switches can help the player, whereas with the weaker, oblivious adversary of prediction from experts, an increased switching budget only aids the player up to a certain point. Since one usually assumes the strongest adversary that yields sublinear regret, it is the continuity of the action space that permits an adaptive adversary and thus the lack of a phase transition.

Q4: Cover's impossibility result.

A4: We thank Reviewer 1 for pointing out an error in our reference to Cover's impossibility result in Footnote 2. To clarify, Cover showed that in the ordinary setting, any adaptive adversary can force linear regret; we state this on page 2. The analogous impossibility result with a switching constraint was proven by Altschuler et al., which we cite in the next paragraph. The footnote should have referred to this analogous fact instead; we will correct this in the revision.

Q5: The difference between the switching-constrained and switching-cost settings.

A5: In the discrete setting, there is indeed a duality between the two: their minimax rates are within a polylogarithmic factor in T for certain regimes (see Altschuler et al.). We have not previously considered the switching-cost formulation, but of course a K-budgeted mini-batching algorithm achieves switching-cost penalty $O(T/\sqrt{K}) + cK$; optimally setting $K = \left(T/2c\right)^{2/3}$ yields penalty $O(T^{2/3}c^{1/3})$. It is not immediately clear whether this is optimal for certain c or if there is a reduction in the other direction, so we defer this interesting topic to future work. We also thank Reviewer 2 for bringing the Awerbuch et al. paper to our attention, and we will add it to the Related Work in the revision.