We thank the reviewers for their careful reviews! Below we restate the questions in blue.

Review 1. Do the results continue to hold for any arbitrary tie-breaking rules? For example, ties are broken lexicographically. Our results hold for all tie-breaking rules by which any bidder i's allocation is deterministically determined by the comparisons of her bid b_i with the other bidders' bids, and not affected by any other information (such as the specific value of b_i). More formally, the allocation for bidder i should be determined by a vector $\boldsymbol{y}_i \in \{<, =, >\}^{n-1}$, which records the comparison between b_i and every other bidder's bid. Breaking ties lexicographically satisfies these conditions. We remark that ties occur with probability 0 in equilibrium; we discuss tie-breaking only for the sake of utility learning.

Review 2. Clarify how the proofs rely on / differ from those in previous work. Morgenstern and Roughgarden (2016) and Balcan et al. (2018) proposed the following framework for bounding the psuedodimension of a class $\mathcal{F}$ of functions: given samples that are to be shattered and for any (fixed) witnesses, one classifies the functions in $\mathcal{F}$ into categories, so that functions in the same category must output the same label on all the samples; by counting and bounding the number of such categories, one can bound the number of shattered samples. Our proof follows this strategy. To bound the number of categories, the argument makes use of monotonicity of bidding functions, which is specific to our problem.

Relationship between Lemma 3.10 and existing work. Devanur et al. (2016) show the concentration of the expectation of a single function on the empirical product distribution. Lemma 3.10 of our paper generalizes this so that the concentration applies simultaneously to a family of functions. This is more convenient for proving uniform convergence. Our proof follows arguments similar to those of Devanur et al.

Does Theorem 3.3 hold even when the valuations are not independent? The answer is no. When values are correlated, a bidder i's value v_i gives information on the other bidders' values, and so interim utilities must be defined with respect to *conditional* distributions (over $\boldsymbol{v}_{-i}$) that vary for each v_i; this changes the problem substantially. In general, utilities over each conditional distribution require $\Omega(n/\epsilon^2)$ samples to learn; in total, $\Omega(|T_i|n/\epsilon^2)$ samples are needed.

Review 3. (1) How would an agent form a hypothesis about what strategies others will follow? In modern ad auctions (such as oCPA/oCPA ads), bidding decisions are often handed over to the platform, which also estimates the utilities for the bidders. The platform is the entity that is most likely to have access to both value samples and bidding strategies for such estimates.

(2) How to find a candidate BNE to test? One of the consequences of this work is that, with $\tilde{O}(n/\epsilon^2)$ samples, one can not only test whether a strategy profile is at approximate equilibrium, but also compute approximate equilibria (see Corollary 3.14). In particular, one may simply compute an approximate equilibrium on the empirical product distribution on the samples, using an existing algorithm (e.g., Shen et al. 2020). The strategies form an approximate equilibrium on the original distribution as well (see Corollary 3.13).

(3) How [does one estimate the interim regret to a bidder]? As a consequence of our results, it suffices to compute a bidder i's interim regret assuming others' values are drawn from the empirical distribution. Balcan et al. (EC' 19) compute the regret by discretizing bidder i's type space $[0, H]$ into $1 + H/\epsilon$ points, $P = \{0, \epsilon, 2\epsilon, \ldots, H\}$, and computing $\max_{v_i, b_i \in P}\{\mathbf{E}_{\boldsymbol{v}_{-i} \sim \text{emp}}[U_i(v_i, b_i, \boldsymbol{v}_{-i}) - U_i(v_i, v_i, \boldsymbol{v}_{-i})]\}$. This estimate is within $O(\epsilon)$ the actual interim regret ($O(\epsilon)$ discretization error + $O(\epsilon)$ sampling error). In a first price auction, the best response on the empirical distribution can be computed even without discretization. A key observation is that a best-responding bid must equal a value that appears in a sample of an opponent's value — any other bid can be lowered to decrease payment without affecting the allocation. One may therefore enumerate the $m(n-1)$ values to search for a best-responding bid.

Review 4. How to connect the sample complexity bounds derived back to showing how they can be used to inform practical decisions or recommendations in auction systems. Please see our responses to Reviewer 3's questions.

10 - "Connection between this setting and the first price auction." "This setting" refers to auctions with search costs.

55 - ...the line of prior work focused on estimation for revenue maximization while this work focuses on estimating interim utilities, right? Yes, we will clarify this.

249 - Correlation of the empirical distribution. Even though $\boldsymbol{F}$ is a product distribution, the empirical distribution, which is the uniform distribution over samples $\{\boldsymbol{s}^1, \ldots, \boldsymbol{s}^m\}$, is correlated. For example, given two samples $\boldsymbol{s}^1 = (1, 3)$ and $\boldsymbol{s}^2 = (2, 4)$, the empirical distribution is the uniform distribution over $\{(1, 3), (2, 4)\}$, which is correlated. Empp in this example estimates utility on the uniform distribution over $\{(1, 3), (1, 4), (2, 3), (2, 4)\}$, which is a product distribution.

142, 171, 175 - We agree with your comments and will make the suggested changes. Thanks!