Selling to A Sophisticated No-Regret Buyer

Henrique Schechter Vera
Advised by Matt Weinberg

Submitted in partial fulfillment
of the requirements for the degree of
Bachelor of Science in Engineering

Department of Computer Science
Princeton University

April 2023

Abstract

Consider a repeated single item auction with a single buyer who has a value for the item randomly drawn from known distribution $\mathcal{D}$ each round and bids according to an online learning algorithm. "Selling to a No-Regret Buyer" by Braverman et al. presents a strategy for the seller which, whenever the buyer bids according to a mean-based learning algorithm, extracts revenue that is arbitrarily close to the expected welfare. We extend these results to two settings where the bidder does not use a simple mean-based learning algorithm. First, we consider a bidder using a mean-based learning algorithm with recency bias, where the results of recent rounds are weighed more strongly. We show how much revenue the strategy yields as a function of the recency bias factor β. Next, we consider a bidder using a k-switching learning algorithm, where what we define as a *g-mean-based* learning algorithm is given as options all "meta-strategies" which switch bids at most k times. We present a new strategy and show how much revenue it yields as a function of the g for which the learning algorithm is g-mean-based. In both settings, we also determine which parameter values allow the algorithm to be no-regret, and which yield revenue that is arbitrarily close to the welfare.

Acknowledgements

I would like to profoundly thank all who have contributed to my work, directly or indirectly.

Thank you to Matt, for being the best advisor–academic and otherwise–I could have asked for. I am forever grateful for your kindness, patience, enthusiasm, and dedication, all of which have been critical to this thesis.

Thank you to my parents, for their unwavering love and support, for encouraging me throughout my entire academic journey, and for always advocating for my happiness.

Thank you to my sister Lara, for being the incredible role model I look up to and a lifelong best friend.

Thank you to my friends and family in Paraguay, for the warm memories and for helping me get to where I am today.

Thank you to Cindy, for the abundance of love and joy these past two years, and for being my pillar throughout this thesis.

Thank you to the friends who have been with me since day one at Princeton–Creston, Alexis, Daniel, Srija, Samantha, Lauren–for sticking with me until the end.

Thank you to all my roommates, past and present–Jay, Jack, Keso, Caleb, Brandon, Darin–for the daily antics and good-hearted fun.

Thank you to all the friends I've met along the way, in countless other ways–Garcia, Shruti, Cecilia, Joseph, Shivani, Preeti, Zach, Andy, Jojo, Shantanu, Beck, Neo–for making my time at Princeton the richest it could have been, and for all the games, parties, heart-to-hearts, and banter.

Contents

1 Introduction

Consider a bidder deciding how much to bid in an auction. For example, an advertiser must bid in an auction through Google Ads or Microsoft Advertising in order to acquire ad spots in Google Search, Bing, Yahoo!, and DuckDuckGo search engine results pages, as well as in the countless non-search websites, mobile apps, and videos affiliated to these two ads platforms. If the auction follows the format of the truthful Vickrey-Clarke-Groves auction [Vic61; Cla71; Gro73], then the bidder should clearly bid their value. If the auction follows the format of the Generalized First-Price (GFP) or Generalized Second-Price (GSP) auction, as is the case in Google Ads auctions [OS22], then the optimal decision is less clear. Bidders could attempt to compute a Bayes-Nash equilibrium of the associated game and play accordingly, but this is unrealistic because it is computationally expensive and requires accurate priors. Moreover, bidders might not even know the underlying mechanisms of the auction they are participating in.

Consequently, bidders might use online learning algorithms to learn how to bid optimally, for instance by using commercial bid optimizers. We give special attention to bidders who *no-regret learn*, because bidder behavior on Microsoft Advertising is empirically consistent with no-regret learning [NST15], where here no-regret simply means the bidder is at least as happy with his own strategy as with any strategy which constantly bids the same amount. This motivates the question: If a seller knows that buyers are no-regret learning over time, how should they maximize revenue? Even in the case where there is just a single item for sale to a single buyer, this problem is quite interesting—and relevant, as lone bidders in sponsored search auctions are pitted against a seller that adaptively sets reserve prices based on past bids. Braverman et al. [Bra+17] presented a strategy with which the seller can extract (expected) revenue arbitrarily close to the expected welfare if the buyer bids according to any *mean-based* learning algorithm. A mean-based algorithm, intuitively, is one in which the bidder chooses to bid an amount A with low probability if there is any amount B that has historically done better than A, by some threshold.

However, this strategy seems vulnerable. Intuitively, it works by "luring" the bidder into submitting high bids early on by giving away the item for free, and then charging very high prices towards the end. This works well against mean-based algorithms, because if there are bids that have historically performed very well, mean-based algorithms are likely to select them. However, algorithms with *recency bias* counteract this, weighing the results of recent rounds more strongly, and thus more quickly switching out of historically good bids that have begun overcharging. Similarly, no-regret k-switching algorithms, with which the buyer is at least as happy as with any strategy that switches bids at most k times, also pose a problem for this strategy, since strategies which switch bids often are able to switch out of a bid whenever it begins overcharging. It is thus of interest to examine the performance of this strategy in these two settings, to better understand the strength of the strategy.

In this paper, we extend the results from Braverman et al. to these two more complicated types of learning algorithms, building off of prior work Schiffer and Zhang [SZ19]. We first consider a bidder employing a mean-based learning algorithm with recency bias. We show how the regret of the algorithm increases and the revenue yielded under the strategy from Braverman et al. decreases as a function of β. We also provide a range of β which allows the algorithm to remain no-regret and the strategy from Braverman et al. to yield revenue that is arbitrarily close to the welfare. We then consider a bidder using what we call a *g-mean-based* learning algorithm which is given as possible options all "meta-strategies" which switch bids at most k times. Intuitively, a g-mean-based algorithm is one in which the bidder chooses to bid an amount A with probability at most $g(D)$ if there is any amount B that has historically done better than A by a value of D. We determine which values of k allow the algorithm to remain no-regret, when the learning algorithm being used is either the Multiplicative Weights Update (MWU) [AHK12] or Follow the Perturbed Leader (FTPL) [Han58; KV05]. We then present a new strategy similar to the one in Braverman et al. We analyze it for general learning algorithms, showing both the revenue it yields as a function of g and a set of g for which it yields revenue arbitrarily close to the welfare. Finally, we show that strategy yields revenue arbitrarily close to the welfare when the learning algorithm is the Multiplicative Weights Update or Follow the Perturbed Leader.

2 Background

2.1 Related Work

There are several problems that are strongly related, yet different, from ours. One of these is dynamic auctions [Pap+22; ADH16; Mir+16a; Mir+16b; LP17]. Here, as in our setting, a single buyer repeatedly bids with a value drawn from a known distribution over a sequence of rounds, but unlike our setting, the buyer is fully strategic and reasons about how their choices might affect the seller's behavior. Another is the fishmonger problem [DPS14; Imm+17], where a fishmonger tries to sell an identical fresh fish every day to the same buyer via a posted price. One key difference is that here the buyer draws a value once and fixes it, so the seller might try to learn the buyer's value, and the buyer might try to hide it. Also, the literature on this problem looks at perfect Bayesian equilibria, where again the buyer is fully strategic and reasons about how their behavior will affect the seller's.

In our model, although buyers care about the future, they don't reason about how their decisions might affect the seller's decisions in the future. This is a more fitting model for sponsored search auctions, since search engines usually keep their proprietary algorithms for setting reserves based on past data confidential, making it impossible to be fully strategic.

Another line of related work considers the Price of Anarchy of simple combinatorial auctions when bidders no-regret learn [Rou15; ST13; NST15; DS15]. There are three crucial differences between our setting and theirs: they study welfare maximization while we study revenue maximization; they study combinatorial auctions while we study single-item auctions; and they study buyers who no-regret learn because of the strategic behavior of other buyers, with a publicly known auction format, while we study buyers who have to no-regret learn even when they are the sole buyer because the auction format is hidden.

Other work has looked at learning from the point of view of the seller [CR15; DHP15; MR15; MR16; GN16; CD17; Dud+16]. Here, buyers' values are drawn from an unknown distribution, and the seller attempts to learn an approximately optimal auction with as few samples as possible. The buyer participates in only one round. In our setting, the buyer has no information to learn, as the buyer's value distribution is known, and the buyer plays multiple rounds.

Though there is clearly a vast amount of work examining repeated sales in auctions and no-regret learning for buyers and sellers, our line of work is unique in studying how a seller might adapt their strategy when faced with a no-regret buyer. Note this includes not just this paper and Braverman et al.'s "Selling to a No-Regret Buyer", but also Zhang's extension to multiple bidders, "Selling to No-Regret Buyers" [Bra+17; Zha20].

Finally, it is worth mentioning that no-regret learning in online decision problems is a rich, well-studied subfield of algorithmic game theory which helps inform our work. This is of course the case since in our setting, the buyer is no-regret learning in face of an online decision problem. In this paper, we focus on two popular solutions to this problem, the Multiplicative Weights Update method [AHK12] and the Follow the Perturbed Leader algorithm [Han58; KV05]. See survey [BC12] and paper [LZ07] for more details about the multi-armed bandit problem and the contextual bandits problem, respectively, the latter of which we base our model off of.

2.2 Model

We generally follow the model presented in [Bra+17]. We consider a setting with 1 buyer and 1 seller. There are T rounds, and in each round the seller has one item for sale. At the start of each round t, the buyer's value v_t (known only to the buyer) for the item is drawn independently from some distribution $\mathcal{D}$ (known to both the seller and the buyer). For simplicity, we assume $\mathcal{D}$ has a finite support of size m, supported on a finite set $C = \{v_1, ..., v_m\}$ with $0 \leq v_1 < v_2 < ... < v_m \leq 1$. For each $i \in [m]$, v_i has probability q_i of being drawn under $\mathcal{D}$. Note that since the buyer has the additional information of their current value for the item, our setting is *contextual*, where at each round t the buyer learns context v_t.

The seller then presents n options for the buyer, which can be thought of as "possible bids" (we will interchangeably refer to these as options, bids, or arms). Each arm i is labelled with a bid value $b_i \in [0, 1]$, with $b_1 < ... < b_n$. Upon pulling arm i at round t, the buyer receives the item with some allocation probability $a_{i,t}$, and must pay a price $p_{i,t} \in [0, a_{i,t} \cdot b_i]$. These values $a_{i,t}$ and $p_{i,t}$ are chosen by the seller during time t, but remain unknown to the buyer until he plays an arm. Note that all of our strategies for the seller will be *non-adaptive* in the sense that that $a_{i,t}, p_{i,t}$ are set before the first round starts. Now, at each round t, the buyer learns either the values $a_{i,t}, p_{i,t}$ for the arm i he played (in the *bandits* setting), or the values for *all* arms (in the *experts* setting). All our results hold for both settings, except for those pertaining to Multiplicative Weights Update and Follow the Perturbed Leader, which require the experts setting.

In order for the selling strategies to represent sponsored search auctions, we require the allocation/price rules to be monotone. That is, if $i > j$, then for all t, $a_{i,t} \geq a_{j,t}$ and $p_{i,t} \geq p_{j,t}$. In other words, bidding higher should result in a (weakly) higher probability of receiving the item and (weakly) higher expected payment. We'll also require the existence of an arm 0 with bid $b_0 = 0$ and $a_{0,t} = 0$ for all t; i.e., an arm which charges nothing but does not give the item. Playing this arm represents the choice of not participating in the auction. Finally, we will assume our buyer is non-conservative, meaning they are *not* constrained to only submit bids less than their current value for the item. This is crucial to our strategies. We now formalize some concepts:

Definition 1 (Reward)**.** *Let $r_{i,t}(v) = a_{i,t} \cdot v - p_{i,t} \in [-1, 1]$ be the reward from arm i in round t when the buyer has value v for the item. A learning algorithm in the contextual bandits model typically takes as training input each round the tuple $(r_{i,t}(v), i, v)$; for ease of notation, we will assume $r_{i,t}(v)$ contains all this information and conveys it to learning algorithms.*

Definition 2 (No-regret)**.** *An algorithm $\mathcal{A}$ that pulls arm I_t in round t is δ-no-regret if $\mathbb{E}[Reg(\mathcal{A})] \leq \delta$, and no-regret if $\delta = o(T)$, where $Reg(\mathcal{A})$ is defined as $\max_{\pi \in \Pi : C \to [n]} \sum_{t=1}^{T} r_{\pi(v_t),t}(v_t) - \sum_{t=1}^{T} r_{I_t,t}(v_t)$; that is, how much more total reward the buyer could have attained if they had used the optimal policy π mapping values $v_t \in C$ to arms in $[m]$. Note that in the recency bias setting, Π will represent all policies that simply choose a fixed arm, and in k-switching, Π will represent all policies which switch arms at most k times.*

Definition 3 (Welfare)**.** *The welfare, $Val(\mathcal{D})$, is equal to $\mathbb{E}_{v \sim \mathcal{D}}[v]$*

Definition 4 (Mean-Based Learning Algorithm)**.** *Let $\sigma_{i,t}(v) = \sum_{s=1}^{t} r_{i,s}(v)$. An algorithm is γ-mean-based if it is the case that whenever $\sigma_{i,t}(v) < \sigma_{j,t}(v) - \gamma T$, then the probability $p_{i,t}(v)$ that the algorithm pulls arm i on round t if it has context v is at most γ. We say an algorithm is mean-based if it is γ-mean-based for some $\gamma = o(1)$*

As a final note about the model, observe that we could phrase it differently. We frame our problem as a single buyer who repeatedly draws a value independently from $\mathcal{D}$ and online learns with their value as context. However, we could alternatively imagine a population of m different buyers who each have a fixed value v_i, and thus online learn with no context. Each round, exactly one buyer arrives at the auction, each with probability q_i. This is a mathematically equivalent model, and so all of our results hold in this model as well.

2.3 *Selling to a No-Regret Buyer*

Braverman et al. [Bra+17] provided an example of a strategy for the seller such that when the buyer bids using a *mean-based learning algorithm*, the seller extracts revenue that is arbitrarily close to the expected welfare. Specifically, they showed that in the setting we described, the following holds:

Theorem 1. *If the buyer is running a mean-based algorithm, for any constant* $\varepsilon > 0$*, there exists a strategy for the seller which obtains revenue at least* $(1-\varepsilon)Val(\mathcal{D})T - o(T)$*.*

The strategy they design which extracts revenue arbitrarily close to the expected welfare is as follows:

Strategy 1. *If every element in the support of* $\mathcal{D}$ *is at least* $1-\varepsilon$*, then sell the item at price* $1-\varepsilon$*. Otherwise, use the following strategy:*

Define $\rho = min(v_m, 1-\varepsilon/2)$*, and* $\delta = \frac{(1-\rho)}{(1-v_1)}$*. In addition to the zero arm, offer* $n = \frac{log(\varepsilon/2)}{log(1-\delta)}$ *possible options, each with maximum bid value* $b_i = 1$*. Divide the timeline of each arm into three sessions:*

1. $\emptyset$ ***session:*** *For the first* $(1-(1-\delta)^{i-1})T$ *rounds, the seller charges 0 and does not give the item to the buyer, i.e.* $(p_{i,t}, a_{i,t}) = (0, 0)$*.*
2. ***0 session:*** *For the next* $(1-\delta)^{i-1}(1-\rho)T$ *rounds, the seller charges 0 and gives the item to the buyer, i.e.* $(p_{i,t}, a_{i,t}) = (0, 1)$*.*
3. ***1 session:*** *For the final* $(1-\delta)^{i-1}\rho T$ *rounds, the seller charges 1 and gives the item to the buyer, i.e.* $(p_{i,t}, a_{i,t}) = (1, 1)$*.*

Note that this strategy is monotone; if $i > j$*, then* $p_{i,t} \geq p_{j,t}$ *and* $a_{i,t} \geq a_{j,t}$*.*

We also get the following as a product of the proof of Theorem 1 in [Bra+17]:

Corollary 2. *If the buyer is running a* γ*-mean-based algorithm, for any constant* $\varepsilon > 0$*, Strategy 1 obtains revenue that is at least the "Mean-Based Minimum Revenue",* $MBMR(\varepsilon, T, \mathcal{D}) := (1-\varepsilon)Val(\mathcal{D})T - n\gamma T(1 - n\gamma + Val(\mathcal{D})(1-\frac{\varepsilon}{2}))$*.*

3 Recency Bias

We consider a recency-biased bidding algorithm; that is, a bidding algorithm $\mathcal{A}^*$, such that the action on round $t+1$ given past rewards of $\{r_{i,s}(v_s)\}_{s\leq t}$ will be the output of $\mathcal{A}(\{\beta^s r_{i,s}(v_s)\}_{s\leq t})$, where $\mathcal{A}$ is another bidding algorithm and I_s is the arm $\mathcal{A}^*$ pulls at round s.

3.1 Regret Bounds

We begin by finding a relationship between the recency bias factor β and the regret of $\mathcal{A}^*$. We then determine a constraint on β that guarantees $\mathcal{A}^*$ remains no-regret if $\mathcal{A}$ is no-regret. Here we consider regret with respect to the optimal arm, where Π is restricted to policies that map to a single arm.

Theorem 3. *If $\mathcal{A}^*$ is an algorithm using a δ-no-regret bidding algorithm $\mathcal{A}$ with recency bias factor $\beta > 1$, then $\mathcal{A}^*$ is $(\delta + \Delta)$-no-regret, where $\Delta = 2 \cdot (\frac{\beta(\beta^T-1)}{\beta-1} - T)$*

Proof.

$$\begin{aligned}
\mathbb{E}[Reg(\mathcal{A}^*(\{r_{i,s}(v_s)\}_{s\leq t}))] &= \mathbb{E}[\max_{i\in[n]} \sum_{t=1}^{T} r_{i,t}(v_t) - \sum_{t=1}^{T} r_{I_t,t}(v_t)] \\
&\leq \mathbb{E}[\max_{i\in[n]} \sum_{t=1}^{T} \beta^t r_{i,t}(v_t) - \sum_{t=1}^{T} \beta^t r_{I_t,t}(v_t)] + 2\sum_{t=1}^{T}(\beta^t - 1) \\
&\leq \mathbb{E}[Reg(\mathcal{A}(\{\beta^s r_{i,s}(v_s)\}_{s\leq t}))] + 2\sum_{t=1}^{T}(\beta^t - 1) \\
&\leq \delta + 2\sum_{t=1}^{T} \beta^t - 2T \\
&\leq \delta + 2 \cdot (\frac{\beta(\beta^T - 1)}{\beta - 1} - T)
\end{aligned}$$

The first line is the definition of regret. The first inequality comes from the fact that $r_{i,t}(v_t) \leq \beta^t - 1 + \beta^t r_{i,t}(v_t)$ for all $r_{i,t}(v_t) \in [-1, 1]$ (see Lemma 1), which is necessary to account for both positive and negative values of $r_{i,t}(v_t)$. Applying this inequality to each $r_{i,t}(v_t)$ term in the expression gives the resulting inequality by linearity of expectation. ∎

We now prove the lemma used above:

Lemma 1. *$r_{i,t}(v) \leq \beta^t - 1 + \beta^t r_{i,t}(v)$ for all $r_{i,t}(v) \in [-1, 1], \beta > 1$*

Proof. Since $r_{i,t}(v) \in [-1, 1]$, then $r_{i,t}(v) + 1 \geq 0$. Since $\beta > 1$, then $\beta^t > 1$. Thus, we have it that

$$\begin{aligned} r_{i,t}(v) + 1 &\leq \beta^t (r_{i,t}(v) + 1) \\ r_{i,t}(v) + 1 &\leq \beta^t r_{i,t}(v) + \beta^t \\ r_{i,t}(v) &\leq \beta^t - 1 + \beta^t r_{i,t}(v) \end{aligned}$$

■

From the theorem above, we also get the following:

Corollary 4. *If $\mathcal{A}^*$ is an algorithm using a no-regret bidding algorithm $\mathcal{A}$ with recency bias factor $\beta > 1$ and β satisfies $(\frac{\beta(\beta^T - 1)}{\beta - 1} - T) \in o(T)$, then $\mathcal{A}^*$ remains no-regret*

Proof. Since $\mathcal{A}$ is no-regret, then it is δ-no-regret for some $\delta \in o(T)$. Thus, using Theorem 3, we have it that

$$\begin{aligned} \mathbb{E}[Reg(\mathcal{A}^*(\{r_{i,s}(v_s)\}_{s \leq t}))] &\leq \delta + 2 \cdot (\frac{\beta(\beta^T - 1)}{\beta - 1} - T) \\ &:= \delta^* \in o(T) \end{aligned}$$

and so $\mathcal{A}^*$ is no-regret. ■

Theorem 5. *If $\mathcal{A}^*$ is an algorithm using a no-regret bidding algorithm $\mathcal{A}$ with recency bias factor $\beta > 1$ and $\beta \leq (1 + \sigma)^{1/T}$ for some $\sigma \in o(1)$, then $\mathcal{A}^*$ remains no-regret.*

Proof.

$$\begin{aligned} \frac{\beta(\beta^T - 1)}{\beta - 1} - T &= \sum_{t=1}^{T} (\beta^t - 1) \\ &\leq \sum_{t=1}^{T} (\beta^T - 1) \\ &= T(\beta^T - 1) \\ &\leq T\sigma \in o(T) \end{aligned}$$

By Corollary 4, the proof is complete. ■

3.2 Revenue Bounds

As before, we consider a bidding algorithm $\mathcal{A}^*$, such that the action on round $t+1$ given past rewards of $\{r_{i,s}(v_s)\}_{s\leq t}$ will be the output of $\mathcal{A}(\{\beta^s r_{i,s}(v_s)\}_{s\leq t})$. Now, we additionally assume $\mathcal{A}$ is a γ-*mean-based* bidding algorithm.

In this section, we provide lower bounds on the revenue the seller extracts by using Strategy 1. First note that if every element in the support of $\mathcal{D}$ is at least $1-\varepsilon$, then Strategy 1 sells the item at price $1-\varepsilon$. Since $\mathcal{D}$ is supported on $[0,1]$, this ensures a $(1-\varepsilon)$ approximation to the welfare. The rest of the section is dedicated to find lower bounds for the revenue of Strategy 1 when $\mathcal{D}$ is not entirely supported on $[1-\varepsilon, 1]$, and thus assumes this condition. Observe that in this setting $v_1 < 1-\varepsilon/2$, and so because we also know $v_1 < v_m$, we have it that $v_1 < \rho$ and thus $\delta < 1$.

We begin by introducing the following concept to greatly simplify notation:

Definition 5 (β-Weighed Regret Between Arms)**.** *Let* $\sigma^{\beta}_{i,t}(v) = \sum_{s=1}^{t} \beta^s r_{i,s}(v)$. *Then the β-weighed regret between arms i and j at time t is* $Reg^{\beta}_{t,i,j}(v) = \sigma^{\beta}_{i,t}(v) - \sigma^{\beta}_{j,t}(v)$. *This is the difference in cumulative rewards over the first t rounds, weighed by recency bias factor β, between arm i and arm j, given value v. For readability, let* $Reg_{t,i,j}(v) = Reg^{1}_{t,i,j}(v) = \sigma_{i,t}(v) - \sigma_{j,t}(v)$

Lemma 2. *For all* $i,j \in [n]$, $t \in [T]$, $\beta > 1$, *we have it that* $Reg^{\beta}_{t,i,j}(v) \geq Reg_{t,i,j}(v) - 2(\frac{\beta(\beta^T-1)}{\beta-1} - T)$

Proof.

$$\begin{aligned}
Reg^{\beta}_{t,i,j}(v) &= \sum_{s=1}^{t} \beta^s r_{i,s}(v) - \sum_{s=1}^{t} \beta^s r_{j,s}(v) \\
&\geq \sum_{s=1}^{t} r_{i,s}(v) - \sum_{s=1}^{t} r_{j,s}(v) - 2\sum_{s=1}^{t}(\beta^s - 1) \\
&\geq \sum_{s=1}^{t} r_{i,s}(v) - \sum_{s=1}^{t} r_{j,s}(v) - 2\sum_{s=1}^{T}(\beta^s - 1) \\
&= \sum_{s=1}^{t} r_{i,s}(v) - \sum_{s=1}^{t} r_{j,s}(v) - 2(\frac{\beta(\beta^T - 1)}{\beta - 1} - T) \\
&= Reg_{t,i,j}(v) - 2(\frac{\beta(\beta^T - 1)}{\beta - 1} - T)
\end{aligned}$$

The first inequality comes from rewriting Lemma 1 as $\beta^s r_{j,s}(v_i) \geq r_{j,s}(v_i) - (\beta^s - 1)$ for all $r_{i,s}(v_i) \in [-1, 1]$ and applying this inequality to each term in the summation. The second inequality comes from the fact that all the terms in the summation are positive, so t can be increased to T. ■

Lemma 3. *If the seller uses Strategy 1, then for each $j \in \{1, ..., n-1\}$, $j' > j$, $v_i \in \mathcal{D}$, and $\tau \in [A_j, B_j(v_i)]$, we have it that if $Reg^{\beta}_{\tau,j,j'}(v_i) \geq Reg_{\tau,j,j'}(v_i) - \alpha$, then $\sigma^{\beta}_{j,\tau}(v_i) > \sigma^{\beta}_{j',\tau}(v_i) + \gamma T$, where $A_j := (1 - \rho(1 - \delta)^{j-1})T$, $B_j(v_i) := A_j + \frac{min(v_i,\rho)}{1-v_1}(1 - \rho)(1 - \delta)^{j-1}T - \gamma T - \alpha$, and $\alpha > 0$ is any constant*

Proof.

$$\begin{aligned} Reg^{\beta}_{\tau,j,j'}(v_i) &\geq Reg_{\tau,j,j'}(v_i) - \alpha \\ &= \sum_{s=1}^{\tau} r_{j,s}(v_i) - \sum_{s=1}^{\tau} r_{j',s}(v_i) - \alpha \\ &> \gamma T + \alpha - \alpha \\ &= \gamma T \end{aligned}$$

The first line is assumed. The second line is the definition of regret between arms. The third line uses Lemma B.2 in [Bra+17], except we add the term $-\alpha$ to $B_j(v_i)$. Crucially, Lemma B.2 holds after this change because it is still the case that $B_j(v_i) < A_{j+1}$, since $\alpha > 0$, and so we are only making $B_j(v_i)$ smaller. ■

Lemma 4. *If for each $v_i \in \mathcal{D}$, $j \in \{1, ..., n-1\}$, and round $\tau \in [A_j, B_j(v_i)]$, for all $j' > j$, we have it that $\sigma^{\beta}_{j,\tau}(v_i) > \sigma^{\beta}_{j',\tau}(v_i) + \gamma T$, then the expected revenue of the seller is at least $MBMR(\varepsilon, T, \mathcal{D}) - (1 - n\gamma)n\alpha$*

Proof. It follows from the mean-based condition that for all $j : A_j < B_j(v_i)$, in the interval $[A_j, B_j(v_i)]$ the buyer with value v_i will, with probability at least $(1 - n\gamma)$, choose an arm currently in its 1-session (i.e. an arm with label at most j) and hence pay 1 each round. Recall the buyer has value v_i for the item with probability q_i.

Then, the total contribution of the buyer with value v_i to the expected revenue of the seller must be at least

$$\begin{aligned}
& q_i \sum_{j\,:\,A_j<B_j(v_i)} (1-n\gamma)(B_j(v_i)-A_j) \\
\geq\ & q_i \sum_{j=1}^{n} (1-n\gamma)(B_j(v_i)-A_j) \\
=\ & q_i \sum_{j=1}^{n} (1-n\gamma)(\frac{min(v_i,\rho)}{1-v_1}(1-\rho)(1-\delta)^{j-1}T-\gamma T-\alpha) \\
=\ & (1-n\gamma)q_iT(-n\gamma-\frac{n\alpha}{T}+\frac{(1-\rho)min(v_i,\rho)}{1-v_1}\sum_{j=1}^{n}(1-\delta)^{j-1}) \\
=\ & (1-n\gamma)q_iT(-n\gamma-\frac{n\alpha}{T}+\frac{(1-\rho)min(v_i,\rho)(1-(1-\delta)^n)}{\delta(1-v_1)}) \\
=\ & (1-n\gamma)q_iT(-n\gamma-\frac{n\alpha}{T}+min(v_i,\rho)(1-(1-\delta)^n)) \\
=\ & q_iTmin(v_i,\rho)(1-(1-\delta)^n)-q_iT((1-n\gamma)(n\gamma+\frac{n\alpha}{T})+n\gamma\, min(v_i,\rho)(1-(1-\delta)^n)) \\
\geq\ & q_iT(1-\varepsilon/2)^2v_i-q_iT((1-n\gamma)(n\gamma+\frac{n\alpha}{T})+n\gamma\, min(v_i,\rho)(1-(1-\delta)^n)) \\
\geq\ & (1-\varepsilon)q_iv_iT-q_iT((1-n\gamma)(n\gamma+\frac{n\alpha}{T})+n\gamma\, min(v_i,\rho)(1-(1-\delta)^n)) \\
\geq\ & (1-\varepsilon)q_iv_iT-q_iT(n\gamma(1-n\gamma+min(v_i,\rho)(1-(1-\delta)^n)))-(1-n\gamma)q_in\alpha
\end{aligned}$$

The first inequality comes from the fact that the terms where $A_j \geq B_j(v_i)$ have a non-positive contribution to the sum. The second line is the definition of $B_j(v_i)$. The fifth line uses the definition of δ. The third to last line uses the facts that $(1-(1-\delta)^n) = 1-\varepsilon/2$ (since $n = log(\varepsilon/2)/log(1-\delta)$) and $min(v_i,\rho) \geq (1-\varepsilon/2)v_i$ (since if $min(v_i,\rho) \neq v_i$, then $\rho = min(v_m, 1-\varepsilon/2) \geq (1-\varepsilon/2)v_i$, because $v_i \leq 1$ and $v_i \leq v_m$). Finally, the second to last line uses the fact that $(1-\varepsilon/2)^2-(1-\varepsilon) = \frac{\varepsilon^2}{4} \geq 0$ for all ε.

Summing this contribution over all $v_i \in \mathcal{D}$, we have that the expected revenue of the seller is at least

$$\begin{aligned}
& \sum_{i\in[m]} ((1-\varepsilon)q_i v_i T - q_i T(n\gamma(1-n\gamma+min(v_i,\rho)(1-(1-\delta)^n))) - (1-n\gamma)q_i n\alpha) \\
= \;& (1-\varepsilon)T(\sum_{i\in[m]} q_i v_i) - Tn\gamma(1-n\gamma)(\sum_{i\in[m]} q_i) - Tn\gamma(1-(1-\delta)^n)(\sum_{i\in[m]} q_i\, min(v_i,\rho)) \\
& -(1-n\gamma)n\alpha(\sum_{i\in[m]} q_i) \\
\geq \;& (1-\varepsilon)T(\sum_{i\in[m]} q_i v_i) - Tn\gamma(1-n\gamma)(\sum_{i\in[m]} q_i) - Tn\gamma(1-(1-\delta)^n)(\sum_{i\in[m]} q_i\, v_i) \\
& -(1-n\gamma)n\alpha(\sum_{i\in[m]} q_i) \\
= \;& (1-\varepsilon)T(\mathbb{E}_{v\sim\mathcal{D}}[v]) - Tn\gamma(1-n\gamma) - Tn\gamma(1-(1-\delta)^n)(\mathbb{E}_{v\sim\mathcal{D}}[v]) - (1-n\gamma)n\alpha \\
= \;& (1-\varepsilon)\,Val(\mathcal{D})T - n\gamma T(1-n\gamma+Val(\mathcal{D})(1-\frac{\varepsilon}{2})) - (1-n\gamma)n\alpha \\
= \;& MBMR(\varepsilon,T,\mathcal{D}) - (1-n\gamma)n\alpha
\end{aligned}$$

■

Combining Lemmas 3 and 4, we have proven the following:

Theorem 6. *Say $\mathcal{A}$ is a γ-mean-based bidding algorithm. For a bidder using any bidding algorithm $\mathcal{A}^*$ defined as $\mathcal{A}^*(\{r_{i,s}(v_s)\}_{s\le t}) = \mathcal{A}(\{\beta^s r_{i,s}(v_s)\}_{s\le t})$, if $Reg_{t,i,j}(v) \le Reg^{\beta}_{t,i,j}(v) + \alpha$ for all $i,j \in [n], t \in [T]$, and some $\alpha > 0$, then for any constant $\varepsilon > 0$ there exists a strategy for the seller that yields expected revenue that is at least $MBMR(\varepsilon,T,\mathcal{D}) - (1-n\gamma)n\alpha$.*

Combining this with Lemma 2, we also have the following:

Theorem 7. *Say $\mathcal{A}$ is a γ-mean-based bidding algorithm. If the bidder is using any bidding algorithm $\mathcal{A}^*$ defined as $\mathcal{A}^*(\{r_{i,s}(v_s)\}_{s\le t}) = \mathcal{A}(\{\beta^s r_{i,s}(v_s)\}_{s\le t})$, then for any constant $\varepsilon > 0$ there exists a strategy for the seller that yields expected revenue that is at least $MBMR(\varepsilon,T,\mathcal{D}) - 2(1-n\gamma)n(\frac{\beta(\beta^T-1)}{(\beta-1)} - T)$.*

Now, we know that if $\mathcal{A}$ is mean-based, then $\gamma \in o(1)$ and, by Corollary 2, $MBMR(\varepsilon, T, \mathcal{D}) = (1-\varepsilon)Val(\mathcal{D})T - o(T)$. We also showed in Theorem 5 that if $\beta \leq (1+\sigma)^{1/T}$ for some $\sigma \in o(1)$, then $(\frac{\beta(\beta^T-1)}{(\beta-1)} - T) \in o(T)$. Thus, we also have the following:

Theorem 8. *Say $\mathcal{A}$ is a mean-based bidding algorithm. If the bidder is using any bidding algorithm $\mathcal{A}^*$ defined as $\mathcal{A}^*(\{r_{i,s}(v_s)\}_{s\leq t}) = \mathcal{A}(\{\beta^s r_{i,s}(v_s)\}_{s\leq t})$ such that $\beta \leq (1+\sigma)^{1/T}$ for some $\sigma \in o(1)$, then for any constant $\varepsilon > 0$ there exists a strategy for the seller that yields expected revenue that is at least $(1-\varepsilon)Val(\mathcal{D})T - o(T)$.*

4 k-Switching

We consider a bidding algorithm $\mathcal{A}^*$ with k-switching; that is, a bidding algorithm $\mathcal{A}^*$ which takes another bidding algorithm $\mathcal{A}$ and as input gives it the set of meta-arms that consists of all strategies that switch arms at most k times over the T rounds. Though we will also provide revenue bounds for general $\mathcal{A}$, we will focus on the case where $\mathcal{A}$ is either the Multiplicative Weights Update algorithm (MWU) or the Follow the Perturbed Leader algorithm (FTPL). This gives us concrete algorithms to analyze, so we can show that there are commonly used algorithms which can be no-regret under reasonable parameters and which yield maximal revenue in the strategy we present in section 4.3.

We first define some common notation. Let $\mathbb{M}_{k,T}$ be the set of meta-arms that switch arms at most k times over T rounds. Let I_t be the arm the algorithm $\mathcal{A}^*$ chooses in round t, let M_t be the arm the meta-arm $M \in \mathbb{M}_{k,T}$ chooses at time t, and let $r_{M,t}(v_t) = r_{M_t,t}(v_t)$ be the reward arm meta-arm M receives in round t. We can now define $\mathcal{A}^*(\{r_{i,s}(v_s)\}_{s\leq t}, [m]) = \mathcal{A}(\{r_{M,s}(v_s)\}_{s\leq t}, \mathbb{M}_{k,T})$.

We now bound the size of the set of meta-arms, which will be useful for both the regret and revenue bounds:

Lemma 5. *The k-switching input to bidding algorithm $\mathcal{A}$ has $|\mathbb{M}_{k,T}| =: n^{(k)} < (T-1)^{k+1}(n-1)^{k+1}$ arms*

Proof. For each $i \in [0, k]$, we must first choose an arm out of n, then choose an arm out of $n-1$ to switch into (i times), and finally choose during which rounds to execute the switches. Thus, the amount of strategies that switch arms at most k times over T rounds is

$$\begin{aligned}
n^{(k)} &= \sum_{i=0}^{k} n(n-1)^i \cdot \binom{T-1}{i} \\
&< \sum_{i=0}^{k} n(n-1)^i (T-1)^i \\
&= n \cdot \frac{(T-1)^{k+1}(n-1)^{k+1} - 1}{(T-1)(n-1) - 1} \\
&= \frac{(T-1)^{k+1}(n-1)^{k+1} - 1}{(T-1) - \frac{T}{n}} \\
&< (T-1)^{k+1}(n-1)^{k+1}
\end{aligned}$$

∎

4.1 Regret Bounds

The maximum possible number of switches is $T-1$. In the two following theorems, we show that $k \in o(\frac{T}{\ln(T)})$ will keep the k-switching algorithm no-regret, for both Follow The Perturbed Leader and Multiplicative Weights Update. This means we can allow close to the maximal number of switches, despite the number of meta-arms growing exponentially with k, and still remain no-regret. Note that in this setting we look at regret with respect to the optimal policy that switches arms at most k times; in other words, we restrict the set of policies Π to the set of strategies which switch arms at most k times.

Theorem 9. *Say $\mathcal{A}$ is the Multiplicative Weights Update algorithm with regret-minimizing multiplicative update factor η. Any k-switching algorithm $\mathcal{A}^*$ defined as $\mathcal{A}^*(\{r_{i,s}(v_s)\}_{s\leq t}, [m]) = \mathcal{A}(\{r_{M,s}(v_s)\}_{s\leq t}, \mathbb{M}_{k,T})$ is no-regret if $k \in o(\frac{T}{\ln(T)})$*

Proof. The Multiplicative Weights Update algorithm gives the following bound for any $M \in \mathbb{M}_{k,T}$, where $\eta \leq \frac{1}{2}$ is the multiplicative update factor, $m_{i,t}(v_t) = -r_{i,t}(v_t)$ is the cost of arm i at round t, and $m_{M,t}(v_t) = -r_{M,t}(v_t)$ is the cost of meta-arm M at round t:

$$\begin{aligned}
\mathbb{E}[\sum_{t=1}^{T} m_{I_t,t}(v_t)] &\leq \sum_{t=1}^{T} m_{M,t}(v_t) + \eta\sum_{t=1}^{T}|m_{M,t}(v_t)| + \frac{\ln(n^{(k)})}{\eta} \quad \text{[AHK12]} \\
\mathbb{E}[\sum_{t=1}^{T} m_{I_t,t}(v_t)] - \sum_{t=1}^{T} m_{M,t}(v_t) &\leq \eta\sum_{t=1}^{T}|m_{M,t}(v_t)| + \frac{\ln(n^{(k)})}{\eta} \\
\sum_{t=1}^{T} r_{M,t}(v_t) - \mathbb{E}[\sum_{t=1}^{T} r_{I_t,t}(v_t)] &\leq \eta\sum_{t=1}^{T}|m_{M,t}(v_t)| + \frac{\ln(n^{(k)})}{\eta} \\
\mathbb{E}[\sum_{t=1}^{T} r_{M,t}(v_t) - \sum_{t=1}^{T} r_{I_t,t}(v_t)] &\leq \eta\sum_{t=1}^{T}|m_{M,t}(v_t)| + \frac{\ln(n^{(k)})}{\eta} \\
\mathbb{E}[\max_{M\in\mathbb{M}_{k,T}} \sum_{t=1}^{T} r_{M,t}(v_t) - \sum_{t=1}^{T} r_{I_t,t}(v_t)] &\leq \eta\sum_{t=1}^{T}|m_{M,t}(v_t)| + \frac{\ln(n^{(k)})}{\eta} \\
\mathbb{E}[Reg(\mathcal{A}^*(\{r_{i,s}(v_s)\}_{s\leq t}))] &\leq \eta\sum_{t=1}^{T}|m_{M,t}(v_t)| + \frac{\ln(n^{(k)})}{\eta}
\end{aligned}$$

Substituting the upper bound for $n^{(k)}$ from Lemma 5 and the upper bound $|m_{i,t}| = |r_{i,t}| \leq 1$, we get:

$$\mathbb{E}[Reg(\mathcal{A}^*(\{r_{i,s}(v_s)\}_{s\leq t}))] < \eta T + \frac{(k+1)\ln((T-1)(n-1))}{\eta}$$

This is minimized for $\eta = \sqrt{\frac{(k+1)\ln((T-1)(n-1))}{T}}$ with value $2\sqrt{T(k+1)\ln((T-1)(n-1))}$. This is clearly $o(T)$ when $k \in o(\frac{T}{\ln(T)})$. ■

Theorem 10. *Say $\mathcal{A}$ is the Follow the Perturbed Leader algorithm with regret-minimizing decay rate parameter λ. Any k-switching algorithm $\mathcal{A}^*$ defined as $\mathcal{A}^*(\{r_{i,s}(v_s)\}_{s\le t}, [m]) = \mathcal{A}(\{r_{M,s}(v_s)\}_{s\le t}, \mathbb{M}_{k,T})$ is no-regret if $k \in o(\frac{T}{\ln(T)})$*

Proof. The Follow the Perturbed Leader algorithm gives the following bound for any $M \in \mathbb{M}_{k,T}$, where η is the decay rate parameter η of the exponential distribution used to draw perturbances, $m_{i,t} = -r_{i,t}$ is the cost of arm i at round t, and $m_{M,t}(v_t) = -r_{M,t}(v_t)$ is the cost of meta-arm M at round t:

$$\begin{aligned}
\mathbb{E}[\sum_{t=1}^{T} m_{I_t,t}(v_t)] &\le (1+\lambda) \min_{M\in\mathbb{M}_{k,T}} \sum_{t=1}^{T} m_{M,t}(v_t) + \frac{O(\ln(n^{(k)}))}{\lambda} \quad \text{[KV05]} \\
\mathbb{E}[\sum_{t=1}^{T} m_{I_t,t}(v_t)] - \min_{M\in\mathbb{M}_{k,T}} \sum_{t=1}^{T} m_{M,t}(v_t) &\le \lambda \min_{M\in\mathbb{M}_{k,T}} \sum_{t=1}^{T} m_{M,t}(v_t) + \frac{O(\ln(n^{(k)}))}{\lambda} \\
\max_{M\in\mathbb{M}_{k,T}} \sum_{t=1}^{T} r_{M,t}(v_t) - \mathbb{E}[\sum_{t=1}^{T} r_{I_t,t}(v_t)] &\le \lambda \min_{M\in\mathbb{M}_{k,T}} \sum_{t=1}^{T} m_{M,t}(v_t) + \frac{O(\ln(n^{(k)}))}{\lambda} \\
\mathbb{E}[\max_{M\in\mathbb{M}_{k,T}} \sum_{t=1}^{T} r_{M,t}(v_t) - \sum_{t=1}^{T} r_{I_t,t}(v_t)] &\le \lambda \min_{M\in\mathbb{M}_{k,T}} \sum_{t=1}^{T} m_{M,t}(v_t) + \frac{O(\ln(n^{(k)}))}{\lambda} \\
\mathbb{E}[Reg(\mathcal{A}^*(\{r_{i,s}(v_s)\}_{s\le t}))] &\le \lambda \min_{M\in\mathbb{M}_{k,T}} \sum_{t=1}^{T} m_{M,t}(v_t) + \frac{O(\ln(n^{(k)}))}{\lambda} \\
\mathbb{E}[Reg(\mathcal{A}^*(\{r_{i,s}(v_s)\}_{s\le t}))] &\le \lambda \min_{M\in\mathbb{M}_{k,T}} \sum_{t=1}^{T} m_{M,t}(v_t) + \frac{C\cdot\ln(n^{(k)})}{\lambda}, \forall T > T_0
\end{aligned}$$

Here $C > 0$ and $T_0 \ge 0$ are some constants. Substituting the upper bound for $n^{(k)}$ from Lemma 5 and the upper bound $m_{i,t} = -r_{i,t} \le 1$, we get:

$$\mathbb{E}[Reg(\mathcal{A}^*(\{r_{i,s}(v_s)\}_{s\le t}))] < \lambda T + \frac{C(k+1)\ln((T-1)(n-1))}{\lambda}, \forall T > T_0$$

This is minimized for $\lambda = \sqrt{\frac{C(k+1)\ln((T-1)(n-1))}{T}}$ with value $2\sqrt{C(k+1)\ln((T-1)(n-1))T}$. This is clearly $o(T)$ when $k \in o(\frac{T}{\ln(T)})$. ■

4.2 Strategy

Here we introduce the new strategy and provide some important bounds on its parameters:

Strategy 2. *If every element in the support of $\mathcal{D}$ is at least $1-\varepsilon$, then sell the item at price $1-\varepsilon$. Otherwise, use the following strategy:*

Define $\rho = max(1-\frac{\varepsilon}{2}, 1-(1-v_1)(1-(1-\frac{\varepsilon}{4})^{1/k}))$, and $\delta = \frac{(1-\rho)}{(1-v_1)}$. In addition to the zero arm, offer $n = max(\frac{\ln(\varepsilon/4)}{\ln(1-\varepsilon/2)}, 4k\frac{\ln(4/\varepsilon)}{\varepsilon})$ possible options, each with maximum bid value $b_i = 1$. Divide the timeline of each arm into three sessions:

1. *$\emptyset$ **session:** For the first $(1-(1-\delta)^{i-1})T$ rounds, the seller charges 0 and does not give the item to the buyer, i.e. $(p_{i,t}, a_{i,t}) = (0,0)$.*
2. ***0 session:** For the next $(1-\delta)^{i-1}(1-\rho)T$ rounds, the seller charges 0 and gives the item to the buyer, i.e. $(p_{i,t}, a_{i,t}) = (0,1)$.*
3. ***1 session:** For the final $(1-\delta)^{i-1}\rho T$ rounds, the seller charges 1 and gives the item to the buyer, i.e. $(p_{i,t}, a_{i,t}) = (1,1)$.*

Note that this strategy is monotone; if $i > j$, then $p_{i,t} \geq p_{j,t}$ and $a_{i,t} \geq a_{j,t}$.

As with Strategy 1, if every element in the support of $\mathcal{D}$ is at least $1-\varepsilon$, Strategy 2 sells the item at price $1-\varepsilon$, ensuring a $(1-\varepsilon)$ approximation to the welfare (since $\mathcal{D}$ is supported on $[0,1]$). The rest of chapter 4 is dedicated to find lower bounds for the revenue of Strategy 2 when $\mathcal{D}$ is not entirely supported on $[1-\varepsilon, 1]$, and thus assumes this is the case. We begin by showing our parameter values are valid.

Lemma 6. *For $k \geq 0, \varepsilon \in (0,1)$, Strategy 2 has $\delta \in (0,1), \rho \in (0,1], n \geq 2$*

Proof. We first show that $\delta \in (0,1)$ and $\rho \in (0,1]$.
Consider first the case when $\rho = 1-\varepsilon/2$. Note immediately that $\rho \in (1/2,1) \subset (0,1]$. Observe that since $\mathcal{D}$ is not entirely supported on $[1-\varepsilon,1]$, we have it that $v_1 < 1-\varepsilon/2$. Thus, when $\rho = 1-\varepsilon/2$, we have

$$\begin{aligned}\delta &= (1-\rho)/(1-v_1) \\ &= (\varepsilon/2)/(1-v_1) \\ &< 1\end{aligned}$$

Moreover, since $v_1 \geq 0$, then in this case

$$\begin{aligned}\delta &= (\varepsilon/2)/(1-v_1) \\ &\geq \varepsilon/2 \\ &> 0\end{aligned}$$

Now, consider the case where $\rho = 1-(1-v_1)(1-(1-\varepsilon/4)^{1/k})$. We know $(1-\varepsilon/4)^{1/k} \in ((3/4)^{1/k},1)$ and $v_1 \in [0,1]$, so $\rho \leq 1-(1-1)(1-(1-\varepsilon/4)^{1/k}) = 1$ and $\rho > 1-(1-0)(1-(3/4)^{1/k}) = (3/4)^{1/k} \geq 0$. Moreover,

$$\begin{aligned}\delta &= (1-v_1)(1-(1-\varepsilon/4)^{1/k})/(1-v_1) \\ &= 1-(1-\varepsilon/4)^{1/k} \\ &< 1-(3/4)^{1/k} \\ &\leq 1\end{aligned}$$

and

$$\begin{aligned}\delta &= 1-(1-\varepsilon/4)^{1/k} \\ &> 1-(1)^{1/k} \\ &= 0\end{aligned}$$

Thus, in all cases we have $\delta \in (0,1)$ and $\rho \in (0,1]$.
Finally, we show $n \geq 2$. We find a lower bound for n over $\varepsilon \in [0,1]$, which provides a lower bound on n over $\varepsilon \in (0,1)$. We first find the minima of both possible values of n over ε. Both $\frac{\ln(\varepsilon/4)}{\ln(1-\varepsilon/2)}$ and $4k\ \ln(4/\varepsilon)/\varepsilon$ are minimized at $\varepsilon = 1$, with values 2 and $4k\ln(4)$, respectively. The maximum of these two values, which is the resulting value of n, is smallest when $k = 0$, with value 2. Thus, $n \geq 2$. ■

4.3 Revenue Bounds

We now show a lower bound on the revenue this strategy attains for any k-switching algorithm, before then showing a tighter bound for MWU and FTPL. We begin with some helpful lemmas.

Lemma 7. *Under Strategy 2,* $(1-\delta)^k - (1-\delta)^n \leq 1 - \varepsilon/2$

Proof. First consider the case where $1-(1-v_1)(1-(1-\varepsilon/4)^{1/k}) \geq 1-\varepsilon/2$, and so $\rho = max(1-\varepsilon/2, 1-(1-v_1)(1-(1-\varepsilon/4)^{1/k})) = 1-(1-v_1)(1-(1-\varepsilon/4)^{1/k})$ and $\delta = (1-\rho)/(1-v_1) = 1-(1-\varepsilon/4)^{1/k}$. We have it that

$$\begin{aligned}
n &= max(\frac{\ln(\varepsilon/4)}{\ln(1-\varepsilon/2)}, 4k\frac{\ln(4/\varepsilon)}{\varepsilon}) \\
n &\geq 4k\frac{\ln(4/\varepsilon)}{\varepsilon} \\
(\varepsilon n)/(4k) &\geq \ln(4/\varepsilon) \\
\ln(\varepsilon/4) &\geq -(\varepsilon n)/(4k) \\
\varepsilon/4 &\geq e^{-(\varepsilon/4)(n/k)} \\
\varepsilon/4 &\geq (1-\varepsilon/4)^{n/k} \\
(\varepsilon/4)^{1/n} &\geq (1-\varepsilon/4)^{1/k} \\
1-(\varepsilon/4)^{1/n} &\leq 1-(1-\varepsilon/4)^{1/k} \\
1-(\varepsilon/4)^{1/n} &\leq \delta \\
1-\delta &\leq (\varepsilon/4)^{1/n} \\
(1-\delta)^n &\leq \varepsilon/4
\end{aligned}$$

and

$$\begin{aligned}
1-(1-\varepsilon/4)^{1/k} &= \delta \\
1-\delta &= (1-\varepsilon/4)^{1/k} \\
(1-\delta)^k &= 1-\varepsilon/4
\end{aligned}$$

Thus, we have it that

$$\begin{aligned}
(1-\delta)^k-(1-\delta)^n &\geq (1-\varepsilon/4)-(\varepsilon/4) \\
&\geq 1-\varepsilon/2
\end{aligned}$$

Now, consider the case where $1-(1-v_1)(1-(1-\varepsilon/4)^{1/k}) < 1-\varepsilon/2$. In this case, $\rho = max(1-\varepsilon/2, 1-(1-v_1)(1-(1-\varepsilon/4)^{1/k})) = 1-\varepsilon/2$ and

$\delta = (1-\rho)/(1-v_1) = \varepsilon/(2(1-v_1))$. We then have it that

$$\begin{array}{rcl} n &=& max(\frac{\ln(\varepsilon/4)}{\ln(1-\varepsilon/2)}, 4k\,\frac{\ln(4/\varepsilon)}{\varepsilon}) \\ n &\geq& \frac{\ln(\varepsilon/4)}{\ln(1-\varepsilon/2)} \\ n\,\ln(1-\varepsilon/2) &\leq& \ln(\varepsilon/4) \\ (1-\varepsilon/2)^n &\leq& \varepsilon/4 \\ 1-\varepsilon/2 &\leq& (\varepsilon/4)^{1/n} \\ 1-(\varepsilon/4)^{1/n} &\leq& \varepsilon/2 \\ 1-(\varepsilon/4)^{1/n} &\leq& \varepsilon/(2(1-v_1)) \\ 1-(\varepsilon/4)^{1/n} &\leq& \delta \\ 1-\delta &\leq& (\varepsilon/4)^{1/n} \\ (1-\delta)^n &\leq& \varepsilon/4 \end{array}$$

and

$$\begin{array}{rcl} 1-(1-v_1)(1-(1-\varepsilon/4)^{1/k}) &<& 1-\varepsilon/2 \\ (1-v_1)(1-(1-\varepsilon/4)^{1/k}) &>& \varepsilon/2 \\ 1-(1-\varepsilon/4)^{1/k} &>& \varepsilon/(2(1-v_1)) \\ 1-(1-\varepsilon/4)^{1/k} &>& \delta \\ 1-\delta &>& (1-\varepsilon/4)^{1/k} \\ (1-\delta)^k &>& 1-\varepsilon/4 \end{array}$$

Thus, we have it that

$$\begin{array}{rcl} (1-\delta)^k-(1-\delta)^n &>& (1-\varepsilon/4)-(\varepsilon/4) \\ &>& 1-\varepsilon/2 \end{array}$$

In both cases, we have it that $(1-\delta)^k-(1-\delta)^n \geq 1-\varepsilon/2$. ■

Lemma 8. *If the seller uses Strategy 2 (or Strategy 1), then for any* $j \in \{1, ..., n-1\}$*, we have it that any arm* $j' > j$ *cannot begin its* 0*-session before arm* j *begins its* 1*-session.*

Proof. Arm j starts its 1-session at round $A_j := (1 - \rho(1-\delta)^{j-1})T$. Arm $j+1$ begins its 0-session at round $Z_{j+1} := (1-(1-\delta)^j)T$. Recall that we defined $\delta = (1-\rho)/(1-v_1)$, meaning $\rho = 1 - \delta(1-v_1)$. Since $0 \leq v_1 < v_2 < ... < v_m \leq 1$, we have it that $\rho \geq 1 - \delta$. Thus, $\rho(1-\delta)^{j-1} \geq (1-\delta)^j$. Thus, $(1-\rho(1-\delta)^{j-1})T \leq (1-(1-\delta)^j)T$, meaning $A_j \leq Z_{j+1}$ Trivially, $Z_{j+1} \leq Z_{j'}, j' > j$, since the smaller arms start their 0-session the earliest. Thus, $A_j \leq Z_{j'}, j' > j$. ■

We now introduce some notation. Let $M(i)$ represent the meta-arm that solely selects arm i. We represent any meta-arm as $M([S_{t_1}, T_{t_1}, t_1], ..., [S_{t_f}, T_{t_f}, t_f])$, where $[S_t, T_t, t]$ represents a switch from source arm S_t to target arm T_t at time t, and $0 < t_1 < ... < t_f \leq T$. Finally, recall that M_t represents the arm that meta-arm M chooses at round t.

Lemma 9. *If the seller uses Strategy 2, then for each* $j \in \{k+1, ..., n-1\}$*,* $j' > j$*,* $v_i \in \mathcal{D}$*, and* $\tau \in [A_j, B_j(v_i)]$*, we have it that* $\sigma_{M(j),\tau}(v_i) \geq \sigma_{M',\tau}(v_i) + D$ *for any* $D > 0$*, where* $A_j := (1-\rho(1-\delta)^{j-1})T$*,* $B_j(v_i) := A_j + \frac{min(v_i,\rho)}{1-v_1}(1-\rho)(1-\delta)^{j-1}T - D$*, and* M' *is any meta-arm that chooses arm* j' *in round* τ

Proof. Consider the meta-arm $M^* = M([1,2,A_1],[2,3,A_2],...,[k,j,A_k])$. That is, M^* is the arm that repeatedly pulls the current arm until it reaches its 1-session and then switches to the arm with the next earliest 0-session, until it switches to arm j on its last switch.

Note first that at any round $\tau \geq A_j$, this arm always has at least as much reward as arm $M(j)$. The arm earns more than arm j before switching to j (by pulling arms $< j$, which have earlier and longer 0-sessions), and the same as arm j after switching to j (by pulling arm j). Therefore, by Lemma B.2 in [Bra+17], $\sigma_{M^*,\tau}(v_i) \geq \sigma_{M(j),\tau}(v_i) \geq \sigma_{M(k),\tau}(v_i) + D$ for $k > j$. Crucially, note that Lemma B.2 holds because it does not rely on the values of ρ and n, which are the only things different in Strategy 2.

Now, consider any $M' : M'_t \leq j \implies t > \tau, \forall t$; that is, any meta-arm that does not select any arms $\leq j$ before τ. We will show that $\sigma_{M^*,\tau}(v_i) \geq \sigma_{M',\tau}(v_i) + D$ for any such M'. Note that of all arms $> j$, $j+1$ is the smallest and thus the one with the earliest and largest 0-session. Note also that since $B_j < A_{j+1}$, arm $j+1$ won't yet have entered its 1-session by τ. Thus, for all rounds in $[0, \tau]$, exclusively pulling arm $j+1$ is weakly better than pulling any possible sequence of arms $> j$, and so $\sigma_{M(j+1),\tau}(v_i) \geq \sigma_{M',\tau}(v_i)$. Thus, $\sigma_{M^*,\tau}(v_i) \geq \sigma_{M(j+1),\tau}(v_i) + D \geq \sigma_{M',\tau}(v_i) + D$.

Now, the only meta-arms left to consider are those M' for which $\exists M'_t : M'_t \leq j, t \leq \tau$; that is, all meta-arms where there is at least one arm $\leq j$ pulled before τ. By construction, the best such arms (which also pull an arm $l > j$ on round τ) are of the form $M'_* = M([1, 2, A_1], [2, 3, A_2], ..., [k, l, A_k]), l > j$. In other words, those which do exactly as M^*, repeatedly pulling the current arm until it reaches its 1-session and then switching to the arm with the next earliest 0-session, except they switch to an arm $l > j$ instead of arm j on their last switch. Until time A_k, M^* and M'_* choose the same arms and have the same reward. By Lemma B.2, arm j will have at least D more total reward than arm l in round τ. Now, by Lemma 8, we have it that arms j and l must start their 0-session after A_k. Thus, for each of these two arms, its cumulative rewards in intervals $[0, \tau]$ and $[A_k, \tau]$ are equal. Thus, arm j will earn at least D more reward than arm l between A_k and τ, and so M^* will earn at least D more reward than M'_* between A_k and τ. Thus, for $M' : M'_t \leq j, t \leq \tau$, we have it that $\sigma_{M^*,\tau}(v_i) \geq \sigma_{M'_*,\tau}(v_i) + D \geq \sigma_{M',\tau}(v_i) + D$. Therefore, $\sigma_{M^*,\tau}(v_i) \geq \sigma_{M',\tau}(v_i) + D$ for any meta-arm M' that chooses arm $l > j$ in round τ. ■

We now introduce the notion of a g-mean-based learning algorithm, which generalizes the concept of a mean-based learning algorithm:

Definition 6 (g-Mean-Based Learning Algorithm)**.** *Let $\sigma_{i,t}(v) = \sum_{s=1}^{t} r_{i,s}(v)$. An algorithm (for the contextual bandits problem) is g-mean-based for a function $g(\cdot)$ if it is the case that whenever $\sigma_{i,t}(v) < \sigma_{j,t}(v) - D$, then the probability that the algorithm pulls arm i on round t if it has context v is at most $g(D)$, for any $D > 0$.*

Theorem 11. *Say $\mathcal{A}$ is a g-mean-based learning algorithm. If the bidder is using any k-switching bidding algorithm $\mathcal{A}^*$ defined as $\mathcal{A}^*(\{r_{i,s}(v_s)\}_{s \leq t}, [m]) = \mathcal{A}(\{r_{M,s}(v_s)\}_{s \leq t}, \mathbb{M}_{k,T})$, then for any constant $\varepsilon > 0$ there exists a strategy for the seller that yields expected revenue that is at least $(1 - n^{(k)} g(D)) \cdot ((1 - \varepsilon) Val(\mathcal{D}) T - D(n-k))$, where $n^{(k)} = |\mathbb{M}_{k,T}|$ and $D > 0$.*

Proof. Say the seller uses Strategy 2. From Theorem 11, we have it that for all $j : A_j < B_j(v_i), j > k$, in each round in the interval $[A_j, B_j(v_i)]$ the probability that the buyer with value v_i chooses an arm $l > j$ is at most $g(D)$. Thus, by a union bound over all meta-arms, the probability in each of these rounds that they choose *any* arm $> j$ is at most $n^{(k)} g(D)$.

Thus, over this interval, the buyer with value v_i will, with probability at least $(1-n^{(k)}g(D))$, choose an arm currently in its 1-session (i.e. an arm with label at most j) and hence pay 1 each round. Recall the buyer has value v_i for the item with probability q_i. Then, the total contribution of the buyer with value v_i to the expected revenue of the seller must be at least

$$
\begin{aligned}
& q_i \sum_{j:A_j<B_j(v_i),j>k} (1-n^{(k)}g(D))(B_j(v_i)-A_j) \\
\geq\ & q_i \sum_{j=k+1}^{n} (1-n^{(k)}g(D))(B_j(v_i)-A_j) \\
=\ & q_i \sum_{j=k+1}^{n} (1-n^{(k)}g(D))(\frac{min(v_i,\rho)}{1-v_1}(1-\rho)(1-\delta)^{j-1}T-D) \\
=\ & (1-n^{(k)}g(D))q_iT(-\frac{n-k}{T}D+\frac{(1-\rho)min(v_i,\rho)}{1-v_1}\sum_{j=k+1}^{n}(1-\delta)^{j-1}) \\
=\ & (1-n^{(k)}g(D))q_iT(-\frac{n-k}{T}D+\frac{(1-\rho)min(v_i,\rho)((1-\delta)^k-(1-\delta)^n)}{\delta(1-v_1)}) \\
=\ & (1-n^{(k)}g(D))q_iT(-\frac{n-k}{T}D+min(v_i,\rho)((1-\delta)^k-(1-\delta)^n)) \\
\geq\ & (1-n^{(k)}g(D))q_iT(-\frac{n-k}{T}D+v_i(1-\varepsilon/2)((1-\delta)^k-(1-\delta)^n)) \\
\geq\ & (1-n^{(k)}g(D))q_iT(-\frac{n-k}{T}D+v_i(1-\varepsilon/2)(1-\varepsilon/2)) \\
\geq\ & (1-n^{(k)}g(D))q_iT(-\frac{n-k}{T}D+v_i(1-\varepsilon))
\end{aligned}
$$

The first inequality comes from the fact that the terms where $B_j(v_i) \leq A_j$ have a non-positive contribution to the sum. The second line is the definition of $B_j(v_i)$. The fourth to last line uses the definition of δ. The third to last line uses the fact that $min(v_i,\rho) \geq (1-\varepsilon/2)v_i$ (since if $min(v_i,\rho) \neq v_i$, then $\rho = max(1-\varepsilon/2, 1-(1-v_1)(1-(1-\varepsilon/4)^{1/k})) \geq 1-\varepsilon/2 \geq (1-\varepsilon/2)v_i$, because $v_i \leq 1$). The second to last line uses Lemma 7. The last line uses the fact that $(1-\varepsilon/2)^2-(1-\varepsilon) = \frac{\varepsilon^2}{4} \geq 0$ for all ε.

Summing this contribution over all $v_i \in \mathcal{D}$, we have that the expected revenue of the seller is at least

$$
\begin{aligned}
& \sum_{i\in[m]} (1-n^{(k)}g(D))q_iT(-\frac{n-k}{T}D+v_i(1-\varepsilon)) \\
= \; & ((1-n^{(k)}g(D))(1-\varepsilon)T)\sum_{i\in[m]} q_iv_i - (1-n^{(k)}g(D))(n-k)DT\sum_{i\in[m]} q_i \\
= \; & ((1-n^{(k)}g(D))(1-\varepsilon)T)\mathbb{E}_{v\sim\mathcal{D}}[v] - (1-n^{(k)}g(D))(n-k)D(1) \\
= \; & (1-n^{(k)}g(D))\cdot((1-\varepsilon)Val(\mathcal{D})T - D(n-k))
\end{aligned}
$$

■

We now show that this bound improves for a subset of possible g.

Theorem 12. *Say $\mathcal{A}$ is a g-mean-based learning algorithm for some $g(D) \in O(e^{-\xi D})$ for some $\xi \in \omega(\frac{\ln(T)}{T})$. If the bidder is using any k-switching bidding algorithm $\mathcal{A}^*$ defined as $\mathcal{A}^*(\{r_{i,s}(v_s)\}_{s\leq t},[m]) = \mathcal{A}(\{r_{M,s}(v_s)\}_{s\leq t},\mathbb{M}_{k,T})$, then for any constant $\varepsilon > 0$ there exists a strategy for the seller that yields expected revenue that is at least $(1-o(1))(1-\varepsilon)Val(\mathcal{D})T - o(T)$.*

Proof. We know $\exists B>0, D_0>0, \xi\in\omega(\frac{\ln(T)}{T})$ such that $g(D)\leq Be^{-\xi D}, \forall D\geq D_0$. We also know by Lemma 5 that $n^{(k)} < (T-1)^{k+1}(n-1)^{k+1}$. Let $D = \frac{(k+1)\ln((T-1)(n-1))}{\xi} + \frac{T}{\ln(T)}$ (though we could replace $\frac{T}{\sqrt{\ln(T)}}$ with anything in $\omega(\sqrt{\frac{T}{\ln(T)}})\cap o(T)$). Then we have

$$
\begin{aligned}
& \lim_{T\to\infty} \ln(n^{(k)}g(D)) \\
< \; & \lim_{T\to\infty} \ln((T-1)^{k+1}(n-1)^{k+1})Be^{-\xi(\frac{(k+1)\ln((T-1)(n-1))}{\xi}+\frac{T}{\ln(T)})} \\
= \; & \lim_{T\to\infty} (k+1)\ln((T-1)(n-1)) - \xi(\frac{(k+1)\ln((T-1)(n-1))}{\xi} + \frac{T}{\ln(T)}) + ln(B) \\
= \; & \lim_{T\to\infty} (k+1)\ln((T-1)(n-1)) - (k+1)\ln((T-1)(n-1)) - \xi\frac{T}{\ln(T)} + ln(B) \\
= \; & \lim_{T\to\infty} -\xi\frac{T}{\ln(T)} + ln(B) \\
= \; & -\infty
\end{aligned}
$$

$$
\implies \quad \lim_{T\to\infty} n^{(k)}g(D) = 0
$$

Thus, the quantity $(1 - n^{(k)}g(D))$ goes to 1.

Now, notice that

$$\begin{aligned}
\xi \in \omega(\frac{\ln(T)}{T}) &\implies \frac{\ln(T)}{T} \in o(\xi) \\
&\implies \lim_{T\to\infty} \frac{\frac{\ln(T)}{T}}{\xi} = 0 \\
&\implies \frac{\ln(T)}{\xi} \in o(T) \\
&\implies \frac{(k+1)\ln((T-1)(n-1))}{\xi} \in o(T)
\end{aligned}$$

Moreover, it's clear that $\frac{T}{\ln(T)} \in o(T)$, and thus $D \in o(T)$, which clearly ensures $B_j(v_i) > A_j$. Combining all this with Theorem 11, we get that Strategy 2 yields revenue at least

$$\begin{aligned}
& (1 - n^{(k)}g(D)) \cdot ((1-\varepsilon)Val(\mathcal{D})T - D(n-k)) \\
= \ & (1 - o(1))(1-\varepsilon)Val(\mathcal{D})T - o(T)
\end{aligned}$$

■

We now show two functions $g(D)$ for which MWU and FTPL are g-mean-based, and then use them to show Theorem 12 applies to MWU and FTPL.

Theorem 13. *The Multiplicative Weights Update algorithm with multiplicative update factor η is g-mean-based with $g(D) := \frac{1}{1+e^{\eta(D-2)}}$.*

Proof. Consider two arms i^*, j^* such that $\sigma_{i^*,t} < \sigma_{j^*,t} - D$ for some D. We will show that the probability that algorithm $\mathcal{A} = FTPL$ pulls arm i^* on round t is at most $g(D) = \frac{1}{1+e^{\eta(D-2)}}$.

The Multiplicative Weights Update algorithm chooses an arm i at time t proportional to its weight $w_{i,t}$. That is, the probability that the algorithm $\mathcal{A} = MWU$ selects arm i at time t is $P[\mathcal{A}(\{r_{i,s}(v_s)\}_{s<t}) = i] = \frac{w_{i,t}}{\sum_{j=1}^{n} w_{j,t}}$. Since the weight at time $t+1$ is defined as $w_{i,t+1} = w_{i,t}(1 - \eta m_{i,t}) = w_{i,t}(1 + \eta r_{i,t})$, we have it that for any arm i, at time t', $w_{i,t'} = \prod_{t=1}^{t'-1}(1 + \eta r_{i,t})$. Since $\eta r_{i,t} \in [-\frac{1}{2}, \frac{1}{2}]$, then $(1 + \eta r_{i,t}) \approx e^{\eta r_{i,t}}$. Thus, $w_{i,t'} \approx \prod_{t=1}^{t'-1} e^{\eta r_{i,t}} = e^{\eta \sum_{t=1}^{t'-1} r_{i,t}} = e^{\eta \sigma_{i,t'-1}}$. In the remainder of this proof, we will use this approximation as equality.

Note that since $r_{i,s} \in [-1,1]$ for all i, s, we have it that $\sigma_{i^*,t-1} < \sigma_{j^*,t-1} - (D-2)$. Thus, $w_{j*,t} = e^{\eta \cdot \sigma_{j^*,t-1}} > e^{\eta(\sigma_{i^*,t-1}+(D-2))}$. Finally, we then have:

$$\begin{aligned}
P[\mathcal{A}(\{r_{i,s}(v_s)\}_{s<t}) = i^*] &= \frac{w_{i^*,t}}{\sum_{i=1}^{n} w_{i,t}} \\
&\leq \frac{w_{i^*,t}}{w_{i^*,t} + w_{j^*,t}} \\
&= \frac{e^{\eta\sigma_{i^*,t-1}}}{e^{\eta\sigma_{i^*,t-1}} + e^{\eta\sigma_{j^*,t-1}}} \\
&< \frac{e^{\eta\sigma_{i^*,t-1}}}{e^{\eta\sigma_{i^*,t-1}} + e^{\eta(\sigma_{i^*,t-1}+(D-2))}} \\
&= \frac{e^{\eta\sigma_{i^*,t-1}}}{e^{\eta\sigma_{i^*,t-1}} + e^{\eta\sigma_{i^*,t-1}} e^{\eta(D-2)}} \\
&= \frac{1}{1 + e^{\eta(D-2)}}
\end{aligned}$$

■

Theorem 14. *The Follow the Perturbed Leader algorithm with perturbations that are exponentially distributed with decay rate parameter λ is g-mean-based with $g(D) := \frac{e^{-\lambda(D-2)}}{2}$.*

Proof. Consider two arms i^*, j^* such that $\sigma_{i^*,t} < \sigma_{j^*,t} - D$ for some D. We will show that the probability that algorithm $\mathcal{A} = FTPL$ pulls arm i^* on round t is at most $g(D) = \frac{e^{-\lambda(D-2)}}{2}$.

By definition of FTPL, the output of the algorithm at time t will be $\mathcal{A}(\{r_{i,s}(v_s)\}_{s<t}) = \underset{i\in[n]}{argmax}\,(\sigma_{i,t-1} + \phi_{i,t})$, where $\phi_{i,t}$ are i.i.d. random variables drawn from the exponential distribution $d\mu(x) = \lambda x e^{-\lambda x}$. Note that since $r_{i,s} \in [-1,1]$ for all i, s, we have it that $\sigma_{i^*,t-1} < \sigma_{j^*,t-1} - (D-2)$. Thus, we have it that:

$$\begin{aligned}
P[\mathcal{A}(\{r_{i,s}(v_s)\}_{s<t}) = i^*] &\leq P[\sigma_{i^*,t-1} + \phi_{i^*,t} >= \sigma_{j^*,t-1} + \phi_{j^*,t}] \\
&\leq P[\phi_{i^*,t} - \phi_{j^*,t} >= \sigma_{j^*,t-1} - \sigma_{i^*,t-1}] \\
&\leq P[\phi_{i^*,t} - \phi_{j^*,t} > D-2] \\
&\leq P[\phi_{i^*,t} > \phi_{j^*,t} + (D-2)]
\end{aligned}$$

Where the first inequality comes from the fact that for arm i^* to be chosen in round t, it must at least have a larger value of $\sigma_{i,t-1} + \phi_{i,t}$ than j^*.

Now, since $\phi_{i,s}$ are exponential random variables, we can compute the quantity on the right-hand side directly:

$$\begin{aligned} P[\phi_{i^*,t} > \phi_{j^*,t} + (D-2)] &= \int_0^\infty \lambda \cdot e^{-\lambda x} \cdot e^{-\lambda((D-2)+x)} dx \\ &= \int_0^\infty \lambda \cdot e^{-\lambda(D-2)} \cdot e^{-2\lambda x} dx \\ &= \frac{e^{-\lambda(D-2)}}{2} \cdot \int_0^\infty 2\lambda e^{-2\lambda x} dx \\ &= \frac{e^{-\lambda(D-2)}}{2} \cdot (-e^{-2\lambda x}|_0^\infty) \\ &= \frac{e^{-\lambda(D-2)}}{2} \cdot (1 - e^{-\lambda\infty}) \\ &= \frac{e^{-\lambda(D-2)}}{2} \end{aligned}$$

■

We now apply Theorem 12 to the Follow the Perturbed Leader and Multiplicative Weights Update algorithms, and get the following:

Theorem 15. *Say $\mathcal{A}$ is either the Follow the Perturbed Leader algorithm or the Multiplicative Weights Update algorithm with regret-minimizing parameters. If the bidder is using any k-switching bidding algorithm $\mathcal{A}^*$ defined as $\mathcal{A}^*(\{r_{i,s}(v_s)\}_{s\le t}, [m]) = \mathcal{A}(\{r_{M,s}(v_s)\}_{s\le t}, \mathbb{M}_{k,T})$, then for any constant $\varepsilon > 0$ there exists a strategy for the seller that yields expected revenue that is at least $(1-o(1))(1-\varepsilon)Val(\mathcal{D})T - o(T)$.*

Proof. We know from Theorems 14 and 15 that $g(D) = \frac{1}{1+e^{\eta(D-2)}} < e^{-\eta(D-2)}$ under MWU and $g(D) = \frac{e^{-\lambda(D-2)}}{2} < e^{-\lambda(D-2)}$ under FTPL, and we know from Theorems 9 and 10 that the regret-minimizing parameter values for these algorithms are $\eta = \sqrt{\frac{(k+1)\ln((T-1)(n-1))}{T}}$, $\lambda = \sqrt{\frac{C(k+1)\ln((T-1)(n-1))}{T}}$, both of which are in $\omega(\frac{\ln(T)}{T})$, since $\lim_{T\to\infty} \frac{\sqrt{\frac{ln(T)}{T}}}{\frac{ln(T)}{T}} = \lim_{T\to\infty} \sqrt{\frac{T}{ln(T)}} = \infty$. Thus, by Theorem 12, we get revenue that is at least $(1-o(1))(1-\varepsilon)Val(\mathcal{D})T - o(T)$. ■

Finally, we conclude by proving that Strategy 2 and Strategy 1 cannot yield maximal revenue for $k \geq n-1$, to motivate some future work.

Theorem 16. *Say the buyer is bidding according to a no-regret k-switching algorithm with $k \geq n-1$. Then, Strategies 1 and 2 are unable to extract revenue that is arbitrarily close to the welfare.*

Proof. By Lemma 8, we know an arm $j > j'$ cannot start its 0-session before j' begins its 1-session. Thus, each arm only begins its 0-session after the arm immediately smaller than it finishes its 1-session (and thus also its 0-session). Thus, no two arms are in their 0-session at the same time, and so to obtain the total amount of rounds during which there is an arm in a 0-session, we can simply sum the amount of rounds each arm is in its 0-session. Formally, the family of sets $\{\{t \in [T] : \text{arm } j \text{ is in its 0-session during } t\} : j \in [n]\}$ is pairwise disjoint, meaning the size of the union of all sets in the family is simply the sum of their sizes. This means that the total amount of rounds during which there is an arm offering the item at no cost is

$$\begin{aligned}
\sum_{i=1}^{n}(1-\delta)^{i-1}(1-\rho)T &= \frac{(1-(1-\delta)^n)}{\delta}(1-\rho)T \\
&= (1-(1-\delta)^n)(1-v_1)T \\
&> (1-(1-\delta)^n)\varepsilon T \\
&> 0
\end{aligned}$$

Where in the first inequality we use the fact that $v_1 < 1-\varepsilon$ because $\mathcal{D}$ is not entirely supported on $[1-\varepsilon, 1]$, and in the second we use that $\delta < 1$ in Strategy 1 and in Strategy 2 (the latter of which we showed in Lemma 6).

Now, because a given arm's 0-session consists of a continuous sequence of rounds, and none of these sequences intersect, then a meta-arm that switches arms $n-1$ times would be able to extract utility $(1-(1-\delta)^n)(1-v_1)Val(\mathcal{D})T$ by repeatedly pulling the current arm until it reaches its 1-session and then switching to the arm with the next earliest 0-session. This is the meta-arm $M^* = M([1,2,A_1],[2,3,A_2],...,[n-1,n,A_{n-1}])$. Thus, for $k \geq n-1, \exists M \in \mathbb{M}_{k,T} : \sum_{t=1}^{T} r_{M,t} = (1-(1-\delta)^n)(1-v_1)Val(\mathcal{D})T$. Now, let T^0 represent the set of rounds for which the buyer pulls an arm in its 0-session and let T^1 represent the set of rounds for which the buyer pulls an arm in its 1-session. Note $T^0 \cap T^1 = \emptyset$.

Then, the utility of the buyer is

$$
\begin{aligned}
\mathbb{E}[\sum_{t\in T^0} v_t + \sum_{t\in T^1}(v_t-1)] &= \mathbb{E}[\sum_{t=1}^{T} a_{I_t,t}\cdot v_t - p_{I_t,t}] \\
&= \mathbb{E}[\sum_{t=1}^{T} r_{I_t,t}(v_t)] \\
&= \mathbb{E}[\max_{M\in\mathbb{M}_{k,T}} \sum_{t=1}^{T} r_{M,t}(v_t)] - \mathbb{E}[Reg(\mathcal{A}^*(\{r_{i,s}(v_s)\}_{s\leq t}))] \\
&\geq \mathbb{E}[\max_{M\in\mathbb{M}_{k,T}} \sum_{t=1}^{T} r_{M,t}(v_t)] - o(T) \\
&\geq (1-(1-\delta)^n)(1-v_1)Val(\mathcal{D})T - o(T)
\end{aligned}
$$

Where here we use the definitions of $r_{i,t}$ and no-regret, and the reward of meta-arm M^* which we found above. Now we show that $|T^0| + |T^1| \leq (1-(1-\delta)^n)(1-v_1)T + o(T)$. We already showed above that $|T^0| \leq (1-(1-\delta)^n)(1-v_1)T$. For that value of $|T^0|$, we can use the inequality above to get

$$
\begin{aligned}
\mathbb{E}[\sum_{t\in T^0} v_t + \sum_{t\in T^1}(v_t-1)] &= (1-(1-\delta)^n)(1-v_1)Val(\mathcal{D})T + \mathbb{E}[\sum_{t\in T^1}(v_t-1)] \\
\mathbb{E}[\sum_{t\in T^1}(v_t-1)] &\geq -o(T)
\end{aligned}
$$

Since $\mathbb{E}[v_t-1] < 0$, this can only hold if $|T_1| \in o(T)$. We then have it that the value of the buyer is

$$
\begin{aligned}
\mathbb{E}[\sum_{t=1}^{T} a_{I_t,t}\cdot v_t] &= \mathbb{E}[\sum_{t\in T^0} v_t + \sum_{t\in T^1} v_t] \\
&= |T^0|Val(\mathcal{D}) + |T^1|Val(\mathcal{D}) \\
&\leq (1-(1-\delta)^n)(1-v_1)Val(\mathcal{D})T + o(T)
\end{aligned}
$$

Combining the bounds for the utility and value of the buyer, we have it that the seller gets revenue

$$
\begin{aligned}
\mathbb{E}[\sum_{t=1}^{T} p_{I_t,t}] &= \mathbb{E}[\sum_{t=1}^{T} a_{I_t,t}\cdot v_t - (a_{I_t,t}\cdot v_t - p_{I_t,t})] \\
&= \mathbb{E}[\sum_{t=1}^{T} a_{I_t,t}\cdot v_t] - \mathbb{E}[r_{I_t,t}(v_t)] \\
&\leq o(T)
\end{aligned}
$$

This means that over a long horizon, the seller is able to extract at most 0 revenue per round on average, which is obviously much less than extracting revenue equal to the welfare each round. ■

5 Conclusion

We showed that when the buyer applies recency bias to a γ-mean-based no-regret bidding algorithm $\mathcal{A}$ with recency bias factor β, then the expected regret won't increase by more than $H(\beta, T) := 2(\frac{\beta(\beta^T-1)}{\beta-1} - T)$. Moreover, the revenue of the seller under Strategy 1 won't decrease by more than $(1 - n\gamma)nH(\beta, T)$. As a result, if we restrict β so that $\beta \leq (1+\sigma)^{1/T}$ for some $\sigma \in o(1)$, the bidder will remain no-regret, and Strategy 1 will be able to extract revenue that is arbitrarily close to the welfare. This is already interesting because we found values of β that are usable for the buyer (by maintaining no-regret) and some values of β for which Strategy 1 yields maximal revenue, but it is particularly exciting because the entire set of β values which we found to satisfy one of these also satisfies the other. As a sidenote, we remark that $\beta \leq 1 + \frac{\upsilon}{T}$ for some $\upsilon \in o(1)$ implies that $\beta \leq (1 + \sigma)^{1/T}$ for some $\sigma \in o(1)$, and so this is an alternative (but more restrictive) constraint on β that yields the same bounds.

We also showed that k-switching algorithms using one of two common learning algorithms, Multiplicative Weights Update and Follow the Perturbed Leader, are able to allow $k \in o(\frac{T}{\ln(T)})$, almost the maximum value of k, and remain no-regret. This is remarkable because this is the case despite the number of meta-arms growing exponentially with k. We then presented Strategy 2, and lower bounded the revenue it attains. We began with a general lower bound of $(1 - n^{(k)}g(D))((1 - \varepsilon)Val(\mathcal{D})T - D(n - k))$ when the buyer is k-switching using any g-mean-based algorithm. We then found that Strategy 2 extracts revenue that is arbitrarily close to the welfare when the buyer uses a g-mean-based learning algorithm with $g(D) \in O(e^{-\xi D})$ for some $\xi \in \omega(\frac{\ln(T)}{T})$. Next, we provided practical relevance to that bound by showing it applies to both Multiplicative Weights Update and Follow the Perturbed Leader. This is especially important because our bound relied on the g-mean-based property with exponentially decreasing $g(D)$, which is a much stronger condition than the mean-based property. Moreover, since we had found that these two algorithms specifically were able to remain no-regret for reasonably large k, then they represent two examples of algorithms for k-switching which both allow the buyer to remain no-regret and yield revenue that is arbitrarily close to the welfare.

Intuitively, Strategy 2 uses more arms than Strategy 1 (ensuring there's significantly more than k) and spends less time luring the bidder with 0-sessions (and more time overcharging them). Unfortunately, it does this in part by utilizing k in its initialization of ρ and n, which means the strategy requires the seller to have some knowledge about k. More importantly, if the seller does not have perfect information about k, they might not set n to be sufficiently large and let $k \geq n - 1$. As we showed in Theorem 16, in this case the seller would not be able to attain revenue greater than $o(T)$, which is negligible. Further work could search for auctions that are able to attain more revenue when $k \geq n - 1$, perhaps by somehow incentivizing the buyer to return to arms they have already switched out of, so that $n - 1$ switches are no longer sufficient for the buyer.

References

[AHK12] Sanjeev Arora, Elad Hazan, and Satyen Kale. "The Multiplicative Weights Update Method: a Meta-Algorithm and Applications". In: *Theory of Computing* 8.6 (2012), pp. 121–164. DOI: 10.4086/toc.2012.v008a006. URL: https://theoryofcomputing.org/articles/v008a006.

[ADH16] Itai Ashlagi, Constantinos Daskalakis, and Nima Haghpanah. "Sequential Mechanisms with ex-post Participation Guarantees". In: *CoRR* abs/1603.07229 (2016). arXiv: 1603.07229. URL: http://arxiv.org/abs/1603.07229.

[Bra+17] Mark Braverman et al. "Selling to a No-Regret Buyer". In: *CoRR* abs/1711.09176 (2017). arXiv: 1711.09176. URL: http://arxiv.org/abs/1711.09176.

[BC12] Sébastien Bubeck and Nicolò Cesa-Bianchi. "Regret Analysis of Stochastic and Nonstochastic Multi-armed Bandit Problems". In: *CoRR* abs/1204.5721 (2012). arXiv: 1204.5721. URL: http://arxiv.org/abs/1204.5721.

[CD17] Yang Cai and Constantinos Daskalakis. "Learning Multi-item Auctions with (or without) Samples". In: *CoRR* abs/1709.00228 (2017). arXiv: 1709.00228. URL: http://arxiv.org/abs/1709.00228.

[Cla71] Edward H. Clarke. "Multipart Pricing of Public Goods". In: *Public Choice* 11 (1971), pp. 17–33. ISSN: 00485829, 15737101. URL: http://www.jstor.org/stable/30022651 (visited on 04/10/2023).

[CR15] Richard Cole and Tim Roughgarden. "The Sample Complexity of Revenue Maximization". In: *CoRR* abs/1502.00963 (2015). arXiv: 1502.00963. URL: http://arxiv.org/abs/1502.00963.

[DS15] Constantinos Daskalakis and Vasilis Syrgkanis. "Learning in Auctions: Regret is Hard, Envy is Easy". In: *CoRR* abs/1511.01411 (2015). arXiv: 1511.01411. URL: http://arxiv.org/abs/1511.01411.

[DHP15] Nikhil R. Devanur, Zhiyi Huang, and Christos-Alexandros Psomas. "The Sample Complexity of Auctions with Side Information". In: *CoRR* abs/1511.02296 (2015). arXiv: 1511.02296. URL: http://arxiv.org/abs/1511.02296.

[DPS14] Nikhil R. Devanur, Yuval Peres, and Balasubramanian Sivan. "Perfect Bayesian Equilibria in Repeated Sales". In: *CoRR* abs/1409.3062 (2014). arXiv: 1409.3062. URL: http://arxiv.org/abs/1409.3062.

[Dud+16] Miroslav Dudık et al. "Oracle-Efficient Learning and Auction Design". In: *CoRR* abs/1611.01688 (2016). arXiv: 1611.01688. URL: http://arxiv.org/abs/1611.01688.

[GN16] Yannai A. Gonczarowski and Noam Nisan. "Efficient Empirical Revenue Maximization in Single-Parameter Auction Environments". In: *CoRR* abs/1610.09976 (2016). arXiv: 1610.09976. URL: http://arxiv.org/abs/1610.09976.

[Gro73] Theodore Groves. "Incentives in Teams". In: *Econometrica* 41.4 (1973), pp. 617–631. ISSN: 00129682, 14680262. URL: http://www.jstor.org/stable/1914085 (visited on 04/10/2023).

[Han58] James Hannan. "Approximation to Bayes Risk in Repeated Play". In: *Contributions to the Theory of Games (AM-39), Volume III*. Princeton: Princeton University Press, 1958, pp. 97–140. ISBN: 9781400882151. DOI: doi:10.1515/9781400882151-006. URL: https://doi.org/10.1515/9781400882151-006.

[Imm+17] Nicole Immorlica et al. "Repeated Sales with Multiple Strategic Buyers". In: *Proceedings of the 2017 ACM Conference on Economics and Computation*. EC '17. Cambridge, Massachusetts, USA: Association for Computing Machinery, 2017, pp. 167–168. ISBN: 9781450345279. DOI: 10.1145/3033274.3085130. URL: https://doi.org/10.1145/3033274.3085130.

[KV05] Adam Kalai and Santosh Vempala. "Efficient algorithms for online decision problems". In: *Journal of Computer and System Sciences* 71.3 (2005). Learning Theory 2003, pp. 291–307. ISSN: 0022-0000. DOI: https://doi.org/10.1016/j.jcss.2004.10.016. URL: https://www.sciencedirect.com/science/article/pii/S0022000004001394.

[LZ07] John Langford and Tong Zhang. "The Epoch-Greedy Algorithm for Contextual Multi-Armed Bandits". In: *Proceedings of the 20th International Conference on Neural Information Processing Systems*. NIPS'07. Vancouver, British Columbia, Canada: Curran Associates Inc., 2007, pp. 817–824. ISBN: 9781605603520.

[LP17] Siqi Liu and Christos-Alexandros Psomas. “On the Competition Complexity of Dynamic Mechanism Design”. In: *CoRR* abs/1709.07955 (2017). arXiv: 1709.07955. URL: http://arxiv.org/abs/1709.07955.

[Mir+16a] Vahab Mirrokni et al. “Dynamic Auctions with Bank Accounts”. In: *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*. 2016, pp. 387–393. URL: http://www.ijcai.org/Proceedings/16/Papers/062.pdf.

[Mir+16b] Vahab S. Mirrokni et al. “Optimal dynamic mechanisms with ex-post IR via bank accounts”. In: *CoRR* abs/1605.08840 (2016). arXiv: 1605.08840. URL: http://arxiv.org/abs/1605.08840.

[MR15] Jamie Morgenstern and Tim Roughgarden. “The Pseudo-Dimension of Near-Optimal Auctions”. In: *CoRR* abs/1506.03684 (2015). arXiv: 1506.03684. URL: http://arxiv.org/abs/1506.03684.

[MR16] Jamie Morgenstern and Tim Roughgarden. “Learning Simple Auctions”. In: *29th Annual Conference on Learning Theory*. Ed. by Vitaly Feldman, Alexander Rakhlin, and Ohad Shamir. Vol. 49. Proceedings of Machine Learning Research. Columbia University, New York, New York, USA: PMLR, 23–26 Jun 2016, pp. 1298–1318. URL: https://proceedings.mlr.press/v49/morgenstern16.html.

[NST15] Denis Nekipelov, Vasilis Syrgkanis, and Eva Tardos. “Econometrics for Learning Agents”. In: *Proceedings of the Sixteenth ACM Conference on Economics and Computation*. EC ’15. Portland, Oregon, USA: Association for Computing Machinery, 2015, pp. 1–18. ISBN: 9781450334105. DOI: 10.1145/2764468.2764522. URL: https://doi.org/10.1145/2764468.2764522.

[OS22] Michael Ostrovsky and Andrzej Skrzypacz. “Pure-Strategy Equilibrium in the Generalized First-Price Auction”. In: *Available at SSRN 4250550* (2022).

[Pap+22] Christos Papadimitriou et al. “On the complexity of dynamic mechanism design”. In: *Games and Economic Behavior* 134 (2022), pp. 399–427. ISSN: 0899-8256. DOI: https://doi.org/10.1016/j.geb.2022.01.024. URL: https://www.sciencedirect.com/science/article/pii/S0899825622000306.

[Rou15] Tim Roughgarden. "The Price of Anarchy in Games of Incomplete Information". In: *ACM Trans. Econ. Comput.* 3.1 (Mar. 2015). ISSN: 2167-8375. DOI: 10.1145/2737816. URL: https://doi.org/10.1145/2737816.

[SZ19] Ben Schiffer and Shirley Zhang. *Maximum Revenue Auctions for Non-Mean-Based Bidding Algorithms*. Unpublished student project for COS521. 2019.

[ST13] Vasilis Syrgkanis and Eva Tardos. "Composable and Efficient Mechanisms". In: *Proceedings of the Forty-Fifth Annual ACM Symposium on Theory of Computing*. STOC '13. Palo Alto, California, USA: Association for Computing Machinery, 2013, pp. 211–220. ISBN: 9781450320290. DOI: 10.1145/2488608.2488635. URL: https://doi.org/10.1145/2488608.2488635.

[Vic61] William Vickrey. "Counterspeculation, Auctions, and Competitive Sealed Tenders". In: *The Journal of Finance* 16.1 (1961), pp. 8–37. ISSN: 00221082, 15406261. URL: http://www.jstor.org/stable/2977633 (visited on 04/10/2023).

[Zha20] Shirley Zhang. "Selling to No-Regret Buyers". Senior thesis. Princeton University, 2020. URL: http://arks.princeton.edu/ark:/88435/dsp014j03d2694.