Robust Repeated Auctions under Heterogeneous Buyer Behavior

Shipra Agrawal* Constantinos Daskalakis† Vahab Mirrokni‡ Balasubramanian Sivan§

Abstract

We study revenue optimization in a repeated auction between a single seller and a single buyer. An essential difference between a repeated auction setting and a single-shot interaction is that in the former, the seller can extract much higher revenue, close to the entire surplus of the buyer. However, in order to extract such a high revenue, even the simplest dynamic mechanisms in the literature make several strong assumptions about the buyer behavior. In particular, the widely used solution concept of dynamic incentive compatibility requires that the buyer make his current round decision with an infinite lookahead, namely, taking into account the consequences of his choices today on the utility of all the future rounds. This requires the buyer to completely understand and trust the seller's mechanism, and further believe that a) the interaction will last, and b) the seller will stick to his announced mechanism, *for all future rounds*. What if a buyer is not informed of and/or lacks trust in the seller's mechanism? Such a buyer could:

1. express his limited trust on the seller by evaluating his decisions today only based on its effect on the utility of k future rounds (a k-lookahead buyer).
2. completely disregard the seller's description of the mechanism, and instead make his decisions through his favorite learning algorithm using only past feedback from the mechanism.

Can we design mechanisms which are robust against such buyer behaviors?

In this paper, we answer this question by designing a simple state-based mechanism that is simultaneously near optimal against a k-lookahead buyer for all k, a buyer who is a no-regret learner, and a buyer who is a policy-regret learner. That is, against each kind of buyer our mechanism gets a constant fraction of the optimal revenue possible against that buyer. We complement our positive results with almost tight impossibility results showing that the revenue approximation tradeoffs achieved by our mechanism for different k's are near optimal.

*Columbia University. sa3205@columbia.edu.
†Massachusetts Institute of Technology. costis@csail.mit.edu.
‡Google Research. mirrokni@google.com.
§Google Research. balusivan@google.com.

1 Introduction

Developing a theory of repeated auctions that outlines the boundaries for what is and is not possible is of both scientific and commercial significance. On the application side, it is partly motivated by online sale of display ads in ad exchanges via repeated auctions. An essential difference that sets apart the repeated/dynamic setting from its one-shot counterpart is the significantly higher revenue that is achievable in the former. The key reason for this difference is simply that bundling may increase revenue, and repeated interactions provide ample opportunities to bundle across time.

In its gross form, a dynamic mechanism that bundles across time could simply demand the buyer to pay her entire surplus for T rounds, save some small ϵ, upfront for the promise of getting the item for free in all future rounds. Such a mechanism exploits concentration bounds in a way that one-shot mechanisms cannot. Indeed, a risk-neutral buyer with no outside option would have no choice but to accept this offer to get a utility of $\Theta(\epsilon)$, with high probability, or else get 0 utility. Such mechanisms that threaten buyers to get either tiny or 0 utility have several drawbacks, the most prominent being that they force the buyer to make a huge payment upfront which is unappealing. This motivated a string of recent work [ADH16, MLTZ16a, BML16, MLTZ17] proposing mechanisms that satisfy *per round ex-post individual rationality*, i.e. that the buyer's utility is non-negative in every round under his optimal strategy, rather than *interim individual rationality,* which only requires that the buyer's long-term expected utility is non-negative.

Nevertheless, the ability of these dynamic mechanisms to extract high revenue depends crucially on several non-trivial assumptions:

1. The buyer completely understands the seller's mechanism. In particular, he understands, and can optimally respond to the conseqeunces of his actions today on his utility k rounds later, for all k;

2. The buyer believes that the interaction with the seller will last for all future rounds;

3. The buyer believes that the seller will stick to her proposed mechanism for all future rounds.

In particular, the notion of 'infinite look-ahead buyers' which is baked into the widely used concept of dynamic incentive-compatibility, requires that the buyer's action in every round takes into account the consequences of his action on his utility in *all future rounds*, thus relying on all of the above assumptions. There are important practical reasons invalidating these assumptions. Firstly, the buyer may not be fully informed about and/or trust all the details of the seller's mechanism. Furthermore, cognitive/computational limitations or uncertainty about future may prevent buyers from being infinite lookahead. In the context of online advertising, for example, given the number and variety of display ad exchanges in the market, with credibility levels all across the spectrum, the buyers often don't trust that the seller will faithfully implement the announced mechanism [Exc].

As a result, the seller often faces a buyer population that employs a variety of strategies, beyond perfectly rational infinite lookahead utility maximization, in order to maximize their *perceived* utility. Such a buyer could

1. be myopic or more generally, have a limited lookahead, i.e., evaluate his decisions today only based on their effect on the utility of k future rounds (a k-lookahead buyer).

2. be a learner, i.e., completely disregard the seller's description of the mechanism, and instead make his decisions through his favorite learning algorithm using only his observed feedback so far.

In face of such heterogeneous behaviors, the revenue optimal solution for seller is to have a tailored mechanism for each buyer behavior. However, there are strong reasons precluding the implementation of different mechanisms, each targeting a specific buyer behavior. Such discriminative targeting may be legally infeasible, and it may also be practically infeasible, as it could be hard for the seller to identify a buyer's response behavior. The latter may not even be well-defined, as buyers may change their response strategy across time. These observations motivates us to we ask the following question.

Can we design mechanisms which are robust against heterogeneous buyer behaviors?

Specifically, we seek a *single* mechanism that gets near optimal revenue simultaneously against buyers with different lookahead and learning behaviors, i.e., against each kind of buyer, obtains a constant fraction of the optimal revenue achievable by mechanisms tailored for that specific buyer.

Our Setting and Research Questions We study a repeated interaction between a single seller and a single buyer over multiple rounds. At the beginning of each round $t = 1, 2, \ldots$, there is a single fresh good for sale whose private value $v_t \in V$ for the buyer is drawn from a publicly known distribution[1] F with finite expectation μ. The buyer observes the valuation v_t and makes a bid b_t. The good for sale in round t has to be either allocated to the buyer or discarded immediately (i.e., not carried forward). The buyer's valuations are additive across rounds.

Our goal is to investigate the sensitivity of revenue extraction to variations in both the lookahead attitude and learning behavior of the buyers. On lookahead attitudes, by current understanding the notion of forward-looking buyer is a key driver of increase in revenue in dynamic mechanisms, but is a strong assumption like *infinite* lookahead really necessary? We ask the following questions.

1. Is the infinite lookahead assumption really necessary to get a revenue close to the full surplus? If buyers are forward looking, i.e., k-lookahead for finite $k \geq 1$, but not infinite lookahead, what loss should the seller suffer because of this?

2. How does a mechanism that gets close to the full surplus as revenue with forward-looking buyers, fare against myopic (0-lookahead) buyers? Can we design a *robust* mechanism that *simultaneously* achieves a high revenue against all lookahead levels, including myopic buyers?

On learning behavior, along similar lines, we question the necessity of assuming a completely informed buyer — we ask how well can a dynamic mechanism fare against a buyer using reasonably sophisticated learning algorithms to decide bids based on the past observations. Formally, we model learning behavior of buyers using the popular concept of no-regret learning, and allow different levels of learning sophistication by considering buyers who minimize simple regret vs. policy regret (defined in section 2).

Our Contributions In this paper, we answer the above questions by formalizing the notion of robustness to the buyer's lookahead attitude. In particular, we make the following contributions:

1. We introduce a **novel revenue optimization framework** to study robust dynamic mechanisms. In this framework, a desirable mechanism is required to simultaneously achieve a high revenue against a range of buyer behaviors, namely, (a) buyers with different lookahead attitudes (myopic, k-lookahead, infinite lookahead), and (b) learning buyers using different learning strategies (no-regret learning, policy-regret learning).

[1] As we explain later, for our positive results, the seller only needs to know the mean μ and not the whole distribution F.

2. We provide **impossibility results on the achievable revenue tradeoffs** in our framework, showing that, for all $\epsilon \in [0,1]$, no mechanism that is non-payment forceful (see definition in Section 2) and obtains an average per round revenue of $(1-\epsilon)\mu$ against an infinite lookahead buyer, can simultaneously get average per round revenue more than $2\epsilon \mathrm{Rev}^{\mathrm{Mye}}$ against a myopic buyer, where $\mathrm{Rev}^{\mathrm{Mye}}$ is the optimal revenue in a single-item auction against a single buyer with distribution F. See Theorem 4.5.

3. We present a **simple state-based mechanism** (named $M(\epsilon, \rho, p)$) with the following desirable properties; see Theorem 4.3.
   Robustness: It simultaneously earns an average per round revenue of at least (i) $(1-\epsilon)\mu$ against a forward-looking buyer (k-lookahead for any $k \geq 1$), (ii) $\frac{\epsilon}{2}\mathrm{Rev}^{\mathrm{Mye}}$ against a myopic buyer, (iii) $(1-\epsilon)\mu$ against a policy-regret learner, (iv) $\frac{\epsilon}{2}\mathrm{Rev}^{\mathrm{Mye}}$ against a no-regret learner[2], for any parameter $\epsilon \in (0,1]$.
   Near-optimality: Even if a mechanism were tailored to a specific behavior, the average per round revenue achievable would be bounded by: $\mathrm{Rev}^{\mathrm{Mye}}$ for a myopic buyer [Mye81], and by μ for any forward-looking or learning buyer (the revenue upper bound of μ is trivial as it is the entire surplus). Thus, our mechanism achieves a constant factor approximation to the optimal revenue achievable in each individual setting, by choosing say $\epsilon = \frac{1}{2}$. Further, as demonstrated by our impossibility result, the specific *revenue tradeoff* against myopic and forward looking buyers achieved by our mechanism is also near-optimal.
   Individual Rationality: Our mechanism is interim individually rational (IIR) for a k-lookahead buyer for every $k \geq 0$; see the discussion above or Section 2 for a definition. Moreover, it is "almost per-round ex-post IR" for k-lookahead buyers, with large enough k (see Section 2.4 for a precise definition). For buyers who are learning, per-round ex-post IR is not the right measure because no-regret-learning algorithms should be explorative, and will venture into negative utility strategies as well. A meaningful measure is whether at the end of the game, the buyer's learning strategy achieves a non-negative, or close to non-negative, overall utility. Accordingly, we show that under mild assumptions on the class of experts, any no-regret learning, or no-policy-regret learning algorithm is guaranteed to get a buyer utility no smaller than $-o(T)$.

The above results answer the question of whether the assumption of infinite lookahead is indispensable for earning the full surplus as revenue. We show that a much weaker assumption of forward-looking buyer is enough. Our results characterize precisely the loss that the seller must suffer in order for her revenue to be robust against both forward-looking and myopic buyers. Further, our impossibility result demonstrates that robustness is an appropriate lens to analyze dynamic mechanisms, as it rules out unappealing mechanisms that extract unreasonably high seller revenue while providing close-to-zero utility to the buyer: any mechanism that extracts almost full surplus for the seller against forward-looking buyers (i.e., $\epsilon \simeq 0$) would have close to 0 revenue against myopic buyers, and thus is not robust. In particular, it should also be appreciated that the ability of mechanisms to "threaten" the buyers or "promise" things in the future is non-existent when the buyer is myopic, and very limited when the buyer is 1-lookahead or k-lookahead for small k. Moreover, strategems that are tailored to a certain kind of lookahead behavior may fail for different lookahead behavior. Our mechanisms have to walk a thin line to simultaneously extract revenue from buyers with different lookahead attitudes.

[2]While the tradeoff for myopic vs k lookahead buyer is $\frac{\epsilon}{2}\mathrm{Rev}^{\mathrm{Mye}}$ vs $(1-\epsilon)\mu$, the tradeoff that we can obtain for no-regret vs no-policy-regret learner is actually slightly better: $\epsilon\mathrm{Rev}^{\mathrm{Mye}}$ vs $(1-\epsilon)\mu$. But when we want a single mechanism to get all four simultaneously, we have to settle for the $\frac{\epsilon}{2}\mathrm{Rev}^{\mathrm{Mye}}$ vs $(1-\epsilon)\mu$ tradeoff for no-regret vs no-policy-regret as well

Similarly, our results also answer the question of whether the buyer has to trust the seller's words completely in order for the seller to extract a very high revenue. We show that the buyer can learn a great deal himself through a no-policy-regret learning algorithm and against such a buyer, the seller can earn a very high revenue (almost the entire surplus). But, such sophisticated buyer learning behavior is not necessary for a high revenue. We show that even if the buyer follows a simple no-regret learning strategy, we can get a constant fraction of $\mathrm{Rev}^{\mathrm{Mye}}$.

2 Our framework

In this section, we present the main components of our framework, including a *state-based mechanism design setting*, *formal models for buyers' heterogeneous lookahead and learning behaviors*, and a *revenue optimization objective* to calibrate the mechanisms against different behaviors.

2.1 Setup

We study a repeated interaction between a single seller and a single buyer for a finite number of rounds T. At the beginning of each round $t = 1, 2, \ldots, T$, there is a single fresh good for sale whose private value $v_t \in V$ for the buyer is drawn from a publicly known distribution F with bounded support and expectation μ. The buyer observes his value v_t at the beginning of the round and makes a bid b_t. The good for sale in round t has to be either allocated to the buyer or discarded immediately (i.e., not carried forward). The buyer's valuations are additive across rounds.

Round outcome. The outcome of the game in round t is a pair (x_t, p_t), where $x_t \in \{0, 1\}$ indicates whether or not the buyer received the good and $p_t \in \mathbb{R}$ is the payment made by the buyer to the seller. For the outcome pair (x_t, p_t) in round t, the linear utility u_t of the buyer in round t is given by $u_t = v_t x_t - p_t$, and the seller's revenue is given by p_t.

State-based mechanism. A mechanism M is defined as a 5-tuple $M = (\mathcal{S}, Q, x, p, s_1)$, where

1. $\mathcal{S}$ is the state space over which the mechanism operates. The state space $\mathcal{S}$ can be finite dimensional or countably infinite dimensional. The cardinality of $\mathcal{S}$ can be finite, countably infinite or uncountably infinite.

2. $x : \mathcal{S} \times \mathbb{R}^+ \to \boldsymbol{\Delta}^{\{0,1\}}$ is the (randomized) allocation function, which at the beginning of round t receives as input the state $s_t \in \mathcal{S}$ in round t, the bid $b_t \in \mathbb{R}^+$ made by the bidder in round t, and outputs the allocation $x_t \in \{0, 1\}$ for the bidder in round t, where $x_t \sim x(s_t, b_t)$.

3. $p : \mathcal{S} \times \mathbb{R}^+ \times \{0, 1\} \to \mathbb{R}$ is the payment function, which at the beginning of round t takes as input the state $s_t \in \mathcal{S}$ in round t, the bid $b_t \in \mathbb{R}^+$ made by the bidder in round t, and the allocation x_t sampled as above, and outputs the payment p_t for the bidder in round t, i.e., $p_t = p(s_t, b_t, x_t)$. We restrict our mechanisms to require that the payment p_t is always non-negative, and 0 when the bid is 0, i.e.,

$$p(s_t, b_t, x_t) \geq 0, \text{ and } p(s_t, 0, x_t) = 0, \text{ for any } s_t, b_t, x_t. \tag{1}$$

   This ensures that the mechanisms are *non-payment forceful*, as they cannot be forced to pay the buyer in any state and they cannot force payments out of buyers who bid 0.

4. $Q : \mathcal{S} \times \mathbb{R}^+ \times \{0, 1\} \times \mathbb{R} \rightarrow \mathbf{\Delta}^{\mathcal{S}}$ is a state-transition function that takes as input at the end of round t the state $s_t \in \mathcal{S}$, the bid $b_t \in \mathbb{R}^+$ of the bidder in round t, the allocation $x_t \in \{0, 1\}$ and the payment $p_t \in \mathbb{R}$, and outputs the distribution of next state s_{t+1} for round $t + 1$, i.e., $s_{t+1} \sim Q(s_t, b_t, x_t, p_t)$. [3]

5. $s_1 \in \mathcal{S}$ is the starting state.

2.2 Lookahead behaviors

ℓ-lookahead utility. We will define lookahead attitudes of buyers using the concept of ℓ-lookahead utility. Assuming there are at least ℓ remaining rounds following round t, the buyer computes ℓ-lookahead utility of a bid b in round t as the expected utility over the current round plus the maximum expected utility obtainable over the next ℓ rounds. More precisely, at round t, given the current state s_t and valuation v_t, for any ℓ, the buyer's expected ℓ-lookahead utility for bid b, assuming $T \geq t + \ell$, is defined as:

$$U^t_\ell(s_t, v_t, b) \quad := \quad \mathrm{E}\left[v_t\, x_t(b) - p_t(b) + \sup_{b'} U^{t+1}_{\ell-1}(s_{t+1}, v_{t+1}, b') \middle| s_t, v_t\right], \tag{2}$$

$$U^t_0(s_t, v_t, b) \quad := \quad \mathrm{E}\left[v_t\, x_t(b) - p_t(b) \middle| s_t, v_t\right], \tag{3}$$

where $x_t(b) \sim x(s_t, b), p_t(b) = p(s_t, b, x_t(b)),\ s_{t+1} \sim Q(s_t, b, x_t(b), p_t(b))$. And, the expectation was taken over the random values $v_{t+1} \sim F$ and any randomization in the mechanism.

Buyer lookahead behaviors. We define the following types of buyers with different lookahead attitudes:[4]

- k-lookahead buyer: A *k-lookahead buyer* is a buyer who, in every round t, picks his bid b_t to maximize $\min\{k, T - t\}$-lookahead utility, i.e.,

  $$b_t = \max_{b \in \mathbb{R}^+} U^t_{\min\{k, T-t\}}(s_t, v_t, b) \tag{4}$$

  We refer to the bid b_t computed above as a **k-lookahead optimal bid**. (Note that in general, the maximizer in the abvoe equation may not exist, see technical remark 2.) Two special cases of k-looahead buyers are:

  - Myopic buyer: We refer to a 0-lookahead buyer as a *myopic buyer*.
  - Infinite-lookahead buyer: If a mechanism lasts for T rounds, we refer to a k-lookahead buyer with $k \geq T - 1$ as an *infinite-lookahead buyer*.

- Forward-looking buyer: A *forward looking buyer* maximizes k_t-lookahead utility for some $k_t \in \{1, \ldots, T - t\}$, at every time t. Thus, a forward-looking buyer may use different lookaheads at different time steps, but always looks ahead at least one step.

[3]Given that our price function is a deterministic function of s_t, b_t, x_t we could have also suppressed p_t from the arguments of Q.

[4]See Technical Remark 2 for some comments about this definition.

Technical Remark. To be completely formal, the definition of k-lookahead buyer given above requires that the maximizer in (4) exists. There are mechanisms in which this maximizer does not exist. For instance, consider a single-state mechanism that offers the item at a price of ϵ, whenever the bid is $1 - \epsilon < 1$, and does not offer the item when the bid is 1. This mechanism has no optimal k-lookahead bid for any k. Such mechanisms are undesirable as they make it difficult for the buyers to decide what bid to use, and therefore for the sellers to understand what revenue to expect, even if they know their buyer's lookahead attitude. For this reason, the mechanisms that we construct are such that there always exists an optimal k-lookahead bid. On the other hand, our impossibility results hold even if optimal lookahead policies do not exist as long as a k-lookahead buyer is assumed to use a good enough approximately optimal bid, which *is* guaranteed to exist given our assumption about the boundedness of the support of F.

2.3 Learning behaviors

No-regret learning. To formally model a broad class of buyer learning behaviors, we use the concept of *no-regret learning*, a widely studied solution concept in the context of T round online prediction problem with advice from N experts. In this problem, at every round $t = 1, \ldots, T$, an adversary picks reward $\mathbf{g_t} = \{g_{1,t}, \ldots, g_{N,t}\}$ where $g_{i,t}$ is the reward associated with expert i. The learner needs to pick an expert $i_t \in [N]$, and observe reward $g_{i_t,t}$. Regret in time T is defined as

$$\text{Regret}(T) = \max_{i \in [N]} \sum_{t=1}^{T} g_{i,t} - \sum_{t=1}^{T} g_{i_t,t}. \tag{5}$$

A *no-regret online learning algorithm* for this problem uses the past observations $g_{i_1,1}, \ldots, g_{i_{t-1},t-1}$ to make the decision i_t in every round t such that $Regret(T) \leq o(T)$. When the number of experts N is finite, there are efficient and natural algorithms (e.g., EXP3 algorithm based on multiplicative weight updates) that achieve $O(\sqrt{NT \log N})$ regret.

Note that the regret compares the total reward achieved by the learner with the reward of best *single* expert in hindsight. Furthermore, even if the adversary is adaptive (i.e., generates $\mathbf{g_t}$ adaptively based on $i_1, \ldots, i_t$), the performance of the best expert is being evaluated over the sequence of inputs $\mathbf{g_1}, \ldots, \mathbf{g_T}$ produced by the adversary in response to the learner's decision, and not those that *would be produced* if this expert was used in all rounds. This is an important distinction between the above definition of regret, for which efficient online learning algorithms like EXP3 are known, vs. the more sophisticated '*policy regret*' which we define next.

Policy regret learning. A no-policy-regret learning algorithm is a more sophisticated learner based on the definition of policy regret from [ADT12]. Such an algorithm faces an adaptive adversary, and achieves $o(T)$ *policy regret*, defined as:

$$\text{Policy-regret}(T) = \max_{j_1, \ldots, j_T \in \mathcal{C}_T} \sum_{t=1}^{T} g_{j_t,t}(j_1, \ldots, j_t) - \sum_{t=1}^{T} g_{i_t,t}(i_1, \ldots, i_t). \tag{6}$$

where $\mathcal{C}(T)$ is some benchmark class of deterministic sequences of experts of length T, and the $g_{j_t,t}$ and $g_{i_t,t}$ have been explicitly written as a function of past decisions to indicate adaptive adversarial response to the *sequence* of choices so far. A special case is where $\mathcal{C}(T)$ is the class of *single* expert sequences, i.e.,

$$\text{Policy-regret}(T) = \max_{i} \sum_{t=1}^{T} g_{i,t}(i, \ldots, i) - \sum_{t=1}^{T} g_{i_t,t}(i_1, \ldots, i_t). \tag{7}$$

Buyer learning behaviors. We consider a buyer who only gets to observe whether the current state is good ($s_t \in \perp$) or bad ($s_t \notin \perp$), and the valuation v_t, before making the bid, and the outcome (allocation, price) after making the bid, but does not know (or does not trust) anything else about the seller's mechanism. Using these observations, the buyer is trying to decide bids b_t, using a learning algorithm under the experts learning framework described above. We formalize the notion of different levels of learning sophistication among buyers by considering two classes of learners:

- *No-regret learner:* Such a buyer considers, as experts, a finite collection $\mathcal{E}$ of mappings from the state information ($s_t = \perp$ or $s_t \neq \perp$) and valuation v_t to a bid, i.e., set of experts

$$\mathcal{E} = \{f : [\perp, \not\perp] \times \bar{V} \to \bar{V}\}; \quad (8)$$

  where $\bar{V}$ is an discretized (to arbitrary accuracy) version of V, in order to obtain a finite set of experts. The buyer uses a no-regret learning algorithm to decide which expert $f_t \in \mathcal{E}$ to use in round t to set $b_t = f_t(s_t, v_t)$. The adversarial reward at time t is given by the buyer's t^{th}-round utility, determined by the seller's mechanism's output, i.e., on making bid $b_t = f_t(s_t, v_t)$ in round t, the reward is given by buyer's utility

$$u_t(b_t) := \mathrm{E}[v_t x_t - p_t | s_t, v_t, b_t]$$

  so that for a no-regret learning buyer (refer to (5)),

$$\text{Regret}(T) = \max_{f \in \mathcal{E}} \sum_{t=1}^{T} u_t(f(s_t, v_t)) - \sum_{t=1}^{T} u_t(b_t) = o(T) \quad (9)$$

  Here, we slightly abused the notation to define f as a function of s_t, v_t, where as technically it is only a function of $f(1(s_t = \perp), v_t)$, that is, it only uses whether $s_t = \perp$ or $s_t \neq \perp$.

- *Policy-regret learner:* This more sophisticated buyer uses a no-*policy-regret* learning algorithm. Following (7), the important distinction from the definition of regret in the previous paragraph is that now the total utility of best expert must be evaluated over the *trajectory of states achieved by the expert*. To make explicit the dependence of t^{th} round utility on past decisions through the state at time t, let us denote the utility in round t as $u(b_t, s_t)$ Then, following (7), policy-regret of such a buyer is given by:

$$\text{Policy-Regret}(T) = \max_{f \in \mathcal{E}} \sum_{t=1}^{T} u_t(f(s'_t, v_t), s'_t) - \sum_{t=1}^{T} u_t(b_t, s_t), \quad (10)$$

  where $s'_1, \ldots, s'_T$ is the (possibly randomized) trajectory of states that would be observed on using the expert to decide the bids in all rounds. We define a no-policy-regret learning buyer as those for which the above quantity is guaranteed to be $o(T)$.

  While constructing such a no-policy-regret learner is difficult in general, for our proposed stochastic state-based mechanisms and the above special case of single expert sequences, this is achievable by some simple learning strategies. In fact, a simple buyer learning strategy that will work for our mechanism to achieve no-policy-regret with high probability, is to explore each possible bid for some time and then use the best single bid for the rest of the time steps.

2.4 Individual rationality

For k-lookahead buyers. We define a mechanism to be *interim individually rational (IIR)* for a k-lookahead buyer iff at any time t, state s_t, valuation v_t, and bid b_t that is k-lookahead optimal at time t, we have that:

$$U^t_{\min\{k,T-t\}}(s_t, v_t, b_t) \geq 0.$$

Note that the non-payment-forceful condition 3 in state-based mechanism's definition 2 guarantees that IIR is satisfied for all k.

We also define a mechanism to be *per round ex-post individually rational* for a k-lookahead buyer iff for any time t, state s_t, valuation v_t, and bid b_t that is k-lookahead optimal at time t, we have that:

$$\mathrm{E}_{x_t \sim x(s_t,b_t)}[x_t \cdot v_t - p_t(s_t, b_t, x_t)|v_t, s_t] \geq 0. \tag{11}$$

While we do not impose ex-post individual rationality as a hard constraint on our mechanisms, our constructions will satisfy *almost* per-round ex-post individual rationality for bidders that are k-lookahead with large enough k. A mechanism is *almost per-round ex-post individual rational* if there exist bids $b_1, \ldots, b_T$ that satisfy (11), and which, with high probability, are both k-lookahead near-optimal (utility-wise) at the respective time steps t for the buyer, and guarantee near-optimal average T round revenue for the seller, assuming T is large enough.[5]

For learning buyers. As discussed in the introduction, per-round individual rationality is not a meaningful measure of IR for learning buyers. A natural measure of IR is after the end of T rounds, whether the no-regret strategy followed by the buyer will result in ex-post non-negative utility for the buyer. That is, whether $\sum_{t=1}^{T} u_t(b_t) \geq 0$, where $u_t(\cdot)$ is the round t utility as defined in the previous section.

3 Our Mechanisms: $M(\epsilon, \rho)$ and $M(\epsilon, \rho, p)$

We prove our positive results by proposing a class of mechanisms $M(\epsilon, \rho, p)$ parametrized by ϵ, ρ and p. For illustration purposes, we also define a simpler mechanism $M(\epsilon, \rho)$ parameterized just by ϵ and ρ.

For both these mechanisms, the state-space is $\mathbb{R} \times \mathbb{N}$. The current state will be denoted by a pair $s_t = (\bar{b}, n)$, roughly corresponding to the current average of bids in the rounds where allocation was 1, and the number of those rounds, respectively.

A high level description of mechanisms $M(\epsilon, \rho)$ and $M(\epsilon, \rho, p)$ is as follows. Mechanism $M(\epsilon, \rho)$ accepts the current bid (makes an allocation of 1) if the buyer is in a 'good state', that is, if the average of bids accepted in the past is above the threshold of $(1 - \epsilon)\mu$. Otherwise, it rejects the bid (makes no allocation), and with probability ρ, transfers the buyer in a good state. Mechanism $M(\epsilon, \rho, p)$ behaves the same way as $M(\epsilon, \rho)$ when the buyer is in a good state. But in the bad state, if the buyer's current bid b_t is above p, then with probability ρ, it accepts the bid and transfers the buyer to a good state.

Below are the precise definitions.

Definition 1 (Mechanism $M(\epsilon, \rho)$)

- **State Space $\mathcal{S}$:** *The state space is $\mathcal{S} = \mathbb{R} \times \mathbb{N}$. We represent a state $s \in \mathcal{S}$ by a pair $(\bar{b}, n) \in \mathbb{R} \times \mathbb{N}$. We refer to states $s = (\bar{b}, n)$ with $\bar{b} \geq (1 - \epsilon)\mu$ as 'good states', and all the other states as 'bad states'. Abusing notation a little, if s is a good state we evaluate $s \neq \perp$ as true, and if s is a bad state we evaluate $s = \perp$ as true.*

[5]The constant lower bounds on k and T for this guarantee only depend on the bounded on the support of F and its variance, and the approximation guarantees required for the near-optimality of the k-lookahead utility and the near-optimality of the revenue.

- **Starting state** s_1**:** *The mechanism starts in the borderline good state, i.e.,*

$$s_1 = ((1-\epsilon)\mu, 0).$$

- **Allocation rule** $x(s_t, b_t)$**:** *Given current state* s_t*, bid* b_t*, this mechanism allocates if and only if the current state is a good state. The allocation* $x_t \sim x(s_t, b_t)$ *is determinstic here, and is given by*

$$x_t = \begin{cases} 1, & \text{if } s_t \neq \perp, \\ 0, & \text{otherwise.} \end{cases}$$

- **Payment rule** $p(s_t, b_t, x_t)$**:** *This is a first price mechanism, i.e.,*

$$p(s_t, b_t, x_t) = \begin{cases} b_t, & \text{if } x_t = 1, \\ 0, & \text{otherwise.} \end{cases}$$

- **State-transition function** $Q(s_t, b_t, x_t, p_t)$**:** $Q(s_t, b_t, x_t, p_t)$ *provides the distribution of next state* s_{t+1}*. Let* $s_t = (\bar{b}, n)$*. Then, in this mechanism, if* $x_t = 1$*, the next state* $s_{t+1} \sim Q(s_t, b_t, x_t, p_t)$ *is deterministic, and is given by the new average bid:*

$$s_{t+1} = (\frac{\bar{b}n + b_t}{n+1}, n+1)$$

Otherwise, if $x_t = 0$ *(then it must also be the case that* $s_t = \perp$*), then,*

$$s_{t+1} = \begin{cases} ((1-\epsilon)\mu, n), & \text{with probability } \rho, \\ s_t, & \text{with probability } 1-\rho. \end{cases}$$

Definition 2 (Mechanism $M(\epsilon, \rho, p)$**)** *Mechanism* $M(\epsilon, \rho)$ *has exactly the same state-space, starting state, and payment rule as the mechanism* $M(\epsilon, \rho)$ *defined above. The allocation rule and state-transition function are slightly different, as defined below.*

- **Allocation rule** $x(s_t, b_t)$**:** *Given current state* s_t*, bid* b_t*, this mechanism always allocates in a good state, otherwise, it allocates with probability* ρ *if the bid* b_t *is above the price* p*. That is,*

$$x(s_t, b_t) = \begin{cases} \text{Bernoulli}(1, 0), & \text{if } s_t \neq \perp, \\ \text{Bernoulli}(\rho, 1-\rho), & \text{if } s_t = \perp \text{ and } b_t \geq p, \\ \text{Bernoulli}(0, 1), & \text{otherwise.} \end{cases}$$

- **State-transition function** $Q(s_t, b_t, x_t, p_t)$**:** *In this mechanism, next state* $s_{t+1} \sim Q(s_t, b_t, x_t, p_t)$ *is a deterministic function of* $s_t = (\bar{b}, n)$ *and* x_t*.*

$$s_{t+1} = \begin{cases} \left(\frac{\bar{b}n + b_t}{n+1}, n+1\right), & \text{if } s_t \neq \perp, \\ ((1-\epsilon)\mu, n+1), & \text{if } s_t = \perp \text{ and } x_t = 1, \\ s_t, & \text{otherwise.} \end{cases}$$

4 Revenue Tradeoffs

Notation. We will use the shorthand $\mathrm{Rev}_k^M \equiv \mathbb{E}\left[\frac{\sum_{t=1}^T p_t}{T}\right]$ to denote the expected average revenue for mechanism M starting at state s_1 facing a k-lookahead buyer in all rounds $t = 1, \ldots, T$, with the understanding that the expectation symbol denotes expectation over all trajectories that could result from the randomness in the buyer's values and bids, the allocation in each round, and the state transitions. We similarly define Rev_+^M to be the revenue against any forward-looking buyer.

4.1 Warmup: Revenue tradeoffs of mechanism $M(\epsilon, \rho)$

In this section, as an illustrative example, we first analyze the revenue achieved by mechanism $M(\epsilon, \rho)$ against different kinds of buyers. This mechanism turns out to be almost too good to be true for forward-looking buyers : in Lemma 4.2, we prove that using this mechanism against such buyers, the seller can achieve $(1-\epsilon)\mu$ revenue for arbitrarily small ϵ by using a small $\rho \le \epsilon$. Given the definition of mechanism $M(\epsilon, \rho)$, observe that a very small ρ effectively leads to a threat-based mechanism: a mechanism that threatens to put a buyer indefinitely into a bad state if the buyer's average bid ever goes below $(1-\epsilon)\mu$. Therefore, if the buyer is forward-looking, she will never bid in a way to take her average bid below this threshold, thus giving a high revenue to the seller. However, if the buyer is myopic (possibly because she doesn't believe the threat), this mechanism completely fails : in Lemma 4.1, we show that against myopic buyers, $M(\epsilon, \rho)$ will result in 0 revenue for the seller.

This illustrates the significance of our requirement that the mechanism should achieve a nontrivial revenue against both myopic and forward-looking buyers: this requirement rules out such undesirable mechanisms like those described above. In next section, we propose a simple mechanism that can achieve close to optimal revenue tradeoff against all these different types of buyers.

Lemma 4.1 $\mathrm{Rev}_0^{M(\epsilon,\rho)} = 0$*, i.e.,* $M(\epsilon, \rho)$ *fails against a myopic buyer.*

Proof: Given s_t, v_t at time t, the myopic buyer make a bid b_t that maximizes 0-lookahead expected utility, given by

$$U_t^0(s_t, v_t, b_t) = \mathrm{E}[v_t x_t - p_t | v_t, s_t]$$

where x_t, p_t are given by the allocation and payment rules of the mechanism, as a function of the bid b_t and s_t. The expectation is over randomization in these rules. Now, by definition of $M(\epsilon, \rho)$ mechanism, if $s_t \neq \perp$, then $x_t = 1$ and $p_t = b_t$, irrespective of the bid value, so that $U_t^0 = v_t - \mathrm{E}[b_t]$, which is maximized by picking $b_t = 0$. If $s_t = \perp$, then $x_t = 0, p_t = 0$ in $M(\epsilon, \rho)$, making the utility $U_t^0 = 0$, irrespective of the bid value.

Therefore, a dominant strategy for the myopic buyer is to always bid 0, irrespective of the current state and valuation. Since the payment rule is first price rule (i.e., $p_t = b_t$ when $x_t = 1$, and 0 otherwise), this makes the seller's revenue 0. ∎

Lemma 4.2 *For any* $\epsilon > 0$*,* $M(\epsilon, \rho)$ *with* $\rho \le \epsilon$ *achieves a seller's revenue of* $\mathrm{Rev}_+^{M(\epsilon,\rho)} \ge (1-\frac{1}{T})(1-\epsilon)\mu$.

We omit the proof of above lemma, which follows the same lines of argument as the proof of a similar bound for mechanism $M(\epsilon, \rho, p)$, stated as part (c) in Theorem 4.3.

4.2 Main results: Revenue tradeoffs of mechanism $M(\epsilon, \rho, p)$

Here, we present our main result that the proposed mechanism $M(\epsilon, \rho, p)$ *simultaneously* achieves near-optimal revenue in face of a wide variety of buyer behaviors, namely, different *lookahead attitudes*: myopic, forward-looking, k-lookahead, and infinite lookahead, and broad classes of *learning behaviors*, formally modeled as no-regret learners and policy regret learners. Refer to Section 2 for formal definitions of these behaviors. The proof of this theorem is provided in Appendix A.

Theorem 4.3 *The mechanism $M(\epsilon, \rho, p^*)$ with $p^* = \arg\max_p p(1 - F(p))$ and $\rho \le \frac{\epsilon}{2-\epsilon}$, satisfies the following properties:*

(a) The optimal k-lookahead bid exists for all $k \ge 0, 0 \le t \le T$.

(b) The revenue against any myopic buyer is bounded as

$$\mathrm{Rev}_0^{M(\epsilon,\rho,p^*)} \ge \frac{\rho}{\rho+1} R_{mye} - \frac{1}{T},$$

where R_{mye} is the one round Myerson revenue, namely, $R_{mye} = p^*(1 - F(p^*))$.

(c) The revenue against any forward-looking buyer is bounded as

$$\mathrm{Rev}_+^{M(\epsilon,\rho,p^*)} \ge (1 - \frac{1}{T})(1-\epsilon)\mu.$$

(d) The revenue against any k-lookahead buyer (for $k \ge 1$) is bounded as

$$\mathrm{Rev}_k^{M(\epsilon,\rho,p^*)} \ge (1 - \frac{1}{T})(1-\epsilon)\mu.$$

(e) The revenue against a no-regret learner using class $\mathcal{E}$ of experts (refer to (8)*) is bounded as*

$$\mathrm{Rev}_{LS}^{M(\epsilon,\rho,p^*)} \ge \frac{\rho}{\rho+1} \mathrm{Rev}^{\mathrm{Mye}} - o(1).$$

(f) The revenue against a policy-regret learner with benchmark class $\mathcal{C}_T$ containing all single expert sequences is bounded as

$$\mathrm{Rev}_{LP}^{M(\epsilon,\rho,p^*)} \ge (1-\epsilon)\mu - o(1).$$

(g) Against all types of buyers – myopic, forward-looking and k-lookahead – the mechanism is IIR. For large enough constant k, the mechanism also satisfies almost per-round ex-post IR *(see section 2.4). Further, against no-regret and no-policy-regret learning buyers, the mechanism guarantees an almost non-negative ex-post utility, namely at least $-o(T)$.*

Corollary 4.4 *For large T, for any $\epsilon \in (0, 1)$, the mechanism $M(\epsilon, \rho, p^*)$ with $p^* = \arg\max_p p(1-F(p))$, with $\rho = \frac{\epsilon}{2-\epsilon}$, simultaneously achieves an average per round revenue of at least (i) $(1-\epsilon)\mu$ against a forward-looking buyer, (ii) $\frac{\epsilon}{2}\mathrm{Rev}^{\mathrm{Mye}}$ against a myopic buyer, (iii) $(1-\epsilon)\mu$ against a policy-regret learner, (iv) $\frac{\epsilon}{2}\mathrm{Rev}^{\mathrm{Mye}}$ against a no-regret learner.*

4.3 Impossibility results for achievable revenue tradeoffs

Theorem 4.5 *There exists a bounded, monotone hazard rate distribution F such that, for all $\epsilon \in [0, 1]$, no mechanism that is non-payment forceful (see equation* (1) *in the definition of state-based mechanism in Section 2) can simultaneously have expected per round revenue at least $\epsilon \cdot \mathrm{Rev}^{\mathrm{Mye}}$, against a myopic buyer with per round value distribution F, and expected per round revenue at least $(1 - \frac{\epsilon}{2}) \cdot \mu$, against an infinite look-ahead buyer with per round distribution F, where, as usual, $\mathrm{Rev}^{\mathrm{Mye}} = \max_x x(1 - F(x))$ and $\mu = E_{x \sim F}[x]$.*

Main Ideas We show our impossibility result by showing that there are distributions (even exponential distribution conditioned to lie in $[0, 1]$ works) for which any mechanism that gets $\Omega(\epsilon) \cdot \mathrm{Rev}^{\mathrm{Mye}}$ average per round revenue against a myopic buyer, will also necessarily offer her a utility that is $\Omega(\epsilon) \cdot \mu$ on average. But simultaneously getting a revenue of $(1 - \epsilon)\mu$ against the infinite lookahead buyer means that an infinite lookahead buyer is being left with at most $\epsilon\mu$ utility. If a myopic buyer is able to earn strictly more than $\epsilon\mu$ utility, so can an infinite lookahead buyer. This caps the utility achievable by a myopic buyer, which we use to cap the revenue achievable against a myopic buyer. Surprisingly the same argument doesn't work if we do a infinite vs 1-lookahead comparison, i.e., while the result may seem intuitive in hindsight, it is definitely not.

5 Other related work

There are several streams of literature in dynamic mechanism design. We begin with the stream that is closest to our work.

Optimal dynamic mechanisms. [PPPR16] show that the optimal deterministic dynamic mechanism satisfying ex-post IR constraints even in a single buyer 2 rounds setting, when the values are correlated is NP-hard. I.e., buyer learns his value of each round when it begins, and both buyer and seller know the distribution from which these values are drawn. They show that the optimal deterministic mechanism when the rounds have independent valuations can be computed in polynomial time. The optimal randomized mechanism even with correlated valuations can be computed in polynomial time. [MLTZ16a] study the single seller single buyer setting and show that with the IIR constraint, a very simple class of mechanisms called bank-account mechanisms that maintain a single scalar variable as state already obtain a significantly higher revenue and welfare compared to the single shot optimal. [ADH16] and [MLTZ16b] characterize the optimal ex-post IR mechanisms and consider approximations thereof via simple mechanisms (mainly in the single seller, single buyer setting, but their results also extend to the multi-bidder case) that again hold a single scalar variable as state. [BML16], consider a single seller single buyer setting and show that the seller can earn almost the entire surplus as revenue, even after imposing per round ex-post IR requirements and martingale utilities for the buyer. [MLTZ17] study oblivious dynamic mechanism design, namely, one where the seller is not aware of the future distributions of the buyer, and just the distribution for this round: they show that even with just this information, one can construct an ex-post IR dynamic incentive compatible mechanism that gets a $\frac{1}{5}$ of the optimal dynamic mechanism that knows all future distributions.

Mechanism design for buyers with evolving values. Another major focus area in dynamic mechanism design is one where buyers experience the same or related good repeatedly over time, and their value for the good evolves with time/usage. Initiated by the work of Baron and Besanko [BB84] there is a large body of work [Bes85, Bat05, CL00, BS14, ES07] that study optimizations in the presence of evolving values. Recent works include those by [BS15, AS13], where they consider general models where value evolution could depend on the action of the mechanism. [KLN13, PST14] study revenue optimal dynamic mechanism design where the buyer's value evolves based on signals that she receives each period. [CDKS16] study martingale value evolution for the buyer and show that simple constant pricing schemes followed by a free trial earns a constant fraction of the entire surplus.

Bargaining, durable goods monopolist and Coase conjecture. There is a large body of literature in economics that studies settings where the value is initially drawn from a distribution, but in subsequent rounds, the value remains the same, i.e., there is not a fresh draw in every round. This setting can be motivated based on several applications including bargaining, durable goods monopoly and behavior based discrimination. See [FVB06] for an excellent survey and references there in for an overview of this area.

Dynamically arriving and departing agents. Yet another body of work that comes under the umbrella of dynamic mechanisms is one where agents arrive and depart dynamically. Naturally focus is quite different from what we do in this paper.

Lookahead Search. The study of k-lookahead search can be viewed in the context of *bounded rationality*, as pioneered by Herb Simon [Sim55]. He argued that, instead of optimizing, agents may apply a class of heuristics, termed satisfying heuristics in decision making, A natural choice of such heuristics is restricting the search space of best-response moves. Lookahead search in decision-making has been motivated and examined in great extent by the artificial intelligence community [Nau83, dKaS92, SKN09]. Lookahead search is also related to the sequential thinking framework in game theory [SW94]. More recently, [MTV12] study the quality of equilibrium outcomes for look-ahead search strategies for various classes of games. They observe that the quality of resulting equilibria increases in generalized second-price auctions, and duopoly games, but not in other classes of games. No prior work studies dynamic mechanisms that are robust against various lookahead search strategies.

6 Discussion and further remarks

Applications to online advertising. On the practical side, the study of robust dynamic mechanisms in this paper is motivated by the online sale of display ads in ad exchanges via repeated auctions. Display ad exchanges are online market places for trading ad impressions in real time that account for a significant and increasing fraction of overall sales of Internet display ads. At the heart of these exchanges is a repeated interaction between a seller and a buyer who bids in real time, with hundreds of thousands of repeated auctions per day, stretching over several weeks/months. While new dynamic mechanisms can be studied to deploy in such advertising exchanges, ensuring that such strategies are robust against various responses from advertising agencies is very important.

Practicality of single buyer setting. While analyzing a single buyer mechanism is a natural first step in studying the landscape of what is and isn't possible, we claim that this investigation is already practically relevant for two reasons. Firstly, most of the billions of auctions run in display ad exchanges everyday are very thin, i.e., involve very few buyers, sometimes just one. In such mechanisms the buyer gets a large part of the surplus, and it becomes important in these auctions to devise methods to move some surplus from the buyer to the seller, without also reducing the buyer surplus to too little. Our results and mechanisms will provide insight and guidance in these situations. Secondly, even when there are multiple buyers, in many situations the revenue is primarily driven by a single buyer, where the ideas in our paper become relevant.

Independence of valuations across rounds. Our setting also assumes independence of valuations across rounds, which is used crucially in our proofs. When interacting with the same buyer repeatedly, can independence across rounds be possible? Indeed it is contrived when the same good is being sold across several rounds. But in the online advertising application, for example, every round's good is an opportunity to show an ad impression to a user who is identified by a vector of features. As a result, there might be little or no correlation between users across the rounds. Based on this observation, independence is a natural first-step assumption in these settings.

Further, our mechanisms depend only on a single parameter of the buyer's valuation distribution, namely its mean μ. In this sense, they are quite robust to assumptions on the buyer's valuation distribution as well.

Finite vs. infinite number of rounds. While it is cleanest to present all our results in an infinite rounds setting, the definition of seller revenue, which is the limit of the average revenue across T rounds, as $T \to \infty$, would be more involved to define: in particular, the limit need not exist for all mechanisms. Still our bounds on the revenue of our constructed mechanisms would hold under $\liminf$, and our impossibility results would hold under appropriate use of $\liminf$ and $\limsup$. To simplify exposition, we focus on the finite horizon case where the number of rounds T is large, and we study all lookaheads from $k = 0$ to T (the T-lookahead buyer is effectively the infinite lookahead buyer). The finiteness costs us a $(1 - \frac{1}{T})$ or $1 - o(1)$ factor in our revenue bounds (refer to Theorem 4.3). As T grows, $1 - o(1)$ factor becomes effectively 1, and irrelevant in our characterization results.

References

[ADH16] Itai Ashlagi, Constantinos Daskalakis, and Nima Haghpanah. Sequential mechanisms with ex-post participation guarantees. In *Proceedings of the 2016 ACM Conference on Economics and Computation, EC '16, Maastricht, The Netherlands, July 24-28, 2016*, pages 213–214, 2016.

[ADT12] Raman Arora, Ofer Dekel, and Ambuj Tewari. Online bandit learning against an adaptive adversary: from regret to policy regret. In *ICML*. icml.cc / Omnipress, 2012.

[AS13] Susan Athey and Ilya Segal. An efficient dynamic mechanism. *Econometrica*, 81(6):2463–2485, 2013.

[Bat05] Marco Battaglini. Long-term contracting with markovian consumers. *American Economic Review*, 95(3):637–658, 2005.

[BB84] David P. Baron and David Besanko. Regulation and information in a continuing relationship. *Information Economics and Policy*, 1(3):267 – 302, 1984.

[Bes85] David Besanko. Multi-period contracts between principal and agent with adverse selection. *Economics Letters*, 17(12):33 – 37, 1985.

[BML16] Santiago Balseiro, Vahab S. Mirrokni, and Renato Paes Leme. Dynamic mechanisms with martingale utilities. 2016.

[BS14] Dirk Bergemann and Philipp Strack. Dynamic revenue maximization: A continuous time approach. 2014.

[BS15] Dirk Bergemann and Philipp Strack. Dynamic revenue maximization: A continuous time approach. *Journal of Economic Theory*, 159, Part B:819 – 853, 2015. Symposium Issue on Dynamic Contracts and Mechanism Design.

[CDKS16] Shuchi Chawla, Nikhil R. Devanur, Anna R. Karlin, and Balasubramanian Sivan. Simple pricing schemes for consumers with evolving values. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10-12, 2016*, pages 1476–1490, 2016.

[CL00] Pascal Courty and Hao Li. Sequential screening. 67(4):697–717, 2000.

[dKaS92] J. de Kleer and O. Raiman andMark Shirley. One step lookahead is pretty good. *Readings in Model-Based Diagnosis*, pages 138–142,, 1992.

[ES07] Peter Eso and Balzs Szentes. Optimal information disclosure in auctions and the handicap auction. *Review of Economic Studies*, 74(3):705–731, 2007.

[Exc] Transparency in programmatic advertising. `https://auditedmedia.com/blog/transparency-programmatic-advertising-its-two-way-street`.

[FVB06] Drew Fudenberg and J Miguel Villas-Boas. Behavior-based price discrimination and customer recognition. *Handbook on economics and information systems*, 1:377–436, 2006.

[KLN13] Sham M Kakade, Ilan Lobel, and Hamid Nazerzadeh. Optimal dynamic mechanism design and the virtual-pivot mechanism. *Operations Research*, 61(4):837–854, 2013.

[MLTZ16a] Vahab S. Mirrokni, Renato Paes Leme, Pingzhong Tang, and Song Zuo. Dynamic auctions with bank accounts. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI 2016, New York, NY, USA, 9-15 July 2016*, pages 387–393, 2016.

[MLTZ16b] Vahab S. Mirrokni, Renato Paes Leme, Pingzhong Tang, and Song Zuo. Optimal dynamic mechanisms with ex-post IR via bank accounts. *CoRR*, abs/1605.08840, 2016.

[MLTZ17] Vahab S. Mirrokni, Renato Paes Leme, Pingzhong Tang, and Song Zuo. Oblivious dynamic mechanism design. 2017.

[MTV12] Vahab S. Mirrokni, Nithum Thain, and Adrian Vetta. A theoretical examination of practical game playing: Lookahead search. In *Algorithmic Game Theory - 5th International Symposium, SAGT 2012*, pages 251–262, 2012.

[Mye81] Roger B. Myerson. Optimal auction design. *Math. Oper. Res.*, 6(1):58–73, February 1981.

[Nau83] Dana S. Nau. Decision quality as a function of search depth on game trees. *J. ACM*, 30(4):687–708, 1983.

[PPPR16] Christos Papadimitriou, George Pierrakos, Christos-Alexandros Psomas, and Aviad Rubinstein. On the complexity of dynamic mechanism design. In *Proceedings of the Twenty-seventh Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '16, pages 1458–1475, Philadelphia, PA, USA, 2016. Society for Industrial and Applied Mathematics.

[PST14] Alessandro Pavan, Ilya Segal, and Juuso Toikka. Dynamic mechanism design: A myersonian approach. *Econometrica*, 82(2):601–653, 2014.

[Sim55] Herbert A. Simon. A behavioral model of rational choice. 69(1):99–118, 1955.

[SKN09] E. Sefer, U. Kuter, and D. Nau. Real-time a* search with depth-k lookahead. In *Proceedings of the International Symposium on Combinatorial Search*, 2009.

[SW94] Dale Stahl and Paul Wilson. Experimental evidence on players' models of other players. *Journal of Economic Behavior & Organization*, 25(3):309–327, 1994.

A Proof of Theorem 4.3

Proof: Properties (a) and (b) for myopic buyers are proven in Lemma A.1. Property (a) for k-lookahead optimal bids for $k \geq 1$ is proven in Lemma A.3.

Property (c) follows almost immediately from Lemma A.3, where we show that for any $k \geq 1$ and time t, such that $s_t \neq \perp$, any optimal k-lookahead bid is such that the next state s_{t+1} is deterministically a good state, i.e., $s_{t+1} \neq \perp$. Since mechanism $M(\epsilon, \rho, p)$ starts in a good state ($s_0 \neq \perp$), as a consequence of this lemma, a forward-looking buyer using optimal k_t-lookahead optimal bid for $k_t \geq 1$ for all t, will remain in good states for all the T rounds. By the definition of a good state, this implies that the average bid for the first $T-1$ rounds is at least $(1-\epsilon)\mu$. In the T-th round, the buyer will behave as a myopic buyer, and even if we count that revenue as 0, the T-round revenue is at least $(T-1)(1-\epsilon)\mu$, and the average revenue in the property (c) immediately follows.

Property (d) is a direct corollary of property (c).

Properties (e) and (f) are proven in Lemma A.6 and Lemma A.7 respectively.

For property (g), observe that in mechanism $M(\epsilon, \rho, p)$, the payment is equal to the bid whenever allocation is 1, and 0 otherwise, therefore, IIR property holds trivially. For a myopic buyer, as we show in Lemma A.1, the optimal myopic bid is 0 in a good state, and p^* in a bad state whenever $v_t \geq p^*$. This bidding strategy is ex-post IR. For a forward-looking buyer, a consequence of Lemma A.3 is that the k-lookahead buyer always remains in good state, and the k-lookahead optimal bid at any time t with $s_t \neq \perp$ should be such that the next state is a good state. An ex-post IR strategy that can achieve this is to bid $v_t(1-\epsilon)$ at all time steps t. This strategy will keep the buyer in good state for large t with high probability (as long as ϵ is large enough), and further achieve near-optimal average buyer utility (close to $\epsilon\mu$).

For a learning buyer, the fact that $\sum_{t=1}^{T} u_t(b_t) \geq -o(T)$ follows immediately from the no-regret and no-policy regret guarantees (refer to (9) and (10)) since the benchmark class considered (E) contains an expert that can bid 0 at all time steps to get ≥ 0 total utility. ■

Lemma A.1 *The mechanism $M(\epsilon, \rho, p^*)$ satisfies the following properties against a myopic buyer:*

(a) An optimal myopic bid exists for all time steps t.

(b) The average per round revenue against a myopic buyer is bounded as:

$$\mathrm{Rev}_0^{M(\epsilon,\rho,p^*)} \geq \frac{\rho}{\rho+1} R_{mye} - \frac{1}{T},$$

where R_{mye} is the one round Myerson revenue, namely, $R_{mye} = \max_p p(1-F(p)) = p^*(1-F(p^*))$.

Proof:
Buyer's optimal myopic bid: Let us first understand a myopic buyer's bidding strategy under mechanism $M(\epsilon, \rho, p^*)$. At any time t, given s_t, v_t, the myopic buyer makes the bid b_t that maximizes 0-lookahead

expected utility, given by $U_0^t(s_t, v_t, b_t) = \mathrm{E}[x_t v_t - p_t | v_t, s_t]$. Now, consider two cases:
(a) $s_t \neq \perp$: in this case, by definition of allocation rule in $M(\epsilon, \rho, p^*)$, $x_t = 1, p_t = b_t$, so that $U_t^0 = \mathrm{E}[v_t - b_t]$, irrespective of the value of bid b_t. Therefore, the (unique) utility maximizing strategy for the buyer is to bid 0, i.e.,

$$b_t = 0 \text{ when } s_t \neq \perp.$$

(b) $s_t = \perp$: in this case, by definition of allocation rule in $M(\epsilon, \rho, p^*)$, with probability ρ, $x_t = 1, p_t = b_t$ if $b_t \geq p^*$, so that $U_t^0 = \mathrm{E}[(v_t - b_t)1(b_t \geq p^*)]$. Therefore, the whenever $v_t \geq p^*$, the (unique) utility maximizing strategy for the buyer is to bid p^*, otherwise $x_t = 0$ and any bid less than p^* (including 0) is optimal, i.e.,

$$\begin{array}{c} b_t = p^*, x_t = 1 \text{ when } s_t = \perp, v_t \geq p^*, \text{ and} \\ b_t < p^*, x_t = 0 \text{ when } s_t = \perp, v_t < p^* \end{array}$$

Seller's revenue The seller's expected revenue is $\mathrm{E}[\frac{1}{T}\sum_{t=1}^T p_t]$. Now, for mechanism $M(\epsilon, \rho, p^*)$, $p_t = b_t$ whenever $x_t = 1$ and 0 otherwise. For a myopic buyer as described above, $\mathrm{E}[p_t | s_t \neq \perp] = 0$, $\mathrm{E}[p_t | s_t = \perp] = p^* \Pr(v_t \geq p^*) = p^*(1 - F(p^*))$. Substituting:

$$\begin{aligned} \mathrm{Rev}_0^{M(\epsilon,\rho,p^*)} &= \mathrm{E}[\frac{1}{T}\sum_{t=1}^T p_t] \\ &= \frac{1}{T}\sum_{t=1}^T \mathrm{E}[p_t | s_t \neq \perp] \Pr(s_t \neq \perp) + \mathrm{E}[p_t | s_t = \perp] \Pr(s_t = \perp) \\ &= p^*(1 - F(p^*)) \frac{1}{T}\sum_{t=1}^T \Pr(s_t = \perp). \end{aligned}$$

Here, $\sum_{t=1}^T \Pr(s_t = \perp)$ is the expected number of times bad state is visited in the T time steps. Now, by definition, mechanism $M(\epsilon, \rho, p^*)$ starts in good state $s_0 = ((1 - \epsilon)\mu, 0) \neq \perp$. A myopic buyer will bid 0 in this state which will get accepted (see the above discussion in optimal myopic bid), and she will immediately go to the bad state $((1 - \epsilon)\frac{\mu}{n+1}, 0) = \perp$. Transfer from bad state to the borderline good state $s_0 = ((1 - \epsilon)\mu, 0)$ happens with probability $\rho \Pr(v_t \geq p^*) = \rho(1 - F(p^*))$. Again, in s_0 the myopic bidder will bid 0 and immediately transfer back to a bad state. Therefore, the sequence of states takes the form $\not\perp \perp\perp\perp\perp\perp \not\perp \perp\perp\perp \not\perp \perp\perp \ldots$, i.e., sequence of bad states interspersed with *single* good states. The expected length of a subsequence $\not\perp \perp^+$ is $1 + \frac{1}{\rho(1-F(p^*))}$, with $\frac{\rho(1-F(p^*))}{1+\rho(1-F(p^*))} \geq \frac{\rho}{\rho+1}$ fraction of bad states. Accounting for the interruption in the last $\not\perp \perp^+$ sequence due to end of time horizon T, we have that the expected number of steps in a bad state is at least

$$T\frac{\rho(1 - F(p^*))}{1 + \rho(1 - F(p^*))} - 1 \geq T\frac{\rho}{\rho + 1} - 1.$$

Substituting, we get:

$$\mathrm{Rev}_0^{M(\epsilon,\rho,p^*)} \geq p^*(1 - F(p^*)) \left(\frac{\rho}{\rho + 1} - \frac{1}{T}\right)$$

■

Lemma A.2 *Under mechanism $M(\epsilon, \rho, p)$ with $\rho \leq \epsilon$, at any time $t < T$, an optimal 1-lookahead bid exists, and is such that the next state s_{t+1} is deterministically a good state, i.e., $s_{t+1} \neq \perp$.*

Proof: By definition, an optimal 1-lookahead bid b_t (if exists) maximizes 1-lookahead expected utility, i.e.,

$$b_t = \arg\max_b U_1^t(s_t, v_t, b)$$

In mechanism $M(\epsilon, \rho, p)$, for any t such that $s_t \neq \perp$, the allocation and payment are always $x_t = 1, p_t = b_t$. Therefore, using recursive relation between U_k^t and U_{k-1}^{t+1},

$$U_1^t(s_t, v_t, b_t) = \mathrm{E}[(v_t - b_t) + \sup_{b'} U_0^{t+1}(s_{t+1}, v_{t+1}, b')|v_t, s_t] \tag{12}$$

where $s_{t+1} \sim Q(s_t, b_t, 1, b_t)$. Now, for mechanism $M(\epsilon, \rho, p)$, bids get accepted in bad state with at most ρ probability, therefore,

$$\mathrm{E}[\sup_{b'} U_0^{t+1}(s_{t+1}, v_{t+1}, b')|s_{t+1} = \perp] \leq \rho\mu,$$

where as

$$\mathrm{E}[\sup_{b'} U_0^{t+1}(s_{t+1}, v_{t+1}, b')|s_{t+1} \neq \perp] = \mu$$

which can be achieved by $b' = 0$. Also, for $M(\epsilon, \rho, p)$, if s_t is a good state, then depending on the bid, the next state s_{t+1} will be deterministically bad or good state. For any bid b_t such that s_{t+1} is a bad state, substituting above in (12), we have that 1-lookahead utility is at most

$$v_t + \rho\mu,$$

where as if s_{t+1} is a good state, then $U_1^t(s_t, v_t, b_t) \geq (v_t - b_t) + \mu$. This is maximized by the minimum bid required to keep s_{t+1} as good state. In fact, in any good state s_t, the bid $b_t = (1-\epsilon)\mu$, always ensures s_{t+1} is a good state, and makes the 1-lookahead utility atleast

$$v_t + \epsilon\mu$$

Therefore, if $\epsilon > \rho$, then there exists at least one bid such that $s_{t+1} \neq \perp$ with strictly better 1-lookahead utility than any other bid such that $s_{t+1} = \perp$. This proves that any 1-lookahead optimal bid will have the stated property.

■

Lemma A.3 *Under mechanism $M(\epsilon, \rho, p)$ with $\rho \leq \frac{\epsilon}{2-\epsilon}$, for any $k \geq 1$ and time t, such that $s_t \neq \perp$, an optimal k-lookahead bid exists, and is such that the next state s_{t+1} is deterministically a good state, i.e., $s_{t+1} \neq \perp$.*

Proof: We prove by induction. In Lemma A.2, this property was proven for 1-lookahead policy. Assume this is true for $1, \dots, k-1$, then we prove for k.

By definition, a k-lookahead optimal bid (if exists) maximizes the k-lookahead utility. In mechanism $M(\epsilon, \rho, p)$, if $s_t \neq \perp$, then $x_t = 1, p_t = b_t$, and depending on the value of bid b_t, the next state s_{t+1} is either bad or a good state deterministically. Suppose for contradiciton that the k-lookahead optimal bid b_t is such that s_{t+1} is a bad state. Then, we will show that there exists a bid b'_t that achieves strictly better k-lookahead utility.

Consider the bidding strategy that bids k-lookahead optimal bid b_t at time t, $k-1$-lookahead optimal bid b_{t+1} at time $t+1$, and so on. Let $\tau \in [1,k]$ be a random variable defined as the minimum of k and the number of steps it takes to reach a good state under this strategy, when starting from the bad state s_{t+1} at time $t+1$, i.e., minimum τ such that $s_{t+\tau+1} \neq \perp$ or $\tau = k$. Now, for $i = 1, \ldots, \tau$, let A_{t+i} be the event that a Bernoulli(ρ) coin toss is a success. In $M(\epsilon, \rho, p)$ mechanism, in a bad state s_{t+i}, nothing gets added to the utility if A_{t+i} is false, and at most v_{t+i} gets added to the utility if A_{t+i} is true. Therefore, the contribution to the utility in steps $t, t+1 \ldots, t+\tau$ is upper bounded by

$$v_t + \sum_{i=1}^{\tau} v_{t+i} 1(A_{t+i}).$$

Therefore,

$$\begin{aligned} U_k^t(s_t, v_t, b_t) \quad &\leq \quad v_t + \mathrm{E}[\sum_{i=1}^{\tau} v_{t+i} I(A_{t+i}) \\ &\quad +1(\tau < k)\mathrm{E}[\sup_b U_{k-\tau-1}^{t+\tau+1}(s_{t+\tau+1}, v_{t+\tau+1}, b)|s_{t+\tau+1}]|s_t, v_t] \end{aligned} \tag{13}$$

Next, we compare the above upper bound on utility achieved by b_t to the k-lookahead utility achieved by $b_t' = (1-\epsilon)\mu$ at time t. To lower bound this k-lookahead utility, we consider the utility of the following bidding strategy starting from $b_t' = (1-\epsilon)\mu$ at time t. Let τ' be a random variable which given s_t, v_t, has the same distribution as the random variable τ defined above. Then,

- in steps t to $t + \min\{\tau', k-1\}$, bid $(1-\epsilon)\mu$
- if $\tau' = k$, bid 0 at time $t + \tau'$
- if $\tau' < k$, use $k-j-1$ lookahead optimal bid starting for time $t+j+1$ for $j = \tau', \tau'+1, \ldots, k-1$.

Then, the k-lookahead utility for bid $b_t' = (1-\epsilon)\mu$ can be lower bounded by the utility of the above strategy

$$\begin{aligned} U_k^t(s_t, v_t, b_t') \quad &\geq \quad v_t - (1-\epsilon)\mu + \mathrm{E}[\sum_{i=1}^{\tau'} v_{t+i} - (1-\epsilon)\tau' + 1(\tau' = k)(1-\epsilon)\mu \\ &\quad +1(\tau' < k)\mathrm{E}[\sup_b U_{k-\tau'-1}^{t+\tau'+1}(s_{t+\tau'+1}, v_{t+\tau'+1}, b)|s_{t+\tau'+1}]|s_t, v_t] \end{aligned} \tag{14}$$

Now, we show that the last term from (14) dominates the last term from (13). Since τ and τ' have the same distribution, in fact it suffices to compare only the expected sup utility terms for each i.

Note that by definition of τ, when $\tau = i$, the state s_{t+i+1} reached in (13) is a borderline good state, i.e., $s_{t+i+1} = ((1-\epsilon)\mu, n)$ for some n. Also, the bidding strategy used to obtain (14) is such that it doesn't leave the good state until at least time $t + \tau' + 1$. Therefore, when $\tau' = i$ the state $s_{t+i+1} \neq \perp$. Now, using Lemma A.4 for $k-i-1$, we have for all i, $s \neq \perp$,

$$\mathrm{E}[\sup_b U_{k-i-1}^{t+i+1}(s_{t+i+1}, v_{t+i+1}, b | s_{t+i+1} = s] \geq \mathrm{E}[\sup_b U_{k-i-1}^{t+i+1}(s_{t+i+1}, v_{t+i+1}, b | s_{t+i+1} = \pm] \tag{15}$$

Therefore, we derive that the last term in (14), is greater than or equal to the corresponding term in (13).

Using this observation, and subtracting (13) from (14), we can bound the total difference (denoted as Δ) in k lookahead utilities of b_t and b'_t as

$$\begin{aligned} \Delta &:= U_k^t(s_t, v_t, b'_t) - U_k^t(s_t, v_t, b_t) \\ &\geq \mathrm{E}[\sum_{i=1}^{\tau'} v_{t+i} - (\tau' + 1)(1-\epsilon)\mu + I(\tau' = k)(1-\epsilon)\mu | s_t, v_t] - \mathrm{E}[\sum_{i=1}^{\tau} v_{t+i} I(A_{t+i}) | s_t, v_t] \end{aligned}$$

Since τ and τ' have the same distribution given s_t, v_t, we can replace τ' by τ in above:

$$\Delta \geq \mathrm{E}[\sum_{i=1}^{\tau} v_{t+i} - (\tau + 1)(1-\epsilon)\mu + I(\tau = k)(1-\epsilon)\mu | s_t, v_t] - \mathrm{E}[\sum_{i=1}^{\tau} v_{t+i} I(A_{t+i}) | s_t, v_t]$$

Combining the first and last term in above, we get $\sum_{i=1}^{\tau} v_{t+i} I(\overline{A_{t+i}})$. Now, $v_{t+i} I(\overline{A_{t+i}}) - \mu(1-\rho)$, $i = 1, 2, \ldots$ form a martingale, and τ is a finite stopping time ($\tau \leq k$), therefore, by Wald's equation,

$$\mathrm{E}[\textstyle\sum_{i=1}^{\tau} v_{t+i} I(\overline{A_{t+i}}) | s_t, v_t] = \mathrm{E}[\tau | s_t, v_t] \mathrm{E}[v_{t+1} I(\overline{A}_{t+1}) | s_t, v_t] = \mathrm{E}[\tau | s_t, v_t] \mu(1-\rho)$$

In the last expression we used that A_{t+1} and v_{t+1} are independent, given s_t, v_t. Substituting, we obtain, (in below we drop the conditional on s_t, a_t for notatioal brevity)

$$\begin{aligned} \Delta &\geq \mathrm{E}[\tau]\mu(1-\rho) - (\mathrm{E}[\tau] + 1)(1-\epsilon)\mu + \Pr(\tau = k)(1-\epsilon)\mu \\ &= \mathrm{E}[\tau](\epsilon - \rho)\mu - (1-\epsilon)\mu + \Pr(\tau = k)(1-\epsilon)\mu \\ &= \mathrm{E}[\tau](\epsilon - \rho)\mu - \Pr(\tau < k)(1-\epsilon)\mu \end{aligned}$$

Now, let X be a geometric random variable with success probability ρ, then τ stochastically dominates $\min\{X, k\}$. And, from Lemma A.5

$$\begin{gathered} \mathrm{E}[\tau] \geq \mathrm{E}[\min\{X, k\}] = \tfrac{1}{\rho} \Pr(X < k) + \Pr(X \geq k), \\ \Pr(\tau < k) \leq \Pr(X < k) = 1 - (1-\rho)^{k-1} \end{gathered}$$

The proof is completed by the following algebraic manipulations:

$$\begin{aligned} \Delta &\geq \mathrm{E}[\tau](\epsilon - \rho)\mu - \Pr(\tau < k)(1-\epsilon)\mu \\ &= \frac{1}{\rho} \Pr(X < k)(\epsilon - \rho)\mu + \Pr(X \geq k)(\epsilon - \rho)\mu - \Pr(X < k)(1-\epsilon)\mu \\ &= \frac{(\epsilon - \rho)\mu}{\rho} - (1-\epsilon)\mu + \Pr(X \geq k)(-\frac{(\epsilon - \rho)\mu}{\rho} + (\epsilon - \rho)\mu + (1-\epsilon)\mu) \\ &= \frac{\epsilon\mu}{\rho} - (2-\epsilon)\mu + (1-\rho)^{k-1}\mu(2 - \rho - \frac{\epsilon}{\rho}) \end{aligned}$$

We are given that $\rho \leq \frac{\epsilon}{(1-\epsilon)}$. Consider two cases: $2 - \rho - \frac{\epsilon}{\rho} > 0$ and $2 - \rho - \frac{\epsilon}{\rho} \leq 0$. In the first case, the second term above is positive so that $\Delta_{k+1} > \frac{\epsilon\mu}{\rho} - (2-\epsilon)\mu \geq 0$, because $\rho \leq \frac{\epsilon}{(2-\epsilon)}$. In the second case, Δ_{k+1} is minimized for $k = 1$, i.e., when $\Delta_{k+1} = \Delta_2 = \frac{\epsilon\mu}{\rho} - (2-\epsilon)\mu + \mu(2 - \rho - \frac{\epsilon}{\rho}) = (\epsilon - \rho)\mu > 0$.

This proves that $U_k^t(s_t, v_t, b'_t) - U_k^t(s_t, v_t, b_t) = \Delta > 0$ when $\rho \leq \frac{\epsilon}{(2-\epsilon)}$, proving a contradiction that b_t is not k-lookahead optimal. Thus, the k-lookahead optimal bid if exists will ensure that $s_{t+1} \neq \perp$.

In fact, by induction optimal $k-1$-lookahead bid exists, so that the optimal k-lookahead bid for any t such that $s_t \neq \perp$ is given by:

$$b_t := \arg \max_{b : (\bar{b}n + b)/(n+1) \geq (1-\epsilon)\mu} \mathrm{E}[v_t - b + \max_{b'} U_{k-1}^{t+1}(s_{t+1}, v_{t+1}, b') | s_t, v_t],$$

which by applying this lemma for $k-1, k-2, \ldots$ can be derived to be the minimum bid that would keep s_{t+1} as a good state. ■

Lemma A.4 *Under mechanism $M(\epsilon, \rho, p)$ with $\rho \leq \frac{\epsilon}{2-\epsilon}$, an optimal k-lookahead bid b_t at time t, when starting from any good state $s_t = s \neq \perp$, would achieve at least as much utility as when starting from a borderline state $s_t = s' = ((1-\epsilon)\mu, n)$. That is,*

$$U_k^t(s, v_t, b_t) \geq U_k^t(s', v_t, b_t), \forall s \neq \perp, s' = ((1-\epsilon)\mu$$

Proof: Consider the case when the starting state is a borderline state $s' = ((1-\epsilon)\mu, n)$. Opening up the recursive definition of k-lookahead utility, we obtain the following expression in terms of bids $b_{t+1}, \ldots, b_{t+k-1}$ which are optimal $k-1, k-2, \ldots, 1$ lookahead bids respectively.

$$U_k^t(s', v_t, b_t) \quad := \quad \mathrm{E}[\sum_{\tau=t}^{t+k} v_\tau x(s_\tau, b_\tau) - p(s_\tau, b_\tau, x_\tau) | s_t = s', v_t]$$

Using Lemma A.3, the optimal k-lookahead bid for any $k \geq 1$ is such that the next state is a good state, so that if the starting state s' is a good state, then so are the states $s_\tau, \tau = t+1, \ldots, t+k$ in the above expression. This further implies that if the starting state is a borderline state $s' = (1-\epsilon)\mu, n$, then the sum of bids $b_t, b_{t+1}, \ldots, b_{t+k-1}$ must be atleast $(1-\epsilon)\mu k$. Since in good state, the allocation is always 1 and the payment is equal to the bid, we obtain the following upper bound on the utility:

$$\begin{aligned} U_k^t(s', v_t, b_t) \quad &= \quad \mathrm{E}[\sum_{\tau=t}^{t+k} v_\tau - b_\tau | s_t, v_t] \\ &\leq \quad v_t + \mathrm{E}[\sum_{i=1}^{k} v_{t+i}] - k(1-\epsilon)\mu. \end{aligned}$$

Now, on starting from another good state, say $s = (\bar{b}, n') \neq \perp$, since $\bar{b} \geq (1-\epsilon)\mu$, the sum of bids $b_t, b_{t+1}, \ldots, b_{t+k-1}$ needs to be *at most* $(1-\epsilon)\mu$ to remain in a good state, and $b_{t+k} = 0$ as the optimal myopic bid (for good state) will be used in this last step. Therefore, for any $s \neq \perp$,

$$U_k^t(s, v_t, b_t) \geq v_t + \mathrm{E}[\sum_{i=1}^{k} v_{t+i}] - k(1-\epsilon)\mu \geq U_k^t(s', v_t, b_t)$$

■

Lemma A.5 *Let X be a geometric random variable with success probability ρ, then*

$$E[\min\{X, k\}] = \frac{1}{\rho} \Pr(X < k) + \Pr(X \geq k)$$

Proof:

$$
\begin{aligned}
\mathrm{E}[\min\{X,k\}] &= \mathrm{E}[XI(X<k)] + \Pr(X \geq k)k \\
&= \sum_{j=1}^{k-1}(1-\rho)^{j-1}\rho j + k(1-\rho)^{k-1} \\
&= \mathrm{E}[X] - \sum_{j=k}^{\infty}(1-\rho)^{j-1}\rho j + k(1-\rho)^{k-1} \\
&= \mathrm{E}[X] - (1-\rho)^{k-1}\sum_{j=1}^{\infty}(1-\rho)^{j-1}\rho(j+k-1) + k(1-\rho)^{k-1} \\
&= \mathrm{E}[X] - (1-\rho)^{k-1}\mathrm{E}[X] - (1-\rho)^{k-1}\sum_{j=1}^{\infty}(1-\rho)^{j-1}\rho(k-1) + k(1-\rho)^{k-1} \\
&= \frac{1}{\rho}(1-(1-\rho)^{k-1}) - (1-\rho)^{k-1}(k-1) + k(1-\rho)^{k-1} \\
&= \frac{1}{\rho}(1-(1-\rho)^{k-1}) + (1-\rho)^{k-1} \\
&= \frac{1}{\rho}\Pr(X<k) + \Pr(X \geq k)
\end{aligned}
$$

■

Lemma A.6 (Claim (e) of Theorem 4.3) *Against a buyer who is no-regret learner for the class $\mathcal{E}$ of experts (refer to Equation* (8) *and* (9)*), the mechanism $M(\epsilon,\rho,p)$ achieves a revenue of atleast*

$$\mathrm{Rev}_{LS}^{M(\epsilon,\rho,p^*)} \geq \frac{\rho}{\rho+1}\mathrm{Rev}^{\mathrm{Mye}} - o(1).$$

Proof: Let us first consider trajectories of states of form $\not\perp \perp\perp\perp\perp\perp \not\perp \perp\perp\perp \not\perp \perp\perp \ldots$, i.e., sequence of bad states interspersed with *single* good states.

Consider the bidding function $f(s_t,v_t)$ such that $f(s_t,v_t) = p^*$ when $s_t = \perp$ and $v_t \geq p^*$, and 0 otherwise. This is (arbitrarily close to) one of the experts in the class $\mathcal{E}$ of experts that the buyer is using. In the mechanism $M(\epsilon,\rho,p)$, with probability ρ, the bid p^* made in a bad state will get accepted, to earn utility $u_t(f(s_t,v_t)) = \rho(v_t - p^*)$ for the buyer. Therefore, for any sequence of states and valuations, the first term in the regret definition (9) is at least

$$\sum_{t=1}^{T} u_t(f(s_t,v_t)) \geq \rho \sum_{t:s_t=\perp, v_t \geq p^*}(v_t - p^*) + \sum_{t:s_t \neq \perp} v_t$$

Since the buyer is using a no-regret learning algorithm, she must be achieving a utility that is within $o(T)$ of the above utility. Now, the maximum utility in good states is v_t, and in mechanism $M(\epsilon,\rho,p)$, the buyer cannot not make any positive utility in bad states where $v_t \leq p^*$. Therefore, the buyer can afford to lose at most $o(T)$ of the bad state auctions where $v_t \geq p^*$. This means that any no-regret learning buyer must bid $b_t \geq p^*$ in all except $o(T)$ of the time steps with $s_t \in \perp, v_t \geq p^*$. Let B be the number of bad states, and B' be the number of those bad states where $v_t \geq p^*$ and the bidder bids at least p^*, then,

$$\mathrm{E}[B'] \geq \mathrm{E}[B](1 - F(p^*)) - o(T)$$

Also, let G be the number of good states. Then, due to the construction of the mechanism $M(\epsilon, \rho, p)$, the buyer could have obtained $G-1$ good states (all except the first good state), only by winning $G-1$ bad state auctions. And, since bad state auctions can be won only with probability ρ, we have that $\mathrm{E}[G] \leq \mathrm{E}[B]\rho + 1$. This gives,

$$\mathrm{E}[B] \geq \frac{1}{\rho+1}(T-1),$$

Combining above observations

$$\mathrm{E}[B'] \geq \mathrm{E}[B](1-F(p^*)) - o(T) \geq \frac{1}{\rho+1}(1-F(p^*))(T-1) - o(T).$$

Therefore, the total expected revenue of seller is atleast

$$\rho p^* \mathrm{E}[B'] \geq \frac{\rho}{\rho+1} p^*(1-F(p^*))(T-1) - o(T) = \frac{\rho}{\rho+1}\mathrm{Rev}^{\mathrm{Mye}} T - o(T)$$

Now, consider sequences of states with more than one consecutive good states, e.g., $\not\perp$ $\perp\perp\perp\perp$ $\not\perp\not\perp\not\perp$ $\perp\perp\perp\perp$ $\not\perp\not\perp$ ... etc. Then, in a subsequence of consecutive good states the bid average over all good states except the first one (call them trailing states) must be at least $(1-\epsilon)\mu$, so that the buyer makes at most $\bar{v} - (1-\epsilon)\mu$ utility, where $\bar{v}$ denotes the average valuation over those states. On the other hand, the above expert f, which bids 0 in good states, makes $\bar{v}$ utility in each of those trailing states (in hindsight). Further, in the bad states, and in rest of the (non-trailing) good states, f is achieving the best possible utility. Therefore, given the no-regret condition, the number of trailing states can be at most $o(T)$ and do not effect the revenue calculations above. ■

Lemma A.7 (Claim (f) of Theorem 4.3) *Against a buyer who is a policy regret learner for a class $\mathcal{C}$ of sequences containing all sequences of single experts (refer to Equation* (10)*), the mechanism $M(\epsilon, \rho, p)$ achieves a revenue of at least*

$$\mathrm{Rev}_{LP}^{M(\epsilon,\rho,p^*)} \geq (1-\epsilon)\mu - o(1).$$

Proof: The sequence of constant bid $(1-\epsilon)\mu$ would keep a buyer always in good state $(1-\epsilon)\mu$, and achieve a utility of $\epsilon\mu$. Therefore, if the class $\mathcal{C}_T$ contains this sequence, the policy-regret learning buyer must achieve $\epsilon\mu - o(T)$ utility. Now, bad states can achieve at most $\rho\mu$ utility on average, therefore, of $\rho < \epsilon - o(1)$, this implies that the bad states can be at most $o(T)$. This implies that the trailing good states are atleast $T - o(T)$. Since the buyer needs to maintain a bidding average of at least $(1-\epsilon)$ over the trailing good states, this gives revenue of atleast $(1-\epsilon)\mu T - o(T)$. ■

B Proof of Theorem 4.5

Proof: We will take our cumulative density function F to be the exponential distribution of rate 1, conditioned on being less than 1. In particular, F is supported on $[0,1]$ and defined as follows $F(x) = \frac{e}{e-1} \cdot (1-e^{-x}), x \in [0,1]$. It is easy to see that $\mathrm{Rev}^{\mathrm{Mye}} = \frac{e^{W(1)}+W(1)-2}{e-1} \approx 0.19$ (where $W(\cdot)$ is the product-log function), $\mu = \frac{e-2}{e-1} \approx 0.42$, and that F is a monotone hazard rate distribution.

Now let $M = (\mathcal{S}, Q, x, p, s_1)$ be a mechanism as defined in Section 2. Consider a myopic buyer being at state $s_t \in \mathcal{S}$ of the mechanism at time t. Given his realized value $v_t \sim F$ and facing the (randomized) allocation rule $x(s_t, \cdot)$ and price rule $p(s_t, \cdot, \cdot)$ of the mechanism in state s_t, he would map his value v_t to some bid b_t to maximize his expected utility $\mathrm{E}_{x \sim x(s_t, b_t)}[x] \cdot v_t - \mathrm{E}_{x \sim x(s_t, b_t)}[p(s_t, b_t, x)]$.

For any state $s \in \mathcal{S}$, let us denote by $b_s : \mathbb{R} \to \Delta^{\mathbb{R}^+}$ the (potentially randomized) mapping from realized value to bid under which the revenue of mechanism M against a myopic buyer is at least $\epsilon \cdot \mathrm{Rev}^{\mathrm{Mye}}$. Note that this mapping only depends on s for a myopic buyer. We do not need to require that for all s, v, $b_s(v)$ is optimal. The only assumption that we will make is that, for all s, v, v', the distribution over bids $b_s(v)$ does not result in worse utility for a buyer with value v compared to the distribution $b_s(v')$. Given this definition, let us also define the effective allocation probability and effective price functions, $\hat{x} : \mathcal{S} \times \mathbb{R} \to [0, 1]$ and $\hat{p} : \mathcal{S} \times \mathbb{R} \to \mathbb{R}$ respectively, as follows:

$$\forall s, v : \hat{x}(s, v) = \mathrm{E}_{b \sim b_s(v), x \sim x(s,b)}[x] \quad \text{and} \quad \hat{p}(s, v) = \mathrm{E}_{b \sim b_s(v), x \sim x(s,b)}[p(s, b, x)].$$

Via standard argumentation, for all $s \in \mathcal{S}$, $\hat{x}(s, \cdot)$ and $\hat{p}(s, \cdot)$ satisfy the incentive compatibility constraint that:

$$\forall v, v' : \hat{x}(s, v) \cdot v - \hat{p}(s, v) \geq \hat{x}(s, v') \cdot v - \hat{p}(s, v').$$

Moreover, given that M is non-payment forceful, for all $s \in \mathcal{S}$, we get that

$$\hat{p}(s, 0) = 0.$$

Using Myerson's payment identity, it is standard to argue that any mechanism $(\hat{x}(s, \cdot), \hat{p}(s, \cdot))$ satisfying the above constraints can be implemented as a distribution over take-it-or-leave-it offers of the item at different prices. That is, there exists a distribution G_s over prices such that the expected revenue and expected buyer utility resulting from $(\hat{x}(s, \cdot), \hat{p}(s, \cdot))$ can be written as:

$$\mathrm{Rev}^s_{\mathrm{myop}} = \mathrm{E}_{v \sim F}[\hat{p}(s, v)] \equiv \mathrm{E}_{v \sim F, p \sim G_s}[p \cdot 1_{v \geq p}] \stackrel{*}{=} \frac{e}{e-1} \mathrm{E}_{p \sim G_s}[p \cdot e^{-p}];$$
$$\mathrm{Ut}^s_{\mathrm{myop}} = \mathrm{E}_{v \sim F}[\hat{x}(s, v) \cdot v - \hat{p}(s, v)] \equiv \mathrm{E}_{v \sim F, p \sim G_s}[(v - p) \cdot 1_{v \geq p}] \stackrel{*}{=} \frac{e}{e-1} \mathrm{E}_{p \sim G_s}[e^{-p}].$$

(In the above, the equalities $\stackrel{*}{=}$ follow by plugging in for F the distribution defined above.) So, in particular, since it can be assumed without loss of generality that G_s is supported on $[0, 1]$, it follows that $\mathrm{Ut}^s_{\mathrm{myop}} \geq \mathrm{Rev}^s_{\mathrm{myop}}$. To summarize, in any state $s \in \mathcal{S}$, a myopic buyer makes at least as high a utility as he pays a payment.

Now consider a mechanism that has expected per round revenue against a myopic buyer at least $\epsilon \cdot \mathrm{Rev}^{\mathrm{Mye}}$. It follows from the above derivation that it should also then give expected per round utility at least $\epsilon \cdot \mathrm{Rev}^{\mathrm{Mye}}$ to a myopic buyer. As an infinite look-ahead buyer (aiming to maximize his utility) can certainly pretend to be myopic, this means that the mechanism must give expected per round utility at least $\epsilon \cdot \mathrm{Rev}^{\mathrm{Mye}}$ to an infinite look-ahead buyer. Hence, the mechanism can make at most $\mu - \epsilon \cdot \mathrm{Rev}^{\mathrm{Mye}} \geq (1 - \frac{\epsilon}{2}) \cdot \mu$ revenue from an infinite look-ahead buyer. ■