Privacy Can Arise Endogenously in an Economic System with Learning Agents

Nivasini Ananthakrishnan ✉
University of California, Berkeley, CA, USA

Tiffany Ding ✉
University of California, Berkeley, CA, USA

Mariel Werner ✉
University of California, Berkeley, CA, USA

Sai Praneeth Karimireddy ✉
University of California, Berkeley, CA, USA

Michael I. Jordan ✉
University of California, Berkeley, CA, USA

Abstract

We study price-discrimination games between buyers and a seller where privacy arises endogenously – that is, utility maximization yields equilibrium strategies where privacy occurs naturally. In this game, buyers with a high valuation for a good have an incentive to keep their valuation private, lest the seller charge them a higher price. This yields an equilibrium where some buyers will send a signal that misrepresents their type with some probability; we refer to this as *buyer-induced privacy*. When the seller is able to publicly commit to providing a certain privacy level, we find that their equilibrium response is to commit to ignore buyers' signals with some positive probability; we refer to this as *seller-induced privacy*. We then turn our attention to a repeated interaction setting where the game parameters are unknown and the seller cannot credibly commit to a level of seller-induced privacy. In this setting, players must learn strategies based on information revealed in past rounds. We find that, even without commitment ability, seller-induced privacy arises as a result of reputation building. We characterize the resulting seller-induced privacy and seller's utility under no-regret and no-policy-regret learning algorithms and verify these results through simulations.

2012 ACM Subject Classification Theory of computation → Theory and algorithms for application domains

Keywords and phrases Privacy, Game Theory, Online Learning, Price Discrimination

Digital Object Identifier 10.4230/LIPIcs.FORC.2024.9

Related Version *Full Version*: `https://arxiv.org/abs/2404.10767`

Funding TD acknowledges support from the National Science Foundation Graduate Research Fellowship Program under grant no. 2146752, SPK is partially supported by a Mobility Fellowship by the Swiss National Science Foundation. We also acknowledge support from the European Union (ERC-2022-SYG-OCEAN-101071601).

Acknowledgements We thank Alireza Fallah and Stephen Bates for helpful discussions.

1 Introduction

The question of how to define and preserve privacy in the age of machine learning has been a topic of ongoing debate in the computer science and policy communities [11]. The widely accepted theoretical framework of differential privacy [8] formalizes privacy as the ability to withstand membership inference attacks. That is, differential privacy ensures that the output of a computation obfuscates whether a particular data point was present in the input.

However, the practical implementations of differential privacy has been fraught with challenges. There has been significant debate around how to interpret the key privacy parameter ε and how to choose it [21]. This is especially true when data is continuously collected from users (what does it mean to have a guarantee of $\varepsilon = 1$ *per data point* when a user's data is continuously collected?) This has also led to controversies where companies have claimed their algorithms are private, when in fact the chosen ε value confers negligible protection [24]. Further complicating matters, there are multiple variants and extensions of differential privacy – e.g. (ε, δ)-DP [8], Reyni-DP [19], Gaussian-DP [7], etc. – each with different parameters and interpretations.

Perhaps more fundamentally, a growing body of work argues that the public's understanding of privacy is drastically different from differential privacy [23, 18]. While differential privacy focuses on membership inference, privacy is more commonly understood to mean the prevention of the platform using one's data in ways that are misaligned with the individual's interests, such as price discrimination or other exploitative practices.

This work seeks to provide a new perspective on privacy that bridges the gap between the theoretical computer science view and the public's intuitive understanding. We develop a game-theoretic model of privacy that allows us to analyze the effect of privacy choices on all the stakeholders. Additionally, the framework shows how to derive *optimal* privacy mechanisms that balance the gain in privacy with loss of accuracy in order to maximize net utility. In our model, a "principal" (e.g., a platform or seller) can observe signals from "agents" (e.g., users or buyers) and use this information to maximize its own profit, while the agents have an incentive to obfuscate their data to prevent exploitation. We focus on a price-discrimination setting involving interactions between buyers and sellers.

We show that "buyer-induced privacy" behavior, which resembles randomized response, arises endogenously as an equilibrium strategy. Furthermore, we find that the seller is often better off *committing* to not observing the agents' data at all ("seller-induced privacy"), as the revenue loss from buyer-induced privacy can be substantial. Finally, we extend our analysis to a dynamic setting where the seller is a learning agent who interacts with multiple buyers over time. We demonstrate how a simple external auditing mechanism can implement the sellers's commitment to privacy and lead to an equilibrium with endogenously arising privacy-preserving behavior.

Our results provide a new framework for understanding privacy that encompasses both the theoretical guarantees of differential privacy and the practical, user-centric notion of privacy. By modeling privacy as an emergent property of an economic system, we hope to offer insights that can inform the design of privacy-preserving platforms and policies.

Motivating example. In the absence of regulation, online retailers may price discriminate based on information they have collected about past purchases of the customers. Some customers may be willing to pay more for a good than others, perhaps due to innate preferences for certain types of good or because they have more disposable income. The retailer wants to identify customers with higher valuations and charge them higher prices in order to maximize their revenue.

Since customers are aware of the potential for price discrimination, they may engage in evasive action to protect their privacy. Customers may avoid choosing goods that signal their true preferences for less consequential purchases, e.g., a high-income customer choosing between an expensive water bottle that is slightly better than a cheaper option may opt to buy the cheaper bottle in an attempt to obscure their income status. This evasive action imposes a cost on the customer, who misses out on buying their truly preferred product, and also on the retailer, who would have preferred to sell the more expensive product.

What are the behaviors that arise at equilibrium? What if the seller can credibly commit to not price discriminate? How do these behaviors change in more realistic settings where game parameters are not known and strategies must be learned based on past interactions? These are questions we answer in this paper.

1.1 Preview of contributions

We introduce a price-discrimination game in Definition 1 that involves buyers of two types – one with a high valuation and one with a low valuation of an item. A seller may potentially track buyers' signals that reveal their valuations. We characterize the perfect Bayes Nash equilibrium of this game in Theorem 2 and show that a buyer-induced privacy mechanism emerges in the equilibrium. That is, the buyer with a high valuation, with some probability, chooses an evasive action to appear to have a low valuation.

We then introduce commitment ability for the seller wherein a seller can commit to not track buyers' signals with some probability. In the price-discrimination game with commitment, the equilibrium response (Corollary 5) results in seller-induced privacy, which obviates the need for buyer-induced privacy. That is, with some probability, the seller chooses to commit to respect privacy and voluntarily does not track signals. Due to this privacy commitment from the seller, it is optimal for buyers to truthfully report their type. We call this seller-induced privacy the "commitment strategy" and denote the resulting utility $\mathbb{U}_1^*$.

In Section 3, we remove the seller's commitment ability but give buyers access to the seller's historical pricing. We model this as a repeated interaction between a seller and buyers with each buyer participating in only one round. The pricing history is used by buyers to construct the seller's "reputation" (i.e., an estimate of the probability of price discrimination), which buyers then use to inform their signaling strategy. We model the buyers as using a reputation construction procedure that satisfies a consistency condition given in Definition 8, which requires that the reputation is able to differentiate between sellers employing price-discriminating strategies and non-price-discriminating strategies. In Proposition 10, we show the existence of such a reputation mechanism using the available history. We show that consistent reputation can yield seller-induced privacy (i.e., ignoring signals), depending on the model of the seller; we consider no-regret and no-policy-regret sellers. Our findings are:

1. With a no-regret seller, there could be no seller-induced privacy. That is, the seller can use signals and price discriminate in every round and still be no-regret (Proposition 13).
2. Regret minimization achieves strictly less average utility (asymptotically) than $\mathbb{U}_1^*$ (Proposition 14).
3. Employing the commitment strategy in every round is a no-policy-regret algorithm for the seller (Proposition 20).
4. Employing the commitment strategy in every round ensures the seller (asymptotically) an average utility of $\mathbb{U}_1^*$. This the highest possible average utility achievable (asymptotically) in the repeated interaction (Proposition 21).

1.2 Related work

Our work sits at the intersection of many areas, ranging from classical economics to online learning.

There is a vast literature on *privacy* in computer science studying mechanisms for notions of privacy such as differential privacy [8]. The mechanisms arising in our setting resemble mechanisms in these works. We observe local privacy (buyer-induced privacy) where users add noise to their data. We also observe central privacy (seller-induced privacy) where the platform ensures similar outcomes for different user data.

Literature in economics studies the economic implications of enacting privacy mechanisms (see [1] for a survey). Within this body of work, there is a literature on privacy and *price discrimination* (e.g., [2, 5, 20, 12]). We build on this work and extend to a setting that relaxes common-prior assumptions for buyers and sellers so that players must now devise strategies based on what they learn from repeated interactions.

In these repeated interactions, we observe the emergence of a *reputation-based privacy mechanism*. This reputation, learned by buyers based on previous interactions, takes the place of the prior that is used in the single-interaction game. There are numerous papers in economics on reputation focusing on sellers' reputations for the quality of the proffered good [15, 22, 9]. We focus on seller's reputation for enacting price discrimination and analyze how this arises in an online learning framework.

We also study the differences in behavior that arise from seller *commitment*, which has been studied in [14], [2], [12] and [16]. We show that even without commitment, similar behavior can arise through repeated interactions where reputation substitutes for the role of commitment.

Finally, we draw upon work on *online learning* and *repeated games*. There are a number of papers [4, 6, 13, 10] on repeated interactions between a principal and an agent where the agent chooses actions based on evolving beliefs about the principal's actions. In our setting, we interpret the evolving beliefs as the reputation of the principal. Our setting differs in two ways. The first is that the principal's actions are not revealed at the end of the round. Instead partial information about the action, depending on the agent's response, is revealed. The second is that our results hold for weaker conditions on the agent's beliefs compared to previous work.

2 A Price-Discrimination Game

We formulate price discrimination as a sequential, incomplete-information game between n buyers and a seller.

▶ **Definition 1** (PD game). *The price-discrimination game with parameters $n, \alpha, \mu, \overline{\theta}, \underline{\theta}, c_B, c_S$, denoted the $(n, \alpha, \mu, \overline{\theta}, \underline{\theta}, c_B, c_S)$-*PD game*, has the following extensive-form representation.*

1. ***Nature's move.*** *The game begins with Nature assigning types to each participant according to random draws. For $i \in [n]$, the type for buyer i is $\theta_i \in \{\underline{\theta}, \overline{\theta}\}$, representing their valuation of the item being sold, with $\underline{\theta} < \overline{\theta}$. A buyer is type $\overline{\theta}$ with probability μ and type $\underline{\theta}$ with probability $1 - \mu$. The seller's type χ is either* signal aware *($\chi = 1$) or* signal blind *($\chi = 0$). The seller is signal aware with probability α and signal blind with probability $1 - \alpha$.*
2. ***Signaling stage.*** *Based on their assigned type θ_i, each buyer signals $s_i \in \{\underline{s}, \overline{s}\}$. Signaling one's true type ($\underline{s}$ for type $\underline{\theta}$ and $\overline{s}$ for type $\overline{\theta}$) incurs no cost, whereas signaling a mismatched type, referred to as "evasion," imposes a cost c_B on the buyer and a cost c_S on the seller.*[1]

[1] We can more generally allow for each type of buyer impose a different evasion cost (e.g., if a $\overline{\theta}$-buyer evades, the costs are $\bar{c}_B, \bar{c}_S \in \mathbb{R}$, and if a $\underline{\theta}$-buyer evades, the costs are $\underline{c}_B, \underline{c}_S \in \mathbb{R}$. However, as we later

3. ***Pricing decision.*** *The seller chooses a price p_i to set for buyer i. The information the seller can use to set the prices depends on the type of seller. A signal-aware seller can set prices depending on the signals sent by the buyers, that is, they can set one price for all buyers that signaled $\underline{s}$ and a different price for all buyers that signaled $\overline{s}$. A signal-blind seller must set the same price for all buyers since they have no information to distinguish buyers.*
4. ***Purchase decisions.*** *Each buyer, based on the price p_i set for them and their valuation θ_i, makes a choice $b_i \in \{0, 1\}$, to purchase the item ($b_i = 1$) or not ($b_i = 0$).*
5. ***Utilities.*** *All players receive their respective utilities. Each buyer's positive utility is zero if they do not buy the item and the difference between their valuation and price otherwise. If they took evasive action in the signaling stage, their negative utility is equal to their cost of evasion c_B. That is, buyer i's utility is*

$$u_B(\theta_i, s_i, p_i, b_i) = (\theta_i - p_i)b_i - c_B e(\theta_i, s_i)$$

where $e(\theta_i, s_i) = \mathbb{1}\{(\theta_i = \underline{\theta} \wedge s_i = \overline{s}) \vee (\theta_i = \overline{\theta} \wedge s_i = \underline{s})\}$ indicates evasion or not. The seller's overall utility is the sum of utilities $u_S(\theta_i, s_i, p_i, b_i)$ from their interactions with each buyer. The positive utility due to buyer i is the revenue p_i if buyer i buys and zero otherwise. If the buyer took evasive action in the signaling stage, the seller incurs negative utility c_S. That is, the seller's utility is

$$u_S\left((\theta_i, s_i, p_i, b_i)_{i=1}^n\right) = \sum_{i=1}^n u_S(\theta_i, s_i, p_i, b_i) = \sum_{i=1}^n p_i b_i - c_S e(\theta_i, s_i).$$

Mixed strategies. For simplicity of presentation, our game definition is stated in terms of pure strategies (i.e., players take deterministic actions). However, we can more generally allow players to employ mixed strategies. A *mixed strategy* for a player is a distribution over allowed actions conditioned on the information available when taking the action: buyer i's mixed signaling strategy induces a conditional distribution over signals $\pi_i^s(\cdot|\theta_i) \in \Delta(\{\underline{s}, \overline{s}\})$; the seller's mixed pricing strategy induces conditional distributions $\pi^p(\cdot|\underline{s}, \chi)$, $\pi^p(\cdot|\overline{s}, \chi)$ over positive reals with the constraint $\pi^p(\cdot|s = \underline{s}, \chi = 0) = \pi^p(\cdot|s = \overline{s}, \chi = 0)$; finally, each buyer i's mixed buying strategy induces conditional distribution $\pi_i^b(\cdot|\theta_i, p_i) \in \Delta(\{0, 1\})$.

Let $\pi = (\pi^s, \pi^p, \pi^b)$ denote a mixed strategy profile. π, along with the probability of player types described in Step 1 of Definition 1 (which we will denote $p(\chi)$ and $p(\theta_i)$) induce a distribution over action profiles with the probability of an action profile $(\chi, (\theta_i, s_i, p_i, b_i)_{i=1}^n)$ given by

$$\mathbb{P}\left(\chi, (\theta_i, s_i, p_i, b_i)_{i=1}^n\right) = p(\chi) \prod_{i=1}^n p(\theta_i)\pi_i^s(s_i|\theta_i)\pi^p(p_i|\theta_i, \chi)\pi_i^b(b_i|\theta_i, p_i). \tag{1}$$

Given a mixed strategy profile π, we will denote the expected utility for the seller and buyer i by

$$U_S(\pi) = \mathbb{E}\left[u_S\left((\theta_i, s_i, p_i, b_i)_{i=1}^n\right)\right] \qquad \text{and} \qquad U_B^i(\pi) = \mathbb{E}\left[u_B\left(\theta_i, s_i, p_i, b_i\right)\right],$$

where the expectation is over the joint distribution in (1).

show, the only costs that are relevant are the evasion costs associated with the $\overline{\theta}$-seller, because the $\underline{\theta}$ seller will never choose to evade, so we can think of $c_B = \bar{c}_B$ and $c_S = \bar{c}_S$.

Solution concept. We study the *perfect Bayes Nash equilibrium (PBNE)*. Mixed strategies of players constitute a PBNE if the following conditions hold: (1) sequential rationality, meaning that each player's strategy constitutes a best response to their beliefs about the other players' types and strategies, given the history of the game up to the point of choosing the action and (2) consistency of beliefs, meaning that players' beliefs about other players' types are updated following Bayes' rule.

The following theorem characterizes the PBNE of the price-discrimination game described in Definition 1.

▶ **Theorem 2.** *An $(n, \alpha, \mu, \overline{\theta}, \underline{\theta}, c_B, c_S)$-PD game has the following unique perfect Bayes Nash equilibrium. Define $\Delta\theta = \overline{\theta} - \underline{\theta}$.*

(a) *Buyers with type $\theta_i = \underline{\theta}$ will signal $s_i = \underline{s}$.*

(b) *Buyers with type $\theta_i = \overline{\theta}$ will signal*

$$s_i = \begin{cases} \underline{s} \text{ w.p. } q^* & \text{if } \alpha > c_B/\Delta\theta \\ \overline{s} & \text{otherwise.} \end{cases} \qquad \text{where } q^* = \min\left\{1, \frac{(1-\mu)\underline{\theta}}{\mu\Delta\theta}\right\}$$

(c) *The signal-aware seller sets price*

$$p^*_{\text{signalaware}}(s) = \begin{cases} \underline{\theta} & \text{if signal } s = \underline{s} \text{ is observed} \\ \overline{\theta} & \text{if signal } s = \overline{s} \text{ is observed.} \end{cases}$$

(d) *The signal-blind seller sets price*

$$p^*_{\text{signalblind}} = \begin{cases} \underline{\theta} & \text{if } \underline{\theta} \geq \mu\overline{\theta} \\ \overline{\theta} & \text{otherwise.} \end{cases}$$

(e) *Buyer i buys the good if and only if their price p_i is at most their value, so*

$$b_i = \mathbb{1}\{\theta_i \leq p_i\}.$$

The proof is given in Appendix A.1.

▶ Remark 3 (Buyer-induced privacy). The $\overline{\theta}$-buyers' equilibrium response can be interpreted as a privacy-protecting mechanism. This type of buyer is vulnerable to price discrimination, so rather than always signaling their true type, they may choose to randomize their signal. More specifically, if the cost of evasion is very high, the $\overline{\theta}$-buyer will tell the truth, but if the evasion cost is low enough, the $\overline{\theta}$-buyer can receive a reduction in price that is higher than their evasion cost. In the latter case, the $\overline{\theta}$-buyer must then choose the maximum evasion probability q^* such that it is still in the seller's best interest to take the the buyer's signal at face value. We call this randomization "buyer-induced privacy."

Theorem 2 tells us that strategic behavior can only happen if $c_B < \Delta\theta$ (otherwise, we can never have $\alpha > c_B/\Delta\theta$, so buyers will always signal truthfully). For the rest of the paper, we will focus on this setting.

▶ **Assumption 1.** *In all following results, we assume $c_B < \Delta\theta$.*

A natural next question is how each player's utility is affected by the game parameters. In particular, we focus on the effect of α, due to its connection to privacy. In Figure 1, we visualize the utilities of the seller and $\overline{\theta}$-buyers as α varies from 0 to 1. Observe that the seller's utility increases for α less than some threshold value α^*, whose exact value we

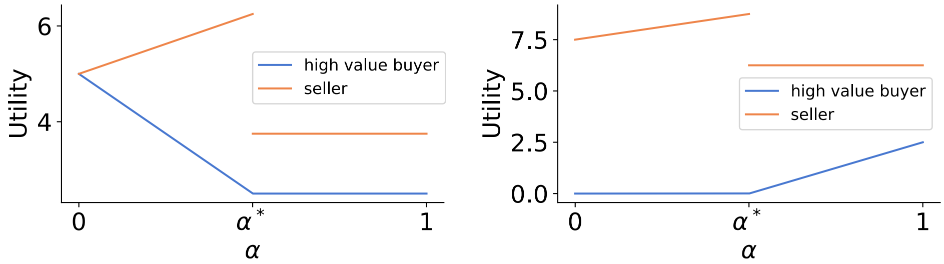

Figure 1 Plots of the $\overline{\theta}$-buyer and seller utilities as a function of α in the $\underline{\theta} \geq \mu\overline{\theta}$ setting (left) and the $\underline{\theta} < \mu\overline{\theta}$ setting (right).

give in the corollary below. This corresponds to the set of PD-games where the buyer's equilibrium response is truthful. Beyond α^*, the $\overline{\theta}$-buyers' equilibrium response changes to being strategic and the seller's utility drops. We formalize the ordering of utilities in the following corollary.

▶ **Corollary 4.** *(Order of utilities) Fix $n, \mu, \overline{\theta}, \underline{\theta}, c_B, c_S$ and let $u_S(\alpha), u_B(\alpha)$ denote the seller's and $\overline{\theta}$-buyers' equilibrium utilities of the $(n, \alpha, \mu, \overline{\theta}, \underline{\theta}, c_B, c_S)$-PD game. $u_S(\cdot)$ is maximized at $\alpha^* = c_B/\Delta\theta$, and the equilibrium utilities for the settings where the seller is always signal blind ($\alpha = 0$), is always signal aware ($\alpha = 1$), and is signal aware with probability α^* ($\alpha = \alpha^*$) have the following orderings:*

(a) *When $\underline{\theta} \geq \mu\overline{\theta}$,*

$$u_S(\alpha^*) > u_S(0) > u_S(1) \qquad \text{and} \qquad u_B(0) > u_B(1) = u_B(\alpha^*).$$

(b) *When $\underline{\theta} < \mu\overline{\theta}$,*

$$u_S(\alpha^*) > u_S(0) > u_S(1) \qquad \text{and} \qquad u_B(1) > u_B(0) = u_B(\alpha^*).$$

$\underline{\theta}$-buyers always receive a utility of zero, regardless of the value of α.

2.1 Price discrimination with seller commitment

A key takeaway from Corollary 4 is that the seller's utilities are dependent on the value of α, and if the seller could choose a value of α, they would want to choose $\alpha = \alpha^*$ to maximize their utility. Suppose we are now in a setting where the seller is able to choose and publicly commit to an α. As a motivating example, suppose that the seller must go through a data broker to access signals, and the data broker publishes trusted summaries of what fraction of buyers the seller requests data on. In such a setting, where α is chosen by the seller instead of treated as given, we arrive at the following equilibrium.

▶ **Corollary 5.** *(Equilibrium of price-discrimination game with commitment) When the seller has commitment power (i.e., is able to credibly communicate to sellers that they will not price discriminate with some probability), the perfect Bayes Nash equilibrium of the PD-game consists of the following strategies:*

(a) *The seller commits to not price-discriminating (by playing $p^*_{\text{signalblind}}$ from Theorem 2) with probability $1 - \alpha^*$, where $\alpha^* = c_B/\Delta\theta$.*

(b) *All buyers always signal truthfully.*

The buyers' buying decisions are the same as in Theorem 2.

Proof. (a) follows directly from Corollary 4, which tells us that the seller's utility is maximized at α^*, and (b) comes from applying Theorem 2 with $\alpha = \alpha^*$. ◀

▶ Remark 6. Commitment ability allows the seller to achieve a higher utility by providing seller-induced privacy. This seller-induced privacy obviates the need for buyers to take evasion action to create buyer-induced privacy, which benefits the seller. We use $\mathbb{U}_1^*$ to refer to the seller's maximum achievable equilibrium utility in the single interaction price discrimination game with commitment. This utility is achieved when the seller plays the strategy given in Corollary 5.

3 Repeated Interactions

In the previous section, we saw the emergence of seller-induced privacy when the seller has commitment ability. If possible, the seller would commit to providing seller-induced privacy (by ignoring signals with probability $1 - \alpha^*$, as in Corollary 5), thereby limiting the extent of price discrimination performed by the seller. However, these results hinge on the buyer believing that the α stated by the seller truly corresponds to the probability of price discrimination. Without this credible commitment from the seller, the story becomes more complicated.

In this section, we study whether seller-induced privacy can still arise in the absence of such commitment ability, through the development of a reputation based on the seller's historical pricing. We ask the question of how the extent of privacy and resulting utilities differ under reputation-based privacy versus commitment-based privacy. We model the seller as making pricing decisions using an online learning algorithm and show how different models such as *no-regret* and *no-policy-regret* lead to different answers to this question.

In the repeated interaction setting, we also relax the assumptions that the distribution μ over agent types and the probability α that the seller looks at the agent's signal are publicly known. Rather than playing the single-interaction equilibrium strategies, which require full knowledge of game parameters, the players now have to learn strategies online based on past interactions.

3.1 Setup

We consider repeated interactions between a seller and buyers where a new batch of buyers is drawn at each round. We call this as the *repeated PD protocol.* Each round is similar to the one-shot PD-game from Definition 1 but with the following differences: (1) There is one fixed seller throughout all rounds. (2) When players choose actions, they not only have access to information from the current round (as was the case in the one-shot PD game) but also some information from previous rounds. Specifically, at round t, the seller has access to $((s_i^\tau, p_i^\tau)_{i=1}^n)_{\tau=1}^{t-1}$, the signals they observed and the prices they set in previous rounds, and each buyer i has access to $(((\theta_i^\tau, s_i^\tau, p_i^\tau)_{i=1}^n)_{\tau=1}^{t-1})$, the buyer types, signals, and prices of all buyers from previous rounds. This modeling of the buyers' access is appropriate in settings where buyer information is pooled either through crowd-sourcing or by an auditing entity and made available to buyers. (3) The parameter μ (the probability of a type-$\overline{\theta}$ buyer) is not known to the seller. (4) The probability that the seller will price discriminate is not known to buyers, as was assumed in the one-shot PD game; rather, buyers must estimate this probability based on past rounds. We write out the repeated interaction protocol in detail in Appendix B.

3.2 Model of the buyers

Since each buyer participates in only one round of the repeated PD protocol, the equilibrium response is still appropriate to model the buyer's response. However, in the repeated interaction setting, we no longer assume the buyers hold a static, prior belief about the probability of a signal-aware seller. Instead, buyers have evolving beliefs based on the seller's interactions with past buyers.

Some specific buyer strategies we will refer to are π^s_{truthful}, which corresponds to always signaling truthfully, and $\pi^s_{\text{strategic}}$, which corresponds to signaling $\underline{s}$ with probability q^* (as defined in Theorem 2) and signaling $\overline{s}$ with probability $1 - q^*$. We consider the following model of buyer behavior.

▶ **Definition 7** (Consistent belief based equilibrium responding (CBER) buyers)**.** *Consistent belief based equilibrium responding buyers (or CBER-buyers) form a sequence of beliefs $(\hat{\alpha}_t)_{t=1}^T$ satisfying a consistency property defined below and at round t, choose the corresponding equilibrium strategy (from Theorem 2) of the PD-game with $\alpha = \hat{\alpha}_t$. That is, $\underline{\theta}$-buyers always signal truthfully, and $\overline{\theta}$-buyers signal truthfully (play π^s_{truthful}) if $\hat{\alpha}_t \leq \alpha^*$ and signal the opposite type with probability q^* otherwise (play $\pi^s_{\text{strategic}}$).*

We now explain the consistency property. Given a sequence of seller mixed strategies action profiles that induce the sequences of distributions $(\pi^p_t(\cdot|s = \overline{s}))_{t=1}^T$ and $(\pi^p_t(\cdot|s = \underline{s}))_{t=1}^T$ indicating price distributions at each round for signals $\underline{s}, \overline{s}$ respectively, define α_t to be

$$\alpha_t = \mathbb{P}_{\overline{P} \sim \pi^p_t(\cdot|s=\overline{s}), \underline{P} \sim \pi^p_t(\cdot|s=\underline{s})} \left[\overline{P} \neq \underline{P}\right].$$

That is, α_t denotes the probability of a different price for $\overline{s}$ compared to $\underline{s}$ at round t. The probability here is over the randomness due to the seller's mixed strategy at round t. α_t is a measure of extent of price discrimination by the seller at round t.

▶ **Definition 8** (Consistent sequence)**.** *Let $\bar{\alpha}_T = (1/T)\sum_{t=1}^T \alpha_t$. We say a sequence of estimators $(\hat{\alpha}_t)_{t=1}^T$ is* consistent *if $\lim_{T\to\infty} |\mathbb{E}[\hat{\alpha}_T] - \bar{\alpha}_T| = 0$, where the expectation is taken over the randomness of the history $H_T = ((\theta^t_i, s^t_i, p^t_i)_{i=1}^n)_{t=1}^{T-1}$ used to construct $\hat{\alpha}_T$.*

A useful implication of consistency is that $\hat{\alpha}_T$ converges pointwise to $\bar{\alpha}_T$.

▶ **Lemma 9.** *If $(\hat{\alpha}_t)_{t=1}^T$ is a consistent sequence of beliefs, then for any $\epsilon < 0$ and $\delta > 0$, there exists some positive integer N such that for all $T > N$, we have $\mathbb{P}\left[|\hat{\alpha}_T - \bar{\alpha}_T| \geq \epsilon\right] \leq \delta$.*

Proof. Due to consistency and the definition of limits, there exists N such that for all $T > N$, we have $|\mathbb{E}[\hat{\alpha}_T] - \bar{\alpha}_T| \leq \delta\epsilon$. Thus, for $T > N$, we can apply Markov's inequality to get $\mathbb{P}(|\hat{\alpha}_T - \bar{\alpha}_T| \geq \epsilon) \leq (|\mathbb{E}[\hat{\alpha}_T] - \bar{\alpha}_T|)/\epsilon \leq \delta\epsilon/\epsilon = \delta$. ◀

The following proposition and associated proof provide an algorithm to construct a consistent sequence of estimators $(\hat{\alpha}_t)_{t=1}^T$.

▶ **Proposition 10** (Existence of consistent sequence)**.** *Assume that buyers equilibrium-respond to $\hat{\alpha}_t$ at each round t. Then, for any sequence of seller actions, there exists a sequence of estimators $(\hat{\alpha}_t)_{t=1}^T$ that is consistent.*

Proof sketch: Since there are multiple buyers at each round, we can infer whether the seller is price discriminating or not by comparing the prices charged to a buyer who signals $\underline{s}$ and a buyer who signals $\overline{s}$. However, only some rounds are informative about price discrimination; in rounds where all buyers send the same signal, we are not able to determine if the seller had a price discriminatory pricing policy in place. The consistent estimator $\hat{\alpha}_t$ we consider is the fraction of past rounds where price discrimination is observed, normalized to account for the probability that a round is likely to be informative about price discrimination. We show that $E[\hat{\alpha}_t] = (1/t)\sum_{\tau=1}^{t-1}\alpha_\tau$, which implies that $\lim_{T\to\infty}|\mathbb{E}[\hat{\alpha}_T] - \bar{\alpha}_T| = \lim_{T\to\infty}\left|(1/T)\sum_{t=1}^{T-1}\alpha_t - (1/T)\sum_{t=1}^{T}\alpha_t\right| = \lim_{T\to\infty}\alpha_T/T = 0$. See Appendix C.1 for the full proof.

3.3 Model of the seller

Since the seller does not a priori know the distribution over buyer types and is engaged in multiple rounds of the repeated interaction, modeling the seller's response by the one-shot equilibrium from Theorem 2 is not reasonable. Instead, we consider the seller as optimizing various common objectives of repeated interactions such as regret minimization and policy-regret minimization.

The seller's mixed strategy at a given round is a pair of probability distributions $\pi_t^p = (\pi_t^p(\cdot|\overline{s}), \pi_t^p(\cdot|\underline{s}))$. Let Π denote the set of possible mixed strategies. For rational sellers, we can focus on distributions supported only on $\{\underline{\theta}, \overline{\theta}\}$ without loss of generality. Prices supported on $\{\underline{\theta}, \overline{\theta}\}$ maximize seller revenue in each round. The seller's effect on future rounds is also not affected by limiting the support. This is because the parameters α_t that the buyers' consistent estimator estimates treats *any* difference in prices as indicating price discrimination, so all price differences are treated the same.

Some specific seller strategies we will refer to are π_{PD}^p and π_{noPD}^p. The former is the "always-price-discriminating strategy," with $\pi_{\mathrm{PD}}^p(\overline{\theta}|\overline{s}) = \pi_{\mathrm{PD}}^p(\underline{\theta}|\underline{s}) = 1$. The latter is the "never-price-discriminating strategy," with $\pi_{\mathrm{noPD}}^p(\underline{\theta}|\underline{s}) = \pi_{\mathrm{noPD}}^p(\underline{\theta}|\overline{s}) = 1$ if $\underline{\theta} \geq \mu\overline{\theta}$ and $\pi_{\mathrm{noPD}}^p(\overline{\theta}|\underline{s}) = \pi_{\mathrm{noPD}}^p(\overline{\theta}|\overline{s}) = 1$ otherwise.

3.3.1 Regret-minimizing seller

The first seller model we consider is a regret-minimizing seller.

▶ **Definition 11** (Seller's regret). *Given a sequence of mixed strategy profiles* $\{\pi_t\} = \{(\pi_t^s, \pi_t^p, \pi_t^b)\}_{t=1}^T$, *the* seller's average regret *is*

$$R_T^S(\{\pi_t\}_{t=1}^T) = \frac{1}{T}\left[\max_{\pi^{p*}\in\Pi}\sum_{t=1}^T U_S(\pi_t^s, \pi_t^{p*}, \pi_t^b) - \sum_{t=1}^T U_S(\pi_t^s, \pi_t^p, \pi_t^b)\right].$$

▶ **Definition 12** (No-regret algorithm). *Let $\mathcal{A}_B$ be an algorithm employed by the buyer in the repeated PD protocol. A seller algorithm $\mathcal{A}_S$ in the repeated PD protocol is a* no-regret algorithm *for the seller given $\mathcal{A}_B$ if the sequence of mixed strategies $(\pi_t)_{t=1}^T$ generated by the interaction between $\mathcal{A}_B$ and $\mathcal{A}_S$ has seller's average regret that is sublinear in the number of rounds. That is, $R_T^S((\pi_t)_{t=1}^T) \in o(1)$.*

We will denote by $(\pi_t)_{t=1}^T$ the sequence of random variables denoting the players' mixed strategies in each round. Our results analyze the asymptotic convergence of average seller utility. We say that the average seller utility *asymptotically converges* to some value v if $\lim_{T\to\infty}\mathbb{E}\left[(1/T)\sum_{t=1}^T U_S(\pi_t)\right] = v$. We write $U_S(\pi^p)$ and $U_S(\pi^s, \pi^p)$ when it is clear what the other arguments are.

If the seller employs a no-regret algorithm, then the seller could end up always price-discriminating i.e., no seller-induced privacy. This is stated below.

▶ **Proposition 13.** *(Always price-discriminating is regret minimizing) Given CBER-buyers, the seller algorithm that always employs the price-discrimination strategy i.e., $\pi^p_t = \pi^p_{\text{PD}}$ for all timesteps t is a no-regret algorithm for the seller. The seller's average utility asymptotically converges to a value at most $u_S(1)$, where $u_S(1)$ is the seller's equilibrium utility in the single-interaction PD-game with $\alpha = 1$.*

Proof sketch. The strategy of CBER-buyers in each round is either π^s_{truthful} or $\pi^s_{\text{strategic}}$. For both these buyer responses, the seller's optimal strategy is to always price discriminate, as shown in the computation of the seller's equilibrium response in the proof of Theorem 2. In other words, the seller incurs zero regret in each round by always price-discriminating.

Next, we analyze the seller's average utility. Note that when $\pi^p_t = \pi^p_{\text{PD}}$, the probability of seeing different prices for different signals is $\alpha_t = 1$, so $\bar{\alpha}_t = 1$ for all t. By Lemma 9, $\hat{\alpha}_t$ becomes greater than α^* eventually (where α^* is as defined in Corollary 5), which causes $\overline{\theta}$-buyers to play $\pi^s_{\text{strategic}}$. In other words, eventually the seller and buyers will all be playing their equilibrium strategies for the PD-game with $\alpha = 1$, so their average utilities will converge to the corresponding equilibrium utilities. See Appendix C.2 for the full proof. ◀

The next proposition tells us that regret minimization necessarily causes the seller to achieve a worse expected average utility that the optimal utility they can achieve in the single interaction setting.

▶ **Proposition 14** (Regret minimization is inherently at odds with achieving $\mathbb{U}^*_1$)**.** *Given CBER-buyers, for any no-regret seller algorithm, the seller's average utility asymptotically converges to strictly less than $\mathbb{U}^*_1$.*

Proof sketch. Define $\mathcal{T} = \{t \in [T] : \hat{\alpha}_t \le \alpha^*\}$ to be the set of rounds where $\overline{\theta}$-buyers' signaling strategy is π^s_{truthful}. In all other rounds, their signaling strategy is $\pi^s_{\text{strategic}}$. Define $\beta = (1/T)\sum_{t\in\mathcal{T}} \alpha_t$ to be a measure of simultaneous truthfulness from buyers and price-discrimination by the seller. Our proof involves the following parts. We outline the parts and state them as lemmas here and prove them in Appendix C.3

1. Obtaining $\mathbb{U}^*_1$ requires the buyers to be truthful strictly more than α^* fraction of rounds.
   ▶ **Lemma 15.** $\lim_{T\to\infty} |\mathcal{T}|/T \le \alpha^*$ *implies that* $\lim_{T\to\infty} \left(\sum_{t=1}^{T} U_S(\pi_t)\right)/T < \mathbb{U}^*_1$.
2. The no regret property requires that the seller price discriminates in most rounds where buyers are truthful. So β is close to $|\mathcal{T}|/T$.
   ▶ **Lemma 16.** $\lim_{T\to\infty} |\mathcal{T}|/T \le \lim_{T\to\infty} \sum_{t\in\mathcal{T}} \alpha_t/T$.
3. There is a limit on simultaneous price-discrimination and truthful signaling due to the buyers' consistent beliefs. That is, β converges to at most α^*.
   ▶ **Lemma 17.** $\lim_{T\to\infty} \sum_{t\in\mathcal{T}} \alpha_t/T \le \alpha^*$.

From Lemmas 16, 17, $\lim_{T\to\infty} |\mathcal{T}|/T \le \alpha^*$. Lemma 15 shows that this means average seller utility is strictly less than $\mathbb{U}^*_1$. ◀

3.3.2 Policy-regret-minimizing seller

As we have seen, regret minimization does not guarantee that the seller achieves higher than price-discrimination utility. On the other hand, if we model the seller as minimizing policy regret [3], the seller *necessarily* achieves utility that is higher than the utility achieved by the naive strategy of always price discriminating.

▶ **Definition 18** (Seller's policy regret). *Consider a buyer algorithm $\mathcal{A}_B$ and a seller algorithm $\mathcal{A}_S$. Let $(\pi_t(\mathcal{A}_B, \mathcal{A}_S))_{t=1}^T$ be the sequence of mixed strategies generated by the interaction between $\mathcal{A}_B$ and $\mathcal{A}_S$. Given a sequence of mixed strategies $(\pi_t)_{t=1}^T$, the* seller's average policy regret *of $(\pi_t)_{t=1}^T$ relative to a buyer algorithm $\mathcal{A}_B$ and a baseline class $\mathbb{A}_S$ of seller algorithms is*

$$PR_T^S\left((\pi_t)_{t=1}^T; \mathcal{A}_B, \mathbb{A}_S\right) = \max_{\mathcal{A}_S \in \mathbb{A}_S} \frac{1}{T}\sum_{t=1}^T U_S(\pi_t(\mathcal{A}_B, \mathcal{A}_S)) - \frac{1}{T}\sum_{t=1} U_S(\pi_t)$$

▶ **Definition 19** (No-policy-regret algorithm). *Let $\mathcal{A}_B$ be an algorithm employed by the buyer in the repeated PD protocol. An algorithm $\mathcal{A}_S$ is a* no-policy-regret algorithm *for the seller given $\mathcal{A}_B$ and relative to a class of seller algorithms $\mathbb{A}_S$ if the sequence of mixed strategies $(\pi_t(\mathcal{A}_B, \mathcal{A}_S))_{t=1}^T$ generated by the interaction between $\mathcal{A}_B$ and $\mathcal{A}_S$ satisfies $PR_T^S((\pi_t(\mathcal{A}_B, \mathcal{A}_S); \mathcal{A}_B, \mathbb{A}_S)_{t=1}^T) \in o(1)$.*

Consider a baseline class $\mathbb{A}_S^{MS}$ consisting of seller algorithms that employ the same mixed strategy in each round, that is, $\pi_t^p(\cdot|\overline{s})$ is the same distribution for all t and similarly for $\pi_t^p(\cdot|\underline{s})$.

▶ **Proposition 20** (Policy-regret-minimizing seller achieves $\mathbb{U}_1^*$). *Given CBER-buyers, if the seller achieves sub-linear policy regret relative to $\mathbb{A}_S^{MS}$, then the seller's average utility asymptotically converges to at least $\mathbb{U}_1^*$.*

Proof sketch. Under the conditions of this proposition, the seller's utility must, by definition of policy regret, approach a utility at least as high (or better) than the utility of any strategy in $\mathbb{A}_S^{MS}$ as $T \to \infty$. Recall that $\mathbb{U}_1^*$ is the seller utility achieved in the PD game when $\alpha = \alpha^*$. Consider the PD game that results in a seller utility of at least $\mathbb{U}_1^* - \epsilon$, which is achieved by the seller price-discriminating with probability $\tilde{\alpha} < \alpha^*$. Then the repeated-interaction strategy of always price-discriminating with probability $\tilde{\alpha}$ has an average expected utility of at least $\mathbb{U}_1^* - \epsilon$ (this must be true due to the consistency of buyer beliefs; see the full proof in Appendix C.4 for details). Taking ϵ to 0 gives the desired result. ◀

Combining the previous result with the following result tells us that a no-policy regret seller's algorithm will cause the seller's average utility to asymptotically converge to *exactly* $\mathbb{U}_1^*$. In fact, this result tells us the stronger result that there does not exist *any* seller algorithm that can achieve utility higher than $\mathbb{U}_1^*$.

▶ **Proposition 21.** *Given CBER-buyers, for any seller algorithm, the seller's average utility asymptotically converges to at most $\mathbb{U}_1^*$.*

Proof sketch. This proof is similar to the argument of the proof of Proposition 14 and the full proof is in Appendix C.4. The key ideas again are that for high seller utility, there must be sufficiently many rounds where simultaneously, the seller price discriminates and the buyer reports truthfully. Since the buyers' belief estimators are consistent, this cannot be the case. The difference between the average seller utility and $\mathbb{U}_1^*$ is a constant times the following quantity: $\frac{1}{T}\sum_{t\in\mathcal{T}}(\pi_t^p(\overline{\theta}|\overline{s}) - \pi_t^p(\underline{\theta}|\underline{s})) - \alpha^*$, where $\mathcal{T}$ is the set of rounds where the buyer signals truthfully. Lemma C.1, 17 (from the proof of Proposition 14) show that the consistency property implies that this difference converges to most zero. ◀

4 Experiments

In this section, we simulate the *repeated PD protocol* with $\mu = 0.5$, $\underline{\theta} = 5$, $\overline{\theta} = 15$, $c_B = c_S = 5$, and $n = 10$ and empirically verify our theoretical claims from Section 3. We report the convergence of buyer and seller utilities, seller actions, and buyer estimators. The seller and buyer algorithms we consider are described below. Code is available at `https://github.com/nivasini/PrivacyDynamics`.

4.1 Algorithms

Seller.

1. **Signal-blind seller.** The seller plays the regret-minimizing Exp3 algorithm (specifically Exp3-IX in Chapter 12 of [17]). At round t the seller sets a price $p_t \in \{\underline{\theta}, \overline{\theta}\}$ according to the algorithm's current sampling distribution, charges p_t to all buyers and updates the sampling distribution based on the resulting average utility from the buyers' purchase decisions.
2. **Signal-aware seller.** The seller plays a contextual version of Exp3, which we call CExp3, in which the algorithm maintains two sampling distributions over prices $\{\underline{\theta}, \overline{\theta}\}$, conditioned on the received signal, $\underline{s}$ or $\overline{s}$. At each round, the seller samples once from each distribution and charges one price $\underline{p}_t$ to all buyers who signal $\underline{s}$ and $\overline{p}_t$ to all buyers who signal $\overline{s}$. Depending on the sampling distributions, $\underline{p}_t$ and $\overline{p}_t$ may or may not be equal.
3. **Stackelberg equilibrium seller.** The seller commits to an $\alpha^* = {}^{c_B}/_{\Delta\theta}$ level of price-discrimination, i.e., they play the $(\alpha = 1)$-PD equilibrium strategy (Theorem 2) with probability α^* and the $(\alpha = 0)$-PD equilibrium with probability $1 - \alpha^*$.

CBER-Buyer. Using a sequence of consistent estimators $\{\hat{\alpha}_\tau\}_{\tau=1}^{t-1}$ (Def. 8) to estimate the seller's probability of price-discrimination at each round, each buyer plays the $(\alpha = \hat{\alpha}_\tau)$-PD equilibrium strategy. For our simulations, buyers use the estimator described in (3) to estimate the seller's probability of price discrimination at each round. All buyers in a single round use the same estimator.

4.2 Discussion

Convergence of Utilities. Figure 2 shows convergence of seller and buyer utilities for each of the seller's algorithms played against a CBER-buyer. As expected, when a seller plays Exp3 (which ignores signals) against a CBER-buyer, the players' utilities converge to the $(\alpha = 0)$-PD equilibrium utility (Theorem 2). When the seller plays CExp3 (which observes signals) against a CBER-buyer, the seller's utility converges to the $(\alpha = 1)$-PD equilibrium utility. Given our experiment parameters, multiple different distributions $\pi_t^p(\cdot|s = \underline{s})$ reward the seller equivalently, while some are more favorable for the buyer than others. Therefore, while the seller's utility will always converge to $(\alpha = 1)$-PD, the buyer's utility may converge to something less than $(\alpha = 1)$-PD. Finally, when the seller plays the Stackelberg equilibrium against a CBER-buyer, the players' utilities converge to the $(\alpha = \alpha^*)$-PD equilibrium utility.

Consistency of $\hat{\alpha}$. Figure 3 illustrates the consistency of the buyer's estimator ((3)). Our simulations show that the buyer's estimate $\hat{\alpha}_t$ of the seller's probability of price discrimination converges to 0 against a seller playing Exp3, to 0.5 against a seller playing α^*-PD (where $\alpha^* = {}^{c_B}/_{\Delta\theta} = 0.5$ given our simulation parameters), and to higher-than-0.5 against a seller playing CExp3. Importantly, $\hat{\alpha}_t$ aligns with the seller's true average probability of price-discrimination, $\bar{\alpha}_t$, giving empirical evidence for Lemma 9.

Convergence of Seller Actions. In Figure 4, we track the cumulative proportion of the seller's price-discriminatory vs. non-price-discriminatory actions. Specifically, we track four seller actions: 1) charging a high price regardless of signal, 2) charging a low price regardless of signal, 3) charging a high price for a low signal and low price for a low signal (PD), and 4) charging a low price for a high signal and a high price for a low signal (reversePD). Given our

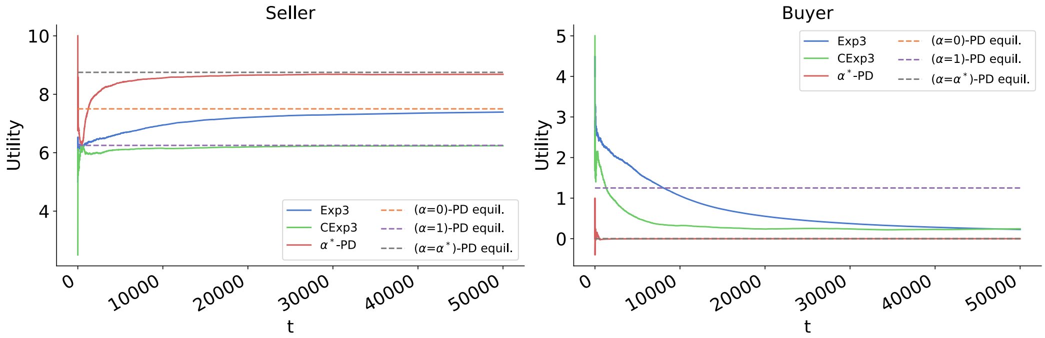

Figure 2 Convergence of seller and buyer utilities for various algorithms. $\underline{\theta} < \mu\overline{\theta}$ with our experiment parameters, so the buyer's $(\alpha = 0)$-PD and $(\alpha = \alpha^*)$-PD utilities are the same (see Corollary 4).

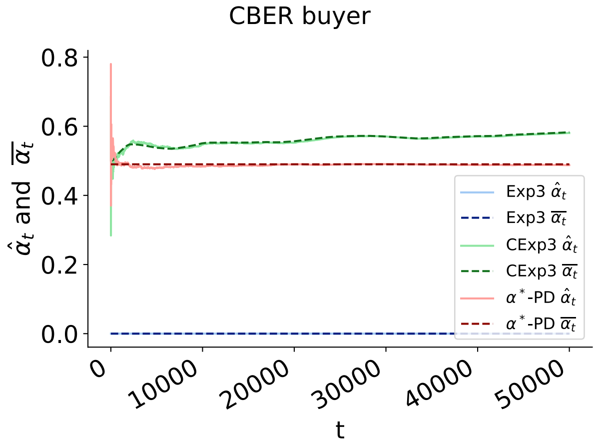

Figure 3 $\hat{\alpha}_t$ and $\bar{\alpha}_t$ over time when seller is playing Exp3, CExp3, or α^*-PD. In all cases, $\hat{\alpha}$ is a consistent estimator of the seller's true probability of price discrimination.

parameter values for these simulations (i.e. $\underline{\theta} < \mu\overline{\theta}$ and $\alpha^* = 0.5$), in equilibrium we would expect that, for each batch of n buyers at a single round: 1) a signal-blind seller sets a high price for all n buyers, 2) a signal-aware seller sets a high price for high-signal buyers and a low price for low-signal buyers, and 3) a α^*-PD seller sets a high price for all high-signal buyers and low price for all low-signal buyers with probability 0.5 and sets a high price for all n buyers with probability 0.5. Figure 4 gives empirical evidence for this intuition.

Biased $\hat{\alpha}$. In realistic settings, the buyer may not have a consistent estimate of price discrimination and instead only have access to a biased $\hat{\alpha}$. Figure 5 examines whether a seller can benefit from non-consistency in the buyer's estimate. The y-axis of the figure tracks the seller's cumulative average utility after 20,000 rounds of interaction with CBER-buyers. We partition the interval $[-1, 1]$ into twenty segments γ_i of width 0.1, and the buyers use estimator $\hat{\alpha}_t + \epsilon_t$, where $\epsilon_t \sim \text{Unif}(\gamma_i)$. The plot then tracks the seller's cumulative average utility after 20,000 rounds of interaction with buyers for each biase interval γ_i. If $\hat{\alpha}_t + \epsilon_t$ is less than 0 or greater than 1, we clip it at those values respectively. In all cases, the seller is hurt by a $\overline{\theta}$-buyer who overestimates the probability of price discrimination (high values of

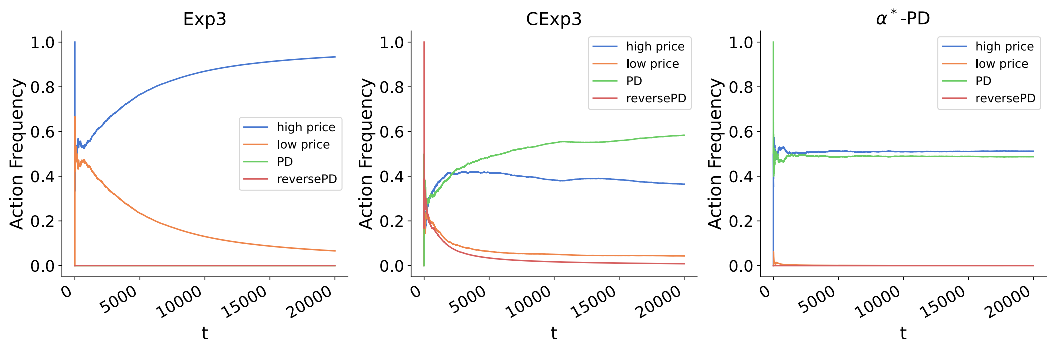

Figure 4 Relative frequency of actions for the seller playing Exp3, CExp3 and α^*-PD. The number of PD and reversePD actions for the Exp3 seller are both 0, as is expected.

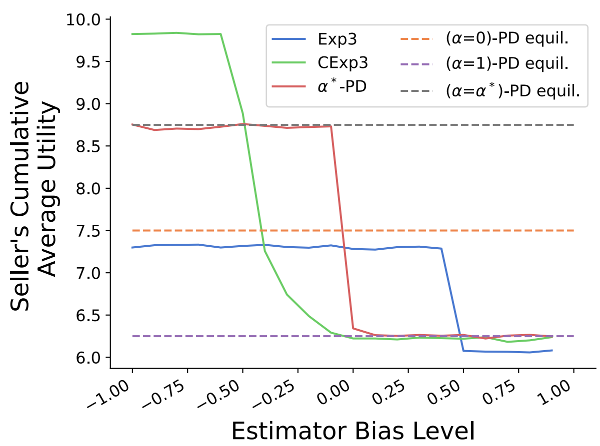

Figure 5 Cumulative average utility of the seller playing against CBER-buyers using biased $\hat{\alpha}$'s.

ϵ_t) and is thus more likely to evade, costing the seller the evasion cost. Against a buyer who underestimates the probability of price discrimination (low values of ϵ_t), neither the Exp3 nor α^*-PD seller gains utility, since the equilibrium behavior of the buyer with consistent $\hat{\alpha}_t$ aligns with the no-price-discrimination equilibrium (see Figure 2). By contrast, the CExp3 seller benefits from a buyer who underestimates the probability of price discrimination, since the seller benefits from discriminatory pricing without incurring the evasion cost. Against a CBER-buyer with consistent estimates, this advantage is impossible at equilibrium.

5 Conclusion

Since the type and level of privacy desired generally depends on the utilities of stakeholders and forms of interaction among them, we propose a game theoretic framework for privacy in this paper. We analyzed the perfect Bayes Nash equilibrium in a single-interaction setting as well as no-regret and no-policy-regret dynamics emerging over repeated interactions. In both these settings, we show how the different components of the game – utilities, actions and information sets (information available to players when choosing actions) impact the privacy levels that emerge.

Our results shed light on the impacts of different privacy-related interventions – we showed that enabling a seller to credibly commit to privacy (e.g., through privacy legislation like the GDPR) or revealing the seller's past behavior (e.g., through privacy auditing) can surprisingly improve their utility. Thus, we believe our framework can be used to help analyze and craft privacy policies.

References

1 Alessandro Acquisti, Curtis Taylor, and Liad Wagman. The economics of privacy. *Journal of Economic Literature*, 54(2):442–492, 2016.

2 Alessandro Acquisti and Hal Varian. Conditioning prices on purchase history. *Marketing Science*, 2004.

3 Raman Arora, Michael Dinitz, Teodor V. Marinov, and Mehryar Mohri. Policy regret in repeated games. *Advances in Neural Information Processing Systems*, 2020.

4 Modibo K. Camara, Jason D. Hartline, and Aleck Johnsen. Mechanisms for a no-regret agent: Beyond the common prior. *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, 2020.

5 Vincent Conitzer, Curtis Taylor, and Liad Wagman. Hide and seek: Costly consumer privacy in a market with repeated purchases. *Marketing Science*, 31(2):277–292, 2012.

6 Yuan Deng, Jon Schneider, and Balasubramanian Sivan. Strategizing against no-regret learners. *Advances in Neural Information Processing Systems*, 32, 2019.

7 Jinshuo Dong, Aaron Roth, and Weijie J Su. Gaussian differential privacy. *arXiv preprint arXiv:1905.02383*, 2019.

8 Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3–4):211–407, 2014.

9 Jeffrey Ely and Juuso Valimaki. Bad reputation. *The Quarterly Journal of Economics*, 118(3):785–814, 2003.

10 Dean P Foster and Rakesh V Vohra. Calibrated learning and correlated equilibrium. *Games and Economic Behavior*, 21(1-2):40, 1997.

11 John Bellamy Foster and Robert McChesney. Surveillance capitalism. *Monthly review*, 66(3):1–31, 2014.

12 Drew Fudenberg and J Miguel Villas-Boas. Behavior-based price discrimination and customer recognition. *Handbook on Economics and Information Systems*, 1:377–436, 2006.

13 Nika Haghtalab, Chara Podimata, and Kunhe Yang. Calibrated Stackelberg games: Learning optimal commitments against calibrated agents. *Advances in Neural Information Processing Systems*, 36, 2024.

14 Oliver D Hart and Jean Tirole. Contract renegotiation and coasian dynamics. *The Review of Economic Studies*, 55(4):509–540, 1988.

15 Johannes Horner. Reputation and competition. *American Economic Review*, 92(3):644–663, 2002.

16 Shota Ichihashi. Online privacy and information disclosure by consumers. *American Economic Review*, 110(2):569–595, 2020.

17 Tor Lattimore and Csaba Szepesvari. *Bandit Algorithms*. Cambridge University Press, 2020.

18 Katrina Ligett and Kobbi Nissim. We need to focus on how our data is used, not just how it is shared. *Communications of the ACM*, 66(9):32–34, 2023.

19 Ilya Mironov. Rényi differential privacy. In *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*, pages 263–275. IEEE, 2017.

20 Rodrigo Montes, Wilfried Sand-Zantman, and Tommaso Valletti. The value of personal information in markets with endogenous privacy. *Center for Economic and International Studies*, 13(352), 2015.

21 Kobbi Nissim, Aaron Bembenek, Alexandra Wood, Mark Bun, Marco Gaboardi, Urs Gasser, David R O'Brien, Thomas Steinke, and Salil Vadhan. Bridging the gap between computer science and legal approaches to privacy. *Harvard Journal of Law & Technology*, 31:687, 2017.

22 Carl Shapiro. Premiums for high quality products are returns to reputation. *Quarterly Journal of Economics*, 98(4):659–679, 1983.

23 Alicia Solow-Niederman. Information privacy and the inference economy. *Northwestern University Law Review*, 117:357, 2022.

24 Jun Tang, Aleksandra Korolova, Xiaolong Bai, Xueqiang Wang, and Xiaofeng Wang. Privacy loss in Apple's implementation of differential privacy on macOS 10.12. *arXiv preprint arXiv:1709.02753*, 2017.

A Proofs from Section 2

A.1 Proof of Theorem 2

Proof. Part (a) comes from the fact that $\underline{\theta}$ buyers have no reason to pretend to have a higher valuation for the good than they actually do. Part (e) comes from the fact that buyers are utility maximizing.

Part (c) comes from the following reasoning: since signal blind sellers cannot see the buyers' signals, they must choose one price to set for all buyers. The seller wants to maximize their revenue, so they would ideally want to set the highest price that the buyer is willing to pay ($\overline{\theta}$ for $\overline{\theta}$-buyers and $\underline{\theta}$ for $\underline{\theta}$-buyers). However, the seller does not know the type of the buyer; all they know is the probability μ that the buyer is $\overline{\theta}$. The seller has to make a decision between charging $\overline{\theta}$ or $\underline{\theta}$. If the seller charges $\underline{\theta}$, both $\underline{\theta}$ and $\overline{\theta}$ agents would be willing to buy, so the expected revenue is $\underline{\theta}$. If the seller charges the higher price $\overline{\theta}$, only $\overline{\theta}$ agents would be willing to buy, so the expected revenue is $\mu\overline{\theta}$, which corresponds to

$$p^*_{\text{signalblind}} = \begin{cases} \underline{\theta} & \text{if } \underline{\theta} \leq \mu\overline{\theta} \\ \overline{\theta} & \text{if } \underline{\theta} > \mu\overline{\theta}. \end{cases}$$

Part (c) and (d) come from the following best-response arguments. Our goal is to show $p^*_{\text{signalaware}}$ is a best response given q^* and vice versa, where

$$p^*_{\text{signalaware}}(s) = \begin{cases} \underline{\theta} & \text{if } s = \underline{s} \text{ is observed} \\ \overline{\theta} & \text{if signal } s = \overline{s} \text{ is observed.} \end{cases} \qquad \text{and} \qquad q^* = \min\left\{1, \frac{(1-\mu)\underline{\theta}}{\mu\Delta\theta}\right\}$$

What is the signal aware seller's best response after seeing $\overline{s}$? From part (a), we know that $\underline{\theta}$ buyers never signal $\overline{\theta}$, so the seller knows that a $\overline{s}$ signal implies that the buyer is type $\overline{\theta}$ and should therefore set a price of $\overline{\theta}$ after seeing $\overline{s}$, i.e., $p^*_{\text{signalaware}}(\overline{s}) = \overline{\theta}$.

What is the signal aware seller's best response after seeing $\underline{s}$? In order for $p^*_{\text{signalaware}}$ to be a best response, it must maximize the seller's expected utility, where the expectation is over the seller's posterior belief over the buyer's type given that they have signaled $\underline{s}$. Given probability q^* that the $\overline{\theta}$ buyer sends signal $\underline{s}$, the seller's posterior belief $\hat{\mu}$ that the buyer is type $\overline{\theta}$ is

$$\hat{\mu} = \mathbb{P}(\theta = \overline{\theta} | s = \underline{s}) = \frac{\mathbb{P}(s = \underline{s} | \theta = \overline{\theta})\mathbb{P}(\theta = \overline{\theta})}{\mathbb{P}(s = \underline{s} | \theta = \overline{\theta})\mathbb{P}(\theta = \overline{\theta}) + \mathbb{P}(s = \underline{s} | \theta = \underline{\theta})\mathbb{P}(\theta = \underline{\theta})} = \frac{q^*\mu}{q^*\mu + 1 - \mu}.$$

Let $f(p)$ denote the seller's expected utility from charging price p after observing signal $\underline{s}$, so

$$f(p) = \begin{cases} p - \hat{\mu} q^* c_S & \text{if } p < \underline{\theta} \\ \hat{\mu} p - \hat{\mu} q^* c_S & \text{if } p \in [\underline{\theta}, \overline{\theta}]. \end{cases}$$

In order for $p^*_{\text{signalaware}}(\underline{s})$ to be a best response, it must be the value that maximizes f:

$$p^*_{\text{signalaware}}(\underline{s}) = \max_p f(p) = \begin{cases} \underline{\theta} & \text{if } q^* \leq \min\left\{1, \frac{(1-\mu)\underline{\theta}}{\mu \Delta \theta}\right\} \\ \overline{\theta} & \text{else.} \end{cases} = \underline{\theta},$$

where the last equality comes from the choice of q^*. This shows that $p^*_{\text{signalaware}}(\underline{s}) = \underline{\theta}$ is a best response for the seller. We now turn our attention to the $\overline{\theta}$-buyer.

What is the optimal probability q^* of evasion for the $\overline{\theta}$-buyer? Let $g(q)$ denote the the expected utility for the $\overline{\theta}$ buyer when they evade with probability q, given that the seller is playing $p^*_{signalblind}$ if they are signal blind and $p^*_{signalaware}$ if they are signal aware, so

$$\begin{aligned} g(q) = &\ \mathbb{P}(\text{seller is signal blind})(\overline{\theta}\text{-buyer utility if seller plays } p^*_{\text{signalblind}}) \\ &+ \mathbb{P}(\text{seller is signal aware})(\overline{\theta}\text{-buyer utility if seller plays } p^*_{\text{signalaware}}). \end{aligned} \tag{2}$$

- If $1 \leq {}^{(1-\mu)\underline{\theta}}/_{\mu\Delta\theta}$, this implies that $\underline{\theta} \geq \mu\overline{\theta}$, so (2) simplifies to

  $$u_B = (1-\alpha)\Delta\theta + (\alpha\Delta\theta - c_B)q.$$

- If ${}^{(1-\mu)\underline{\theta}}/_{\mu\Delta\theta} \leq 1$, this implies $\theta < \mu\overline{\theta}$, so (2) simplifies to

  $$u_B = \begin{cases} (\alpha\Delta\theta - c_B)q & \text{if } q \leq {}^{(1-\mu)\underline{\theta}}/_{\mu\Delta\theta} \\ -c_B q & \text{else.} \end{cases}$$

Combining everything, we see that the $\overline{\theta}$-buyer's optimal probability of evasion is q^* as written in the theorem statement. ◀

B Repeated PD Protocol

The detailed algorithm is described in the arXiv version.

C Proofs from Section 3

C.1 Proof of Proposition 10

Proof. For each round t, let $I_t = \mathbb{1}\left\{\exists i \text{ s.t. } s_i^t = \overline{s} \text{ and } \exists j \text{ s.t. } s_j^t = \underline{s}\right\}$ be an indicator for whether both types of signals are observed at round t, i.e., whether round t is "informative" about if there is price discrimination. For rounds t with $I_t = 1$, we additionally define the following random variables: $\overline{P}_t = p_i^t$ for the smallest $i \in [N]$ such that $s_i^t = \overline{s}$; $\underline{P}_t = p_j^t$ for the smallest $j \in [N]$ such that $s_j^t = \underline{s}$; and $X_t = \mathbb{1}\left\{\overline{P}_t \neq \underline{P}_t\right\}$, an indicator for observed price discrimination. Note that the choice to define $\overline{P}_t$ and $\underline{P}_t$ to correspond to the *smallest* index satisfying the corresponding condition is simply for concreteness; we could equivalently sample uniformly from the set of indices satisfying the condition.

Recall that $H_t = ((\theta_i^\tau, s_i^\tau, p_i^\tau)_{i=1}^n)_{\tau=1}^{t-1}$ is the history known by buyers at the beginning of round t. Consider the following estimator:

$$\hat{\alpha}_t = \frac{1}{t} \sum_{\tau=1}^{t-1} \frac{X_\tau I_\tau}{\mathbb{E}[I_\tau | H_\tau]} \tag{3}$$

The expectation $\mathbb{E}[I_\tau|H_\tau]$ is over the randomness at round τ. Note that $\hat{\alpha}_t$ is computable based on the history H_t, because $\mathbb{E}[I_\tau|H_\tau]$ is computable for any $\tau < t$. We will now show that $\hat{\alpha}_t$ satisfies Definition 8. We start by computing the expectation of $\hat{\alpha}_t$:

$$\begin{aligned}
\mathbb{E}[\hat{\alpha}_t] &= \mathbb{E}\left[\frac{1}{t}\sum_{\tau=1}^{t-1}\frac{X_\tau I_\tau}{\mathbb{E}[I_\tau|H_\tau]}\right] \\
&= \frac{1}{t}\sum_{\tau=1}^{t-1}\mathbb{E}\left[\frac{X_\tau I_\tau}{\mathbb{E}[I_\tau|H_\tau]}\right] && \text{linearity of expectation} \\
&= \frac{1}{t}\sum_{\tau=1}^{t-1}\mathbb{E}\left[\mathbb{E}\left[\frac{X_\tau I_\tau}{\mathbb{E}[I_\tau|H_\tau]}\middle| H_\tau\right]\right] && \text{tower rule} \\
&= \frac{1}{t}\sum_{\tau=1}^{t-1}\mathbb{E}\left[\frac{\mathbb{E}[X_\tau I_\tau|H_\tau]}{\mathbb{E}[I_\tau|H_\tau]}\right]
\end{aligned}$$

Observe that X_τ and I_τ are independent given H_τ. To see why, note that the randomness in $X_\tau|H_\tau$ comes only from the randomness in the seller's mixed strategy at round τ, whereas the randomness in $I_\tau|H_\tau$ comes only from the randomness in the buyers mixed strategy at round τ. The mixed strategies are fixed given H_τ, and the additional randomness is independent. Thus,

$$\begin{aligned}
&= \frac{1}{t}\sum_{\tau=1}^{t-1}\mathbb{E}\left[\frac{\mathbb{E}[X_\tau|H_\tau]\mathbb{E}[I_\tau|H_\tau]}{\mathbb{E}[I_\tau|H_\tau]}\right] \\
&= \frac{1}{t}\sum_{\tau=1}^{t-1}\mathbb{E}[\mathbb{E}[X_\tau|H_\tau]] \\
&= \frac{1}{t}\sum_{\tau=1}^{t-1}\mathbb{E}\left[\mathbb{E}[\mathbb{1}\left\{\overline{P}_\tau \neq \underline{P}_\tau\right\}|H_\tau]\right] && \text{by definition of } X_\tau
\end{aligned}$$

Since $\overline{P}_\tau|H_\tau \sim \pi_t^p(\cdot|s=\overline{s})$ and $\underline{P}_\tau|H_\tau \sim \pi_t^p(\cdot|s=\underline{s})$ by definition of the game, we have

$$= \frac{1}{t}\sum_{\tau=1}^{t-1}\alpha_\tau$$

Finally, plugging in the above expression with $t = T$ into the criterion for consistency, we have

$$\lim_{T\to\infty}\left|\mathbb{E}[\hat{\alpha}_T] - \frac{1}{T}\sum_{t=1}^{T}\alpha_t\right| = \lim_{T\to\infty}\left|\frac{1}{T}\sum_{t=1}^{T-1}\alpha_t - \frac{1}{T}\sum_{t=1}^{T}\alpha_t\right| = \lim_{T\to\infty}\frac{\alpha_T}{T} = 0$$

as desired. The last equality comes from the fact that α_T is a probability, so it is bounded between 0 and 1 for all T. ◀

C.2 Proof of Proposition 13

Proof. First, we will show that always price-discriminating ($\pi_t^p = \pi_{\text{PD}}^p$ for all $t \in [T]$) is no-regret against CBER-buyers. For CBER-buyers, their strategy π_t^s at each round t is either π_{truthful}^s or $\pi_{\text{strategic}}^s$. For both these buyer responses, the seller's optimal strategy is to always price discriminate as shown in the computation of the seller's equilibrium response in the proof of Theorem 2. In other words, the seller incurs zero regret in each round and thus zero average regret.

Next, we will analyze the seller's average utility. Note that when $\pi_t^p = \pi_{\text{PD}}^p$, the probability of seeing different prices for different signals is $\alpha_t = 1$, so $(1/t)\sum_{\tau=1}^{t} \alpha_\tau = 1$ for all t. By the consistency property, $\hat{\alpha}_t$ becomes greater than α^* eventually (where α^* is as defined in Corollary 5) and the buyer plays $\pi_{\text{strategic}}^s$. In other words, eventually the seller and buyers will all be playing their equilibrium strategies for the PD-game with $\alpha = 1$, so their average utilities will converge to the corresponding equilibrium utilities. We make this argument formal below.

Define $\kappa < \infty$ to be the maximum utility that can be achieved by a seller in any round. The finiteness of κ is guaranteed by definition of the seller's utility function. Define $A_T = \{\exists t > \sqrt{T} \text{ s.t. } \hat{\alpha}_t > \alpha^*\}$ and let $A_T^C = \{\hat{\alpha}_t > \alpha^* \text{ for all } t > \sqrt{T}\}$ denote the complement. Let $\gamma_T = \mathbb{P}(A_T)$ and $1 - \gamma_T = \mathbb{P}(A_T^C)$ denote the corresponding probabilities. Then, we can decompose the expected average seller's utility as

$$\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} U_S(\pi_t)\right] = \gamma_t \mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} U_S(\pi_t)\middle| A_T\right] + (1-\gamma_t)\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} U_S(\pi_t)\middle| A_T^C\right]. \tag{4}$$

The first term of (4) is trivially upper bounded by $\gamma_T \kappa$.

To bound the second term of (4), first note that for any round t where $\hat{\alpha}_t > \alpha^*$, the buyer's strategy will be equivalent to their equilibrium strategy with $\alpha = 1$. Thus, the best utility that the seller can achieve for those rounds is $u_S(1)$. It follows that under the condition that $\hat{\alpha}_t > \alpha^*$ for every $t > \sqrt{T}$, we have

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T} U_S(\pi_t) &= \frac{1}{T}\sum_{t=\sqrt{T}}^{T} U_S(\pi_t) + \frac{1}{T}\sum_{t=1}^{\sqrt{T}} U_S(\pi_t) \\
&\leq \frac{1}{T}\sum_{t=\sqrt{T}}^{T} u_S(1) + \frac{1}{T}\sum_{t=1}^{\sqrt{T}} \kappa \\
&= \frac{T-\sqrt{T}}{T} u_S(1) + \frac{\sqrt{T}\kappa}{T} \\
&\leq u_S(1) + \frac{\kappa - u_S(1)}{\sqrt{T}}.
\end{aligned}$$

Plugging back into (4), we get

$$\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} U_S(\pi_t)\right] \leq \gamma_t \kappa + (1-\gamma_t)\left(u_S(1) + \frac{\kappa - u_S(1)}{\sqrt{T}}\right).$$

By the consistency property (Lemma 9), we know $\lim_{T\to\infty} \gamma_T = 0$, which yields the stated asymptotic bound on the seller's average utility. ◀

C.3 Missing Proofs of Lemmas in Proof of Proposition 14

Proof of Lemma 15. Based on the utility orderings from Corollary 4, note the following ordering of seller utilities for different combinations of buyer and seller policies:

$$U_S(\pi_{\text{truthful}}^s, \pi_{\text{PD}}^p) > U_S(\pi_{\text{truthful}}^s, \pi_{\text{noPD}}^p) > U_S(\pi_{\text{strategic}}^s, \pi_{\text{PD}}^p) > U_S(\pi_{\text{strategic}}^s, \pi^p),$$

where π^p is any other pricing strategy besides π^p_{PD} and π^p_{noPD}. We can then write

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T} U_S(\pi_t) &\le \frac{1}{T}\sum_{t\in\mathcal{T}} U_S(\pi^s_{\text{truthful}}, \pi^p_{\text{noPD}}) + \sum_{t\in[T]\setminus\mathcal{T}} U_S(\pi^s_{\text{strategic}}, \pi^p_{\text{PD}}) \\ &= \frac{|\mathcal{T}|}{T} U_S(\pi^s_{\text{truthful}}, \pi^p_{\text{noPD}}) + \left(1 - \frac{|\mathcal{T}|}{T}\right) U_S(\pi^s_{\text{strategic}}, \pi^p_{\text{PD}})\end{aligned}$$

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} U_S(\pi_t) \le U_S(\pi^s_{\text{truthful}}, \pi^p_{\text{noPD}}) \lim_{T\to\infty}\frac{|\mathcal{T}|}{T} + U_S(\pi^s_{\text{strategic}}, \pi^p_{\text{PD}}) \left(1 - \lim_{T\to\infty}\frac{|\mathcal{T}|}{T}\right)$$

Since $U_S(\pi^s_{\text{truthful}}, \pi^p_{\text{PD}}) > U_S(\pi^s_{\text{strategic}}, \pi^p_{\text{PD}})$, the above upper bound on the limit of the average utility is increasing as $\lim_{T\to\infty}|\mathcal{T}|/T$ is increasing. When $\lim_{T\to\infty}|\mathcal{T}|/T \le \alpha^*$,

$$\begin{aligned}&\le \alpha^* U_S(\pi^s_{\text{truthful}}, \pi^p_{\text{PD}}) + (1-\alpha^*) U_S(\pi^s_{\text{strategic}}, \pi^p_{\text{PD}}) \\ &< \alpha^* U_S(\pi^s_{\text{truthful}}, \pi^p_{\text{PD}}) + (1-\alpha^*) U_S(\pi^s_{\text{truthful}}, \pi^p_{\text{PD}}) = \mathbb{U}^*_1\end{aligned}$$

◀

Proof of Lemma 16. Let R^S_T denote the average seller utility in the T rounds. Since the seller is no regret, $\lim_{T\to\infty} R^S_T = 0$.

Consider the regret due to the seller deviating to π^p_{PD} in each round. The gain in utility in each round due to this deviation is non-negative since π^p_{PD} is the best-response to both possible buyer strategies $\pi^s_{\text{truthful}}, \pi^s_{\text{strategic}}$. We can then lower bound the regret by considering regret accumulated in rounds where $\hat{\alpha}_t \le \alpha^*$. In such rounds, all buyers are truthful, so whenever the seller does not charge a buyer the price corresponding to their signal type, they incur regret. The probability that the seller observes $\overline{s}$ but charges $\overline{\theta}$ is $\mu\pi^p_t(\underline{\theta}|\underline{s})$, and this yields a loss of utility of $\Delta\theta$, because the buyer is type $\overline{\theta}$. Similarly, the probability that the seller observes $\underline{s}$ but charges $\overline{\theta}$ is $(1-\mu)\pi^p_t(\overline{\theta}|\underline{s})$, and this yields a loss of utility of $\underline{\theta}$, since the buyer is type $\underline{\theta}$.

$$\begin{aligned}R^S_T &\ge \frac{1}{T}\sum_{t:\hat{\alpha}_t\le\alpha^*} \mu\Delta\theta\pi^p_t(\underline{\theta}|\overline{s}) + (1-\mu)\underline{\theta}\pi^p_t(\overline{\theta}|\underline{s}) \\ &\ge \frac{1}{T}\sum_{t:\hat{\alpha}_t\le\alpha^*} \kappa(1 - (\pi^p(\overline{\theta}|\overline{s}) - \pi^p(\overline{\theta}|\underline{s}))) \qquad \text{where } \kappa := \min\{\mu\Delta\theta, (1-\mu)\underline{\theta}\}\end{aligned}$$

$$\implies \qquad \frac{1}{T}\sum_{t\in\mathcal{T}} \left(\pi^p_t(\overline{\theta}|\overline{s}) - \pi^p_t(\overline{\theta}|\underline{s})\right) \ge \frac{|\mathcal{T}|}{T} - \frac{R^S_T}{\kappa}$$

The above inequality shows that $|\mathcal{T}|/T$ is bounded above by some measure of simultaneous truthfulness and price discrimination. Each quantity $\pi^p_t(\overline{\theta}|\overline{s}) - \pi^p_t(\overline{\theta}|\overline{s})$ is a measure of price-discrimination in each round and is related to α_t as described in the following lemma.

▶ **Lemma C.1.** *When seller pricing strategies are supported on* $\{\overline{\theta}, \underline{\theta}\}$, $\alpha_t \ge \pi^p_t(\overline{\theta}|\overline{s}) - \pi^p_t(\overline{\theta}|\overline{s})$

Proof of Lemma C.1. Since seller pricing strategies are supported on $\{\overline{\theta}, \underline{\theta}\}$, α_t which is the probability of seeing different prices for different signals is

$$\begin{aligned}\alpha_t &= \pi^p_t(\overline{\theta}|\overline{s})\pi^p_t(\underline{\theta}|\underline{s}) + \pi^p_t(\underline{\theta}|\overline{s})\pi^p_t(\overline{\theta}|\underline{s}) \\ &= \pi^p_t(\overline{\theta}|\overline{s})(1 - \pi^p_t(\overline{\theta}|\underline{s})) + (1 - \pi^p_t(\overline{\theta}|\overline{s}))\pi^p_t(\overline{\theta}|\underline{s}) \\ &= \pi^p_t(\overline{\theta}|\overline{s}) + \pi^p_t(\overline{\theta}|\underline{s}) - 2\pi^p_t(\overline{\theta}|\overline{s})\pi^p_t(\overline{\theta}|\underline{s}) \\ &= \pi^p_t(\overline{\theta}|\overline{s}) - \pi^p_t(\overline{\theta}|\underline{s}) + 2\pi^p_t(\overline{\theta}|\underline{s})(1 - \pi^p_t(\overline{\theta}|\overline{s})) \\ &\ge \pi^p_t(\overline{\theta}|\overline{s}) - \pi^p_t(\overline{\theta}|\underline{s})\end{aligned}$$

◀

By inequality 1, Lemma C.1, and since $\lim_{T\to\infty} R_T^S/\kappa = 0$, $\lim_{T\to\infty} \frac{|\mathcal{T}|}{T} \leq \lim_{T\to\infty} \frac{1}{T}\sum_{t\in\mathcal{T}} \alpha_t$. ◀

Proof of Lemma 17. Consider the last index $t^* \in \mathcal{T}$. Let us consider two cases. The first case is $\lim_{T\to\infty} t^*/T < \alpha^*$. Then, $\sum_{t\in\mathcal{T}} \alpha_t/T \leq |\mathcal{T}|/T \leq t^*/T$. This implies $\lim_{T\to\infty} \sum_{t\in\mathcal{T}} \alpha_t/T \leq \alpha^*$. In the second case, $\lim_{T\to\infty} t^* = \infty$. Consider $\bar{\alpha}_{t^*} = \frac{1}{t^*}\sum_{t\leq t^*} \alpha_t \geq \sum_{t\in\mathcal{T}} \alpha_t/T$. By the consistency property, $\lim_{T\to\infty} \bar{\alpha}_{t^*} = \hat{\alpha}_{t^*}$. $\hat{\alpha}_{t^*} \leq \alpha^*$ since $t^* \in \mathcal{T}$. ◀

C.4 Proofs of Propositions 20, 21

Please see arXiv version for proofs.