

MIT Open Access Articles

Robust Repeated Auctions under Heterogeneous Buyer Behavior

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Agrawal, Shipra, Daskalakis, Constantinos, Mirrokni, Vahab S. and Sivan, Balasubramanian. 2018. "Robust Repeated Auctions under Heterogeneous Buyer Behavior."

As Published: 10.1145/3219166.3219234

Publisher: Association for Computing Machinery (ACM)

Persistent URL: <https://hdl.handle.net/1721.1/137491>

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Terms of use: Creative Commons Attribution-Noncommercial-Share Alike



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. Traditionally, the design of repeated auctions requires strong modeling assumptions about the bidder behavior, such as it being myopic, infinite lookahead, or some specific form of learning behavior. Is it possible to design mechanisms which are simultaneously optimal against a multitude of possible buyer behaviors? We answer this question by designing a simple state-based mechanism that is simultaneously approximately 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. Against each type of buyer our mechanism attains a constant fraction of the optimal revenue attainable against that type of buyer. We complement our positive results with almost tight impossibility results, showing that the revenue approximation tradeoffs achieved by our mechanism for different lookahead attitudes are near-optimal.

*IEOR Department, Columbia University. sa3305@columbia.edu.

[†]EECS, Massachusetts Institute of Technology. costis@csail.mit.edu.

[‡]Google Research, New York. mirrokni@google.com.

[§]Google Research, New York. 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 future rounds, save some small ϵ , upfront for the promise of getting the item for free in all future rounds. A risk-neutral buyer would have no choice but to accept this offer to get expected utility of ϵ 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 [Ashlagi et al. \[2016\]](#), [Mirrokni et al. \[2016a\]](#), [Balseiro et al. \[2017\]](#), [Mirrokni et al. \[2017\]](#) 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 (IIR)*, 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 consequences 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 the 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 [Kantor \[2016\]](#).

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 the 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 approximately optimal revenue simultaneously against buyers with different lookahead and learning behaviors, i.e., against each type of buyer, obtains a constant fraction of the optimal revenue achievable by mechanisms tailored to that specific type of buyer.

Our setting and main results. We study a repeated interaction between a single seller and a single buyer over multiple rounds. At the beginning of each round $t = 1, 2, \dots, T$, there is a single fresh good for sale whose private value $v_t \in V$ for the buyer is drawn from a publicly known distribution¹ 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 the ‘learning behavior’ of the buyer. To this end, we introduce a novel framework for characterizing revenue tradeoffs of dynamic mechanisms under heterogeneous buyer behavior. We formalize variations in forward-planning attitude (or lack thereof) of the buyer by considering a range of lookahead levels: from myopic to limited $k \geq 1$ lookahead to infinite lookahead. We model learning behavior of the buyer 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. Formal definitions of these different buyer behavior are provided in Section 2. We characterize a mechanism as robust if it simultaneously achieves near-optimal revenue for different buyer behaviors.

Definition 1 *We call a mechanism (α, β) -robust against lookahead attitudes if, for any per-round valuation distribution F , it simultaneously achieves an expected average (over T rounds) revenue of at least*

- $\alpha \cdot \text{Rev}^{\text{Mye}} - o(1)$ against any myopic buyer, and
- $\beta \cdot \mu - o(1)$ against any k -lookahead buyer with $k \geq 1$,

where $\text{Rev}^{\text{Mye}} = \max_x x(1 - F(x))$, $\mu = E_{x \sim F}[x]$, and $o(1)$ are functions that go to 0 with T .

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

Definition 2 We call a mechanism (α, β) -robust against learning behaviors if, for any per-round valuation distribution F , it simultaneously achieves an expected average (over T rounds) revenue of at least

- $\alpha \cdot \text{Rev}^{\text{Mye}} - o(1)$ against any no-simple-regret learner, and
- $\beta \cdot \mu - o(1)$ against any no-policy-regret learner,

where Myerson revenue Rev^{Mye} and mean μ are defined as in the previous definition, and $o(1)$ are functions that go to 0 with T .

Our main contribution is a simple mechanism that can be tuned robust against different lookahead attitudes, as well as different learning behaviors. Our base mechanism will additionally satisfy interim individual rationality (IIR) and non-payment forcefulness, as defined in Section 2.4. We will subsequently modify our base mechanism to improve its individual rationality properties.

Theorem 1.1 For any $\epsilon \in (0, 1)$, there exists a non-payment forceful and IIR mechanism that is $(\frac{\epsilon}{2}, 1 - \epsilon)$ -robust against lookahead attitudes.

Theorem 1.2 For any $\epsilon \in (0, 1)$, there exists a non-payment forceful and IIR mechanism that is $(\frac{\epsilon}{1+\epsilon}, 1 - \epsilon)$ -robust against learning behaviors.

Note that our robust mechanism navigates a tradeoff between revenue achieved for different types of buyers, i.e., a tradeoff between α and β . Upfront, it is not obvious if α must be decreased to increase β . We prove an impossibility result showing that any dynamic mechanism must face such a tradeoff. Further, our impossibility result provides a quantitative limitation on the revenue tradeoff achievable, demonstrating that our mechanism achieves the best tradeoff possible within a constant factor.

Theorem 1.3 There exists a regular and decreasing hazard rate distribution F such that, for all $\epsilon \in [0, 1]$ and $\delta > 1 - \frac{\epsilon}{2}$, there is no mechanism that is non-payment forceful and (ϵ, δ) -robust against lookahead attitudes.

The proofs of Theorems 1.1–1.2 are provided in Section 4 and Appendix B, while the proof of Theorem 1.3 is provided solely in Appendix B.

Injecting Ex-Post Individual Rationality The base mechanism provided thus far is robust with respect to different buyer behaviors, but only satisfies Interim Individual Rationality. Is it possible to strengthen this mechanism to also satisfy the stringent notion of per-round ex-post Individual Rationality, discussed earlier in the introduction and defined formally in Section 2.4? As it turns out, per-round ex-post IR is too strong a requirement when it comes to extracting revenue that is close to full surplus from forward-looking buyers with limited lookahead. Specifically, we show that there exist valuation distributions for which the revenue achieved by any per round ex-post IR mechanism is exponentially smaller compared to full surplus for finite lookahead buyers.

Theorem 1.4 There exists a valuation distribution F such that any non-payment forceful and per-round ex-post IR mechanism can achieve expected average (over rounds) revenue of at most $\log(k\mu) + 1$ from a k -lookahead buyer with any finite $k \geq 1$.

In view of this limitation for finite lookahead buyers, we turn our attention to myopic buyers and infinite lookahead buyers, providing a modification to our base mechanism that guarantees per-round ex-post IR for these types of buyers. To be precise, for infinite lookahead buyers our mechanism guarantees per-round ex-post IR with high probability, i.e., there exists a strategy for the buyer that with high probability is both optimal and satisfies the per-round ex-post IR conditions. We provide the modification to our mechanism and establish its properties in Appendix A, establishing the following:

Theorem 1.5 *For any $\epsilon \in (0, 1)$, there exists a non-payment forceful and IIR mechanism that is $(\frac{\epsilon}{2}, 1 - \epsilon)$ -robust against lookahead attitudes, ex-post IR against a myopic buyer, and with a high probability ex-post IR against an infinite lookahead buyer.*

The per-round ex-post IR property, even if only guaranteed with high probability for infinite lookahead buyers, is a strong requirement that successfully eliminates advanced selling or threat-based mechanisms. For example, let us consider the advanced selling mechanism that in every round asks the buyer to pre-pay for tomorrow’s good a price equal to its expected value, for the guarantee that this good will be allocated tomorrow. This mechanism gets expected average per round revenue equal to the mean of the distribution, but it violates high probability ex-post IR, since a forward-looking buyer will choose to pre-pay for tomorrow’s good in every round. This results in realized utility of $v - \mu$, which may be negative with constant probability.

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 $t = 1, 2, \dots, T$, 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 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 $u_t(b_t) = v_t x_t - p_t$, and the seller’s revenue is given by p_t .

Dynamic state-based mechanism. We consider dynamic mechanisms, where the allocation and payment (x_t, p_t) , in a round t , may depend on the current state s_t , in addition to the bid b_t made by the buyer. Such a dynamic 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. $x : \mathcal{S} \times \mathbb{R}^+ \rightarrow \Delta^{\{0,1\}}$ 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. $p : \mathcal{S} \times \mathbb{R}^+ \times \{0, 1\} \rightarrow \mathbb{R}$ 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., $p_t = p(s_t, b_t, x_t)$.
4. $Q : \mathcal{S} \times \mathbb{R}^+ \times \{0, 1\} \times \mathbb{R} \rightarrow \Delta^{\mathcal{S}}$ 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., $s_{t+1} \sim Q(s_t, b_t, x_t, p_t)$.²
5. $s_1 \in \mathcal{S}$ is the starting state.

Bidding strategy. We consider multiple types of buyer behaviors, namely, different lookahead attitudes and learning behaviors. Therefore, the buyer’s ‘optimal’ bidding strategy is not necessarily the one that maximizes linear utility $u_t(b) = v_t x_t - p_t$ in round t . Instead the choice of bids $\{b_t\}$ is determined by the behavioral setting, and we refer to them as ‘behaviorally optimal’ choice of bids.

2.2 Lookahead behaviors

ℓ -lookahead utility. We will define lookahead attitudes of buyers using the concept of ℓ -lookahead utility, i.e., the total utility over the current and next ℓ rounds. Assuming there are at least ℓ remaining rounds following round t , the buyer evaluates a bid b in round t by computing its 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:

$$U_\ell^t(s_t, v_t, b) := \mathbb{E}[u_t(b) + \sup_{b_{t+1}, \dots, b_{t+\ell}} \sum_{j=1}^{\ell} u_{t+j}(b_{t+j})] \quad (1)$$

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

$$U_0^t(s_t, v_t, b) := \mathbb{E} \left[v_t x_t(b) - p_t(b) \middle| s_t, v_t \right], \quad (3)$$

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.

²Given 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 .

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

- *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., a behaviorally-optimal bid for such a buyer in round t is given by

$$b_t \in \arg \max_{b \in \mathbb{R}^+} U_{\min\{k, T-t\}}^t(s_t, v_t, b) \quad (4)$$

We refer to the bid b_t computed above as a ***k*-lookahead optimal bid** (Note that in general, the maximizer in the above equation may not exist, see the technical remark below). Two special cases of *k*-lookahead 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 $k = T - 1$ as an *infinite-lookahead buyer*.
- Forward-looking buyer: A *forward looking buyer* maximizes k_t -lookahead utility for some $k_t \in \{1, \dots, T-t\}$, 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.

We say that a buyer is using a **behaviorally-optimal policy** if the buyer’s bid b_t satisfies (4) for $k = k_t$ at all time steps t .

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.

2.3 Learning behaviors

Background on 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 $t = 1, \dots, T$, an adversary picks reward $\mathbf{g}_t = \{g_{1,t}, \dots, g_{N,t}\}$ where $g_{i,t}$ is the reward associated with expert i . The learner needs to pick an expert $i_t \in [N]$ to obtain reward $g_{i_t,t}$. Regret in time T is defined as

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

A *no-regret online learning algorithm* for this problem uses the past observations $g_{i_1,1}, \dots, g_{i_{t-1},t-1}$ to make the decision i_t in every round t such that $\text{Regret}(T) \leq o(T)$. 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 ‘regret’ compares the total reward achieved by the learner to 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, \dots, i_t), in this simple-regret framework, the performance of the best expert is evaluated over the sequence of inputs $\mathbf{g}_1, \dots, \mathbf{g}_T$ 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.

Background on policy regret learning. A no-policy-regret learning algorithm is a more sophisticated learner based on the definition of policy regret from Arora et al. [2012]. Such an algorithm faces an adaptive adversary, and achieves $o(T)$ *policy regret*, defined as:

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

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, so that

$$\text{Policy-regret}(T) = \max_i \sum_{t=1}^T g_{i, t}(i, \dots, i) - \sum_{t=1}^T g_{i_t, t}(i_1, \dots, i_t). \quad (7)$$

Our characterization of 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

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

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

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

For a no-regret learning buyer (refer to (5)), a *behaviorally-optimal policy* is any bidding strategy such that for the trajectories of bids, states and valuations $s_1, v_1, b_1, \dots, s_T, v_T, b_T$, generated by this policy, we have in hindsight,

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

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 $f(\mathbb{1}(s_t = \perp), v_t)$, that is, it only uses whether $s_t = \perp$ or $s_t \neq \perp$.

- *No-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, \dots, s'_T is the (possibly randomized) trajectory of states that would be observed on using the expert to decide the bids in all rounds. For a no-policy-regret learning buyer, a behaviorally optimal policy is any bidding strategy such that for resulting trajectory of bids and states $s_1, b_1, \dots, s_T, b_T$, 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 Desirable properties of a mechanism

Non-payment forceful. A mechanism is *non-payment forceful* if the payment p_t is always non-negative, and 0 when the bid is 0, i.e.,

$$p(s_t, b_t, x_t) \geq 0, \text{ and } p(s_t, 0, x_t) = 0, \text{ for any } s_t, b_t, x_t. \quad (11)$$

Such a mechanism has the desirable property that it cannot be forced to pay the buyer in any state and it cannot force payments out of buyers who bid 0.

Interim Individually Rational (IIR) A mechanism is defined to be *interim individually rational (IIR)* iff at any time t , given any history and current valuation v_t , there exists a bid b_t such that the buyer's expected perceived utility is non-negative. For a k -lookahead buyer, this means that given any state s_t , valuation v_t , for any bid b_t that is k -lookahead optimal at time t , we have

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

Note that since the buyer's utility is a non-negative aggregate (possibly over multiple future time steps) of per time-step buyer utility, the non-payment-forceful condition guarantees IIR.

An interim-individually rational mechanism allows take-it-or-leave-it offer based mechanisms like the following: "pay $E[v]$ today to get the item tomorrow." A forward-looking buyer would find this offer attractive, hence the seller would extract the full surplus, $E[v]$. An unsettling feature of the afore-described mechanism is that, for some realizations of v , the buyer ends up with negative utility. In particular, while this mechanism is interim Individually Rational (IR), it is not 'ex-post IR'.

Ex-post IR This requires that the total utility of a rational buyer at the end of T rounds is non-negative in hindsight. That is, there exists some behaviorally optimal policy, such that all trajectories of bids $b_t, t = 1, \dots, T$ (and implicitly the trajectories of states and valuations $s_1, \dots, s_T, v_1, \dots, v_T$) generated by this policy satisfy

$$\sum_{t=1}^T u_t(b_t) \geq 0.$$

Here, $u_t(b_t) = \mathbb{E}_{x_t \sim x(s_t, b_t)}[x_t \cdot v_t - p_t(s_t, b_t, x_t) | v_t, s_t, b_t]$ is the round t utility defined in the previous section as the buyer’s value from that round’s allocation minus the payment.

Per-round Ex-post IR We also consider a stronger version of ex-post IR considered in some previous works [Ashlagi et al. \[2016\]](#), [Mirrokni et al. \[2016b\]](#) : at each round, the utility of the buyer, defined to be his value from that period’s allocation minus the buyer’s payment, must be non-negative. More precisely, we define a mechanism to be *per round ex-post individually rational* if there exists some behaviorally optimal policy such that every trajectory of bids b_1, \dots, b_T (and implicitly states s_1, \dots, s_T and valuations v_1, \dots, v_T) generated by this policy satisfies

$$\forall t, u_t(b_t) \geq 0. \tag{12}$$

Per-round ex-post IR is a stronger requirement than ex-post IR, and clearly implies ex-post IR. It is also referred to as stage-wise ex-post IR [Mirrokni et al. \[2016b\]](#).

3 Our Mechanism

We prove our positive results by proposing a simple state-based mechanism $M(\epsilon, \rho, p)$ parameterized by three parameters p (threshold price), ϵ , and ρ . A high level description of the mechanism is as follows. The mechanism uses roughly the average of buyer’s past accepted bids to decide the buyer’s state. The mechanism can be in two types of states: ‘good state’ if the current average is above $(1 - \epsilon)\mu$, and ‘bad state’ otherwise. In a good state, the mechanism always accepts the buyer’s bid (irrespective of the bid value), with payment equal to bid. The state transitions from a ‘good state’ to a ‘bad state’ if the average of accepted bids falls below $(1 - \epsilon)\mu$. In a bad state, if the buyer’s bid is above the threshold p , then with probability ρ , the bid is accepted and the buyer is transferred to a good state. Any bid below p is rejected in a bad state.

Below are the precise definitions.

Definition 3 (Mechanism $M(\epsilon, \rho, p)$) • **State Space \mathcal{S} :** *The state space is $\mathcal{S} = \mathbb{R} \times \mathbb{N}$. We represent a state $s \in \mathcal{S}$ by pair $(\bar{b}, n) \in \mathbb{R} \times \mathbb{N}$, where \bar{b} represents an average of n past bids. 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 bad state we say $s = \perp$, otherwise, we say $s \neq \perp$. Further, we refer to any state of form $s_t = ((1 - \epsilon)\mu, 0)$ as a ‘borderline’ good state.

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

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

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

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

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

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

Note that the payment is always smaller than bid, with 0 payment no allocation. By definition, this mechanism is non-payment forceful in all states.

- **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 $s_t = (\bar{b}, n)$. In this mechanism, the state effectively remains the same if $x_t = 0$. Otherwise, it transitions either (from good state) to a state with updated average bid, or (from bad state) to a borderline state.

$$s_{t+1} = \begin{cases} (\bar{b}, n), & \text{if } x_t = 0 \\ \left(\frac{\bar{b} + b_t}{n+1}, n+1\right), & \text{if } s_t \neq \perp, x_t = 1 \\ ((1-\epsilon)\mu, 0), & \text{if } s_t = \perp, x_t = 1, \end{cases}$$

4 Revenue tradeoffs

In this section, we prove revenue guarantees for mechanism $M(\epsilon, \rho, p)$ against buyers with different lookahead attitudes and learning behaviors. Specifically, we demonstrate that with appropriate parameter settings, this mechanism achieves the results stated in Theorem 1.1 and Theorem 1.2.

In below, $p^* := \arg \max_p p(1 - F(p))$ and $\text{Rev}^{\text{Mye}} := p^*(1 - F(p^*))$ denote the Myerson price and Myerson optimal revenue respectively. And, $\text{Rev}^{M(\epsilon, \rho, p)}$ denotes the expected value of average revenue of mechanism $M(\epsilon, \rho, p)$ over T rounds, i.e., $\text{Rev}^{M(\epsilon, \rho, p)} := \frac{1}{T} \sum_{t=1}^T p_t$. Order notation $o(1)$ and $o(T)$ will be used denote asymptotic order with respect to T .

4.1 Revenue against diverse lookahead attitudes

Theorem 1.1 can be obtained as a corollary of the following proposition by substituting $\rho = \frac{\epsilon}{2-\epsilon}$.

Proposition 4.1 *The mechanism $M(\epsilon, \rho, p)$ with parameters $p = p^*, \rho \leq \frac{\epsilon}{2-\epsilon}$ achieves*

(a) *revenue $\text{Rev}^{M(\epsilon, \rho, p)} \geq \frac{\rho}{\rho+1} \text{Rev}^{\text{Mye}} - o(1)$ against myopic buyers, while achieving*

(b) *revenue $\text{Rev}^{M(\epsilon, \rho, p)} \geq (1-\epsilon)\mu - o(1)$ against k -lookahead buyers for any $k \geq 1$.*

Proof Sketch: Here, we provide a proof sketch. A detailed proof is provided in Appendix B.

(a) *Revenue against myopic buyers.* Let us first consider the seller's expected revenue from a myopic buyer over T rounds. By definition, the mechanism starts in the borderline good state

$s_0 = ((1 - \epsilon)\mu, 0) \neq \perp$. Since a myopic buyer is only concerned with immediate utility, she will bid 0 in this state to get allocation with maximum possible utility. However, this will take the average of bids below the boundary of $(1 - \epsilon)\mu$ and the bidder will immediately go to the bad state. In a bad state, such a buyer will bid p^* whenever $v_t \geq p^*$. Therefore, the buyer will return back to the borderline good state s_0 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 buyer will roughly spend $\ell = \frac{1}{\rho(1-F(p^*))}$ time steps in bad state for every visit to a good state. The expected number of such good state-bad state visit cycles is roughly $T/(\ell + 1)$, with seller's revenue of p^* (from bad states) in every cycle. This gives the expected revenue in T rounds as roughly $\frac{Tp^*}{\ell+1} \geq \frac{\rho}{\rho+1}p^*(1 - F(p^*)) = \alpha \text{Rev}^{\text{Mye}}$. An extra $-\frac{1}{T} = -o(1)$ term is obtained due to possible interruption of the last cycle.

(b) *Revenue against k -lookahead buyers.* For k -lookahead buyers for any $k \geq 1$, the key is to demonstrate that in mechanism $M(\epsilon, \rho, p)$ with $\rho < \frac{\epsilon}{2}$, an optimal k -lookahead bid in a good state guarantees that the next state is also a good state. This is proven in Lemma 4.2. The bound on revenue can then be obtained by observing that starting in a good state, a k -lookahead buyer will remain continuously in good states. And, since the bidder always gets charged her bid in a good state, this implies the seller's average revenue is \bar{b}_T which by definition of good states is at least $\mu(1 - \epsilon)$. Here \bar{b}_T denotes the average of bids b_1, \dots, b_T . ■

The following lemma forms the key to main technical results in this paper.

Lemma 4.2 *Assume $\rho \leq \frac{\epsilon}{2-\epsilon}$. Then, at any time t , given that $s_t \neq \perp$, and b_t is a k -lookahead optimal bid, we have that $s_{t+1} \neq \perp$.*

Proof Sketch: To develop an intuition, let us first consider a simpler version of our mechanism where at every step in the bad state, there is an independent probability ρ of getting an allocation and transferring back to the good state, irrespective of the bid. (Note that this mechanism, that does not charge anything in bad states, will result in 0 seller's revenue against myopic buyers).

Assume that current state s_t is a good state. Suppose for contradiction that b is an optimal bid at time t such that $s_{t+1} = \perp$. Let us compare the k -lookahead utility of this bid to a bid $b' \leq (1 - \epsilon)\mu$ that keeps the bidder in a good state s_{t+1} (note that such a bid always exists). The comparison for 1-lookahead buyer is relatively simple. In both cases, such a buyer will plan to bid 0 in the next round (since next round is the last round in 1-lookahead buyer's planning horizon). When $s_{t+1} \neq \perp$, $x_{t+1} = 1$, and therefore, 1-lookahead utility for b' is at least

$$v_t - (1 - \epsilon)\mu_t + \mathbb{E}[v_{t+1}] = v_t + \epsilon\mu.$$

In comparison, when $s_{t+1} = \perp$, the probability of allocation is ρ , and therefore 1-lookahead utility for b is at most

$$v_t + \rho\mathbb{E}[v_{t+1}] = v_t + \rho\mu.$$

Therefore, if $\rho < \epsilon$, b' is clearly a better choice than the optimal bid b , thus giving a contradiction. For k -lookahead buyers, extending this argument involves coupling the k -length trajectories of two bidding strategies: one that plays optimal bids starting from bid b in s_t vs. a strategy that starts from bid b' in state s_t . By coupling these two trajectories, we aim to compare the k -lookahead utilities of the two bids, in order to demonstrate that playing b in the state s_t is strictly suboptimal.

The coupling point of the two trajectories is the (random) time step $t + \tau + 1$ where the (supposedly) optimal strategy's trajectory (that started with b) comes back to a (borderline) good state for the first time. Define the second bidding strategy in a coupled manner to play bid $b' = (1 - \epsilon)\mu$ repeatedly until $t + \tau$, and optimal bids thereafter. Then, at step $t + \tau + 1$, the second trajectory is also in a good state, but possibly not borderline. To relate the utilities of the two trajectories after this coupling point, we use induction along with an observation that the optimal utility for a trajectory starting at any good state is at least as good as the utility for a trajectory starting at a borderline good state (refer to Claim 1 in appendix). Therefore, from point $t + \tau + 1$ onwards, the second trajectory has better, or at least as good utility, as the first supposedly optimal strategy. It remains to compare the utilities in steps $t, t + 1, \dots, t + \tau$.

By definition of τ , under first supposedly optimal strategy, the states at time steps $t + 1, \dots, t + \tau$ are all bad states. Therefore, the total expected utility is at most

$$v_t + \tau\rho\mu$$

On the other hand, the second strategy of bidding $(1 - \epsilon)\mu$ always stays in good state to get total expected utility of at least $v_t - (1 - \epsilon)\mu + \tau\epsilon\mu$ in these steps. A better bound for the second case can be observed by considering the probability that the 'last' (i.e., k^{th}) round occurs at the end of this interval, that is, $\tau = k$. At the last round, the second strategy bids 0 to obtain full valuation as utility. Therefore, a better lower bound for expected utility is

$$v_t - (1 - \epsilon)\mu + (\tau - 1)\epsilon\mu + \Pr(\tau = k)\mu$$

Here $\Pr(\tau = k) = (1 - \rho)^{k-1}\rho$. Comparing this with the utility of the first strategy, after some careful algebraic manipulations we obtain that under condition $\rho \leq \epsilon/(2 - \epsilon)$, the utility for second strategy is better than optimal; this gives a contradiction.

Now, let's consider the more complicated allocation rule actually used by $M(\epsilon, \rho, p)$ in bad state: the allocation in bad state actually depends on the bid, and the bidder gets an allocation with probability ρ only if the bid is above p . Although this makes the payoff for bidder in bad states worse, and provides more incentive to remain in good state; it also makes analysis trickier since now the event of a bidder transferring from bad state to good state depends on her bid (and therefore her valuation) in addition to the random ρ probability event.

Finally, all the coupling arguments and expected revenue analysis sketched above need to be carried out carefully using martingale analysis and stopping times. The details are provided in Appendix B. ■

4.2 Revenue against buyer learning behaviors

Theorem 1.2 can be obtained as a corollary of the following proposition, by substituting $\rho = \epsilon - o(1)$.

Proposition 4.3 *The mechanism $M(\epsilon, \rho, p)$ with $\rho < \epsilon$ and $p = p^*$ achieves a revenue of at least*

- (a) $\text{Rev}^{M(\epsilon, \rho, p)} \geq \frac{\rho}{\rho+1} \text{Rev}^{\text{Mye}} - o(1)$ against any buyer who is a no-regret learner for the class \mathcal{E} of experts (refer to Equation (8) and (9)), and
- (b) $\text{Rev}^{M(\epsilon, \rho, p)} \geq (1 - \epsilon)\mu - o(1)$ against any buyer who is a policy regret learner for a class \mathcal{C} of sequences containing all sequences of single experts (refer to Equation (10)).

Proof:

(a) **No-regret learners.** Consider a bidding function $f(s_t, v_t)$ defined as $f(s_t, v_t) = p^*$ 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, using the above expert as benchmark, we obtain that 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 achievable in any good state is v_t . And, the buyer cannot make any positive utility in a bad state at time t if $v_t \leq p^*$. Therefore, a no-regret learning buyer cannot afford to lose more than $o(T)$ of the bad state auctions with $v_t > p^*$. This means that any no-regret learning buyer must bid $b_t \geq p^*$ in all but possibly $o(T)$ of the rounds t with $s_t = \perp, v_t \geq p^*$. Let B be a random variable denoting the number of bad states in the state trajectory, and let B' denote the number of those bad states with $v_t \geq p^*, b_t \geq p^*$, then,

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

Also, let G be the number of good states.

To lower bound $\mathbb{E}[B]$, let us first consider state trajectories of form

$$\not\perp \perp \perp \perp \perp \perp \not\perp \perp \perp \perp \not\perp \perp \perp \dots,$$

i.e., lone good states interspersed with sequences of bad states. Then, due to the construction of the mechanism $M(\epsilon, \rho, p)$, the only way the buyer can obtain $G - 1$ good states (all except the first good state), only by winning $G - 1$ bad state auctions. And, since the bad state auctions can be won only with probability ρ , we have that $\mathbb{E}[G] \leq \mathbb{E}[B]\rho + 1$. This gives,

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

Combining the above observations,

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

Therefore, the total expected seller's revenue is at least

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

Now, consider sequences of states with more than one consecutive good states, e.g.,

$$\not\perp \perp \perp \perp \perp \not\perp \not\perp \not\perp \perp \perp \perp \perp \not\perp \not\perp \dots \text{ etc.}$$

Then, in any sub-sequence 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 on average, where \bar{v} denotes the average valuation over the trailing states. On the other hand, the above expert f , which bids 0 in good states, makes an average of \bar{v} utility in those trailing states (in hindsight). Further, in the bad states, and in the 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.

(b) **No-policy-regret learners.** Here, we use the sequence of constant bids $(1-\epsilon)\mu$ as a benchmark expert. This bidding strategy ensures that the mechanism is always in a good state $s_t = ((1-\epsilon)\mu, t)$, and achieves a utility of $\epsilon\mu$. Therefore, since the class \mathcal{C} contains such sequences of single experts, the policy-regret learning buyer must achieve at least $\epsilon\mu - o(T)$ utility. Now, in bad states, the buyer can achieve at most $\rho\mu$ utility on average. Therefore, if $\rho < \epsilon$, then the number of bad states in buyer's state trajectory can be at most $o(T)$. This implies that the trailing good states are at least $T - o(T)$. By definition, the bidding average over the trailing good states must be least $(1 - \epsilon)\mu$, and therefore, the seller's revenue is at least $(1 - \epsilon)\mu T - o(T)$. ■

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. Papadimitriou et al. [2016] 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. Mirrokni et al. [2016a] 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. Ashlagi et al. [2016] and Mirrokni et al. [2016b] 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. Balseiro et al. [2017], 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. Mirrokni et al. [2017] 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 Baron and Besanko [1984] there is a large body of work Besanko [1985], Battaglini [2005], Courty and Li [2000], Bergemann and Strack [2014], Eso and Szentes [2007] that study optimizations in the presence of evolving values. Recent works include those by Bergemann and Strack [2015], Athey and Segal [2013], where they consider general models where value evolution could depend on the action of the mechanism. Kakade et al. [2013], Pavan et al. [2014] study revenue optimal dynamic mechanism design where the buyer's value evolves based on signals that she re-

ceives each period. [Chawla et al. \[2016\]](#) 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 [Fudenberg and Villas-Boas \[2006\]](#) 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 [Simon \[1955\]](#). 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 [Nau \[1983\]](#), [de Kleer and Mark Shirley \[1992\]](#), [Sefer et al. \[2009\]](#). Lookahead search is also related to the sequential thinking framework in game theory [Stahl and Wilson \[1994\]](#). More recently, [Mirrokni et al. \[2012\]](#) 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.

References

- 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.
- 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.
- Susan Athey and Ilya Segal. An efficient dynamic mechanism. *Econometrica*, 81(6):2463–2485, 2013. ISSN 1468-0262.
- Santiago R. Balseiro, Vahab S. Mirrokni, and Renato Paes Leme. Dynamic mechanisms with martingale utilities. In *Proceedings of the 2017 ACM Conference on Economics and Computation, EC '17, Cambridge, MA, USA, June 26-30, 2017*, page 165, 2017.
- David P. Baron and David Besanko. Regulation and information in a continuing relationship. *Information Economics and Policy*, 1(3):267 – 302, 1984.
- Marco Battaglini. Long-term contracting with markovian consumers. *American Economic Review*, 95(3):637–658, 2005.
- Dirk Bergemann and Philipp Strack. Dynamic revenue maximization: A continuous time approach. 2014.

- Dirk Bergemann and Philipp Strack. Dynamic revenue maximization: A continuous time approach. *Journal of Economic Theory*, 159, Part B:819 – 853, 2015. ISSN 0022-0531. Symposium Issue on Dynamic Contracts and Mechanism Design.
- David Besanko. Multi-period contracts between principal and agent with adverse selection. *Economics Letters*, 17(12):33 – 37, 1985.
- 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. doi: 10.1137/1.9781611974331.ch101. URL <http://dx.doi.org/10.1137/1.9781611974331.ch101>.
- Pascal Courty and Hao Li. Sequential screening. 67(4):697–717, 2000.
- J. de Kleer and O. Raiman and Mark Shirley. One step lookahead is pretty good. *Readings in Model-Based Diagnosis*, pages 138–142,, 1992.
- Peter Eso and Balzs Szentes. Optimal information disclosure in auctions and the handicap auction. *Review of Economic Studies*, 74(3):705–731, 2007.
- Drew Fudenberg and J Miguel Villas-Boas. Behavior-based price discrimination and customer recognition. *Handbook on economics and information systems*, 1:377–436, 2006.
- Sham M Kakade, Ilan Lobel, and Hamid Nazerzadeh. Optimal dynamic mechanism design and the virtual-pivot mechanism. *Operations Research*, 61(4):837–854, 2013.
- Susan Kantor. Transparency in programmatic advertising. <https://auditedmedia.com/blog/transparency-programmatic-advertising-its-two-way-street>, 2016.
- 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.
- 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, 2016a.
- 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, 2016b. URL <http://arxiv.org/abs/1605.08840>.
- Vahab S. Mirrokni, Renato Paes Leme, Pingzhong Tang, and Song Zuo. Oblivious dynamic mechanism design. 2017. URL <https://ssrn.com/abstract=2873701>.
- Dana S. Nau. Decision quality as a function of search depth on game trees. *J. ACM*, 30(4):687–708, 1983.

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.

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

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.

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

Dale Stahl and Paul Wilson. Experimental evidence on players' models of other players. *Journal of Economic Behavior & Organization*, 25(3):309–327, 1994. URL <http://EconPapers.repec.org/RePEc:eee:jeborg:v:25:y:1994:i:3:p:309-327>.

A Mechanism for per-round ex-post IR with infinite lookahead buyers

As discussed in the introduction, it is not possible to get any reasonable approximation to μ as revenue against k -lookahead buyers for small k when per-round ex-post IR is imposed. Nevertheless, we show here how to make small modifications to the mechanism discussed in Section 3 to get per-round ex-post IR in all but a constant number of rounds for an infinite lookahead buyer (i.e., a T -lookahead buyer in a T -rounds game).

Let B be the highest value in the support of the buyer value distribution and μ the mean of the distribution. The target-revenue R_{target} be defined as

$$R_{target} = T\mu(1 - \epsilon) - \sqrt{4B\mu\sqrt{T}\ln(T)} - \sqrt{2B\mu\ln(T)} \sum_{j=1}^T \frac{1}{\sqrt{j}} \quad (13)$$

$$= T\mu(1 - \epsilon) - c \cdot \sqrt{B\mu T \ln(T)} \quad (14)$$

Definition 4 (Mechanism $M_\infty(\epsilon, \rho, p)$) 1. **State Space \mathcal{S} :** The state space is $\mathcal{S} = \mathbb{R} \times \mathbb{R} \times \mathbb{N}$. A state $s = (TP, EP, t)$ is a good state if $TP \geq EP$, and bad otherwise. TP stands for the total payment made by the buyer in the good state, and EP stands for the payment the mechanism expects the buyer to have paid and t is the total number of rounds elapsed, including the current round that's beginning.

2. **Starting state s_1 :** The mechanism starts at $t = 0$, $EP = 0$, and $TP = \mu\sqrt{T} + \sqrt{4B\mu\sqrt{T}\ln(T)}$.

I.e., the mechanism the buyer a credit of $\mu\sqrt{T} + \sqrt{4B\mu\sqrt{T}\ln(T)}$ to begin with.

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

That is, $x_t \sim x(s_t, b_t)$, where

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

4. **Payment rule** $p(s_t, b_t, x_t)$: This is a first price mechanism, with a small change:

$$p(s_t, b_t, x_t) = \begin{cases} \min(b_t, R_{target} - TP) & \text{if } x_t = 1 \text{ and } TP \leq R_{target}, \\ 0, & \text{otherwise.} \end{cases}$$

5. **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 $s_t = (TP, EP, t)$. In this mechanism, the state effectively remains the same if $x_t = 0$. Otherwise, it transitions to a state with updated total payment and expected payment. The total payment naturally increases by p_t . The expected payment increases not by $\mu(1 - \epsilon)$, but has an additional slack, namely, it increases by $\mu(1 - \epsilon) - \sqrt{\frac{2B\mu \ln(T)}{t}}$.

$$s_{t+1} = \begin{cases} (TP, 0, t + 1), & \text{if } TP \geq R_{target} \\ (TP + p_t, EP + \mu(1 - \epsilon) - \sqrt{\frac{2B\mu \ln(T)}{t}}, t + 1), & \text{if } R_{target} > TP \geq EP \\ (TP, EP, t + 1), & \text{if } TP < EP, x_t = 0 \\ (TP, TP, t + 1), & \text{if } TP < EP, x_t = 1, \end{cases}$$

The main difference between this mechanism $M_\infty(\epsilon, \rho, p)$ and the $M(\epsilon, \rho, p)$ defined in Section 3 is that when in the boundary state of $TP = EP$ at the beginning of round t , in M_∞ the agent needs to bid $\mu(1 - \epsilon) - \sqrt{\frac{2B\mu \ln(T)}{t}}$ to stay in a good state, as opposed to $\mu(1 - \epsilon)$ in $M(\epsilon, \rho, p)$. I.e., there is a slack that vanishes with time, i.e., as t gets large. Thus having to bid even lower than what he had to in Section 3 to remain in the good state, it immediately follows that Lemma 4.2 holds for our modified mechanism as well, i.e., any agent in a good state, will continue to be in a good state as it is less costly. This immediately implies identical revenue guarantees for all lookahead buyers, with a $c\mu\sqrt{T}$ loss in revenue due to the vanishing slack this mechanism provides in every round (and also the slack initially).

Truthful bidding will not lead a bidder out of good state w.h.p. Consider a bidder bidding his true value V_t in round t . The probability that, for any r , after r rounds of bidding true value V_t we have $TP < EP$ is at most $\frac{1}{T^c}$ by Chernoff bounds (the slacks are chosen in M_∞ to satisfy this). Therefore, by union bound, the probability that it ever happens in any of T rounds is at most $\frac{T}{T^c} \rightarrow 0$ as $T \rightarrow \infty$. I.e., with a high probability, truthful bidding will never lead a buyer out of good state.

Truthful bidding is optimal w.h.p. Note that the mechanism doesn't accept payment once it has earned a revenue of R_{target} . This R_{target} is computed to be the minimum revenue that the mechanism is guaranteed to extract from a buyer that always lives at the boundary state. Or equivalently $T\mu - R_{target}$ is the maximum utility such a buyer can earn, and this in turn implies that this is the maximum utility an infinite lookahead buyer can earn. To achieve this utility, an infinite lookahead buyer need not shade his bid to constantly live at the boundary state — just

bidding his true value is an equally good strategy (modulo the tiny probability that this will get him to a bad state, where he bids whatever is necessary to maintain him in good state) in terms of utility. Thus, bidding the true value throughout all the T rounds (which is clearly a per-round ex-post IR strategy) is an optimal strategy with probability $1 - \frac{1}{\text{poly}(T)}$.

The above discussion gives us the proof of Theorem 1.5, restated below

Theorem 1.5 *For any $\epsilon \in (0, 1)$, there exists a non-payment forceful and IIR mechanism that is $(\frac{\epsilon}{2}, 1 - \epsilon)$ -robust against lookahead attitudes, ex-post IR against a myopic buyer, and with a high probability ex-post IR against an infinite lookahead buyer.*

B Missing Proofs

Proof of Proposition 4.1 Here, we provide a detailed proof of Proposition 4.1. Lemma B.1 proves the result for myopic buyers. Next, Lemma B.2 and Lemma B.3 together provide a detailed proof for Lemma 4.2, to complete the proof of Proposition 4.1.

Lemma B.1 *The mechanism $M(\epsilon, \rho, p)$ with $p = p^*$ 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:*

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

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 $M(\epsilon, \rho, p^*, R)$. 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 $U_0^t(s_t, v_t, b_t) = \mathbb{E}[x_t v_t - p_t | v_t, s_t]$. Now, consider two cases:

(a) $s_t \neq \perp$: in this case, by definition of allocation rule in $M(\epsilon, \rho, p^*, R)$, $x_t = 1, p_t = b_t$, so that $U_t^0 = \mathbb{E}[v_t - b_t]$, 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 $M(\epsilon, \rho, p^*, R)$, with probability ρ , $x_t = 1, p_t = b_t$ if $b_t \geq p^*$, so that $U_t^0 = \mathbb{E}[(v_t - b_t) \mathbb{1}(b_t \geq p^*)]$. 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.,

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

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

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

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 $M(\epsilon, \rho, p^*, R)$ 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 $\not\perp \perp \perp \perp \perp \not\perp \perp \perp \perp \not\perp \perp \perp \dots$, 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:

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

■

Lemma B.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} ,

$$U_1^t(s_t, v_t, b_t) = E[(v_t - b_t) + \sup_{b'} U_0^{t+1}(s_{t+1}, v_{t+1}, b') | v_t, s_t] \quad (15)$$

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,

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

where as

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

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 (15), 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 at least

$$v_t + \epsilon\mu$$

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. \blacksquare

Lemma B.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 B.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 contradiction 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 $i = 1, \dots, \tau$, 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 $t, t+1, \dots, t+\tau$ is upper bounded by

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

Therefore,

$$\begin{aligned}
U_k^t(s_t, v_t, b_t) &\leq v_t + \mathbb{E}\left[\sum_{i=1}^{\tau} v_{t+i} I(A_{t+i})\right] \\
&\quad + \mathbb{1}(\tau < k) \mathbb{E}\left[\sup_b U_{k-\tau-1}^{t+\tau+1}(s_{t+\tau+1}, v_{t+\tau+1}, b) | s_{t+\tau+1}\right] | s_t, v_t] \quad (16)
\end{aligned}$$

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 $j = \tau', \tau' + 1, \dots, k - 1$.

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

$$\begin{aligned}
U_k^t(s_t, v_t, b'_t) &\geq v_t - (1 - \epsilon)\mu + \mathbb{E}\left[\sum_{i=1}^{\tau'} v_{t+i} - (1 - \epsilon)\tau' + \mathbb{1}(\tau' = k)(1 - \epsilon)\mu\right] \\
&\quad + \mathbb{1}(\tau' < k) \mathbb{E}\left[\sup_b U_{k-\tau'-1}^{t+\tau'+1}(s_{t+\tau'+1}, v_{t+\tau'+1}, b) | s_{t+\tau'+1}\right] | s_t, v_t] \quad (17)
\end{aligned}$$

Now, we show that the last term from (17) dominates the last term from (16). 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 (16) 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 (17) 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 Claim 1 (stated and proved below) for $k - i - 1$, we have for all $i, s \neq \perp$,

$$\mathbb{E}\left[\sup_b U_{k-i-1}^{t+i+1}(s_{t+i+1}, v_{t+i+1}, b) | s_{t+i+1} = s\right] \geq \mathbb{E}\left[\sup_b U_{k-i-1}^{t+i+1}(s_{t+i+1}, v_{t+i+1}, b) | s_{t+i+1} = \pm\right] \quad (18)$$

Therefore, we derive that the last term in (17), is greater than or equal to the corresponding term in (16).

Using this observation, and subtracting (16) from (17), we can bound the total difference (denoted as Δ) in k lookahead utilities of b_t and b'_t as

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

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

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

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

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

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 notational brevity)

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

Now, let X be a geometric random variable with success probability ρ , then τ stochastically dominates $\min\{X, k\}$. And, from Claim 2 (stated and proved below)

$$\begin{aligned} \mathbb{E}[\tau] &\geq \mathbb{E}[\min\{X, k\}] = \frac{1}{\rho} \Pr(X < k) + \Pr(X \geq k), \\ \Pr(\tau < k) &\leq \Pr(X < k) = 1 - (1 - \rho)^{k-1} \end{aligned}$$

The proof is completed by the following algebraic manipulations:

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

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:

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

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

Claim 1 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,

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

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 $b_{t+1}, \dots, b_{t+k-1}$ which are optimal $k - 1, k - 2, \dots, 1$ lookahead bids respectively.

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

Using Lemma B.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 $s_\tau, \tau = t+1, \dots, t+k$ 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 $b_t, b_{t+1}, \dots, b_{t+k-1}$ must be at least $(1-\epsilon)\mu k$. 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:

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

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 $b_t, b_{t+1}, \dots, b_{t+k-1}$ 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 $s \neq \perp$,

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

■

Claim 2 *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:

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

■

PROOF OF THEOREM 1.3: We will take our distribution F to be the Pareto distribution with parameter $\alpha > 2$, supported on $[1, +\infty)$. In particular, the probability density function is $f(x) = \frac{\alpha}{x^{\alpha+1}}$, $x \in [1, +\infty)$. Note that the mean is $\mu = \frac{\alpha}{\alpha-1}$ and the variance is $\sigma^2 = \frac{\alpha}{(\alpha-1)^2(\alpha-2)}$, which are both finite. It is easy to see that $\text{Rev}^{\text{Mye}} = 1$, and that F is a decreasing hazard rate distribution, as well as a regular 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 $\mathbb{E}_{x \sim x(s_t, b_t)}[x] \cdot v_t - \mathbb{E}_{x \sim x(s_t, b_t)}[p(s_t, b_t, x)]$.

To prove our result let us suppose that there exists a collection of functions $b_s : \mathbb{R} \rightarrow \Delta^{\mathbb{R}^+}$, indexed by states $s \in \mathcal{S}$, which map a realized value for a myopic bidder to a (potentially randomized) bid, and which are such that the expected revenue of mechanism M (average over T rounds) against a myopic buyer using these bidding functions is at least $\epsilon \cdot \text{Rev}^{\text{Mye}} - o(1)$, where $o(1)$ is a function that goes to 0 with T . Note that these mappings are only indexed by single states s as they pertain to the behavior of a myopic buyer. Moreover, note that we actually do not need to require that for all s, v , $b_s(v)$ is optimal. The only assumption that we need to 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, $\hat{x} : \mathcal{S} \times \mathbb{R} \rightarrow [0, 1]$ and $\hat{p} : \mathcal{S} \times \mathbb{R} \rightarrow \mathbb{R}$ respectively, as follows:

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

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:

$$\begin{aligned} \text{Rev}_{\text{myop}}^s &= \mathbb{E}_{v \sim F}[\hat{p}(s, v)] \equiv \mathbb{E}_{v \sim F, p \sim G_s}[p \cdot \mathbf{1}_{v \geq p}] \stackrel{*}{=} \mathbb{E}_{p \sim G_s} \left[\frac{1}{p^{\alpha-1}} \right]; \\ \text{Ut}_{\text{myop}}^s &= \mathbb{E}_{v \sim F}[\hat{x}(s, v) \cdot v - \hat{p}(s, v)] \equiv \mathbb{E}_{v \sim F, p \sim G_s}[(v - p) \cdot \mathbf{1}_{v \geq p}] \stackrel{*}{=} \mathbb{E}_{p \sim G_s} \left[\frac{1}{\alpha - 1} \frac{1}{p^{\alpha-1}} \right]. \end{aligned}$$

(In the above, the equalities $\stackrel{*}{=}$ follow by plugging in for F the distribution defined above.) So, in particular, it follows that $\text{Ut}_{\text{myop}}^s = \frac{1}{\alpha-1} \text{Rev}_{\text{myop}}^s$. To summarize, in any state $s \in \mathcal{S}$, a myopic buyer makes utility that is a factor of $\alpha - 1$ smaller than the payment that he makes.

Now recall that our mechanism has expected average per round revenue against a myopic buyer that is at least $\epsilon \cdot \text{Rev}^{\text{Mye}} - o(1)$. It follows from the above derivation that it should also then give expected average per round utility at least $\frac{1}{\alpha-1} \epsilon \cdot \text{Rev}^{\text{Mye}} - o(1)$ 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 average per round utility at least $\frac{1}{\alpha-1} \epsilon \cdot \text{Rev}^{\text{Mye}} - o(1)$ to an infinite look-ahead buyer. Hence, the expected average per round revenue that the mechanism can get from an infinite look-ahead buyer is at most:

$$\begin{aligned} \mu - \frac{1}{\alpha-1} \epsilon \cdot \text{Rev}^{\text{Mye}} + o(1) &= \mu - \frac{1}{\alpha-1} \epsilon \cdot 1 + o(1) && \text{(recalling that } \text{Rev}^{\text{Mye}} = 1) \\ &= \mu - \frac{\alpha}{\alpha-1} \frac{\epsilon}{\alpha} + o(1) \\ &= \left(1 - \frac{\epsilon}{\alpha}\right) \mu + o(1) && \text{(recalling that } \mu = \frac{\alpha}{\alpha-1}) \end{aligned}$$

As α can be made arbitrarily close to 2, the theorem holds. \blacksquare

PROOF OF THEOREM 1.4: We will prove the stated bound using equal revenue distribution. k -lookahead buyers optimize k -lookahead utility, which in round t is given by

$$U_k^t(s_t, v_t, b_t) = (v_t x(s_t, b_t) - p(s_t, b_t)) + \max_{b'} \mathbb{E}_{s', v'} [U_{k-1}^{t+1}(s', b', v') | s_t, b_t]$$

Let us first consider the case when the allocation function is deterministic, i.e., $x(s, b) \in \{0, 1\}$. Now, for any state s , let B_s^1 be the set of bids such that $x(s, b) = 1$. Let V_s^1 be the set of valuations in the support of F such that given any valuation $v \in V_s^1$, there is at least one utility maximizing bid that gets an allocation in state s . More precisely, given a valuation v and state s , let $B_{s,v}^* := \{\arg \max_b U(s, v, b)\}$ denote the set of utility maximizing bids. Then, V_s^1 is defined the

set of valuations v for which $B_s^1 \cap B_{s,v}^*$ is non-empty. We observe that the set of utility maximizing allocating bids is the same for all valuations in V_s^1 . That is, $B_s^1 \cap B_{s,v}^* = B_s^1 \cap B_{s,v'}^* =: B_s^*$, for any v, v' . This is because, for bids b in B_s^1 , $x(s, b) = 1$; therefore for any valuation v , the utility maximizing bids in $B_s^1 \cap B_{s,v}^*$ are given by set $\{\arg \max -p(s, b) + \mathbb{E}_{s', v'}[\max_{b'} U_{k-1}(s', v', b') | s, b]\}$, which does not depend on valuation v .

Now, let $v_{min} = \inf\{v \in V_s^1\}$. Then, for all $b \in B_s^*$, by IR property:

$$0 \leq U_k^t(s, b, v_{min}) = v_{min} - p(s, b) + \mathbb{E}[\max_{b'} U_{k-1}^{t+1}(s', b', v') | s, b] \leq v_{min} - p(s, b) + k\mu$$

where $k\mu$ is an upper bound on the $(k-1)$ -lookahead utility. Thus, rearranging $0 \leq v_{min} - p(s, b) + k\mu$ gives us:

$$p(s, b) \leq v_{min} + k\mu.$$

Further from the per-round ex-post IR property, $p(s_t, b_t) \leq v_t$. Therefore, expected revenue in round t , given state $s_t = s$ is upper bounded by

$$\mathbb{E}[p(s_t, b_t) | s_t = s] \leq \mathbb{E}_{v \in V}[\min\{v_{min} + k\mu, v\} I(v \in V_s^1)] \leq \mathbb{E}_{v \in V}[\min\{v_{min} + k\mu, v\} I(v \geq v_{min})]$$

For equal revenue distribution, this is upper bounded by

$$\int_{v_{min}}^{v_{min} + k\mu} v \frac{1}{v^2} dv + (v_{min} + k\mu) \times \frac{1}{v_{min} + k\mu} = \log(v_{min} + k\mu) - \log(v_{min}) + 1 \leq \log(k\mu) + 1$$

Randomized allocation. The above argument can be extended to $x(s, b) \in [0, 1]$, by replacing set B_s^1 and V_s^1 by B_s^x and V_s^x respectively, defined for every possible value of allocation $x \in [0, 1]$. That is, B_s^x is the set of bids such that $x(s, b) = x$. And, V_s^x is the set of valuations in the support of F such that given any valuation $v \in V_s^x$, there is at least one utility maximizing bid that gets an allocation of x in state s . Then, $B_{s,v}^* \cap B_s^x$ denote all the bids which are utility maximizing and get allocation of x . Using the same argument as above, this set can be shown to be independent of v , i.e., $B_{s,v}^* \cap B_s^x = B_{s,x}^*$.

Now, for any non-empty V_s^x , let $v_{min,x} = \inf\{v \in V_s^x\}$. Then, for all $b \in B_{s,x}^*$, by IR property:

$$0 \leq U_k^t(s, b, v_{min,x}) = xv_{min,x} - p(s, b) + \mathbb{E}[\max_{b'} U_{k-1}^{t+1}(s', b', v') | s, b] \leq xv_{min,x} - p(s, b) + k\mu$$

where $U_{k-1}^* \leq k\mu$ is an upper bound on the $(k-1)$ -lookahead utility. so that

$$p(s, b) \leq xv_{min,x} + k\mu,$$

For $s_t, b_t, x_t = x(s_t, b_t)$, if b_t is an optimal bid, then $V_{s_t}^{x_t}$ must be non-empty, and above inequality can be applied, to get $p(s_t, b_t) \leq x_t v_{min,x_t} + k\mu$. Further from the per-round ex-post IR property, $p(s_t, b_t) \leq v_t$. Then, the above argument can be repeated while replacing v_{min} by v_{min,x_t} , to obtain the same upper bound.

$$\begin{aligned} \mathbb{E}[p(s_t, b_t) | s_t = s] &\leq \mathbb{E}_{v \in V}[\min\{x_t v_{min} + k\mu, v\} I(v \in V_{s_t}^{x_t})] \\ &\leq \mathbb{E}_{v \in V}[\min\{v_{min,x_t} + k\mu, v\} I(v \geq v_{min,x_t})] \end{aligned}$$

For equal revenue distribution, this is bounded by $\log(v_{min,x_t} + k\mu) - \log(v_{min,x_t}) + 1 \leq \log(k\mu) + 1$.

■