Equilibrium and Learning in Fixed-Price Data Markets with Externality

Safwan Hossain
Harvard University
shossain@g.harvard.edu

Yiling Chen
Harvard University
yiling@seas.harvard.edu

Abstract

We propose modeling real-world data markets, where sellers post fixed prices and buyers are free to purchase from any set of sellers they please, as a simultaneous-move game between the buyers. A key component of this model is the negative externality buyers induce on one another due to purchasing similar data, a phenomenon exacerbated by its easy replicability. In the complete-information setting, where all buyers know their valuations, we characterize both the existence and the quality (with respect to optimal social welfare) of the pure-strategy Nash equilibrium under various models of buyer externality. While this picture is bleak without any market intervention, reinforcing the inadequacy of modern data markets, we prove that for a broad class of externality functions, market intervention in the form of a revenue-neutral transaction cost can lead to a pure-strategy equilibrium with strong welfare guarantees. We further show that this intervention is amenable to the more realistic setting where buyers start with unknown valuations and learn them over time through repeated market interactions. For such a setting, we provide an online learning algorithm for each buyer that achieves low regret guarantees with respect to both individual buyers' strategy and social welfare optimal. Our work paves the way for considering simple intervention strategies for existing fixed-price data markets to address their shortcoming and the unique challenges put forth by data products.

1 Introduction

Data plays a central role in the modern economy, and demand for it by individuals, organizations, and businesses has grown substantially due to the increasing value it provides. It has thus become the subject of trading, and designing and understanding markets for data (or more generally, information) has gained traction in research communities in recent years [1, 2, 3, 4, 5, 6]. Most works have focused on the fundamental question of pricing data - to determine the "right" value of data so that it is allocated to those who value it most [3, 4, 5, 6]. The mechanisms proposed in these works for two-sided markets are usually auction-based, and require buyers to report their private values for data [5, 6]. This is challenging, if not impossible, to achieve in practice due to the peculiar characteristics of data. A dataset's value to a buyer largely arises from how it can improve the performance of her model or learning algorithm. With modern machine learning algorithms more or less acting as a black-box, a buyer doesn't know her value for a dataset until she knows its contents and runs her algorithm on it, which cannot take place before a transaction. Furthermore, even if buyers were able to estimate value a priori through aggregate statistics about the dataset that sellers reveal, they may be unlikely to disclose this to an external party as it may leak significant information about their underlying model or use-cases. It is hence not surprising that most data markets in the real world, such as Snowflake Data Marketplace[1] and AWS Data

[1] https://www.snowflake.com/

Exchange[2], take a very simple format: sellers set a fixed price for each dataset and buyers can freely choose which datasets to purchase. The simplicity of fixed-price data markets is not only a huge operational advantage but also arguably necessary when it is impossible to solicit valuation information from buyers. Such markets are the focus of our paper.

Another important component of modeling data markets is accounting for the (usually negative) externality one buyer's purchase decision has on another. In a competitive setting, a buyer's value for data is predicated on the relative advantage training with it provides with respect to their peers. Equivalently, the value of data can depend on whether others have access to similar data. While externality is not unique to data and is present for other products as well, it is especially prominent in data markets due to another salient feature: data can be replicated at a mass scale with zero marginal cost and sold to multiple buyers, a phenomenon that exacerbates any externality datasets induce. Replication need not even be exact: sellers can offer different versions of the same dataset by injecting noise or by interleaving it with something innocuous. Data externality is thus a persistent phenomenon, and by linking a buyer's utility to another's decisions, it turns the data market into a game.

We are thus motivated to study fixed-price data markets between arbitrary buyers and sellers in the presence of externality. Despite their simplicity, which sidesteps many of the concerns above, this model has not been formally analyzed in literature and its welfare and equilibrium properties are unknown. By naturally modeling this setting as a game between the buyers and utilizing the concepts of game theory, we wish to systematically explore this landscape. We precisely ask:

- How well can fixed-price markets serve the purpose of allocating data to buyers, especially with respect to social welfare?
- How should we model buyer externality and how does this affect buyers' behaviour and market performance?
- Can simple market interventions improve the outcome of fixed-price data markets?
- If buyers learn their value for data by repeatedly interacting with a fixed-price data market, can they learn to play optimally and how much social welfare can be achieved?

1.1 Our contributions

We propose modeling a fixed-price data market as a simultaneous game between the buyers, whose utility depends on the gains derived from the purchased data, the externality they suffer due to others' actions, and possibly some cost associated with market intervention by the principal. We consider two broad families of externalities: *independent* and *joint*. Joint externality allows the externality that buyer j exerts on buyer i to depend on both buyers' actions, whereas independent externality assumes it only depends on buyer j's action. We formally define this setup in section 3, and to better understand the pure equilibrium properties of this game, namely its existence and welfare, we start by assuming buyers know their gains and externalities a priori in section 4. Across both externality families, we show that without any market intervention, pure equilibrium either does not exist or has poor social welfare, confirming the present concerns with fixed-price data markets. We then propose a parameterized revenue-neutral intervention that significantly improves the situation. For independent externality, it guarantees a dominant strategy Nash equilibrium whose social welfare approaches the optimal. It also guarantees pure equilibria for the more general joint externality model. While in the worst case, equilibria may be poor, we show that for any

[2]https://docs.aws.amazon.com/data-exchange/index.html

instance in this setting, there always exists an equilibrium with good welfare properties. Focusing on the simpler externality model, section 5 relaxes the main assumption that buyers know their valuations, and models a realistic setting where they repeatedly interact with the market and learn valuations accordingly. We show that the buyer learning problem can be interpreted as a multi-armed bandit instance with exponentially many arms. We adopt a known algorithm for this difficult setting and provide worst-case and average-case regret bounds for both individual buyers' strategy and cumulative social welfare. Lastly, section 6 discusses the implications of our work, its shortcomings, and avenues for further investigation.

2 Related Works

Bergemann and Bonatti [3] provide a broad overview of the growing literature on markets for information and data. Works here examine the question of how to optimally sell information according to some objectives, and take a mechanism design approach, with a focus on the seller's pricing problem. For example, many works consider a monopoly information holder directly selling a private random signal to buyers [7, 1, 2, 8, 9, 10]. The seller decides on a menu of information products (e.g. partially revealing the signal) and an associated price for each product to maximize her profit. Bergemann and Bonatti [11] consider selling cookies, while Mehta et al. [4] study selling a dataset, also in a monopoly seller setting.

When there are multiple sellers and multiple buyers, auction-based mechanisms have also been leveraged to design two-sided data marketplaces [5, 6]. In these works, the marketplaces themselves are not profit-driven but intend to maximally facilitate the matching of data to buyers. Bergemann and Bonatti [3] and Bergemann et al. [12] also study profit-driven data intermediaries who make a bilateral deal to purchase information from data holders and then sell the information to data buyers. Here, the data holders are consumers, and the data buyers are firms that can use purchased information to price discriminate.

While the problem of how to price data and information is fundamentally important, proposed mechanisms in the literature so far have not found their way into real-world data markets. A menu of partially-revealing information products can be too complex and cumbersome for buyers. As we explained at the beginning of this paper, buyers are unlikely to know their valuation for data a priori, which renders mechanisms that solicit buyer valuations (e.g. auctions) impractical. Our work thus takes an orthogonal direction and steps away from the design question of how to sell data and assumes that data sellers have already made decisions on their information products and prices. Motivated by the real-world fixed-price data markets, we take this market mechanism as a given and adopt a game-theoretic approach to understanding buyer behavior and dynamics in such markets.

A key feature of data markets that is central to our model is buyer externality. This has been highlighted and modeled in studying monopoly data selling [13, 14, 15, 10] and data auctions [6] in competitive environments. These works explicitly model downstream competition (e.g. trading in financial markets) among data buyers, and the negative buyer externality that arises therein. In our model, we abstract away the specifics of competitive environments and consider general families of buyer externality functions. There is another type of externality worth mentioning that has been considered in markets for information: the externality among data sellers. The value of one's data decreases when others decide to share their data, due to the social nature of data [12]. Seller externality is an important phenomenon to consider if sellers are modelled strategically, which we leave for future work.

Our work on the online setting relaxes the assumption that buyers know their utility function a priori. This spiritually parallels studying learning agents in other mechanisms such as auctions [16, 17, 18] and peer prediction [19]. We view studying markets with learning agents as a step toward a more realistic evaluation of market performance.

3 Model

We first introduce our model of a fixed-price data market with buyer externalities as an n-player simultaneous game. This model is flexible as it allows us to incorporate market intervention in the form of transaction fees and consider more realistic online settings where buyers start with no knowledge about their utility, but instead learn them through repeated interaction.

Market Structure: Consider n buyers, $\mathcal{B} = \{b_1, \ldots, b_n\}$, and k sellers, $\mathcal{L} = \{\ell_1, \ldots \ell_k\}$, who without loss of generality have one dataset each to sell. We consider a market where buyers are free to purchase from any subset of the sellers, each of whom posts a fixed price for their dataset and cannot refuse to sell. The data offered by the sellers can be arbitrarily correlated, and multiple buyers may buy from the same seller. Let Γ denote the power set of the sellers, and let $\gamma \in \Gamma$ denote a specific subset of sellers (we often refer to it as a *seller set* or *order*). We use subscripts to distinguish between seller sets chosen by various buyers - for example, γ_i refers to the set of buyers chosen by buyer i - and superscripts to distinguish or index between two seller sets irrespective of buyers - for example, γ^1 and γ^2 refers to two distinct seller sets. We consider buyers simultaneously submitting their orders, and use $S = (\gamma_1, \ldots, \gamma_n)$ to denote the purchase orders of all buyers.

Buyer's Gain and Externality: For a buyer i, their utility is predicated on three factors: (1) the value of the increased performance due to purchased data minus the fixed cost paid to sellers, referred to as the *gain*, $g_i(\gamma)$, (2) the negative externality caused by another buyer's purchase decision, $e_{ij}(\cdot)$, and (3) a potential payment (referred to as transaction cost), $\mathcal{T}_i(\mathcal{S})$, charged by the market principal. Without loss of generality, we assume the gain $g_i(\gamma)$ is bounded between $[-1, 1]$, and the "buy nothing" option ($\gamma = \emptyset$) has gain 0. Note a buyer that decides to "buy nothing" and not participate in the market, will still suffer the negative externality induced by others. The externality caused by an arbitrary buyer j's decision, $e_{ij}(\cdot)$, is also assumed to be bounded between $[0, \frac{1}{n-1}]$. This is without loss of generality and ensures that the total externality faced by a buyer is at most 1, and thus comparable to the gain. We consider a special family of externality functions as well as the most general externality functions:

- *Independent:* $e_{ij}(\gamma_j) : \Gamma \to [0, \frac{1}{n-1}]$
- *Joint:*[3] $e_{ij}(\gamma_i, \gamma_j) : \Gamma \times \Gamma \to [0, \frac{1}{n-1}]$.

The first family of functions assumes the externality that buyer i suffers from buyer j's choice depends only on the latter (i.e. buyer i's decision does not affect her externality). The second family relaxes this and allows the externality to depend on both decisions in arbitrary ways. As we will see in Section 4, the structure of the externality function significantly influences our equilibrium analysis.

[3] To further clarify notation, for two specific seller sets γ^1 and γ^2, $e_{ij}(\gamma^1, \gamma^2)$ implies buyer i owning γ^1 and buyer j owning γ_2. So in general, $e_{ij}(\gamma_1, \gamma_2) \neq e_{ij}(\gamma_2, \gamma_1)$, as the ownership is reversed. However, if we write $e_{ij}(\gamma_i, \gamma_j)$ the arguments already imply which buyer owns which set, and thus $e_{ij}(\gamma_i, \gamma_j) = e_{ij}(\gamma_j, \gamma_i)$

Transaction Cost: Lastly, the market principal may choose to intervene in the market through *transaction cost* $\mathcal{T}_i(\mathcal{S})$, which specifies the amount that buyer i needs to pay (or receive) when the set of purchase orders is $\mathcal{S}$. While the choice of $\mathcal{T}_i(\mathcal{S})$ is the market principal's design decision, we highlight that transaction costs that are revenue neutral and does not require revealing sensitive information are desirable. The former ensures the social welfare (sum of all buyer utility) is unaffected by this new term. Interventions that satisfy these desiderata can also be seen as a redistribution based on public information. We now formally define the buyer's utility.

Definition 1 (Utility and Welfare)**.** *For a buyer i with gain function g_i, externality functions $e_{ij}\,\forall\, j$, and transaction cost $\mathcal{T}_i$, we define her utility for a complete order profile $S = (\gamma_1, \ldots, \gamma_n)$ to be:*

$$u_i(S) = g_i(\gamma_i) - \sum_{j \neq i} e_{i,j}(\cdot) - \mathcal{T}_i(S). \tag{1}$$

We define the social welfare of S as the sum of all buyer utility: $\sum_i u_i(S)$.

Game and Solution Concept: We model buyers in the data market as playing a simultaneous-move game with the above utility function. We note that agent i's utility, and thus her best response, doesn't depend on any other agent's gain $g_j(\cdot)$, which may be kept private. We are primarily interested in analyzing the pure-strategy Nash equilibrium of this game under different externality and transaction cost structures. Beyond the existence of such equilibria, we also aim to compare the social welfare at equilibrium to the optimal social welfare. Common notions for comparison include *price of anarchy*, the ratio of the optimal social welfare to the welfare of the worst equilibrium, and *price of stability*, the ratio of the optimal social welfare to the welfare of the best equilibrium [20, 21, 22]. The former captures the worst-case scenario, and the latter the best-case. However, these are both multiplicative measures that make them unamenable to additive notions of regret that are common in the online analysis we do in section 5. As such, we define comparable additive notions called *welfare regret at equilibrium (WRaE)* to characterize the societal cost of rational self-interested behavior which arises in games.

Definition 2 (WRaE)**.** *Let $S^* = \arg\max_S \sum_i u_i(S)$ be the optimal strategy with respect to social welfare and let $\widehat{\mathcal{S}}$ be the set of all equilibrium strategies. We define the worst-casr* ***welfare regret at equilibrium*** *(WRaE) for our game $\mathcal{G}$ as: $sw(S^*) - \min_{\widehat{S} \in \widehat{\mathcal{S}}} sw(\widehat{S})$. Similarly, we define the best-case WRaE as: $sw(S^*) - \max_{\widehat{S} \in \widehat{\mathcal{S}}} sw(\widehat{S})$.*

Online Setting with Unknown Utilities: Section 4 analyzes the standard simultaneous-move setting wherein each buyer knows their utility for all possible action profiles. While this is clean and consistent with game-theory literature, it requires agents to have a priori knowledge about their gain and externality values [23, 24, 25]. As we have discussed, this is a strong assumption for data products since buyers do not generally know the benefit of a dataset until using it with their models, not to mention the externality. Hence in section 5, we extend our basic model to a randomized setting where gain and externality are both random variables with stationary distributions. Buyers do not know any property of this distribution a priori but can learn them through repeated interaction with the market over time. This setting can be interpreted within the canonical multi-armed bandit framework, where the key measure is that of regret [26]. For a complete purchase order S, let $U_i(S)$ denote the random utility buyer i attains, encapsulating the gain, externality, and transaction costs. Let S^* be the set of orders that maximizes social welfare, and let S^t be the set of orders submitted by all buyers at round t. Then we define our key metric in the online setting as follows:

Definition 3. *We define the expected* ***online welfare regret*** *over time* $t = 1, \dots, T$ *as the difference in total utility (across all agents) between the strategy set that maximizes social welfare* S^**, and the strategy taken at some time* t*,* S^t*:* $R_w(T) = \sum_{i=1}^{n} \mathbb{E}[U_i(S^*)] - \mathbb{E}[U_i(S^t)]$.

4 Data Markets Game with Known Utility

In this section, we consider the simultaneous-move data markets game wherein each buyer knows their gain and externality for each seller set a priori. This allows us to disentangle the game/market dynamics from the playing with learning valuations problem explored in section 5. We focus on the existence of pure strategy Nash Equilibrium and the WRaE under the two externality models. We start with the simpler of the two: independent externality.

4.1 Independent externality

In this setting, the externality buyer i faces due to buyer j depends only on the latter's action: $e_{ij}(\gamma_j)$. We first consider this game without any market intervention - $\mathcal{T}_i(\mathcal{S}) = 0$. While this admits a trivial equilibrium for each buyer with each simply choosing the seller set with the highest gain, even the best-case welfare regret of this equilibrium can be maximal. We formalize this below:

Theorem 1. *For the independent externality model with no intervention -* $\mathcal{T}_i(\mathcal{S}) = 0\,\forall i$ *- there is pure strategy Nash Equilibrium equilibrium. However, the b-WRaE of this setting is maximal -* $\Theta(n)$*.*

Proof. Note each buyer i's utility under these conditions is: $g_i(\gamma_i) - \sum_{j \neq i} e_{ij}(\gamma_j)$. The only aspect of this utility that buyer i can affect is γ_i, and thus she has a dominant strategy of choosing the source with the highest gain [4]. It is intuitive that everyone adopting such a strategy will not always lead to good welfare, even in the best case. Note that the maximum b-WRaE possible on any instance is $O(n)$. To see that there is an instance that achieves this, suppose the number of sellers and buyers are equal ($n = k$). For $k \in [1, \dots, n]$, define γ^k as the seller set containing only seller ℓ_k. For a buyer i, let $g_i(\gamma^i) = 1$ and $g_i(\gamma) = 1 - \epsilon$ for any $\gamma \neq \gamma^i$. Further, for any buyer pair i, j, let $e_{ij}(\gamma_j = \gamma^j) = \frac{1}{n-1}$ and $e_{ij}(\gamma_j = \gamma) = 0\,\forall \gamma \neq \gamma^j$. In this instance, the unique dominant strategy/equilibrium is each buyer i selecting γ^i, which results in 0 utility for all buyers. However, if each buyer i selects any other seller set aside from γ^i, they achieve utility $1 - \epsilon$ each. Since this is the only equilibrium in this instance, the b-WRaE is $n - n\epsilon \to n$ as $\epsilon \to 0$. ☐

This result illustrates the very real and unsatisfactory phenomenon that takes place in modern data markets. Buyers, not incentivized to care about their impact on others, make myopic decisions based purely on their own gain, to the detriment of both the individual and the collective. The impact of this is exacerbated by the combinatorial and easily replicable nature of data. We next consider addressing this through a market intervention in the form of transaction cost $\mathcal{T}_i(\mathcal{S})$. Our proposed cost for this setting, which we call the *net externality* (NE) transaction cost, charges each buyer proportional to the net difference in externality they are contributing to and suffering from. We formally define it below:

Definition 4. *Let* $\alpha \in [0, 1]$ *be a hyper-parameter to be selected by the principal. Then for a state/complete set of purchase orders* $\mathcal{S}$*, the net externality (NE) transaction cost for a buyer* i *is*

[4] Note that choosing the empty set and not participating is a valid strategy. Buyer i suffers externality regardless of what she chooses

given by:

$$\mathcal{T}_i^n(S) = \alpha\left(\sum_{j\neq i} e_{ji}(\gamma_i) - \sum_{j\neq i} e_{ij}(\gamma_j)\right) \tag{2}$$

Observe that this quantity can be negative, in which case the principal pays the buyer. Although this may seem unsatisfactory from a market principle perspective, this is just a redistribution and not an actual payment, since the overall transaction structure is revenue-neutral. This also implies that social welfare is unaffected. Secondly, the transaction cost does not depend on the gain values, which buyers may not wish to reveal. Lastly, note that $\alpha = 0$ corresponds to the setting with no intervention. With this in place, our results improve significantly — a dominant strategy still exists for each buyer, but the welfare regret now decreases to 0 linearly as a function of α.

Theorem 2. *Under the proposed transaction cost and independent externality, each buyer selecting* $\gamma_i^* = \arg\max_\gamma g_i(\gamma) - \alpha\sum_{j\neq i} e_{ji}(\gamma)$ *is a dominant strategy. Further, the worst-case WRaE of this equilibrium is* $n(1-\alpha)$ *and we show this to be tight.*

Proof. Expanding the utility function, we have $g_i(\gamma_i) - \sum_{j\neq i} e_{ij}(\gamma_j) - \alpha\left(\sum_{j\neq i} e_{ji}(\gamma_i) - \sum_{j\neq i} e_{ij}(\gamma_j)\right)$. Since a buyer can only influence γ_i, the second and last terms can essentially be ignored, and the buyer has a dominant strategy regardless of how others act. Note that while multiple strategies may lead to the same maximum, the outcome of all those strategies is the same. Regarding WRaE, we first prove the upper bound, before showing a specific instance that achieves this.

Upper bound: For an arbitrary instance, and let $\mathcal{S}^* = (\gamma_1^*, \gamma_2^*, \ldots, \gamma_n^*)$ denote the social optimal, and let $\widehat{\mathcal{S}} = (\widehat{\gamma_1}, \widehat{\gamma_2}, \ldots, \widehat{\gamma_n})$ denote the dominant strategy taken by the buyers. For notational ease, we will use g_i^* and $\widehat{g}_i$ to refer to $g_i(\gamma_i^*)$ and $g_i(\widehat{\gamma_i})$ respectively. We have the following which follows from the optimality of the dominant strategy:

$$\forall \text{ buyers } i, \widehat{g}_i - \alpha\sum_{j\neq i} e_{ji}(\widehat{\gamma_i}) > g_i^* - \alpha\sum_{j\neq i} e_{ji}(\gamma_i^*) \implies \sum_{j\neq i}\left(e_{ji}(\widehat{\gamma_i}) - e_{ji}(\gamma_i)\right) \leq \min\left(1, \tfrac{\widehat{g}_i - g_i^*}{\alpha}\right) \tag{3}$$

Next, we note that the quantity we wish to upper bound is:

$$\sum_{i=1}^{n}\left(g_i^* - \sum_{j\neq i} e_{ij}(\gamma_j^*) - \widehat{g}_i + \sum_{j\neq i} e_{ij}(\widehat{\gamma_j})\right)$$

$$= \sum_{i=1}^{n}\sum_{j\neq i}\left(e_{ij}(\widehat{\gamma_j}) - e_{ij}(\gamma_j^*)\right) - \sum_{i=1}^{n}(\widehat{g}_i - g_i^*) = \sum_{i=1}^{n}\sum_{j\neq i}\left(e_{ji}(\widehat{\gamma_i}) - e_{ji}(\gamma_i^*)\right) - \sum_{i=1}^{n}(\widehat{g}_i - g_i^*) \tag{4}$$

where the last transition is a reordering that holds since we are summing all possible externality pairs for orders $\widehat{\gamma}$ and γ^*. Next, we apply inequality 3 to upper bound equation 4:

$$(4) \leq \sum_{i=1}^{n}\min\left(1, \tfrac{\widehat{g}_i - g_i^*}{\alpha}\right) - (\widehat{g}_i - g_i^*) = \sum_{i=1}^{n}\min\left(1 - (\widehat{g}_i - g_i^*), \tfrac{\widehat{g}_i - g_i^*}{\alpha} - (\widehat{g}_i - g_i^*)\right) \tag{5}$$

Each summand is a min of two values, which is maximum when the two values are equal. Let $u_i = (\widehat{g}_i - g_i^*)$. Thus we have $\forall i, 1 - u_i = \frac{u_i}{\alpha} - u_i \implies u_i = \alpha$, and it follows that:

$$\sum_{i=1}^{n}\min\left(1 - (\widehat{g}_i - g_i^*), \tfrac{\widehat{g}_i - g_i^*}{\alpha} - (\widehat{g}_i - g_i^*)\right) \leq \sum_{i=1}^{n} 1 - \alpha = n(1-\alpha) \tag{6}$$

Lower bound: Consider a setting with only two sellers ℓ_1 and ℓ_2. For each buyer i, we have the following: $\forall i$, $g_i(\emptyset) = 0; g_i(\ell_1) = 0; g_i(\ell_2) = 1 - \alpha - \epsilon; g_i(\ell_1 \cap \ell_2) = 1$ and $\forall i, j$, $i \neq j$, $e_{ij}(\emptyset) = 0; e_{ij}(\ell_1) = 0; e_{ij}(\ell_2) = 0; e_{ij}(\ell_1 \cap \ell_2) = \frac{1}{n-1}$. Regarding optimal social welfare, note that there is no reason to choose $\emptyset$ or ℓ_1; suppose S is such that k buyers choose $\ell_1 \cap \ell_2$ and $n - k$ choose ℓ_2. We have:

$$sw(S) = (n-k)\left[1 - \alpha - \epsilon - \tfrac{k}{n-1}\right] + k\left[1 - \tfrac{k}{n-1}\right] = n - n\alpha - n\epsilon - k\left[\tfrac{n}{n-1} - \alpha - \epsilon\right] \tag{7}$$

As $\epsilon \to 0$, social welfare is maximized when $k = 0$, since $\alpha \in [0, 1]$ and $\frac{n}{n-1} > 1$. Thus the optimal is all agents choosing ℓ_2, for a social welfare of $n(1 - \alpha)$. As for equilibrium, observe that each buyer's dominant strategy is choosing $\ell_1 \cap \ell_2$, which has social welfare 0. Thus the worst-case WRaE $\to n(1 - \alpha)$. □

This is a strong result since the upper bound implies that on *all* instances, any equilibria has regret less than $n(1 - \alpha)$. Crucially, this decreases linearly with α, with $\alpha \to 1$ leading to the equilibrium solution attaining the optimal social welfare. The matching lower bound shows that this is tight. This positive result in the worst-case makes analysis of the best-case redundant and prompts us to consider the more general joint externality model. We take a similar approach there: we first try to understand the picture without intervention and then investigate whether the proposed transaction cost can address the shortcomings.

4.2 Joint Externality

Recall that for joint externality, when buyer i owns γ_i and buyer j owns γ_j, the externality buyer i suffers is $e_{ij}(\gamma_i, \gamma_j)$, and j suffers is $e_{ji}(\gamma_i, \gamma_j)$. This externality can be represented in matrix form. For any buyer pair (i, j) with $i < j$, define a state matrix $M_{ij} \in [0, 1]^{2^k \times 2^k}$ consisting of all possible seller set combinations, with the rows iterating buyer i's (buyer with the smaller index) orders and the columns iterating buyer j's (buyer with the larger index) orders. That is, $M_{ij}[\ell, k]$ represents the state where buyer i orders γ^ℓ and buyer j orders γ^k. Since M_{ij} and M_{ji} contain the same information, we stick to only using M_{ij} with $i < j$. By applying the externalities e_{ij} and e_{ji} element-wise to the state matrix M_{ij}, we get the matrix representation of the externalities, which we represent by $\mathcal{E}_{ij}$ and $\mathcal{E}_{ji}$. To make things clear, if $i < j$, then $\mathcal{E}_{ij}(l, k) = e_{ij}(M_{ij}(l, k)) = e_{ij}(\gamma_i = \gamma^l, \gamma_j = \gamma^k)$ and $\mathcal{E}_{ji}(l, k) = e_{ji}(M_{ij}(l, k)) = e_{ji}(\gamma_j = \gamma^k, \gamma_i = \gamma^l)$. We start with the existence of pure equilibria and in the following theorem prove that in the joint externality setting, not intervening ($\alpha = 0$) or intervening via the NE transaction cost (definition 4) with $\alpha \neq 0.5$, can lead to instances that do not have a pure strategy Nash equilibrium.

Theorem 3. *Consider a 2-buyer game with an arbitrary number of sellers. Then for any value of $\alpha \neq 0.5$ there exist matrices $\mathcal{E}_{12}$, $\mathcal{E}_{21}$ such that there is no pure strategy Nash Equilibrium. Note $\alpha = 0$ corresponds to no intervention.*

Proof. For two seller sets γ^a and γ^b, let the externality involving these be as follows (note $\mathcal{E}_{12}^{ab}$ is a sub-matrix of $\mathcal{E}_{12}$):

$$\mathcal{E}_{12}^{ab} = \begin{bmatrix} e_{12}(\gamma_1 = \gamma^a, \gamma_2 = \gamma^a) = 0.5 & e_{12}(\gamma_1 = \gamma^a, \gamma_2 = \gamma^b) = 0 \\ e_{12}(\gamma_1 = \gamma^b, \gamma_2 = \gamma^a) = 0 & e_{12}(\gamma_1 = \gamma^b, \gamma_2 = \gamma^b) = 0.5 \end{bmatrix}$$

$$\mathcal{E}_{21}^{ab} = \begin{bmatrix} e_{21}(\gamma_1 = \gamma^a, \gamma_2 = \gamma^a) = 0 & e_{21}(\gamma_1 = \gamma^a, \gamma_2 = \gamma^b) = 0.5 \\ e_{21}(\gamma_1 = \gamma^b, \gamma_2 = \gamma^a) = 0.5 & e_{21}(\gamma_1 = \gamma^b, \gamma_2 = \gamma^b) = 0 \end{bmatrix}$$

Let the externality e_{12} and e_{21} be 1 for any other seller set pair not included above — that is, the full externality matrices for this example, $\mathcal{E}_{12}$ and $\mathcal{E}_{21}$, are rank 2. Next, define the gain values as follows: $g_i(\gamma^a) = g_i(\gamma^b) = 1$ for $i \in \{1, 2\}$. Let the gain for all other seller sets be 0. With this, the utility for both buyers in these states can be expressed as (buyer 1 is the row player, and buyer 2 is the column player):

$$\begin{bmatrix} 1-(1-\alpha)0.5, 1-0.5\alpha & 1-0.5\alpha, 1-(1-\alpha)0.5 \\ 1-0.5\alpha, 1-(1-\alpha)0.5 & 1-(1-\alpha)0.5, 1-0.5\alpha \end{bmatrix}$$

Note that if $\alpha > 0.5$, then buyer 2 prefers state $(\gamma_a, \gamma_b) \succ (\gamma_a, \gamma_a)$, and $(\gamma_b, \gamma_a) \succ (\gamma_b, \gamma_b)$. Then for buyer 1 $(\gamma_b, \gamma_b) \succ (\gamma_a, \gamma_b)$, and $(\gamma_a, \gamma_a) \succ (\gamma_b, \gamma_a)$. This is a clockwise cycle and thus no state involving γ_a and γ_b will be a pure-strategy Nash equilibrium, as one of the buyers will always want to change their strategy. Further, as all other states (not involving γ_a or γ_b) have gain 0 and externality 1, none are preferred over states involving γ_a and γ_b. The converse happens if $\alpha < 0.5$ - i.e. a counter-clockwise cycle is induced. Thus, no pure-strategy Nash equilibrium exists in this setting for $\alpha \neq 0.5$. □

This result illustrates that pure strategy equilibrium may not only exist for an intervention-free market, but also for one with the NE transaction cost with $\alpha \neq 0.5$. While disappointing, the case of $\alpha = 0.5$ exhibits some interesting phenomenon that does always guarantee a pure equilibrium. We first show that any game with $\alpha = 0.5$ NE transaction cost can be reduced to a simpler game.

Lemma 1. *Any game with joint externality and NE transaction cost with $\alpha = 0.5$, can be equivalently represented as one where all buyers have symmetric externalities ($e_{ij}(\gamma_i, \gamma_j) = e_{ji}(\gamma_i, \gamma_j)\, \forall i, j$) and no transaction cost is used ($\alpha = 0$).*

Proof. For any instance $\mathcal{I}$ with joint externality function e_{ij}, e_{ji} and $\alpha = 0.5$ NE transaction cost, construct another instance $\widehat{\mathcal{I}}$ where the externality for a pair (i, j) is $\widehat{e}_{ij}(\gamma_i, \gamma_j) = \frac{1}{2}\left(e_{ij}(\gamma_i, \gamma_j) + e_{ji}(\gamma_i, \gamma_j)\right)$, and $\widehat{\alpha} = 0$. Note the externalities in this instance are symmetric: $\widehat{e}_{ij}(\gamma_i, \gamma_j) = \widehat{e}_{ji}(\gamma_i, \gamma_j)$. Next, the utility for any buyer i in this instance is: $u_i(\mathcal{S}) = g_i(\gamma_i) - \sum_{j \neq i} \widehat{e}_{ij}(\gamma_i, \gamma_j) = g_i(\gamma_i) - \frac{1}{2}\sum_{j \neq i} e_{ij}(\gamma_i, \gamma_j) - \frac{1}{2}\sum_{j \neq i} e_{ji}(\gamma_i, \gamma_j)$, which is exactly the same as buyer i's utility in the original instance $\mathcal{I}$. Thus when $\alpha = 0.5$ in the joint externality setting, we can represent it with this simpler instance with symmetric externality and no transaction cost. □

With the above reduction in hand, we can focus our attention on the pure strategy equilibrium and the associated welfare regret for symmetric externality. The results here will carry over to our true setting with arbitrary joint externality and an NE transaction cost with $\alpha = 0.5$. We start by showing the existence of a pure strategy Nash equilibrium, followed by quantifying its worst-case WRaE.

Theorem 4. *For symmetric externality functions, $e_{ij}(\gamma_i, \gamma_j) = e_{ji}(\gamma_i, \gamma_j)$ for all i, j, $i \neq j$, with no transaction cost ($\alpha = 0$), there always exists a pure strategy Nash Equilibrium. Further, this equilibrium can be reached by playing the sequential best response from any initial point.*

Proof. We prove the existence of equilibria, by showing that for any sequential best response update strategy, at each round, only a single buyer updates to their best response while all others remain at their existing strategy, will converge to a pure equilibrium. Without loss of generality, let the update order for buyers be $1, 2, \ldots, n$. We want to show that no best response sequence will result in a cycle. Due to our finite state space, this necessarily means a pure equilibrium is reached.

Suppose by contradiction that there is a cycle in the best response sequence. Let $\gamma_{1,t}$ denote the choice made by buyer 1 at iteration t, and upon playing their best response, they will move to iteration $t+1$: $\gamma_{1,t+1}$. The completion of a single iteration means all buyers have responded. Thus the existence of a cycle of length T implies a transition sequence of the following form: $(\gamma_{1,1},\gamma_{2,1},\ldots,\gamma_{n,1}) \to (\gamma_{1,2},\gamma_{2,1},\ldots,\gamma_{n,1}) \to \cdots \to (\gamma_{1,2},\gamma_{2,2},\ldots,\gamma_{n,2}) \to \cdots \to (\gamma_{1,T-1},\gamma_{2,T-1},\ldots,\gamma_{n,T-1}) \to \cdots \to (\gamma_{1,1},\gamma_{2,1},\ldots,\gamma_{n,1})$. Note the last state (at the end of iteration T) is equal to the first state since we assumed this is a cycle. Also, while every single buyer may not change their decision moving from t to $t+1$, at least one buyer must change to a strictly improving position at any round transition — if not, we are at a pure equilibrium, and best-response terminates. Without loss of generality, let the buyer who strictly changes her mind at every round be buyer 1. Consider the following inequality representing the transition from a round t to $t+1$ for any buyer i (strict inequality for buyer 1):

$$g_i(\gamma_{i,t}) - \sum_{j<i} e_{ij}(\gamma_{i,t},\gamma_{j,t+1}) - \sum_{j>i} e_{ij}(\gamma_{i,t},\gamma_{j,t}) \le g_i(\gamma_{i,t+1}) - \sum_{j<i} e_{ij}(\gamma_{i,t+1},\gamma_{j,t+1}) - \sum_{j>i} e_{ij}(\gamma_{i,t+1},\gamma_{j,t})$$

Summing these n equalities representing the transition between two consecutive rounds, and using the fact that $e_{ij}(\gamma_i,\gamma_j) = e_{ji}(\gamma_i,\gamma_j)$ and that for buyer 1 the above inequality is strict, we have:

$$\sum_{i=1}^{n} g_i(\gamma_{i,t}) - \sum_{i=1}^{n} g_i(\gamma_{i,t+1}) < \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} e_{ij}(\gamma_{i,t},\gamma_{j,t}) - \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} e_{ij}(\gamma_{i,t+1},\gamma_{j,t+1}) \tag{8}$$

Adding up the inequalities for every transition from $1 \to T$, we note a telescoping sum forming. The rightmost terms on either side of the inequality for the $t \to t+1$ transition cancel with the leftmost terms in the $t+1 \to t+2$ transition. Thus, we are left with:

$$\sum_{k=1}^{n} g_k(\gamma_{k,1}) - \sum_{k=1}^{n} g_k(\gamma_{k,T}) < \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} e_{ij}(\gamma_{i,1},\gamma_{j,1}) - \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} e_{ij}(\gamma_{i,T},\gamma_{j,T}) \tag{9}$$

Recall that since this is a T length cycle, it implies $\forall k$, $\gamma_{k,1} = \gamma_{k,T}$. Thus LHS and RHS are both 0 and we are left with $0 < 0$, a contradiction. Thus no sequential best-response path can lead to a cycle, and due to the finite state space, all must converge to a pure strategy Nash Equilibrium. $\square$

Theorem 5. *For symmetric externality functions, the worst-case WRaE is* $\Omega(n)$

Proof. Consider a setting with n buyers and two sellers ℓ_1 and ℓ_2. Thus there are three meaningful possibilities (we ignore the degenerate $\emptyset$ option by assuming it has 0 gain for all buyers) for the buyers: $\ell_1, \ell_2, \ell_1 \cap \ell_2$. Let the gain vector for buyer 1 be: $[g_1(\ell_1) = 1, g_1(\ell_2) = 1, g_1(\ell_1 \cap \ell_2) = 1-4\epsilon]$ and for all other buyers k be: $[g_2(\ell_1) = 0, g_2(\ell_2) = 1, g_2(\ell_1 \cap \ell_2) = 1]$. Consider the following externality matrix between buyer 1 and the other buyers (rows represent the action space of buyer 1), as well as the externality matrix between buyer i and j, $i \neq j \neq 1$ (rows represent the action space of the buyer with the smaller index). The seller sets are ordered $\ell_1, \ell_2, \ell_1 \cap \ell_2$.

$$\mathcal{E}_{k1} = \mathcal{E}_{1k} = \begin{bmatrix} \frac{1}{n-1} & \frac{1}{n-1} & \frac{4\epsilon}{n-1} \\ \frac{1}{n-1} & \frac{1-\epsilon}{n-1} & \frac{1}{n-1} \\ \frac{1}{n-1} & \frac{1}{n-1} & \frac{\epsilon}{n-1} \end{bmatrix} \;;\; \mathcal{E}_{ij} = \mathcal{E}_{ji} = \begin{bmatrix} \frac{1}{n-1} & \frac{1}{n-1} & \frac{1}{n-1} \\ \frac{1}{n-1} & \frac{1-\epsilon}{n-1} & \frac{1}{n-1} \\ \frac{1}{n-1} & \frac{1}{n-1} & 0 \end{bmatrix}$$

We first observe that all buyers choosing ℓ_2 is a pure equilibrium since any buyer unilaterally moving to ℓ_1 or $\ell_1 \cap \ell_2$ will lead her to incur a higher externality without an increase in gain. This equilibrium has total welfare of $n\epsilon$. Next observe that every agent choosing $\ell_1 \cap \ell_2$ is the social optimal with utility $(n-1)(1-\epsilon/(n-1)) + (1-4\epsilon-\epsilon) = n-6\epsilon$. As $\epsilon \to 0$, the welfare regret turns to n as desired. $\square$

While the welfare regret in the above instance is quite poor given our bounded externalities and gains, this is a worst-case result, similar to the price of anarchy. That is, this expresses the welfare regret of the *worst* equilibrium and does not consider other equilibria whose regret may be much better. For example, in the instance outlined in theorem 5, observe that buyer 1 choosing ℓ_1 while everyone else choosing ℓ_3 is another equilibrium. This profile has utility $1 - 4\epsilon$ for buyer 1 as opposed to $1 - 5\epsilon$ if she were to move to $\ell_1 \cap \ell_2$ and 0 if she moves to ℓ_2. For other buyers, moving to another state unilaterally increases externality without any gain. This equilibrium profile has total welfare $(n-1)(1 - 4\epsilon/(n-1)) + (1 - 4\epsilon) = n - 8\epsilon$, which is only 2ϵ worse than the social welfare optimal. This leads us to resolve the best-case WRaE.

Theorem 6. *In the joint externality setting with $\alpha = 0.5$ NE transaction cost, there always exists a pure equilibrium whose welfare regret at equilibrium - WRaE - is at most $n/2$.*

Proof. Consider starting from the social optimal $S^* = (\gamma_1^*, \dots, \gamma_n^*)$, where each buyer attains utility $u_i(S^*) = g_i(\gamma_i^*) - \sum_{j \neq i} e_{ij}(\gamma_i^*, \gamma_j^*)$ for a total welfare of $sw(S^*)$. If the social optimal is an equilibrium, then the best-case welfare regret is 0. If not, then by theorem 4 we know that an equilibrium can be reached by playing the sequential best response (buyers play the best response one by one). Without loss of generality, suppose buyer 1 is unhappy at the social optimal and changes their decision from γ_1^* to γ_1^1 (denotes that this is buyer 1's first best response), and define this new state as $S_{1,1}$ (denotes buyer 1 has played their first best response). Thus:

$$g_1(\gamma_1^1) - \sum_{j \neq 1} e_{1j}(\gamma_1^1, \gamma_j^*) > g_1(\gamma_1^*) - \sum_{j \neq 1} e_{1j}(\gamma_1^*, \gamma_j^*) \tag{10}$$

Define $\Delta g_i(\gamma_i^t, \gamma_i) = g_i(\gamma_i^t) - g_i(\gamma_i)$ and $\Delta e_{ij}(\gamma_i^t, \gamma_i, \gamma_j) = e_{ij}(\gamma_i^t, \gamma_j) - e_{ij}(\gamma_i, \gamma_j)$. With this new notation, we can express equation 10 as $\Delta g_1(\gamma_1^1, \gamma_1^*) > \sum_{j \neq 1} \Delta e_{1j}(\gamma_1^1, \gamma_1^*, \gamma_j)$ and exploiting symmetric externality, the social optimal at $S_{1,1}$ can be succinctly expressed as $sw(S_{1,1}) = sw(S^*) + \Delta g_1(\gamma_1^1, \gamma_1^*) - 2\sum_{j \neq 1} \Delta e_{1j}(\gamma_1^1, \gamma_1^*, \gamma_j)$. This relationship is in fact satisfied between any two consecutive best response steps: let state $S_{i,k}$ where buyer i plays her k^{th} best response occurs right after state $S_{j,\ell}$ where buyer j plays her ℓ^{th} best response. The following two invariants hold (t_j denotes the number of times buyer j has played the best response up to that state):

$$\Delta g_i(\gamma_i^k, \gamma_i^{k-1}) > \sum_{j \neq i} \Delta e_{ij}(\gamma_i^k, \gamma_i^{k-1}, \gamma_j^{t_j}) \bigwedge sw(S_{i,k}) = sw(S_{j,\ell}) + \Delta g_i(\gamma_i^k, \gamma_i^{k-1}) - 2\sum_{j \neq i} \Delta e_{ij}(\gamma_i^k, \gamma_i^{k-1}, \gamma_j^{t_j})$$

Let $S_e = (\gamma_i^e, \dots, \gamma_n^e)$ be the pure equilibrium state reached by this best response cycle starting from social optimal. By repeatedly applying the two invariants above to $sw(S^*)$ and expanding, we can relate $sw(S^e)$ to $sw(S^*)$ as follows:

$$sw(S^e) = sw(S^*) + \sum_{i=1}^{n} \Delta g_i(\gamma_i^e, \gamma_i^*) - 2\sum_{i=1}^{n} \sum_{j \neq i} \Delta e_{ij}(\gamma_i^e, \gamma_i^*, \gamma_j^*) \tag{11}$$

Observe that summing the first invariant across n implies that:

$$\sum_{i=1}^{n} \Delta g_i(\gamma_i^e, \gamma_i^*) \geq \sum_{i=1}^{n} \sum_{j \neq i} \Delta e_{ij}(\gamma_i^e, \gamma_i^*) \tag{12}$$

Let $\sum_i \sum_{j \neq i} \Delta e_{ij}(\gamma_i^e, \gamma_i^*) \triangleq c$, and note that $c \in [0, n]$ due to the boundedness of e_{ij}. Thus we have: $\sum_{i=1}^{n} \Delta g_i(\gamma_i^e, \gamma_i^*) \geq c \implies sw(S^*) \leq n - c$ since maximum utility for any buyer is 1. Thus, WRaE $\leq n - c$. Next, observe by equation 11, $sw(S^e) \geq sw(S^*) + c - 2c = sw(S^*) - c \implies$ WRaE $\leq c$. Combining these, we have that WRaE $= \min(c, n-c) \leq \frac{n}{2}$. $\square$

To summarize these results, while the picture in the joint externality setting is grim without intervention, and even general NE transaction cost struggles, $\alpha = 0.5$ offers hope. It reduces the game to a simpler one based on symmetric externality, a setting that guarantees a pure equilibrium that can be reached by the sequential best response. While in the worst case, this equilibrium can have high welfare regret, there always exists an equilibrium with "reasonable" regret. Observe that the $n/2$ best-case regret in theorem 6 roughly matches the worst-case upper bound for the independent setting (theorem 2) with $\alpha = 0.5$. This suggests $\alpha = 0.5$ is a silver bullet: it imposes a reasonable cost, charging each buyer half of the net externality they induce, while have favourable properties regardless of the externality model of the buyer. Nonetheless, the independent externality setting does offer stronger results: it guarantees a dominant strategy for each buyer regardless of α, and the welfare regret degrades gracefully. As such, we focus on independent externality for the online setting wherein buyers learn their valuation, leaving online analysis of joint externality for future work.

5 Online setting with learned valuation

We now relax the key assumption made in the preceding section: gain and externality values are known by buyers a priori. We now consider a model where the gains and externalities are random variables whose properties are unknown to the buyer. Buyers interact with this market repeatedly and attempt to learn these quantities over time. This repeated interaction is consistent with real-world behaviour since firms consistently purchase fresh data to ensure their models remain state-of-the-art and capture the latest trends. Buyers in this setting face a non-trivial exploration vs exploitation problem: they must balance exploring their options and learning valuations, with acting optimally based on existing knowledge. This dichotomy faced by buyers can be well modeled by casting it as a multi-armed bandit (MAB) problem. Each buyer faces a set of 2^k "arms", representing the different seller combinations available to them, and all n buyers simultaneously "pull an arm" each round by interacting with the market. Each round terminates with buyers obtaining their ordered data and observing the stochastic gain and externality it induces. Although this setting is challenging owing to the countable yet exponentially large number of "arms", we show that buyers adopting existing state-of-the-art bandit algorithms can lead to low regret not only with respect to their best individual strategy but also with regard to the hindsight optimal social welfare. We first formally define the specifics of this online setting.

5.1 Online Model

Buyer Utility: In the online setting, we assume that both the gain and externality are stationary random variables bound to the same range as before: $G_i(\gamma_i) \in [-1, 1], E_{ij}(\gamma_i) \in [0, \frac{1}{n-1}] \forall\, i, j$. We use g_i and e_{ij} to denote samples from these random variables, $\overline{g}_i$ and $\overline{e}_{ij}$ to denote their expected values, and $\widehat{g}_i$ and $\widehat{e}_{ij}$ to denote sample means based on available observations. For the online model, we focus on independent externality (externality buyer i faces due to j, depends only on the latter - $E_{ij}(\gamma_j)$) with NE transaction cost. Due to this independence, buyers always have a dominant strategy: maximize the terms in their utility that depend on their actions. Like buyers, we assume the principal also does not know the expected externalities a priori, and thus applies the NE transaction cost based on the samples E_{ij}: $\alpha\left(\sum_{j\neq i} E_{ji}(\gamma_i) - \sum_{j\neq i} E_{ij}(\gamma_j)\right)$. With this cost, the component of each buyer's random utility that depends on their action takes on a familiar form: $G_i(\gamma_i) - \alpha \sum E_{ji}(\gamma_i) \triangleq U_i^d(\gamma_i)$. We call this the *dominant strategy* DS utility as it is each buyer's dominant strategy to maximize this (see theorem 2). This quantity for each γ is analogous to the "reward" for an arm in a standard MAB setting. We use $\overline{u}_i^d(\gamma_i) = \mathbb{E}[U_i^d(\gamma_i)] = \overline{g}_i(\gamma_i) - \alpha \sum \overline{e}_{ji}(\gamma_i)$ to denote the expected DS utility and $\widehat{u}_i^d(\gamma_i, h) = \frac{1}{h}\left(U_i^d(\gamma_i)\right)$ as the empirical mean of DS utility based on h samples. For brevity, we also write $\widehat{u}_{i,t}^d(\gamma_i)$ to imply the empirical mean with samples collected till time t.

Regret: Online algorithms are measured by their *regret*, the difference in cumulative loss (from rounds $t = 1, \ldots, T$) between the action taken by the algorithm and the optimal action with full information. Each buyer's goal is to effectively learn valuations so that they can play as close as possible to this optimal. This leads us to consider two types of regret. The *dominant strategy regret* characterizes the cumulative difference in DS utility for a single buyer. The *online welfare regret*, briefly mentioned in section 3, captures the cumulative difference in welfare between the social welfare optimal under full information, and the strategy taken by buyers at a round t. We formally define them as follows:

Definition 5. *The expected* ***dominant strategy regret*** *for a buyer i is the difference in the DS utility between their dominant strategy under full information, γ_i^d, and their strategy at time t, γ_i^t:*

$$R_d^i(T) = \mathbb{E}\left[\sum_{t=1}^{T} U_i^d(\gamma_i^d) - U_i^d(\gamma_i^t)\right] = \sum_{t=1}^{T}\left(g_i(\gamma_i^d) - \alpha \sum_{j \neq i} e_{ji}(\gamma_i^d) - g_i(\gamma_i^t) + \alpha \sum_{j \neq i} e_{ji}(\gamma_i^t)\right) \tag{13}$$

Definition 6. *The expected* ***online welfare regret*** *is the difference in total utility (across all agents) between the strategy that maximizes social welfare, $S^* = (\gamma_1^*, \ldots, \gamma_n^*)$, and the strategy taken at time t, $S^t = (\gamma_1^t, \ldots, \gamma_n^t)$. Recall that since NE transaction cost is revenue-neutral, social welfare is independent of this quantity, and can thus this regret be expressed as:*

$$R_w(T) = \sum_{t=1}^{T}\sum_{i=1}^{N}\left(g_i(\gamma_i^*) - \sum_{j \neq i} e_{ij}(\gamma_j^*) - g_i(\gamma_i^t) + \sum_{j \neq i} e_{ij}(\gamma_j^t)\right) \tag{14}$$

In this work, we provide upper bounds for these regret metrics, which can be stated in a few ways. If the regret bounds are specific to an instance, also known as instance dependent, we refer to them with $R_w(T, \mathcal{I})$, $R_d^i(T, \mathcal{I})$. If the regret bound holds for all instances, i.e. instance independent, we use $R_w(T)$ and $R_d^i(T)$. Occupying a middle ground is Bayesian regret which is the expected regret over all instances: $\mathbb{E}_{\mathcal{I}}[R_w(T, \mathcal{I})]$ and $\mathbb{E}_{\mathcal{I}}[R_d^i(T, \mathcal{I})]$. Bayesian regret is dependent on the distribution over the instances, which we shortly clarify.

Multi-armed bandits: Multi-armed bandit is a very general framework to model the exploration-vs-exploitation trade-off in online decision making [26]. A principal must repeatedly select from a set of possible arms which yield stochastic rewards from an unknown stationary distribution. Their objective is to explore and learn the optimal arm, without incurring too much regret in the process. Mapped to our setting, each buyer is essentially faced with an MAB instance over the possible seller combinations and aims to choose seller sets with high DS utility. The buyer's objective can analogously be viewed as minimizing their DS regret (definition 5). We want to understand algorithms to achieve this and quantify how the resulting strategy affects online welfare regret. We borrow standard terminology and definitions from bandit literature. We refer to $\Delta_i(\gamma_i^t) = \overline{u}_i^d(\gamma_i^d) - \overline{u}_i^d(\gamma_i^t)$ as the expected DS utility gap of a seller set γ_i^t with respect to the dominant strategy optimal γ_i^d for a buyer i. Note that $\Delta_i(\gamma) \geq 0 \ \forall \gamma$. Further, we use $n_t(\gamma_i)$ to denote the number of times set γ_i has been selected by buyer i over the rounds.

Sellers in a metric space The key challenge facing buyers is the exponential number of seller sets: 2^k possible options for a buyer to consider and learn. Common bandit algorithms like Upper Confidence Bound (UCB) still work but can be unsatisfactory due to this large number of arms [27]. However, we can exploit a natural assumption in our setting to seek better bounds. Order the sellers $1, \ldots, k$, and note any seller set γ can be represented as a k bit string, with bit i signifying whether seller ℓ_i is included or not. With this representation, for two seller sets γ^1 and γ^2, define $D_h(\gamma^1, \gamma^2)$ as the Hamming distance between γ^1 and γ^2. Hamming distance counts the number of bit positions in which the two binary strings differ. While it takes on integer values in $[0, \ldots, k]$, we normalize this so that the range is $[0, \ldots, 1]$ in increments of $\frac{1}{k}$. In any case, if γ^1 and γ^2 are close in Hamming distance, then the seller sets consist of roughly the same sellers. Thus it stands to reason that the gain and externalities a buyer experiences from these two seller sets will not be drastically different. We formalize this with the following *metric* property: for each buyer i and any pair of seller sets γ^1 and γ^2, we assume that $|\overline{g}_i(\gamma^1) - \overline{g}_i(\gamma^2)| \leq \lambda_g D_h(\gamma^1, \gamma^2)$ and $|\overline{e}_{ij}(\gamma^1) - \overline{e}_{ij}(\gamma^2)| \leq \lambda_e D_h(\gamma^1, \gamma^2)$ for all j. Without loss of generality, we fix $\lambda_g = 1$ and $\lambda_e = \frac{1}{n-1}$ to be consistent with previous ranges. Thus for two seller-sets γ^1, γ^2: $\overline{u}_i^d(\gamma^1) - \overline{u}_i^d(\gamma^1) \leq 2D_h(\gamma^1, \gamma^2)$. While the gains and externalities faced by a buyer are expected to obey the metric property, no assumptions are made *across* different buyers - i.e. no relation assumed between $\overline{g}_i(\gamma^1)$ and $\overline{g}_j(\gamma^2)$. We now define the Bayesian Regret over such metric instances, as well as formally introduce the notion of covering.

Definition 7. *Let $\mathcal{F}$ denote the set of all instances that satisfy the metric property for gains and externalities. Consider the uniform distribution over all instances in this family, Unif($\mathcal{F}$), and let I be one such instance. Then the Bayesian dominant strategy regret over this distribution is given by $\mathbb{E}_{I \sim \mathit{Unif}(\mathcal{F})}[R_d^i(T, I)]$. Similarly, the Bayesian online welfare regret is given by: $\mathbb{E}_{I \sim \mathit{Unif}(\mathcal{F})}[R_w(T, I)]$*

Definition 8. *Given a metric space (X, d) and $r > 0$, an element $x \in X$ is* ***covered*** *by $x' \in X$ if $d(x, x') \leq r$. A set $S \subseteq X$ is called an r-**cover** if every $x \in X$ is covered by some $x' \in S$.*

5.2 Algorithm

There is a range of work investigating bandit algorithms with arms in continuous space [28, 29, 30]. In many of these setting, a common trick is to perform an $\epsilon-$cover over the arm-space and then run a standard algorithm like UCB over the elements in the cover [28, 29]. While this uniform covering approach is promising in Euclidean space where the number of arms is infinite and covering bounds are tight, it is less attractive in Hamming space. Further, it can be wasteful since it discretizes areas of low and high reward with equal density. Kleinberg et al. [31] address this using an adaptive approach called the *zooming* algorithm, which effectively discretizes a region proportional to its reward. The algorithm maintains a set of *active* arms, each with an associated confidence radius. Any arm that falls within the confidence radius of an active arm (under the given metric) is said to be *covered* by it. At any given round, it maintains an invariant that that all possible arms are covered by some active arm by updating the active set as needed. The algorithm selects an arm from the active set at each round using the UCB rule: choose the arm with the highest empirical mean reward plus confidence radius. This confidence interval is inversely proportional to the number of times an arm is chosen, which effectively discretizes regions with high rewards more finely. The zooming algorithm is state-of-the-art for bandits in metric spaces, and Kleinberg et al. [31] illustrates its optimality under certain conditions.

Due to its strengths, we consider each buyer adopting the zooming algorithm. While the algorithm requires only slight modifications from the original, obtaining meaningful and interpretable regret bounds requires more careful analysis. The value of a choice γ depends on both the gain and externality which leads us to use a stronger concentration result. More importantly, distance in Hamming space is quantized: that is, the distance between two elements can only take on k distinct values. This means that any two elements are at least $\frac{1}{k}$ apart and there are a large number of elements that are the same distance apart and thus cannot be strictly compared. Beyond addressing these issues, we further give a Bayesian regret bound over all metric instances to assess its performance in the average case. The techniques used to address these issues may be of independent interest in bandit settings with countable but exponential arms. Lastly, we not only show that buyers can get close to their dominant strategy (i.e. minimize dominant strategy regret), but also that this process leads to high social welfare (i.e. also minimizes online welfare regret). We begin by presenting a key definition below and then formally stating algorithm 1 which outlines the zooming algorithm for each buyer.

Definition 9. *We define the confidence interval buyer i has for seller set γ at time t by* $c_t^i(\gamma) = \sqrt{\frac{12 \log T}{n_t^i(\gamma)+1}}$*, where $n_t^i(\gamma)$ is the number of times buyer i has selected seller set γ till time t.*

Algorithm 1 Zooming algorithm for buyer i

Active set $\mathcal{A}_i$ randomly initialized with a single seller set
Confidence radius for each $\gamma \in \mathcal{A}_i$ is $c_t^i(\gamma) = \sqrt{\frac{12 \log T}{n_t^i(\gamma)+1}}$
for $t = 1, \dots, T$ **do**
    `// An active seller-set` $\gamma' \in \mathcal{A}_i$ `covers a set` γ `if` $D_h(\gamma, \gamma') \leq c_t^i(\gamma')$
    if there are γ not covered (under Hamming distance) by seller-sets in $\mathcal{A}_i$
        Pick any uncovered choice and append it to $\mathcal{A}_i$
    End
    Define $\text{UCB}_i(\gamma) = \widehat{u}_i^d(\gamma) + 2c_t^i(\gamma)$ for each active choice γ
    Select the arm with the highest UCB_i value.
End

5.3 Dominant strategy regret

This section provides one of the key result in our online setting: bounding the dominant strategy regret of the zooming algorithm for each buyer i, both in the worst case and in expectation. The broad flow of this proof roughly mirrors that of Kleinberg et al. [31] can be sketched out as follows. We first show that with high probability, empirically computed DS utility $\widehat{u}_i^d$ is concentrated around the expected DS utility $\overline{u}_i^d$ with high probability. Conditioned on this, we upper bound the Hamming distance between any two active seller sets in $\mathcal{A}_i$, and also bound the number of times an active choice can be selected. These results allow us to compute the regret contributed by each active arm over the course of the algorithm and provide a general regret statement that can be used for both worst-case and expected-case results. We begin with the following two lemmas:

Lemma 2. *Define the event* $\mathcal{E} = \left\{\forall i, \forall \gamma, \forall t \in [1, \dots, T], |\widehat{u}_{i,t}^d(\gamma) - \overline{u}_i^d(\gamma)| \leq c_t^i(\gamma_i)\right\}$ *for any metric instance. Then* $Pr[\mathcal{E}] \geq 1 - \frac{2}{T^2}$*.*

Proof. Consider a buyer i and a seller set γ. If it is not in the active set or has never been selected, the lemma trivially holds since $n_i^t(\gamma) = 0$, implying $c_t^i(\gamma) > 2$. If γ has been selected at least once, denote the ℓ^{th} realizations by $y_i^\ell = (g_i^\ell(\gamma), e_{1i}^\ell(\gamma), e_{2i}^\ell(\gamma), \dots, e_{ni}^\ell(\gamma))$. Next, note that we can write $\widehat{u}_i^d(\gamma, h) = \widehat{u}_i^d(y_i^1, \dots, y_i^h) = \frac{1}{h}\sum_{\ell=1}^h \left(g_i^\ell(\gamma) - \alpha \sum_{j\neq i} e_{ji}^\ell(\gamma)\right)$ - this is the empirical mean of EDS utility from the first h times γ has been selected by buyer i. Observe that $\widehat{u}_i^d(y_i^1, \ldots, y_i^h)$ satisfies the bounded difference property: $\forall \ell \in [1, \ldots, h], \forall y_i^\ell, \sup_{y_i'^\ell \in [0,1]^n} |\widehat{u}_i^d(\ldots, y_i^\ell, \ldots) - \widehat{u}_i^d(\ldots, y_i'^\ell, \ldots)| \leq \frac{2}{h}$. This allows us to apply McDiarmid's inequality on $\widehat{u}_i^d$ over the random samples our h selections of γ results in, denoted by $Y_i^1, \ldots, Y_i^h$. We slightly abuse notation to write $\widehat{u}_i^d$ as a function of these random variables and note that $\mathbb{E}[\widehat{u}_i^d(Y_1, \ldots, Y_h)] = \overline{g}_i(\gamma) - \alpha \sum_{j \neq i} \overline{e}_{ji}(\gamma) = \overline{u}_i^d(\gamma)$. Thus, by McDiarmid's inequality, we have:

$$\forall i, \forall \gamma, \forall h \ \mathbb{P}\left[|\widehat{u}_i^d(\gamma, h) - \overline{u}_i^d(\gamma)| \leq \sqrt{\tfrac{12 \log(T)}{h+1}}\right] \geq 1 - \tfrac{2}{T^6} \tag{15}$$

Thus for all h, where h is the number of times buyer i selects γ, we have the bad event (no concentration) probability is $\frac{2}{T^6}$. Since $h \leq t$ and $n_t^i(\gamma) \leq t$, we can apply a union bound over all possible h and arrive at (note $\widehat{u}_i^d(\gamma, n_t^i(\gamma))$ is equivalent to $\widehat{u}_{i,t}^d(\gamma)$):

$$\forall i, \forall \gamma, \forall t \ \mathbb{P}\left[|\widehat{u}_{i,t}^d(\gamma) - \overline{u}_i^d(\gamma)| \leq \sqrt{\tfrac{12 \log(t)}{n_t^i(\gamma)+1}}\right] \geq 1 - \tfrac{2}{T^5} \tag{16}$$

By applying a union bound over all $t \in [1, \ldots, T]$, we have that an event defined by $\mathcal{E}_i(\gamma) = \left\{\forall t \, |\widehat{u}_i^d(\gamma) - \overline{u}_i^d(\gamma)| \leq \sqrt{\frac{12 \log(t)}{n_t^i(\gamma)+1}}\right\}$ holds with probability greater than $1 - \frac{2}{T^4}$. We would like to now show that for each buyer, this holds uniformly for all possible seller sets, and not just for a fixed γ. Since the total number of seller sets is exponential, a naive application of union bound provides a poor bound. However, since any inactive seller set trivially satisfies the bound with probability 1, it suffices to consider only the active seller sets for a buyer i, $\mathcal{A}_i$. While this set's size is bounded since we add at most one arm every round, it is random in its composition. We use the same observation made by Kleinberg et al. [31]. For $j \in [1, \ldots, t]$, let Z_i^j denote the j^{th} arm activated by buyer i. Z_i^j is a random variable and $\{Z_i^1, \ldots, Z_i^t\}$ is the set of all activated arms[5]. For a seller set γ, note that the event $\{Z_i^j = \gamma\}$ depends on the outcome of *previously activated sets*, whereas $\mathcal{E}_i(Z_i^j)$ is purely based on the observations derived from the seller set Z_i^j, whatever that happens to be. In other words, the event $\{Z_i^j = \gamma\}$ is independent of $\mathcal{E}_i(Z_i^j)$. Thus we have that $\forall i, \forall Z_i^j$, the clean event holds for the Z_i^j activated arm with the following probability:

$$\mathbb{P}\left[\mathcal{E}_i(Z_i^j)\right] = \sum_\gamma \mathbb{P}[\mathcal{E}_i(Z_i^j) | Z_i^j = \gamma] \, \mathbb{P}[Z_i^j = \gamma] = \sum_\gamma \mathbb{P}[\mathcal{E}_i(\gamma)] \, \mathbb{P}[Z_i^j = \gamma] \leq 1 - \tfrac{2}{T^4} \tag{17}$$

where we note that $\mathbb{P}[\mathcal{E}_i(\gamma)]$ is a constant and can be taken outside the sum. Thus, we have a concentration result for each active seller set. Now we apply a union bound over the whole active set $\mathcal{A}_i$. Noting that $|\mathcal{A}_i| \leq T$, we have that: $\forall i, \mathbb{P}\left[\forall \gamma \in \mathcal{A}_i \, , \mathcal{E}_i(\gamma)\right] \geq 1 - \frac{2}{T^3}$. Lastly, we apply a union bound over all the buyers and assuming that the number of buyers is smaller than T, arrive at the statement of the lemma. $\square$

Lemma 3. *If $\mathcal{E}$ holds, then $\Delta_i(\gamma) \leq 3c_t^i(\gamma)$, $\forall \gamma, i, t$. Further, for two active sets $\gamma^1, \gamma^2 \in \mathcal{A}_i$, we have that $D_h(\gamma^1, \gamma^2) \leq \max\left(\frac{1}{k}, \frac{1}{3}\min(\Delta_i(\gamma^1), \Delta_i(\gamma^2))\right)$.*

Proof. Fix any buyer i, any seller set γ, and a time t. If γ is not in the active set or never selected, then this claim holds trivially since $c_t^i(\gamma) > 2$, allowing us to consider only active arms. Suppose seller set γ was last chosen at some time $s \leq t$. Now consider the optimal dominant strategy for buyer i, γ_i^d, and the two possibilities at time s: (1) Either γ_i^d is already part of the active set, or

[5] If the number of activated arms is less than t, then for $j > |\mathcal{A}_i|$, let Z_i^j be the last arm activated

(2) γ_i^d is covered by some other set $\gamma_i' \in \mathcal{A}_i$ which has confidence radius $c_s^i(\gamma_i') \geq \frac{1}{k}$ (the closest element to γ_i' has to be at least $\frac{1}{k}$ away). Then the following holds at time $s \leq t$ for each case:

$$\begin{aligned} \text{if (1): } \mathrm{UCB}_i(\gamma) > \mathrm{UCB}_i(\gamma_i^d) = \widehat{u}_{i,s}^d(\gamma_i^d) + 2c_s^i(\gamma_i^d) \geq \overline{u}_i^d(\gamma_i^d) \\ \text{if (2): } \mathrm{UCB}_i(\gamma) > \mathrm{UCB}_i(\gamma_i') = \widehat{u}_{i,s}^d(\gamma_i') + 2c_s^i(\gamma_i') \geq \overline{u}_i^d(\gamma_i') + c_s^i(\gamma_i') \geq \overline{u}_i^d(\gamma_i^d) - c_s^i(\gamma_i') \end{aligned} \tag{18}$$

where the last inequality in (1) follows from lemma 2 and the last inequality in (2) follows from the metric property and the covering argument (recall $\overline{u}_i^d(\gamma^1) - \overline{u}_i^d(\gamma^1) \leq 2D_h(\gamma^1, \gamma^2)$). The following upper-bound for $\mathrm{UCB}_i(\gamma)$ holds regardless:

$$\mathrm{UCB}_i(\gamma) = \widehat{u}_{i,s}^d(\gamma) + 2c_s^i(\gamma) \leq \overline{u}_i^d(\gamma) + 3c_s^i(\gamma) = \overline{u}_i^d(\gamma) + 3c_t^i(\gamma) \tag{19}$$

where we use that fact that $c_s^i(\gamma) = c_t^i(\gamma)$ since s is the last time γ was selected. Putting the upper and lower bounds on UCB_i together, we have:

$$\overline{u}_i^d(\gamma) + 3c_t^i(\gamma) \geq \mathrm{UCB}_i(\gamma) \geq \mathrm{UCB}_i(\gamma') \text{ or } \mathrm{UCB}_i(\gamma^*) \geq \overline{u}_i^d(\gamma_i^*) \implies \Delta_i(\gamma) \leq 3c_t^i(\gamma) \tag{20}$$

We now move to the second part of the lemma. First note that by property of the Hamming space, any two active seller sets (in fact any two seller sets) must be at least $\frac{1}{k}$ apart. Consider two active choices γ^1 and γ^2 for buyer i, and suppose γ^1 was activated (at time step t_1) before γ^2 (activated at time-step t_2). The fact that γ^2 was activated implies that it was not covered by γ^1's confidence radius at t_2, and thus $D_h(\gamma^1, \gamma^2) > c_{t_2}^i(\gamma^1) \geq \frac{\Delta_i(\gamma^1)}{3}$. If γ^2 was activated before γ^1, we get the opposite result. Combining the two, we have the $D_h(\gamma^1, \gamma^2) \geq \frac{1}{3}\min(\Delta(\gamma^1), \Delta(\gamma^2))$. Putting this and the Hamming observation of two elements being at least $\frac{1}{k}$ apart, we have the desired result. □

The first lemma illustrates that with high probability, the empirical DS utility is close to its expected value and conditioned on this event, we can relate DS utility gap of an active arm to both the confidence radius and the Hamming distance to another active arm. We now provide the following result for covering in Hamming spaces. With these three components in place, we will prove our general regret bound: theorem 8.

Lemma 4. *Let $\mathbb{F}_2^k$ denote the space of $k-$bit string, and let $\mathcal{N}(\mathbb{F}_2^k, r)$ denote a covering with spheres of Hamming radius r. Then with probability at least $1-\epsilon$, $|\mathcal{N}(\mathbb{F}_2^k, r)| \leq 2^k \left(K + \log\frac{1}{\epsilon}\right) / \binom{k}{rk}$*

Proof. We provide a randomized construction. Independently sample c points in $\mathbb{F}_2^k$, which we denote by $\mathcal{C} = \{z_1, \ldots, z_c\}$. We will choose c such that this is an $r-$covering of $\mathbb{F}_2^k$ with high probability. Let $\mathcal{B}(z_i, r)$ denotes a ball of Hamming radius r centered at z_i. Then for a point $z \in \mathbb{F}_2^k$, the probability that this does not lie in the cover of $\mathcal{C}$ is given by:

$$\mathbb{P}\left[z \notin \bigcup_{i=1}^{c} \mathcal{B}(z_i, r)\right] = \prod_{i=1}^{c} \mathbb{P}\left[z \notin \mathcal{B}(z_i, r)\right] = \left(1 - \frac{\sum_{j=0}^{rk}\binom{k}{j}}{2^k}\right)^c \triangleq \left(1 - \frac{\beta}{2^k}\right)^c \leq 2^{-c\beta 2^{-k}} \triangleq p \tag{21}$$

where the first equation follows from the independence of the $z_i \in \mathcal{C}$ and the last inequality follows from the following relationship: $1 - x \leq 2^{-x}$ for $x \in [0,1]$. We use $\beta = \sum_{j=0}^{r}\binom{k}{j}$ to clean up notation. Therefore the probability that $\mathcal{C}$ is a covering for $\mathbb{F}_2^k$ can be given by:

$$\mathbb{P}[\mathcal{C} \text{ is covering } \mathbb{F}_2^k] = 1 - \mathbb{P}\left[\bigcup_{j=1}^{2^k} z_j \text{ not covered}\right] \geq 1 - \sum_{j=1}^{2^k} \mathbb{P}[z_j \text{ not covered }] \geq 1 - p2^k \tag{22}$$

where we make use of union bound. Observe that, $p2^K = 2^k \cdot 2^{-c\beta 2^{-k}} \triangleq \epsilon$. Thus $\mathcal{C}$ is a covering with probability $1-\epsilon$. The size of this covering, $|\mathcal{C}| = c$, as as a function of ϵ is given by:

$$\epsilon = 2^k \cdot 2^{-c\beta 2^{-k}} \implies \log(\epsilon) = k - c\beta 2^{-k} \implies c = \frac{2^k(k+\log\frac{1}{\epsilon})}{\sum_{j=0}^{rk}\binom{k}{j}} \leq \frac{2^k(k+\log\frac{1}{\epsilon})}{\binom{k}{rk}} \tag{23}$$

□

Theorem 7. *For an instance $\mathcal{I}$ and a buyer i, denote $S_i^{\mathcal{I}}(r)$ as the set of active sellers sets where each element $\gamma \in S_i^{\mathcal{I}}(r)$ satisfies $r \leq \Delta_i(\gamma) \leq 2r$. Then for some $\delta > 0$, the expected dominant strategy regret for every buyer i on $\mathcal{I}$ is given by:*

$$R_d^i(T, \mathcal{I}) \leq \frac{\delta}{k}2T + \sum_{j=-1}^{\log\frac{1}{\delta}} k2^j 108 \log T \left|S_i^{\mathcal{I}}\left(\tfrac{1}{2^j k}\right)\right| + O(2^k(k\log T)^2) \tag{24}$$

Proof. We consider the contribution of seller sets in $S_i^{\mathcal{I}}(r)$ toward the total DS regret suffered by buyer i. By adding this contribution over all values of r, we can express the cumulative regret. For each active seller set $\gamma \in S_i^{\mathcal{I}}(r)$, their contribution toward DS regret is $\Delta_i(\gamma) \cdot n_t^i(\gamma)$. By lemma 3, we have that $\Delta_i(\gamma) \leq 3c_t(\gamma) = 3\sqrt{\frac{12\log T}{n_t^i(\gamma)+1}}$ which implies $n_t(\gamma) \leq \frac{108\log T}{\Delta_i^2(\gamma)}$. Thus, $\Delta_i(\gamma)\cdot n_t^i(\gamma) \leq \frac{108\log T}{\Delta_i(\gamma)} \leq \frac{108\log T}{r}$. Next, assume the clean event $\mathcal{E}$ holds, and consider three ranges of r: (1) $r < \frac{\delta}{k}$, for some $\delta \in (0,1)$, (2) $r \in \left[\frac{\delta}{k}, \frac{2}{k}\right]$, and (3) $r \in \left[\frac{4}{k}, 1\right]$. For (1), we will use a trivial upper bound and will select δ appropriately. For (2), observe that by lemma 3, any two active seller sets satisfy, $D_h(\gamma^1, \gamma^2) \geq \frac{1}{3}\min(\Delta(\gamma^1), \Delta(\gamma^2)) = \frac{c}{3k}$, where $c \leq 2$. In other words, the lower bound for the distance between any active seller is $< \frac{1}{k}$ and thus in the worst case, each seller sets only covers itself. For the elements in (3) however, applying lemma 3 implies that any two active sets are at least $\frac{r}{3}$ apart, $r \geq 4$. The maximum number of active seller sets here is thus upper bounded by the maximal $\frac{r}{3}$ packing of this space, which itself is upper bounded by a minimal $\frac{r}{3}$ covering. For an $S_i^{\mathcal{I}}(r)$, let $\mathcal{N}(S_j^I(r), \frac{r}{3})$, denote such a covering. We can now express the cumulative expected regret for an instance $\mathcal{I}$ using these 3 partitions (r is expressed as powers of 2: $r = 2^j$):

$$R_d^i(T,\mathcal{I}) \leq \frac{\delta}{K}2T + \sum_{j=-1}^{\log\frac{1}{\delta}} k2^j 108\log T \left|S_i^{\mathcal{I}}\left(\tfrac{1}{2^j k}\right)\right| + \sum_{j=-2}^{-\log k} k2^j 108 \log T \left|\mathcal{N}\left(\mathbb{F}_2^k, \left\lfloor\tfrac{1}{2^j 3k}\right\rfloor_{\frac{1}{k}}\right)\right| \tag{25}$$

The second term corresponds (2) and the last term to (3), and $\lfloor\cdot\rfloor_{1/k}$ rounds the argument to the closest smaller multiple of $\frac{1}{k}$. Rewriting term 3 using lemma 4, we state that with probability $1 - \frac{1}{T^2}$:

$$\text{term } 3 \leq \sum_{j=2}^{\log k} \tfrac{k}{2^j} 108 \log T \frac{(k+2\log T)2^k}{\binom{k}{\lfloor 2^j/3\rfloor}} \leq 27k\log T(k+2\log T)2^k \sum_{\ell=1}^{k} \frac{1}{\binom{k}{\ell}} \leq 54k\log T(k+2\log T)2^k$$

where the last transition uses the fact that $\sum_{\ell=1}^k \frac{1}{\binom{k}{\ell}} \leq 2$. With this, we now match the bound in the theorem statement. Denote this by $R(\cdot|\text{clean, covering})$, which captures the two independent conditioning events. The bad events is either lemma 2 not holding (probability $\frac{2}{T^2}$) or the covering not holding (probability $\frac{1}{T^2}$). Thus, the expected regret is dominated by $R(\cdot|\text{clean, covering})$ since: $R_d^i(T,\mathcal{I}) \leq R(\cdot|\text{clean, covering})(1 - \frac{3}{T^2} + \frac{2}{T^4}) + 2T\frac{4}{T^2} \leq R(\cdot|\text{clean, covering}) + \frac{4}{T}$. □

Despite being stated from an instance-dependent perspective, this result is in fact quite general. We can use it to derive an instance-independent worst-case bound, as well as the Bayesian regret over all metric instances. Analogous to how the original zooming algorithm in Euclidean space achieves a worst-case bound equivalent to more naive approaches [26], our worst-case bound matches that of UCB. Although we have a metric space, it is over the Hamming distance which is weaker than say Euclidean distance in that it does not distinguish between a large number of elements. So if we have a large number of seller sets whose gap is less than $\frac{1}{k}$, then for any such active γ, the number of times a buyer may choose it is $n_t^i(\gamma) \geq \mathcal{O}(k^2 \log T)$ and thus its covering radius, $c_t^i(\gamma) \leq \mathcal{O}(\frac{1}{k})$. Thus it can cover no other seller sets, and thus all of them must be activated. In this situation, the metric property is no longer useful, and we devolve into UCB which is optimal upto logarithmic terms [32]. The following corollary formalizes this:

Corollary 1. *Instance independent regret, $R_d^i(T)$, for any buyer using algorithm 1 is $\mathcal{O}(\sqrt{2^k T \log T})$.*

Proof. Starting from the result in theorem 7, we observe that the last term is already instance independent. Next we note that the second term, $\sum_{j=-1}^{\log \frac{1}{\delta}} k 2^j 108 \log T \left| S_i^{\mathcal{I}}\left(\frac{1}{2^j k}\right)\right|$, is increasing in j and maximized when all elements belong to $S_i^I(\delta/k)$:

$$\sum_{j=-1}^{\log \frac{1}{\delta}} k 2^j 108 \log T \left| S_i^{\mathcal{I}}\left(\tfrac{1}{2^j k}\right)\right| \leq k 2^k \tfrac{1}{\delta} 108 \log T \tag{26}$$

Going back to cumulative regret, we now have two terms increasing and decreasing in δ respectively. Their sum is minimized when we set them to be equal:

$$\tfrac{\delta}{k} 2T = k 2^k \tfrac{1}{\delta} 108 \log T \implies 2\delta = \sqrt{\tfrac{k^2 2^k 108 \log T}{T}} \implies R_d^i(T) = \mathcal{O}\left(\sqrt{2^k T \log T}\right) + O(2^k (k \log T)^2)$$

which is dominated by the first term when T is large. □

The worst-case instance formalizes the intuition mentioned and assumes all seller sets have their DS gap in a small region such that no seller-set can cover another. Beyond such pathological instances, covering and zooming are helpful, and we seek to answer the more pertinent question of Bayesian regret. Analogously, what is the average regret we can expect when we uniformly sample from the family of all metric instances, $\mathcal{F}$. We have the following result:

Theorem 8. *The expected regret for any buyer i across all metric instances under a uniform distribution is given by: $\mathbb{E}_{I \sim Unif(\mathcal{F})}[R_d^i(T, I)] = \mathcal{O}(k^2 2^k \log^2 T)$.*

Proof. Since the last term in theorem 7 is instance independent and matches the bound we seek, it suffices to prove an equivalent bound for the first two terms in expectation over all metric instances (we simplify notation by expressing $\mathbb{E}_{\mathcal{I} \in \text{Unif}(\mathcal{F})}$ as $\mathbb{E}_{\mathcal{I}}$):

$$\mathbb{E}_{\mathcal{I}}\left[\tfrac{2\delta T}{k} + \sum_{j=-1}^{\log \frac{1}{\delta}} k 2^j 108 \log T \left| S_i^{\mathcal{I}}\left(\tfrac{1}{2^j k}\right)\right|\right] \leq \tfrac{2\delta}{k} T + O(k 2^k \log T) + k 108 \log T \sum_{j=1}^{\log \frac{1}{\delta}} 2^j \mathbb{E}_{\mathcal{I}}\left[\left| S_i^{\mathcal{I}}\left(\tfrac{1}{2^j k}\right)\right|\right] \tag{27}$$

We seek a bound on the expected number of seller sets whose gap is between $\frac{1}{2^j k}$ and $\frac{2}{2^j k}$, with $j \geq 1$. For each seller set γ, let the random variable $Y_\gamma(\frac{1}{2^j k})$ indicate whether $\Delta(\gamma) \in [\frac{1}{2^j k}, \frac{2}{2^j k}]$.

Thus, $|S_i^I(\frac{1}{2^j k})| = \sum_\gamma Y_\gamma(\frac{1}{2^j k})$. Now for some $\ell \in [1,\dots,k]$, consider all γ that are exactly $\frac{\ell}{k}$ Hamming distance away from dominant strategy optimal for buyer i, γ_i^d (Note this is what $\Delta_i(\cdot)$ is computed with respect to). Denote the set of all such γ, $\mathcal{S}_i^\ell$, and note that given γ_i^d, $\mathcal{S}_i^\ell$ is not random. We can write $\mathbb{E}_\mathcal{I}[|S_i^I(\frac{1}{2^j k})|] = \sum_{\gamma_i^d} \mathbb{E}_\mathcal{I}[|S_i^I(\frac{1}{2^j k})| \big| \gamma_i^d]\,\mathbb{P}(\gamma_i^d)$, and observe that $\mathbb{E}_\mathcal{I}[|S_i^I(\frac{1}{2^j k})\big|\gamma_i^d] = \sum_{\ell=1}^k \sum_{\gamma\in\mathcal{S}^\ell} \mathbb{E}_\mathcal{I}[Y_\gamma(\frac{1}{2^j k})\big|\gamma_i^d]$. Next, since we are concerned with $j \geq 1$, it implies $\frac{1}{2^j k} < \frac{1}{k}$; this, along with all instances satisfying the metric property, leads to the following observation: $\mathbb{E}_\mathcal{I}[Y_\gamma(1/2^j k)|\gamma \in \mathcal{S}_\ell, \gamma_i^d] \leq \mathbb{E}_\mathcal{I}[Y_\gamma(1/2^j k)|\gamma \in \mathcal{S}_1, \gamma_i^d]$. Intuitively, metric property implies that γ that are $\frac{\ell}{k}$ away can be more spread out DS utility gap than those that are only $\frac{1}{k}$ away. Thus:

$$\sum_{\gamma\in\mathcal{S}^\ell} \mathbb{E}_\mathcal{I}\left[Y_\gamma\left(\tfrac{1}{2^j k}\right) \,\middle|\, \gamma_i^d\right] = \mathbb{E}_\mathcal{I}\left[Y_\gamma\left(\tfrac{1}{2^j k}\right) \,\middle|\, \gamma \in \mathcal{S}_\ell, \gamma_i^d\right] |\mathcal{S}_\ell| \leq \mathbb{E}_\mathcal{I}\left[Y_\gamma\left(\tfrac{1}{2^j k}\right) \,\middle|\, \gamma \in \mathcal{S}_1, \gamma_i^d\right] |\mathcal{S}_\ell| \tag{28}$$

Any γ that is Hamming distance $\frac{1}{k}$ from the optimal γ^*, must satisfy $\Delta_i(\gamma) \leq \frac{2}{k}$. Since we are considering all metric instances as uniformly likely and $\frac{1}{2^j k} \leq \frac{1}{2k}$, $\mathbb{P}[Y_\gamma(\frac{1}{2^j k}) = 1|\gamma \in \mathcal{S}_1]$ is directly proportional to the width of that region $(\frac{1}{2^j k})$ with respect to $\frac{2}{k}$ and is equal to $\frac{1}{2\cdot 2^j}$, which implies:

$$\sum_{\gamma\in\mathcal{S}^\ell} \mathbb{E}_\mathcal{I}\left[Y_\gamma\left(\tfrac{1}{2^j k}\right) \,\middle|\, \gamma_i^d\right] \leq \tfrac{1}{2^j}\binom{k}{\ell} \implies \mathbb{E}_\mathcal{I}\left[\left|S_i^I\left(\tfrac{1}{2^j k}\right)\right| \,\gamma_i^d\right] = \sum_{\ell=1}^k \sum_{\gamma\in\mathcal{S}^\ell} \mathbb{E}_\mathcal{I}\left[Y_\gamma\left(\tfrac{1}{2^j k}\right) \,\middle|\, \gamma_i^d\right] \leq \sum_{\ell=1}^k \tfrac{1}{2^j}\binom{k}{\ell} = \tfrac{2^k}{2^j}$$

Thus it follows that $\mathbb{E}_\mathcal{I}[|S_i^I(\frac{1}{2^j k})|] = \sum_{\gamma_i^d} \mathbb{E}_\mathcal{I}[|S_i^I(\frac{1}{2^j k})|\big|\gamma_i^d]\,\mathbb{P}(\gamma_i^d) \leq \frac{2^k}{2^j}$. By plugging this upper bound into the simplified expression for expected regret (equation 27), and focusing on the two terms that depend on δ, we have:

$$\tfrac{2\delta}{k}T + k108\log T \sum_{j=1}^{\log\frac{1}{\delta}} 2^j \mathbb{E}_\mathcal{I}\left[\left|S_i^\mathcal{I}\left(\tfrac{1}{2^j k}\right)\right|\right] \leq \tfrac{2\delta}{k}T + \log(\tfrac{1}{\delta})k2^k 108\log T \tag{29}$$

As we have increasing and decreasing terms in δ, the minimum is reached by equating the two. The equation above however does not have a clean analytic solution. However, we can take a first-order Taylor approximation of $\log\frac{1}{\delta}$ and solve accordingly, yielding $\delta = \frac{k2^k 108\log T}{T}$. Using this value of δ, the total expected regret when uniformly sampling overall metric instances is given by:

$$\mathbb{E}_\mathcal{I}[R(T,I)] = \mathcal{O}(2^k\log T) + \mathcal{O}(k2^k\log^2 T) + \mathcal{O}(k^2 2^k \log^2 T) = \mathcal{O}(k^2 2^k\log^2 T) \tag{30}$$

□

5.4 Online welfare regret

To summarize our results so far, we consider independent externalities with NE transaction cost for the online setting, which admits a dominant strategy for each buyer. Learning and playing this dominant strategy is each buyer's goal, and under a metric assumption, we consider each buyer using the zooming algorithm (Algorithm 1) to tackle this. We provide both worst-case and average-case regret bounds for each buyer with respect to their effective dominant strategy utility. The key question left is how this affects social welfare. In the offline setting, we observed in theorem 2 that agents playing their dominant strategy leads to $\Theta(n(1-\alpha))$ welfare regret (WRaE). We ponder an analogous question in the online learning setting: what is the cumulative welfare regret when agents learn to play dominant strategy using the zooming algorithm? We now show that this regret can be decomposed into the online dominant strategy regret and the offline WRaE.

Theorem 9. *Online welfare regret, $R_w(T)$, is upper-bounded by: $\mathcal{O}(Tn(1-\alpha)) + \sum_{i=1}^{n} R_d^i(T)$.*

Proof. Let $S^* = (\gamma_1^*, \ldots, \gamma_n^*)$ be the social optimal strategy, S^d the dominant strategy of each buyer, and S^t the strategy taken by buyers at time t. We can add and subtract the welfare at the dominant strategy to the online welfare regret, $R_w(T)$, and express it as:

$$R_w(T) = \sum_{t=1}^{T}\sum_{i=1}^{N}\Big(g_i(\gamma_i^*) - \sum e_{ij}(\gamma_j^*) - g_i(\gamma_i^d) + \sum e_{ij}(\gamma_j^d)\Big) + \Big(g_i(\gamma_i^d) - \sum e_{ij}(\gamma_j^d) - g_i(\gamma_i^t) + \sum e_{ij}(\gamma_j^t)\Big)$$

The first part of the sum, $\sum_{i=1}^{N} g_i(\gamma_i^*) - \sum e_{ij}(\gamma_j^*) - g_i(\gamma_i^d) + \sum e_{ij}(\gamma_j^d)$, is exactly the WRaE quantity that we bounded to $\mathcal{O}(n(1-\alpha))$ in theorem 2. Summing this over T, we have $\mathcal{O}(Tn(1-\alpha))$. For the latter sum, we can rearrange the order of summing externalities and express it as:

$$\sum_{i=1}^{N}\left[\sum_{t=1}^{T} g_i(\gamma_i^d) - \sum_{t=1}^{T}\sum_{j\neq i} e_{ji}(\gamma_i^d) - \sum_{t=1}^{T} g_i(\gamma_i^t) + \sum_{t=1}^{T}\sum_{j\neq i} e_{ji}(\gamma_i^t)\right] \tag{31}$$

To clean up notation, let $g_i^d = \sum_{t=1}^{T} g_i(\gamma_i^d)$, $e_i^d = \sum_{t=1}^{T}\sum_{j\neq i} e_{ji}(\gamma_i^d)$, $g_i^t = \sum_{t=1}^{T} g_i(\gamma_i^t)$ and $e_i^t = \sum_{t=1}^{T}\sum_{j\neq i} e_{ji}(\gamma_i^t)$. In other words, equation 31 can be written as: $\sum_{i=1}^{n} g_i^d - e_i^d - g_i^t + e_i^t$. We note that $(g_i^d - \alpha e_i^d) - (g_i^t - \alpha e_i^t) = R_d^i(T)$. Thus,

$$(g_i^d - g_i^t) - (e_i^d - e_i^t) - \alpha(e_i^d - e_i^t) = R_d^i(T) - (e_i^d - e_i^t) \implies (g_i^d - g_i^t) - (e_i^d - e_i^t) \leq R_d^i(T) + (1-\alpha)|(e_i^d - e_i^t)|$$

Thus, the second part of $R_w(T)$ expression is upper bounded by $\sum_i R_d^i(T) + nT(1-\alpha)$. ☐

Corollary 2. *The expected online welfare regret, $\mathbb{E}_{I\sim Unif(\mathbb{F})}[R_w(T,\mathcal{I})]$ is upper bounded by $\mathcal{O}(Tn(1-\alpha)) + \mathcal{O}(nk^2 2^k \log^2 T)$. Further, we can equate the two by setting $\alpha = \max\left(0, \mathcal{O}\left(1 - \frac{k^2 2^k \log^2 T}{T}\right)\right)$, thus ensuring the online welfare regret is $\mathcal{O}(nk^2 2^k \log^2 T)$.*

We complete this section by noting that our proposed algorithm depends on the time horizon T. While this can be unsatisfactory, one can obtain a corresponding horizon-independent version using the simple *doubling trick* [33]. This entails running the algorithm in multiple phases, and in each phase i, an instance of the algorithm is executed for $T = 2i$ rounds. This procedure would also imply that the proposed bound for α mentioned in corollary 2, would be closer to 0 in early rounds and increase roughly at a rate of $\widetilde{\mathcal{O}}(1/T)$ over the phases. Intuitively, it is capturing the trade-off in the online welfare regret: in early rounds, there is a significant regret due to the inherent learning process and thus α can be small. In later rounds, increasing α minimizes the WRaE that playing the dominant strategy causes. This is a nice property as it does not intervene strongly in early rounds when buyers know little and are simply exploring, and only does so when buyers have learned.

6 Discussion

Despite the flurry of recent research on new market structures for data, real world data markets remain far simpler: sellers post fixed prices and buyers are unfettered in their strategy. Our work fills a gap in the literature by studying this simple market under the unique characteristics of data: unknown valuations and negative externality. The presence of externality allow us to naturally model this market as a simultaneous game between buyers. While this game has poor equilibrium properties by itself, a simple intervention can greatly improve its equilibrium characteristics. For the independent externality model, our intervention is nearly perfect, guaranteeing a dominant strategy whose welfare approaches the optimal. For a richer externality model, it still guarantees a pure equilibrium but with understandably weaker welfare guarantees. Crucially, this intervention is revenue-neutral, does not require access to sensitive information, and a single transaction cost ($\alpha = 0.5$) works well regardless of the externality model. This intervention also fares well in the more realistic setting where buyers do not know valuations, but must learn them through repeated interaction. For independent externality, we prove that buyers can learn to play their dominant strategy, and this still achieves low cumulative social welfare regret. Furthermore, this setting allows the intervention parameter α to gracefully increase over time as buyers learn their valuations.

Our work illustrates that when coupled with simple interventions, fixed-price data markets can be an elegant solution for a challenging product. It also leaves open a number of intriguing questions. We analyze a single transaction cost; it is unclear whether this is optimal or even what the space of such interventions are, and how to optimize within it. Interventions optimizing for non-utilitarian notions of welfare, like egalitarianism or Nash welfare are also an interesting research direction [34]. While our transaction cost need not have access to the buyer's gain values, which may be private, it requires access to externality. In a setting where these too are private, considering elicitation mechanisms for externalities become imperative [35, 36, 37]. Understanding the strategic perspective of sellers and their externality within this context in an interesting avenue. Lastly, extending our online analysis to the more general joint externality setting remains to be done.

References

[1] Dirk Bergemann, Alessandro Bonatti, and Alex Smolin. The design and price of information. *American Economic Review*, 108(1):1–48, 2018.

[2] Moshe Babaioff, Robert Kleinberg, and Renato Paes Leme. Optimal mechanisms for selling information. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, pages 92–109. ACM, 2012.

[3] Dirk Bergemann and Alessandro Bonatti. Markets for information: An introduction. *Annual Review of Economics*, 11:85–107, 2019.

[4] Sameer Mehta, Milind Dawande, Ganesh Janakiraman, and Vijay Mookerjee. How to sell a dataset? pricing policies for data monetization. *Pricing Policies for Data Monetization (February 12, 2019)*, 2019.

[5] Anish Agarwal, Munther Dahleh, and Tuhin Sarkar. A marketplace for data: An algorithmic solution. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, pages 701–726, 2019.

[6] Anish Agarwal, Munther Dahleh, Thibaut Horel, and Maryann Rui. Towards data auctions with externalities. *arXiv preprint arXiv:2003.08345*, 2020.

[7] Péter Eső and Balázs Szentes. The price of advice. *The Rand Journal of Economics*, 38(4): 863–880, 2007.

[8] Yiling Chen, Haifeng Xu, and Shuran Zheng. Selling information through consulting. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2412–2431. SIAM, 2020.

[9] Yang Cai and Grigoris Velegkas. How to sell information optimally: an algorithmic study, 2020.

[10] Alessandro Bonatti, Munther Dahleh, Thibaut Horel, and Amir Nouripour. Selling information in competitive environments. *arXiv preprint arXiv:2202.08780*, 2022.

[11] Dirk Bergemann and Alessandro Bonatti. Selling cookies. *American Economic Journal: Microeconomics*, 7(3):259–94, 2015.

[12] Dirk Bergemann, Alessandro Bonatti, and Tan Gan. The economics of social data. *The RAND Journal of Economics*, 53(2):263–296, 2022.

[13] Anat R Admati and Paul Pfleiderer. A monopolistic market for information. *Journal of Economic Theory*, 39(2):400–438, 1986.

[14] Anat R. Admati and Paul Pfleiderer. Direct and indirect sale of information. *Econometrica*, 58(4):901–928, 1990.

[15] Kostas Bimpikis, Davide Crapis, and Alireza Tahbaz-Salehi. Information sale and competition. *Management Science*, 65(6):2646–2664, 2019.

[16] Avrim Blum, Vijay Kumar, Atri Rudra, and Felix Wu. Online learning in online auctions. *Theoretical Computer Science*, 324, 08 2003. doi: 10.1016/j.tcs.2004.05.012.

[17] Jonathan Weed, Vianney Perchet, and Philippe Rigollet. Online learning in repeated auctions. *CoRR*, abs/1511.05720, 2015. URL `http://arxiv.org/abs/1511.05720`.

[18] Mark Braverman, Jieming Mao, Jon Schneider, and Matt Weinberg. Selling to a No-Regret Buyer. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, pages 523–538, Ithaca NY USA, June 2018. ACM. ISBN 978-1-4503-5829-3. doi: 10.1145/3219166.3219233. URL `https://dl.acm.org/doi/10.1145/3219166.3219233`.

[19] Shi Feng, Fang-Yi Yu, and Yiling Chen. Peer prediction for learning agents. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho, editors, *Advances in Neural Information Processing Systems*, 2022. URL `https://openreview.net/forum?id=JoukmNwGgsn`.

[20] Noam Nisan, Tim Roughgarden, Eva Tardos, and Vijay V Vazirani. *Algorithmic game theory*. Cambridge university press, 2007.

[21] Elias Koutsoupias and Christos Papadimitriou. Worst-case equilibria. In *Stacs*, volume 99, pages 404–413. Springer, 1999.

[22] Andreas S Schulz and Nicolás E Stier Moses. On the performance of user equilibria in traffic networks. In *SODA*, volume 3, pages 86–87, 2003.

[23] Erich Prisner. *Game theory through examples*, volume 46. American Mathematical Soc., 2014.

[24] Felix Munoz-Garcia, Daniel Toro-Gonzalez, Felix Munoz-Garcia, and Daniel Toro-Gonzalez. Pure strategy nash equilibrium and simultaneous-move games with complete information. *Strategy and game theory: Practice exercises with answers*, pages 39–86, 2019.

[25] George J Mailath, Larry Samuelson, and J Swinkels. Extensive form reasoning in normal form games. *Foundations in Microeconomic Theory*, page 451, 1991.

[26] Aleksandrs Slivkins et al. Introduction to multi-armed bandits. *Foundations and Trends® in Machine Learning*, 12(1-2):1–286, 2019.

[27] Peter Auer. Using confidence bounds for exploitation-exploration trade-offs. *Journal of Machine Learning Research*, 3(Nov):397–422, 2002.

[28] Robert Kleinberg and Tom Leighton. The value of knowing a demand curve: Bounds on regret for online posted-price auctions. In *44th Annual IEEE Symposium on Foundations of Computer Science, 2003. Proceedings.*, pages 594–605. IEEE, 2003.

[29] Robert Kleinberg. Nearly tight bounds for the continuum-armed bandit problem. *Advances in Neural Information Processing Systems*, 17, 2004.

[30] Peter Auer, Ronald Ortner, and Csaba Szepesvári. Improved rates for the stochastic continuum-armed bandit problem. In *Learning Theory: 20th Annual Conference on Learning Theory, COLT 2007, San Diego, CA, USA; June 13-15, 2007. Proceedings 20*, pages 454–468. Springer, 2007.

[31] Robert Kleinberg, Aleksandrs Slivkins, and Eli Upfal. Bandits and experts in metric spaces. *Journal of the ACM (JACM)*, 66(4):1–77, 2019.

[32] Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47:235–256, 2002.

[33] Nicolo Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games.* Cambridge university press, 2006.

[34] Hervé Moulin. *Fair division and collective welfare.* MIT press, 2004.

[35] Yiling Chen, Yiheng Shen, and Shuran Zheng. Truthful data acquisition via peer prediction. *Advances in Neural Information Processing Systems*, 33:18194–18204, 2020.

[36] Yuqing Kong, Grant Schoenebeck, Biaoshuai Tao, and Fang-Yi Yu. Information elicitation mechanisms for statistical estimation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 2095–2102, 2020.

[37] Aviv Zohar and Jeffrey S Rosenschein. Mechanisms for information elicitation. *Artificial Intelligence*, 172(16-17):1917–1939, 2008.