Leveraging Reviews: Learning to Price with Buyer and Seller Uncertainty

WENSHUO GUO and NIKA HAGHTALAB
University of California, Berkeley
and
KIRTHEVASAN KANDASAMY
University of Wisconsin, Madison
and
ELLEN VITERCIK
Stanford University

Customers can access hundreds of reviews for a single product in online marketplaces. Buyers often use reviews from other customers that share their type—such as height for clothing or skin type for skincare products—to estimate their values, which they may not know a priori. Customers with few relevant reviews may hesitate to purchase except at a low price, so for the seller, there is a tension between setting high prices and ensuring that there are enough reviews so buyers can confidently estimate their values. Simultaneously, sellers may use reviews to gauge the demand for items they wish to sell. In this work, we study this pricing problem in an online setting where the seller interacts with a set of buyers of finitely many types, one by one, over a series of T rounds. At each round, the seller first sets a price. Then, a buyer arrives and examines the reviews of the previous buyers with the same type, which reveal those buyers' ex-post values. Based on the reviews, the buyer decides to purchase if they have good reason to believe their ex-ante utility is positive. Crucially, the seller does not know the buyer's type when setting the price, nor even the distribution over types. We provide a no-regret algorithm that the seller can use to obtain high revenue. When there are d types, after T rounds, our algorithm achieves a problem-independent $\tilde{O}(T^{2/3}d^{1/3})$ regret bound. However, when the smallest probability $q_{\min}$ that any given type appears is large, specifically when $q_{\min} \in \Omega(d^{-2/3}T^{-1/3})$, the same algorithm achieves a $\tilde{O}(T^{1/2}q_{\min}^{-1/2})$ regret bound. We complement these upper bounds with matching lower bounds in both regimes, showing that our algorithm is minimax optimal up to lower-order terms.

This is a summary of work that won the *Exemplary AI Track Paper Award* at EC'24.

Categories and Subject Descriptors: 500 [**Theory of computation**]: Computational pricing and auctions; 500 [**Theory of computation**]: Online learning theory

General Terms: Algorithms; Economics; Theory

Additional Key Words and Phrases: Pricing, Reviews, Revenue maximization, Online learning

1. INTRODUCTION

The rapid growth of e-commerce has allowed customers to gain insights from thousands of reviews before deciding whether to purchase an item. Customers often use reviews by buyers who share their "type" —such as body type for clothes or skin type for skincare products—to develop high-fidelity estimates of their values

Authors' addresses: {wsguo, nika}@berkeley.edu, nika@berkeley.edu, kandasamy@cs.wisc.edu, vitercik@stanford.edu.

for items, which are quantities they may be uncertain of before purchasing.

When learning from reviews, a customer's purchase decision is not just a function of the item's price but also of how certain the customer is about her valuation, which in turn depends on the earlier sales and reviews of the items. This leads to a tension between setting revenue-optimal prices while ensuring buyers have enough reviews to estimate their values confidently. This tension is perhaps most clear for customers of rare types (for example, particularly tall or short individuals shopping for clothing) who may find only a few reviews from similar customers and, due to risk aversion, may only be willing to buy at relatively low prices.

We introduce a model that simultaneously captures the seller's pricing problem, the buyers' learning problem, and the modus through which the buyers learn: reviews. We study how a seller—uncertain about the buyers' type distribution—can learn to set high-revenue prices when the buyers themselves are unsure about their values and are learning from reviews. Thus, there is information uncertainty on both sides of the market: the seller is uncertain about which buyer will arrive and the buyers' type distribution, but the buyer, who knows their type, is unsure about their *ex-ante* value. Both sides of the market are operating with significantly less information than has historically been assumed in mechanism design.

We study this pricing problem with an online sequential learning model where the seller attempts to sell identical copies of an item to a series of distinct buyers over T timesteps. Each buyer has one of d types drawn from a distribution $\mathcal{P}$, and a buyer of type i has an *ex-ante* value of θ_i for the item. At each timestep t, the seller sets a price p_t. The seller knows the *ex-ante* values $\theta_1, \dots, \theta_d$ and thus has some limited information about the buyers (for example, from market research), but he does not know the buyer's type on each round nor even the distribution $\mathcal{P}$. If a buyer of type i purchases the item, they will leave a review communicating their *ex-post* value for the item, which is a random variable with mean θ_i. To decide whether to purchase, a new buyer evaluates reviews left by buyers of type i who bought the item in the past. Specifically, the buyer at round $t \in [T]$ uses the past reviews to select a threshold τ_t and chooses to buy as long as $p_t \leq \tau_t$. If the buyer's threshold τ_t is too pessimistic—for example, it always equals zero no matter the reviews—then optimizing revenue would be hopeless. In our model, we bound the level of risk-aversion that the buyer can display: we assume that τ_t is at least a lower confidence bound we denote LB_t that equals the average of the reviews left by buyers with the same type, minus an uncertainty term that depends on the number of such reviews. Intuitively, the buyer can be confident that their *ex-ante* value is at least LB_t with high probability, so they always buy if they have good reason to believe that their *ex-ante* utility will be positive.

The *ex-post* value is the actual experience of the buyer and is different from the *ex-ante* value due to exogenous stochastic factors that cannot be known at the time of purchase (for example, manufacturing defects, color on the website not matching the actual color). Hence, when there is complete information, the buyer decides based on their *ex-ante* value. In our problem, the buyer does not even know their *ex-ante* value and uses reviews from previous buyers to estimate it.

We provide a no-regret learning algorithm for the seller that balances setting high-revenue prices with soliciting reviews from rare but high-value customers. The key challenge is that the seller does not know the current buyer's type on each round *a priori*: the prices are anonymous. This means the seller does not know the number of reviews that the buyer will use to construct their value estimate. A buyer on any round could be *(i)* a high-value type, but uncertain of their value since their type has few reviews, and thus may be hesitant to purchase except at a low price, *(ii)* a high-value type, and more confident of their value since their type has many reviews, and thus is willing to purchase at a high price, or *(iii)* a low-value type whom the seller should not target even if they were absolutely sure of their value since it leads to small per-purchase revenue. In the first case, it may be worthwhile to initially set a low price to solicit enough reviews to ensure future purchases at a higher price, winning over these rare but high-value customers. The seller, however, has to decide which buyers to win over without knowing the buyer's type on each round, nor even the distribution over types. He may, therefore, wastefully offer a low price to a buyer who would be willing to buy at a higher price.

2. RELATED WORK

Learning to price when buyers do not know their values. Several papers have studied selling repeatedly to a single buyer while the buyer is learning from their experience [Papadimitriou et al. 2022; Ashlagi et al. 2016; Chawla et al. 2022; Feng et al. 2018; Weed et al. 2016; Kandasamy et al. 2023]. However, buyers on online platforms often do not return repeatedly to buy the same item and can only obtain feedback from previous buyers via reviews. In this paper, we study a setting where the buyers can only learn from past reviews.

Ifrach et al. [2019] consider a similar pricing problem for the seller when the buyers learn from reviews. However, their model is limited to one buyer type, whereas we study the setting with multiple buyer types. Moreover, the seller does not know the frequency of each type and the type of buyer who arrives at each round, which leads to crucial difficulties in our analysis.

Learning to price when buyers know their values. Zhao and Chen [2020] study a setting where the buyers know their values, but the seller does not know the distribution over buyers' values. They present an algorithm that uses buyer reviews to obtain a $\tilde{O}\left(T^{1/2}\right)$ regret bound. In contrast, if the seller only observes purchase decisions and not reviews, Kleinberg and Leighton [2003] provide a $\Omega\left(T^{2/3}\right)$ lower bound. While they show that this bound can be improved to $\tilde{\Theta}\left(T^{1/2}\right)$, it requires additional distributional assumptions.

Selling to no-regret buyers who know their values. When buyers know their values, the buyer may strategically improve their purchase decisions or bidding strategy over repeated interactions to achieve a higher accumulated utility. No-regret learning has been explored as a model of buyer behavior [Braverman et al. 2018; Deng et al. 2019; Nekipelov et al. 2015; Devanur et al. 2014]. In this literature, buyers know their values, whereas we work with buyers who do not and need to estimate their values from historical reviews. This leads to different dynamics.

Buyers' social learning from reviews. Our work is related to a rich literature on buyer behavior and social learning from reviews when buyers do not know their values [Ifrach et al. 2019; Boursier et al. 2022; Han and Anderson 2020; Chamley

2004; Besbes and Scarsini 2018; Bose et al. 2006; Crapis et al. 2017; Kakhbod et al. 2021; Acemoglu et al. 2022]. Much of this research can be categorized into two groups depending on whether the decision model is Bayesian or non-Bayesian. It may be computationally challenging for buyers to compute Bayesian updates, so several papers relax this assumption [Crapis et al. 2017; Besbes and Scarsini 2018]. Besbes and Scarsini [2018], for example, study both Bayesian buyers and buyers with limited rationality who can only observe the average of the past reviews. They analyze the conditions under which buyers can recover a product's true quality based on observed feedback. Unlike our paper, the buyers have private signals about the item, influencing their purchase decisions. Our model can be seen as situated between these two extremes because the purchase decisions depend on the average of the past reviews and the number of those reviews. Moreover, whereas Besbes and Scarsini [2018] analyze risk-neutral buyers, we study a form of risk aversion where buyers may not purchase even if the price is below the average reviews.

Unlike this prior research, we do not assume all buyers share a specific decision policy: we identify a broad family of decision policies under which our results hold. We only require that the buyer purchases the item if the price is sufficiently low.

3. NOTATION AND ONLINE LEARNING SETUP

In our model, an item is sold repeatedly to a sequence of distinct buyers over T rounds. Each buyer has a type $i \in [d]$, and there is an unknown distribution $\mathcal{P}$ over the types $[d]$. We use the notation $q_i = \Pr_{j\sim\mathcal{P}}[j = i]$ and $q_{\min} = \min_{i\in[d]} q_i$.

The *ex-ante* value of a buyer with type $i \in [d]$ is $\theta_i \in [0, 1]$ and their *ex-post* value is drawn from a distribution $\mathcal{D}_i$ with support $[0, 1]$ and mean θ_i, with $\theta_1 \leq \theta_2 \leq \cdots \leq \theta_d$. The seller knows $\theta_1, \dots, \theta_d$ but not $\mathcal{P}, \mathcal{D}_1, \dots, \mathcal{D}_d$. At each round $t \in [T]$:

(1) There is a set σ_{t-1} of reviews describing past buyers' types and *ex-post* values.
(2) The seller first sets a price $p_t \in [0, 1]$.
(3) A buyer arrives with type $i_t \sim \mathcal{P}$. They observe past reviews of buyers with type i_t, i.e $\Phi_{i_t,t} = \{v : (i, v) \in \sigma_{t-1} \text{ and } i = i_t\}$, and decide whether to purchase the item. We describe the buyer's purchasing model in more detail later in this section. The seller is unaware of the buyer's type i_t when they set the price.
(4) If the buyer purchases the item, they pay p_t and leave a review of (i_t, v_t) describing both their type and their *ex-post* value $v_t \sim \mathcal{D}_{i_t}$. In this case, $\sigma_t = \sigma_{t-1} \cup \{(i_t, v_t)\}$, and otherwise, $\sigma_t = \sigma_{t-1}$.

Our assumptions and model reflect practical e-commerce settings. First, quite often, it is reasonable to assume that sellers know customers' *ex-ante* values as they may have inside information. For instance, a skincare product vendor may know that a particular product works better on some skin types. However, buyers may not simply trust the seller if they were to publish this value, as the seller has every incentive to overstate this value to maximize revenue. A buyer would instead decide if a product is suitable for her via independent reviews from other customers. Second, for fairness reasons, in e-commerce platforms, sellers typically have to publish a single price for all customers and cannot sell the item at individualized prices. Third, if a buyer does not purchase an item, they will not leave a review, and the seller has no way of knowing their type or *ex-post* value.

Buyers' purchasing model. At time step t, the agent's purchase decision is defined by a threshold $\tau_t(\sigma_{t-1}, i_t) \geq 0$ that takes as input their type i_t and the reviews left by past agents. Intuitively, $\tau_t(\sigma_{t-1}, i_t)$ represents the agent's estimate of their value θ_{i_t} based on past reviews. The agent purchases the item if $p_t \leq \tau_t(\sigma_{t-1}, i_t)$.

A conservative agent would choose $\tau_t(\sigma_{t-1}, i_t)$ to be low in order to always guarantee that $\tau_t(\sigma_{t-1}, i_t) \leq \theta_{i_t}$, so that they only purchase when their *ex-ante* utility is non-negative. An extreme example of this type of conservatism would set $\tau_t(\sigma_{t-1}, i_t) = 0$, meaning that the agent would only purchase the item if offered for free. Optimizing revenue with such a conservative agent would be hopeless. Therefore, we impose the following natural lower bound on $\tau_t(\sigma_{t-1}, i_t)$:

Definition 3.1. Let Φ_t be the reviews left by agents with type i_t, i.e., $\Phi_t = \{v : (i, v) \in \sigma_{t-1}, i = i_t\}$. Let LB_t be the average minus a standard confidence term:

$$\mathrm{LB}_t = \begin{cases} 0 & \text{if } \Phi_t = \varnothing, \\ \max\left\{ 0,\ \frac{1}{|\Phi_t|} \sum_{v \in \Phi_t} v - \sqrt{\frac{1}{2|\Phi_t|} \ln \frac{t}{\eta}} \right\} & \text{else.} \end{cases}$$

We say that the agent on round t is *η-pessimistic* if, $\tau_t(\sigma_{t-1}, i_t) \geq \mathrm{LB}_t$.

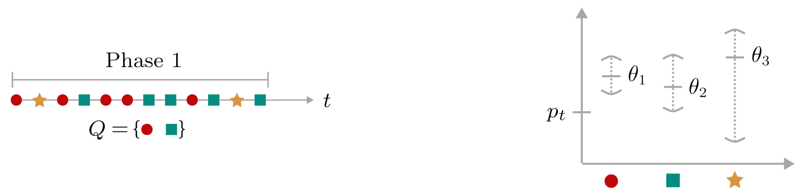
This uncertainty term corresponds to the standard Hoeffding confidence interval. Intuitively, as a buyer sees more reviews from his type, he is more certain about his *ex-ante* value. The $\ln t$ term is necessary to construct a valid confidence interval for an arbitrary algorithm as the data may not be independent: the algorithm's price may depend on previous reviews, which will affect future buyers and reviews. This $\ln t$ term is not fundamental—the lower bound does not use it.

Intuitively, the agents can be confident that *regardless* of the policy used by the seller, with probability $1 - \eta$, for all rounds $t \in [T]$, $\theta_{i_t} \geq \mathrm{LB}_t$. Therefore, if the price is lower than LB_t, an η-pessimistic agent will buy the item as they can be confident, based on past reviews, that their *ex-ante* utility $\theta_{i_t} - p_t$ will be non-negative. This restriction bounds the level of pessimism that the agents can display and thus makes it possible to set reasonable prices.

4. ALGORITHM OVERVIEW AND ANALYSIS

This section describes our algorithm, which has two phases. In the first phase, the algorithm sets a price of 0 for $t_0 = \tilde{\Theta}(T^{1/3} d^{2/3})$ rounds. The agent will buy the item at each round since the price is 0 and leave a review. This allows the algorithm to obtain i.i.d. samples from the type distribution $\mathcal{P}$. In phase 2 (i.e, the remaining $T - t_0$ rounds), the algorithm will ignore types that appeared too rarely during phase 1. Intuitively, customers of these types have a low probability of appearance and thus will have more uncertainty about their values due to fewer reviews. The uncertainty term will cause the lower confidence bound LB_t in Definition 3.1 to be small. As the seller will have to choose a low price to target these customers (even if their ex-ante value is large), they may have to forego higher revenue from more frequent customer types. Therefore, it is not worthwhile for the algorithm to target these customers. We use Q to denote the buyer types that appeared on a sufficiently large fraction of rounds, as in Figure 1a.

To describe the second phase, we use $\mathrm{rev}(p, Q) = p \Pr_{i \sim \mathcal{P}} [\theta_i \geq p \text{ and } i \in Q]$ to denote the expected revenue of a price p restricted to buyers in Q and $p^*(Q) =$



(a) Illustration of the first phase, at the end of which only the red circle and the green square are in Q.

(b) During the second phase, p_t is set low enough to ensure types in S_t will buy.

Fig. 1: Illustration of our algorithm with three types: a red circle, green square, and orange star.

argmax rev(p, Q). In this phase, our algorithm ignores the rare buyers not in Q and aims to set prices that compete with $p^*(Q)$. By competing with $p^*(Q)$, we show that our algorithm also competes with the optimal price p^*.

Observe that $p^*(Q) = \theta_{i_Q}$ for some $i_Q \in Q$. On each round $t > t_0$ of the second phase, our algorithm maintains a set S_t of "active types" such that i_Q is likely in S_t. The algorithm sets the price p_t low enough to ensure that if the current type i_t is in S_t, then the buyer will buy, as in Figure 1b. In particular, we define LB_{it} as the largest price the seller can set to ensure a purchase from a buyer of type i. We then set the price p_t to be the smallest LB_{it} of any type $i \in S_t$.

Next, for each active type $i \in S_t$, the algorithm estimates $\text{rev}(\theta_i, Q)$ along with upper and lower confidence bounds $\widehat{\mu}_{i,t}$ and $\check{\mu}_{i,t}$. The algorithm defines $i_0 = \min \{i \in S_t : \widehat{\mu}_{i,t} \geq \max_{k \in S_t} \check{\mu}_{k,t}\}$ to be the smallest active type such that θ_{i_0} may plausibly be $p^*(Q)$: for all $i < i_0$, the upper confidence bound on $\text{rev}(\theta_i, Q)$ is small ($\widehat{\mu}_{i,t} < \max_{k \in S_t} \check{\mu}_{k,t}$), so it is unlikely that $\theta_i = p^*(Q)$. The algorithm concludes round t by eliminating all types $i < i_0$ from the active set.

Regret definition. We define regret as the difference between:

(1) The algorithm's total expected revenue, and
(2) *(baseline)* The expected revenue of the optimal fixed price if the agents bought whenever their *ex-ante* value was larger than the price.

Under the baseline that we compete with, both the buyer and the seller are equipped with more information than in the learning problem: the seller knows all distributions $\mathcal{P}, \mathcal{D}_1, \ldots, \mathcal{D}_d$ and the buyers know their *ex-ante* values $\theta_1, \ldots, \theta_d$. Therefore, the seller knows *a priori* which customers to target to maximize revenue. Moreover, since the buyers do not need to learn their *ex-ante* values from reviews, the seller can extract higher revenue than they could from uncertain buyers who may only buy when the price is likely lower than their *ex-ante* value. Formally, let $b_t \in \{0, 1\}$ indicate whether the buyer bought on round t and let $p^* = \text{argmax}\, p \Pr_{i \sim \mathcal{P}}[\theta_i \geq p]$ be the price with highest expected revenue if the agents bought whenever their *ex-ante* value was larger than the price. Regret is defined as $\mathbb{E}[R_T] = Tp^* \Pr_{i \sim \mathcal{P}}[\theta_i \geq p^*] - \mathbb{E}[\sum p_t b_t]$.

Regret upper bound and proof overview. We contend with several sources of regret. The first phase of the algorithm, where the item is sold for free, inevitably leads to regret, so it must be made as brief as possible. The algorithm then completely disregards the buyer types that appeared too rarely during that phase. This results in a subset $Q \subseteq [d]$ of buyer types that appear sufficiently often. In the second phase, the algorithm only attempts to optimize revenue with respect to the buyers in Q instead of the entire set $[d]$, which contributes to regret. Finally, the buyers themselves do not know their *ex-ante* values, whereas, under our regret benchmark, buyers buy whenever their *ex-ante* value is larger than the price.

We obtain our final regret bound by analyzing these three sources of error. Our bound depends on the smallest probability that any given type appears, which we denote as $q_{\min}$. If $q_{\min}$ is not tiny—specifically, $q_{\min} > 2d^{-2/3}T^{-1/3}$—then we obtain a regret bound that scales with $\sqrt{T}$, namely $\tilde{O}(T^{1/2}q_{\min}^{-1/2} + T^{1/3}d^{2/3})$. Otherwise, for arbitrary $q_{\min}$, our regret bound scales with $T^{2/3}$ as $\tilde{O}\left(T^{2/3}d^{1/3} + T^{1/3}d^{2/3}\right)$.

Regret lower bound and proof overview. Typical bandit lower bounds rely on hypothesis testing arguments to show that any algorithm would struggle to distinguish between similar problems but with different optimal outcomes. Such an analysis would not capture the main difficulty in our setting: how fast customers estimate their *ex-ante* values from past reviews. Instead, our proof leverages the buyers' uncertainty to establish a $\tilde{\Omega}\left(T^{2/3}d^{1/3}\right)$ worst-case lower bound and a $\tilde{\Omega}(T^{1/2}q_{\min}^{-1/2})$ lower bound when $q_{\min}$ is large. This proves our regret bound is optimal.

Our proof constructs a hard instance where buyer types with a low probability of appearance have comparable *ex-ante* values to types with a high probability of appearance. In each round, an algorithm should decide whether it will target low-probability customers who may be less certain about their value due to fewer reviews and consequently have small LB_t. Keeping prices low to do so leads to low revenue in the current round, but ignoring low-probability customers by choosing a high price risks losing potentially high future revenue. We obtain a tight lower bound by carefully choosing the probability of appearance in our construction.

5. FUTURE DIRECTIONS

Many questions remain open. For example, we assumed that purchases always come with a noisy review. A challenging next step would be to develop pricing strategies when reviews are left with varying probabilities, mimicking real-world buyers.

We studied myopic buyers who make purchase decisions based on value estimates from historical reviews, regardless of the seller's policy. What if the buyers appear over several rounds and behave strategically to lower future prices?

We take a frequentist perspective on this problem. It is also possible to take a Bayesian view of this problem and impose a prior on the *ex-ante* value so that the buyer starts with some prior information. We expect adapting our main proof intuitions to that setting is possible. The main differences would be: *(i)* we would use Bayesian credible intervals instead of frequentist confidence intervals for the η-pessimism definition, *(ii)* we would control the Bayes' risk when estimating the *ex-ante* values instead of frequentist concentration arguments, and *(iii)* our final regret could have a nuanced dependence on this prior which may offer tighter bounds.

Another direction would be to explore the case where the seller does not know the buyers' *ex-ante* values. The key challenge would be related to the regret benchmark: we compete with the optimal price if each buyer knew their *ex-ante* value and bought whenever it was above the price (thus, the buyers are not learning). To compete with this benchmark, we require unbiased estimates of the revenue of different prices if the buyers bought when their *ex-ante* value was above the price. Computing these unbiased estimates is challenging: if a buyer does not buy on a given round, the algorithm does not learn their type, so it cannot tell whether the buyer has a low *ex-ante* value or he has a high value but a low confidence bound. We can circumvent this subtle challenge if the seller knows the buyers' *ex-ante* values. However, this is not possible if the *ex-ante* values are unknown.

REFERENCES

Acemoglu, D., Makhdoumi, A., Malekian, A., and Ozdaglar, A. 2022. Learning from reviews: The selection effect and the speed of learning. *Econometrica 90,* 6, 2857–2899.

Ashlagi, I., Daskalakis, C., and Haghpanah, N. 2016. Sequential mechanisms with ex-post participation guarantees. In *ACM Conference on Economics and Computation (EC)*.

Besbes, O. and Scarsini, M. 2018. On information distortions in online ratings. *Operations Research 66,* 3, 597–610.

Bose, S., Orosel, G., Ottaviani, M., and Vesterlund, L. 2006. Dynamic monopoly pricing and herding. *The RAND Journal of Economics 37,* 4, 910–928.

Boursier, E., Perchet, V., and Scarsini, M. 2022. Social learning in non-stationary environments. In *International Conference on Algorithmic Learning Theory (ALT)*. 128–129.

Braverman, M., Mao, J., Schneider, J., and Weinberg, M. 2018. Selling to a no-regret buyer. In *ACM Conference on Economics and Computation (EC)*.

Chamley, C. 2004. *Rational herds: Economic models of social learning.* Cambridge University Press.

Chawla, S., Devanur, N. R., Karlin, A. R., and Sivan, B. 2022. Simple pricing schemes for consumers with evolving values. *Games and Economic Behavior 134*, 344–360.

Crapis, D., Ifrach, B., Maglaras, C., and Scarsini, M. 2017. Monopoly pricing in the presence of social learning. *Management Science 63,* 11, 3586–3608.

Deng, Y., Schneider, J., and Sivan, B. 2019. Prior-free dynamic auctions with low regret buyers. In *Conference on Neural Information Processing Systems (NeurIPS)*.

Devanur, N. R., Peres, Y., and Sivan, B. 2014. Perfect Bayesian equilibria in repeated sales. In *Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*.

Feng, Z., Podimata, C., and Syrgkanis, V. 2018. Learning to bid without knowing your value. In *ACM Conference on Economics and Computation (EC)*.

Han, S. and Anderson, C. K. 2020. Customer motivation and response bias in online reviews. *Cornell Hospitality Quarterly 61,* 2, 142–153.

Ifrach, B., Maglaras, C., Scarsini, M., and Zseleva, A. 2019. Bayesian social learning from consumer reviews. *Operations Research 67,* 5, 1209–1221.

Kakhbod, A., Lanzani, G., and Xing, H. 2021. Heterogeneous Learning in Product Markets. *Available at SSRN 3961223*.

Kandasamy, K., Gonzalez, J. E., Jordan, M. I., and Stoica, I. 2023. VCG mechanism design with unknown agent values under stochastic bandit feedback. *Journal of Machine Learning Research 24,* 53, 1–45.

Kleinberg, R. and Leighton, T. 2003. The value of knowing a demand curve: Bounds on regret for online posted-price auctions. In *Symposium on Foundations of Computer Science (FOCS)*.

Nekipelov, D., Syrgkanis, V., and Tardos, E. 2015. Econometrics for learning agents. In *ACM Conference on Economics and Computation (EC)*.

Papadimitriou, C., Pierrakos, G., Psomas, A., and Rubinstein, A. 2022. On the complexity of dynamic mechanism design. *Games and Economic Behavior*.

Weed, J., Perchet, V., and Rigollet, P. 2016. Online learning in repeated auctions. In *Conference on Learning Theory (COLT)*.

Zhao, H. and Chen, W. 2020. Stochastic one-sided full-information bandit. In *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD)*.