Selling Data to an Agent with Endogenous Information*

Yingkai Li[†]

November 21, 2021

Abstract

We consider the model of the data broker selling information to a single agent to maximize his revenue. The agent has private valuation for the additional information, and upon receiving the signal from the data broker, the agent can conduct her own experiment to refine her posterior belief on the states with additional costs. In this paper, we show that in the optimal mechanism, the agent has no incentive to acquire any additional costly information under equilibrium. Still, the ability to acquire additional information distorts the incentives of the agent, and reduces the optimal revenue of the data broker. Moreover, we characterize the optimal mechanism when the valuation function of the agent is separable. The optimal mechanism in general may be complex and contain a continuum of menu entries. However, we show that posting a deterministic price for revealing the states is optimal when the prior distribution is sufficiently informative or the cost of acquiring additional information is sufficiently high, and obtains at least half of the optimal revenue for arbitrary prior and cost functions.

1 Introduction

Information can help the decision makers refine their knowledge and make better decisions when there are uncertainty over the environments. The desire for additional information creates opportunities for the data broker to collect payments from the decision makers. There is a large market for selling information such as selling consumer data (e.g., Acxiom and Datalogix), user profiles (e.g., Facebook), credit reports (e.g., Experian, TransUnion), or cookies for web usage (c.f., Bergemann and Bonatti, 2015). Those data brokers extract huge revenue from providing valuable information to the decision makers.

In this paper, we consider the problem of maximizing the revenue of the data broker, where the agent can endogenously acquire additional information. Specifically, there is an unknown state and both the data broker and the agent have a common prior over the set of possible states. The data broker can offer a menu of information structures for revealing the states with associated prices to the agent. Then the agent picks the expected utility maximization entry from the menu, and pays the corresponding price to the data broker. The agent has a private valuation for information and can acquire additional costly information upon receiving the signal from the data broker. The literature has acknowledged the possibility for the agents to conduct their own experiments to be privately informed of the states (e.g., Bergemann, Bonatti, and Smolin, 2018). The distinct feature in our model is that the decision for acquiring additional information is endogenous. Specifically, after receiving the signal from the data broker, the agent can subsequently acquire additional

---

[†]Department of Computer Science, Northwestern University. Email: yingkai.li@u.northwestern.edu

information with costs. For example, the agent is a decision maker who chooses an action to maximize her expected utility based on her posterior belief over the states. The agent will first acquire information from the data broker, and based on her posterior, she can potentially conduct more experiments to refine her belief before taking the action. Another example captured in our model is where the agent is a firm that sells products to consumers, and the information the data broker provides is a market segmentation of the consumers. The firm has a private and convex cost for producing different level of qualities for the product, and the firm can conduct his own experiments (e.g., sending surveys to potential consumers) with additional costs to further segment the market after receiving the information from the data broker.[1] In addition, in our model, we allow the firm to repeat the market research until it is not beneficial to do so, i.e., when the cost of information exceeds the marginal benefits of information. This captures the situation that the firm can decide the date for announcing the product to the market, and before the announcement, the firm sends out surveys to potential consumers each day to learn the segmentation of the markets. At the end of each day, the firm receives an informative signal through the survey, and decides whether to continue the survey in next day, or stop the survey and announce the product with corresponding market prices to the public.

In this model of allowing the agent to acquire information endogenously, we assume that the data broker can provide the same information with lower cost. We argue that this is a natural assumption in reality since data brokers are usually firms that have focused in the relative research area for decades and have the technological innovation advantages to reduce the cost for collecting information.[2] An important observation in our paper is that in the optimal mechanism, after the agent receives the signal from the data broker, she does not have incentive to acquire any additional information with positive cost. Note that this does not imply the equivalence to the setting where the agent cannot acquire additional information. In fact, we show that the ability of acquiring additional information distorts the agent's incentive to purchase information from the data broker, and the optimal revenue of the data broker shrinks as the set of additional experiments available to the agent increases. In additional, this observation is robust in the model of multiple sellers. That is, the agent's ability to acquire additional information originates from the fact that she can acquire information from different data brokers sequentially. We show that under equilibrium, the agent will only acquire information from the firm with lowest cost for providing information.

In this paper, we also provide more structural characterizations of the revenue optimal mechanism under the assumption that the valuation function of the agent is separable, i.e., the value of the agent for any posterior distribution is simply the product of her private type and the value of the posterior distribution. In the examples we provided in previous paragraphs, both the decision maker who chooses an optimal action to maximize her payoff based on the posterior belief and the firm that sells products to consumers to maximize the revenue satisfy the separable condition for their valuation functions. Essentially, the separable condition assumes that the private type of the agent represents her value for additional information, and there is a linear structure on the preference. It excludes the situation where the private type of the agent represents an exogenous private signal correlated with the states (e.g., Bergemann, Bonatti, and Smolin, 2018).

We show that with separable valuations, in the simple case where the agent cannot acquire additional information, selling information reduces to the single-item auction where the quality of the information serves a similar role as the item allocation in the single-item auction. By an

---

[1]Yang (2020) studies a similar model, where the firm cannot conduct his own market research to refine his knowledge. Moreover, the cost function of the firm is linear in Yang (2020), which leads to a qualitatively different result compared to our model. See Section 4.1 for a detailed discussion.

[2]In addition, if the cost of the data broker is always higher than the cost of the agent for acquiring information, it is easy to verify that the revenue of the data broker is zero in this case.

argument similar to Myerson (1981), the optimal mechanism is a simple posted pricing mechanism for revealing full information. When the agent can acquire additional costly information, there exists a threshold type θ^* such that (1) for any type $\theta \geq \theta^*$, the optimal mechanism reveals full information to the agent; and (2) for any type $\theta < \theta^*$, the optimal mechanism may reveal partial information and the individual rational constraint always binds. The first statement is the standard no distortion at the top observation in the optimal mechanism. The second statement suggests that the optimal mechanism may discriminate lower types of the agent by offering the options of revealing partial information to the agent with lower prices. Moreover, the allocations and the prices for those lower types are set such that the agent is exactly indifferent between participation and choosing the outside option (by conducting her own experiments with additional costs). This is a clear distinction from selling information without endogenous information where the individual rational constraint for the lowest type is sufficient for characterizing the optimal mechanism in the latter case.

Our characterizations suggest that the optimal mechanisms for selling information may be complex and contain a continuum of menu entries. A critic on the complex mechanisms is the applicability of the mechanisms in practice compared to the simple but sub-optimal mechanisms. There is a long line of work providing theoretical justifications for simple mechanisms in practice by showing the approximate optimality of simple mechanisms when the optimal mechanisms are complex, e.g., Hartline and Roughgarden (2009); Hartline (2012); Babaioff, Immorlica, Lucier, and Weinberg (2020). In the setting of selling information, a natural candidate for simple mechanisms is posting a deterministic price for revealing full information. Under the assumption of separable valuation, we show that when the prior distribution over states is sufficiently informative, or the cost of information is sufficiently high, posted pricing for revealing full information is the revenue optimal mechanism. Moreover, without any restriction on the prior distribution or the cost function, posted pricing for revealing full information is sub-optimal but achieves at least half of the optimal revenue in the worst case.[3]

We study the comparative statics by comparing the setting with endogenous information to the setting without endogenous information. We show that in the revenue optimal mechanism for an agent with endogenous information, the social welfare is higher since the data broker will provide partial information to the low types in the endogenous information setting. Moreover, the expected revenue of the data broker is smaller, and hence the residual surplus left with the agent is higher. Finally, we apply our characterization of the optimal mechanism to several economic applications, e.g., when the data buyer is a decision maker minimizing the loss of making predictions, or a firm selling products to consumers to maximize profit.

The paper is organized as follows. In Section 2, we formally introduce the model and the assumptions considered in this paper. We provide the characterizations of the optimal mechanisms for general valuation functions in Section 3 and obtain more structure results for separable valuations in Section 4. We conclude our paper in Section 5.

---

[3]Bergemann, Cai, Velegkas, and Zhao (2021) also considers the approximate optimality of posted pricing for revealing full information in the setting where the agent has a private and exogenous signal that is informative about the state. They show that the approximation factor (the multiplicative ratio between the optimal revenue and the revenue from posted pricing) degrades linearly as the number of potential actions for the agent increases. In contrast, we show that pricing for full information is a 2-approximation to the optimal revenue regardless of the number of potential actions if the signals are endogenous.

1.1 Related Work

There is a large literature on selling information to agents with uncertainty over the states. Those papers can be classified into two categories according to the agents' private types. The first category is when the agents' private types represent their willingness to pay for different signal structures (e.g., Yang, 2020; Smolin, 2020; Liu, Shen, and Xu, 2021). In this case, the private types of the agents are assumed to be independent from the realization of the state, and hence the private types do not affect the belief updating process for receiving the signals. The second category is when the agents' private types represent their private signals that are informative about the states (e.g., Admati and Pfleiderer, 1986, 1990; Bergemann, Bonatti, and Smolin, 2018; Bergemann, Cai, Velegkas, and Zhao, 2021). In this case, agents have heterogeneous prior beliefs for the states, and they will update their posteriors accordingly upon receiving the signals from the seller. Note that in both lines of work, the private types of the agents are given exogenously, and hence it is a pure adverse selection model. The distinct feature in our paper compared to those models is that we allow agents to acquire additional information endogenously. Moreover, agents have private preferences for different signal structures. Thus the main focus of our paper is the interaction between adverse selection and moral hazard, and we will provide characterizations of the optimal mechanisms in this setting.

Our paper is relevant to the literature of mechanism design with endogenous information. Crémer and Khalil (1992) considers the model of endogenous information in a contract design model. The main distinction from their model and our paper is the timeline of the agent. In their paper, the agent gather information before signing the contract, while in our model, the agent can observe additional information after the interaction with the data broker. This difference in timeline also distinguishes our model from the literature on auction with endogenous entry (Menezes and Monteiro, 2000) and auction with buyer optimal learning (Shi, 2012; Mensch, 2019), where those papers assume that the agents make the information acquisition decision before interacting with the seller, and the mechanism offered by the seller distorts the agents' incentives to learn their valuation.

The model in this paper also contributes to the broader domain of information design and Bayesian persuasion (e.g., Rayo and Segal, 2010; Kamenica and Gentzkow, 2011), particularly the disclosure of information in auctions. Eső and Szentes (2007) and Li and Shi (2017) consider the setting of selling information to the buyer before the auction to maximize the revenue, and Wei and Green (2020) consider the variant where the buyer can walk away without paying the seller after receiving the information. Bergemann and Bonatti (2015) models the problem of a data provider who sells cookies to match the firms with customers, and Bergemann, Brooks, and Morris (2015) explores the set of possible outcomes can be implemented by in a monopoly model by providing the segmentations of the customers. For a detailed discussion on the information in markets, see the survey of Bergemann and Bonatti (2019).

2 Model

There is a single agent making decisions facing uncertainly over the state space Ω. Let D be the prior distribution over the states. The agent has a private type $\theta \in \Theta$, and type θ is drawn from a commonly known distribution F with density f. The expected utility of the agent given posterior belief $\mu \in \Delta(\Omega)$ is $V(\mu, \theta)$ when her type is θ. We assume that V is monotone in μ in Blackwell order for any type θ. There is a data broker who tries to sell information to the agent to maximize his profit by committing to an information structure that signals the state. Note that an information structure (we will also call this as an experiment) is a mapping $\sigma : \Omega \to \Delta(S)$, where

S is the signal space. Let Σ be the set of all possible experiments.

In this paper, upon receiving a signal $s \in S$, the agent can conduct her own experiment to further refine her posterior belief on the state with additional costs. Let $\hat{\Sigma} \subseteq \Sigma$ be the set of possible experiments can be conducted by the agent. The cost of experiment σ given posterior belief μ of the agent is denoted by $C^A(\sigma, \mu) \geq 0$. Let σ^F be the experiment that reveals full information, i.e., $\sigma^F(\omega) = \omega$ for any $\omega \in \Omega$, and let σ^N be the null experiment that reveals no information with zero cost. In this paper, we assume that $\sigma^N \in \hat{\Sigma}$,[4] and $C^A(\sigma, \mu) > 0$ for any $\sigma \neq \sigma^N$.

A mechanism of the data broker is a menu of experiments and associated prices $\{(\pi_i, p_i)\}$, where π_i is a distribution over experiments. The timeline of the model is illustrated as follows.

1. The data broker commits to a mechanism $\mathcal{M} = \{(\pi_i, p_i)\}$.

2. The agent chooses entry $(\pi, p) \in \mathcal{M}$, and pays price p to the data broker. The experiment σ is realized according to distribution π and then announced publicly to the agent.

3. State $\omega \in \Omega$ is realized according to prior D and the data broker sends signal $s \sim \sigma(\omega)$ to the agent.

4. Upon receiving the signal s, the agent forms posterior belief μ, chooses an experiment $\hat{\sigma} \in \hat{\Sigma}$, and pays cost $C^A(\hat{\sigma}, \mu)$.

5. The agent receives a signal $s \sim \hat{\sigma}(\omega)$, forms refined posterior belief $\hat{\mu}$, and receives expected reward $V(\hat{\mu}, \theta)$.

By the revelation principle, it is without loss to consider the revelation mechanism, that is, the data broker commits to a mapping for information structures $\pi : \Theta \to \Delta(\Sigma)$ and payment rule $p : \Theta \to \mathbb{R}$.[5] By slightly overloading the notation, denote

$$V(\mu, \hat{\Sigma}, \theta) \triangleq \max_{\hat{\sigma} \in \hat{\Sigma}} \mathbf{E}_{\hat{\mu} \sim \hat{\sigma}|\mu}[V(\hat{\mu}, \theta)] - C^A(\hat{\sigma}, \mu)$$

as the maximum utility of the agent given posterior belief μ, the set of possible experiments $\hat{\Sigma}$, and private type θ. Here the notation $\hat{\sigma}|\mu$ represents the distribution over posteriors induced by experiment $\hat{\sigma}$ given the prior belief μ. To simplify the notation, we will use $\mathbf{E}_{\mu \sim \pi(\theta)|D}[\cdot]$ to represent $\mathbf{E}_{\sigma \sim \pi(\theta)}\big[\mathbf{E}_{\mu \sim \sigma|D}[\cdot]\big]$.

Definition 1. *The mechanism $\mathcal{M} = (\pi, p)$ is incentive compatible if for any type $\theta, \theta' \in \Theta$, we have*

$$\mathbf{E}_{\mu \sim \pi(\theta)|D}\Big[V(\mu, \hat{\Sigma}, \theta)\Big] - p(\theta) \geq \mathbf{E}_{\mu \sim \pi(\theta')|D}\Big[V(\mu, \hat{\Sigma}, \theta)\Big] - p(\theta'),$$

and the mechanism (π, p) is individual rational if for any type $\theta \in \Theta$, we have

$$\mathbf{E}_{\mu \sim \pi(\theta)|D}\Big[V(\mu, \hat{\Sigma}, \theta)\Big] - p(\theta) \geq V(D, \hat{\Sigma}, \theta).$$

In this paper, without loss of generality, we focus on mechanisms (π, p) that are incentive compatible and individual rational. The goal of the data broker is to maximize the expected revenue $\mathrm{Rev}(\mathcal{M}) \triangleq \mathbf{E}_{\theta \sim F}[p(\theta)]$.

---

[4]Intuitively, this assumes that the agent can always choose not to make any additional experiment and pays no extra cost.

[5]The agent observes the realized information structure $\sigma \sim \pi(\theta)$.

For any experiment $\hat{\sigma} \in \hat{\Sigma}$ and any mapping $\kappa : S \to \hat{\Sigma}$, let $\hat{\sigma} \circ \kappa$ represent the experiment such that the agent first conducts experiment $\hat{\sigma}$, and conditional on receiving the signal $s \in S$, the agent continues with experiment $\kappa(s)$ to further refine the posterior belief. For any belief μ, let $\hat{\mu}_{\hat{\sigma},s,\mu}$ be the posterior belief of the agent when she conducts experiment $\hat{\sigma}$ and receives the signal s. Throughout this paper, we make the following assumption on the set of possible experiments and the cost function.

Assumption 1. *For any experiment $\hat{\sigma} \in \hat{\Sigma}$ and for any mapping $\kappa : S \to \hat{\Sigma}$, we have $\hat{\sigma} \circ \kappa \in \hat{\Sigma}$. Moreover, for any belief μ, we have*

$$C^A(\hat{\sigma} \circ \kappa, \mu) \leq C^A(\hat{\sigma}, \mu) + \int_\Omega \int_S C^A(\kappa(s), \hat{\mu}_{\hat{\sigma},s,\mu}) \, \mathrm{d}\hat{\sigma}(s|\omega) \, \mathrm{d}\mu(\omega).$$

Intuitively, Assumption 1 assumes that the set of possible experiments is closed under sequential learning, and the cost function exhibits preference for one-shot learning.[6] This captures the scenario where the agent can repeatedly conduct feasible experiments based on her current posterior belief. Next we illustrate several examples that satisfies the above assumptions.

- $\hat{\Sigma}$ is a singleton. In this case, the unique experiment $\sigma^N \in \hat{\Sigma}$ is null experiment with zero cost.

- $\hat{\Sigma}$ is the set of all possible experiments, i.e., $\hat{\Sigma} = \Sigma$. The cost function C^A is the reduction in information cost, i.e., $C^A(\hat{\sigma}, \mu) = H(\mu) - \mathbf{E}_{\hat{\mu} \sim \hat{\sigma}|\mu}[H(\hat{\mu})]$ where H is any concave function. Possible choices of the information cost function H includes the entropy function (e.g., Sims, 2003) or more generally the uniformly posterior separable cost functions (e.g., Bloedel and Zhong, 2020). It is easy to verify that uniformly posterior separable cost functions satisfy Assumption 1.

- $\hat{\Sigma}$ is the set of experiments generated by σ^N and $\hat{\sigma}'$ through sequential learning, where σ^N is the one that reveals no additional information with zero cost, and $\hat{\sigma}'$ is an informative experiment that signals the state with fixed cost, i.e., there exists constant $c > 0$ such that $C^A(\hat{\sigma}', \mu) = c$ for all posterior μ. In this case, the agent can choose experiment $\hat{\sigma}'$ as long as it is beneficial for her given her current belief μ, and in total the agent pays the cost c multiplies the number of times the experiment $\hat{\sigma}'$ is conducted.[7]

Note that although the general results in our paper do not require any additional assumption on the valuation function, we will consider the following class of valuation functions in Section 4 to obtain more structure results on the optimal mechanism. Essentially, we will focus on the setting that the private type of the agent represents her value for acquiring additional information. In particular, the valuation function of the agent is separable.

Definition 2. *The valuation $V(\mu, \theta)$ is* separable *if there exists a function $v(\mu)$ such that $V(\mu, \theta) = v(\mu) \cdot \theta$ for any posterior μ and any type θ.*

Next we introduce two canonical settings that satisfy the separable assumption on the valuation function.

---

[6] Bloedel and Zhong (2020) provide a characterization for the cost function to be indifference for one-shot learning with additional regularity assumptions. In our paper, we only need to assume weak preference for one-shot learning, and the additional regularity assumptions in Bloedel and Zhong (2020) are not essential.

[7] In this case, the agent solves an optimal stopping problem for acquiring additional information. We can also have a continuous time version for acquiring information when the agent has access to signals following a Brownian motion (Georgiadis and Szentes, 2020) or a Poisson process (Zhong, 2017).

Example 1. *Consider the model of a decision maker trying to make a prediction over the finite states Ω. In our paper, the agent is the decision maker who chooses an action from the action space A to maximize her payoff. There are several payoff functions of the decision maker that are commonly considered in the literature.*

- *matching utilities: in this case, the states space and the action space are finite, and $\Omega = A = \{1, \ldots, n\}$. the agent gains positive utility if the chosen action matches the state, i.e., the utility of the agent is $u(a, \omega; \theta) = \theta \cdot \mathbf{1}\,[a = \omega]$, where $\mathbf{1}\,[\cdot]$ is the indicator function and θ is the private type of the agent.[8] Given belief μ, when the agent chooses the action optimally, the expected utility of the agent is $V(\mu, \theta) = \theta \cdot \max_{\omega \in \Omega} \mu(\omega)$. Thus, by letting $v(\mu) = \max_{\omega \in \Omega} \mu(\omega)$, the valuation of the agent is separable and $V(\mu, \theta) = v(\mu) \cdot \theta$ for any posterior μ and type θ.*

- *error minimization: in the case, $\Omega = A \subseteq \mathbb{R}$, and the agent minimizes the square error between the chosen action and the true state, i.e., the utility of the agent is $u(a, \omega; \theta) = -\theta \cdot (a - \omega)^2$. Given belief μ, the optimal choice of the agent is $\mathbf{E}[\omega]$, and the expected utility of the agent is $V(\mu, \theta) = -\theta \cdot \mathrm{Var}(\mu)$, where $\mathrm{Var}(\mu)$ is the variance of distribution μ. Thus, by letting $v(\mu) = -\mathrm{Var}(\mu)$, the valuation of the agent is separable and $V(\mu, \theta) = v(\mu) \cdot \theta$ for any posterior μ and type θ.*

Example 2. *Consider the model of monopoly auction in Mussa and Rosen (1978). In this example, the agent is a firm selling a product to a consumer with private value for different quality levels of the product. The state space $\Omega = \mathbb{R}_+$ represents the space of valuations of the consumers. The firm has private cost parameter c, and the cost for producing the product with quality q is $c \cdot q^2$.[9] Let F_μ and f_μ be the cumulative function and density function given posterior belief μ. Assuming that the distribution μ is regular, i.e., the virtual value function $\phi_\mu(z) = z - \frac{1 - F_\mu(z)}{f_\mu(z)}$ is non-decreasing in z,[10] the optimal mechanism of the firm with cost c is to provide the product with quality $q(z) = \frac{\max\{0, \phi_\mu(z)\}}{2c}$ to the agent with value z, and the expected profit of the firm is*

$$\int_{\mathbb{R}_+} \frac{\max\{0, \phi_\mu(z)\}^2}{4c} \, \mathrm{d}\mu(z).$$

Let $\theta = \frac{1}{c}$ be the private type of the firm, and let $v(\mu) = \frac{1}{4} \int_{\mathbb{R}_+} \max\{0, \phi_\mu(z)\}^2 \, \mathrm{d}\mu(z)$. The valuation function is $V(\mu, \theta) = v(\mu) \cdot \theta$ given any type θ and any belief μ, which satisfies the separable condition.

3 Optimal Mechanism

In this section, we will provide characterizations for the revenue optimal mechanism without any assumption on the valuation function.

Theorem 1. *In the revenue optimal mechanism, the following two properties hold.*

---

[8]This utility function is the matching utility considered in Bergemann, Bonatti, and Smolin (2018). The only difference is that the private type in Bergemann, Bonatti, and Smolin (2018) is the agent's private signal about the state.

[9]Yang (2020) considers a similar setting with linear cost function $c \cdot q$.

[10]Note that the assumption on regularity is not essential for this example. For any distribution μ that is not regular, we can apply the ironing technique in Myerson (1981) to show that the valuation of the agent, i.e., the profit of the firm, is still separable by substituting the virtual value with ironed virtual value.

1. *The agent does not acquire costly information under equilibrium. That is,*

$$\mathbf{E}_{\theta \sim F}\big[\mathbf{E}_{\mu \sim \pi(\theta)|D}\big[C^A(\hat{\sigma}^*_{\mu,\theta}, \mu)\big]\big] = 0$$

   where $\hat{\sigma}^*_{\mu,\theta} \in \arg\max_{\hat{\sigma} \in \hat{\Sigma}} \mathbf{E}_{\hat{\mu} \sim \hat{\sigma}|\mu}[V(\hat{\mu}, \theta)] - C^A(\hat{\sigma}, \mu).$[11]

2. *Revealing full information is in the menu of the optimal mechanism, i.e., there exists a type* $\theta \in \Theta$ *such that* $\pi(\theta) = \sigma^F$.

The proof of Theorem 1 is provided in Appendix A. The second statement of Theorem 1 is the standard no distortion at the top result. As we made no assumption on the type space here, we cannot pin down the type that receives full information. In Section 4, we will provide more structural properties when there is a natural order on the type space, and in that case the highest type will receive full information. What is more interesting is the first statement, where the theorem states that under equilibrium, the agent never (except for a set with measure zero) has incentive to acquire additional costly information after receiving the signal from the data broker. This holds because if the agent with type θ acquires additional information by conducting experiment $\hat{\sigma}_\theta$ with positive cost, the data broker can directly supply this experiment to the agent in the information structure, and increases the payment of type θ by the cost of the experiment $\hat{\sigma}_\theta$. The new mechanism increases the expected revenue of the data broker, and eliminates the incentives for the agent with type θ to further acquire any additional information. In Appendix A, we will formally show that this new mechanism is also incentive compatible and individual rational.[12] Finally, in Appendix C, we show that the observation in Theorem 1 is robust in a common agency model where the agent can repeatedly acquire information from multiple sellers. In that model, the agent will only acquire information from one of the sellers in equilibrium.

Note that although under equilibrium, the agent has no incentive to acquire additional information. The optimal revenue is not equal to the case when the agent cannot acquire additional information. In fact, the ability to potentially acquire additional information distorts the incentives of the agent, and decreases the revenue the seller can extract from the agent. Letting $\text{OPT}(F, \hat{\Sigma})$ be the optimal revenue when the type distribution is F and the set of possible experiments for the agent is $\hat{\Sigma}$, we have the following characterization for the optimal revenue of the data broker, with proof deferred in Appendix A.

Proposition 1. *For any set of experiments* $\hat{\Sigma} \subseteq \hat{\Sigma}' \subseteq \Sigma$, *any type distribution* F, *we have*

$$\text{OPT}(F, \hat{\Sigma}') \leq \text{OPT}(F, \hat{\Sigma}).$$

An immediate implication of Proposition 1 is that the revenue of the data broker is maximized when $\hat{\Sigma} = \{\sigma^N\}$ i.e., the agent cannot acquire additional information.

4 Separable Valuation

In this section, we will obtain more structure results on the optimal mechanism by restricting the type space and the family of valuation functions. We assume that the type space is single dimensional, i.e., $\Theta = [\underline{\theta}, \overline{\theta}] \subseteq \mathbb{R}$. In addition, we assume that the valuation function is separable and satisfies the following convexity assumption.

---

[11]Since $C^A(\hat{\sigma}, \mu) \geq 0$ for any $\hat{\sigma}$ and μ, the agent only acquires costly information for a set with measure zero.

[12]Note that this result relies crucially on Assumption 1. In Appendix D, we will show that if Assumption 1 is violated, it is possible that the agent has strict incentive to acquire additional costly information under equilibrium.

Definition 3. *The separable valuation $V(\mu, \theta)$ is* convex *in μ if there exists a convex function $v(\mu)$ such that $V(\mu, \theta) = v(\mu) \cdot \theta$ for any posterior μ and type θ.*

The definition of separable valuation is introduced in Definition 2, and the convexity in Definition 3 implies that the agent always has higher valuation for Blackwell more informative experiments. The valuation functions illustrated in Example 1 and 2 satisfy both the assumptions of separable valuation and convexity.

Let $\phi(\theta) = \theta - \frac{1-F(\theta)}{f(\theta)}$ be the virtual value function of the agent. Let $\theta^* = \inf_\theta \{\phi(\theta) \geq 0\}$ be the lowest type with virtual value 0. We introduce the following regularity assumption on the type distribution to simplify the exposition in the paper. This assumption is widely adopted in the auction design literature since Myerson (1981).

Assumption 2. *The distribution F is* regular*, i.e., the corresponding virtual value function $\phi(\theta)$ is monotone non-decreasing in θ.*

In this section, we characterize the optimal revenue of the data broker using the envelope theorem (Milgrom and Segal, 2002). For agent with private type θ, the interim utility given revelation mechanism $\mathcal{M} = (\pi, p)$ is

$$U(\theta) = \mathbf{E}_{\mu \sim \pi(\theta)|D} \Big[V(\mu, \hat{\Sigma}, \theta) \Big] - p(\theta) = \mathbf{E}_{\mu \sim \pi(\theta)|D} \Big[v(\mu, \hat{\Sigma}, \theta) \Big] - p(\theta).$$

The derivative of the utility is

$$U'(\theta) = \mathbf{E}_{\mu \sim \pi(\theta)|D} \Big[V_3(\mu, \hat{\Sigma}, \theta) \Big] = \mathbf{E}_{\mu \sim \pi(\theta)|D} \Big[\mathbf{E}_{\hat{\mu} \sim \hat{\sigma}_{\theta,\mu}|\mu} [v(\hat{\mu})] \Big]$$

where $\hat{\sigma}_{\theta,\mu} \in \hat{\Sigma}$ is the optimal experiment for the agent given private type θ and belief μ, and $V_3(\mu, \hat{\Sigma}, \theta)$ is the partial derivative on the third coordinate. Thus, we have

$$U(\theta) = \int_{\underline{\theta}}^{\theta} \mathbf{E}_{\mu \sim \pi(z)|D} \big[\mathbf{E}_{\hat{\mu} \sim \hat{\sigma}_{z,\mu}|\mu} [v(\hat{\mu})] \big] \, \mathrm{d}z + U(\underline{\theta}).$$

Then the revenue of the data broker is

$$\begin{aligned}
\mathrm{Rev}(\mathcal{M}) &= \mathbf{E}_{\theta \sim F} \bigg[\mathbf{E}_{\mu \sim \pi(\theta)|D} \Big[V(\mu, \hat{\Sigma}, \theta) \Big] - \int_{\underline{\theta}}^{\theta} \mathbf{E}_{\mu \sim \pi(z)|D} \big[\mathbf{E}_{\hat{\mu} \sim \hat{\sigma}_{z,\mu}|\mu} [v(\hat{\mu})] \big] \, \mathrm{d}z - U(\underline{\theta}) \bigg] \\
&= \mathbf{E}_{\theta \sim F} \bigg[\mathbf{E}_{\mu \sim \pi(\theta)|D} \Big[V(\mu, \hat{\Sigma}, \theta) - \frac{1 - F(\theta)}{f(\theta)} \cdot \mathbf{E}_{\hat{\mu} \sim \hat{\sigma}_{\theta,\mu}|\mu} [v(\hat{\mu})] \Big] \bigg] - U(\underline{\theta}) \\
&= \mathbf{E}_{\theta \sim F} \bigg[\mathbf{E}_{\mu \sim \pi(\theta)|D} \Big[\phi(\theta) \cdot \mathbf{E}_{\hat{\mu} \sim \hat{\sigma}_{\theta,\mu}|\mu} [v(\hat{\mu})] - C^A(\hat{\sigma}_{\theta,\mu}, \mu) \Big] \bigg] - U(\underline{\theta}),
\end{aligned} \tag{1}$$

where the second equality holds by integration by parts. The next lemma provides sufficient and necessary conditions on the allocations such that the resulting mechanism is incentive compatible and individual rational.

Lemma 1. *An allocation rule π can be implemented by an incentive compatible and individual rational mechanism if and only if for any $\theta, \theta' \in \Theta$,*[13]

$$\int_{\theta'}^{\theta} \mathbf{E}_{\mu \sim \pi(z)|D} \big[\mathbf{E}_{\hat{\mu} \sim \hat{\sigma}_{z,\mu}|\mu} [v(\hat{\mu})] \big] - \mathbf{E}_{\mu \sim \pi(\theta')|D} \big[\mathbf{E}_{\hat{\mu} \sim \hat{\sigma}_{z,\mu}|\mu} [v(\hat{\mu})] \big] \, \mathrm{d}z \geq 0, \tag{IC}$$

$$\int_{\underline{\theta}}^{\theta} \mathbf{E}_{\mu \sim \pi(z)|D} \big[\mathbf{E}_{\hat{\mu} \sim \hat{\sigma}_{z,\mu}|\mu} [v(\hat{\mu})] \big] \, \mathrm{d}z + U(\underline{\theta}) \geq V(D, \hat{\Sigma}, \theta). \tag{IR}$$

---

[13]If $\theta < \theta'$, we use $\int_{\theta'}^{\theta}$ to represent $- \int_{\theta}^{\theta'}$.

The incentive constraint on allocation is similar to the integral monotonicity provided in Yang (2020), where the author considers selling data to an agent without any ability to further acquire information. Note that it is not sufficiently to consider experiments that are Blackwell monotone for designing incentive compatible and individual rational mechanisms.[14] The proof of Lemma 1 is standard, which is deferred in Appendix B.

4.1 Without Endogenous Information

In this section, we characterize the optimal mechanism in the simple case where the agent cannot acquire any additional information, i.e., $\hat{\Sigma} = \{\sigma^N\}$. In this case, we have $C^A(\hat{\sigma}, D) = 0$ for any $\hat{\sigma} \in \hat{\Sigma}$ and any prior D, and $\mathbf{E}_{\hat{\mu} \sim \hat{\sigma}_{\theta}, \mu | \mu}[v(\hat{\mu})] = v(\mu)$. Thus the revenue of any incentive compatible mechanism $\mathcal{M}$ in Equation (1) simplifies to

$$\text{Rev}(\mathcal{M}) = \mathbf{E}_{\theta \sim F}\big[\phi(\theta) \cdot \mathbf{E}_{\mu \sim \pi(\theta)|D}[v(\mu)]\big] - U(\underline{\theta}).$$

Intuitively, in this setting, by viewing $\mathbf{E}_{\mu \sim \pi(\theta)|D}[v(\mu)]$ as a single-dimensional allocation, the incentive constraint simplifies to the monotonicity constraint on quantity $\mathbf{E}_{\mu \sim \pi(\theta)|D}[v(\mu)]$, and hence the allocation in the optimal mechanism is a step function (Myerson, 1981), which in our setting corresponds to revealing full information if $\phi(\theta) \geq 0$ and reveals no information if $\phi(\theta) < 0$. Note that this is equivalent to posting a deterministic price for revealing full information. This intuition is formalized in Proposition 2, with proof deferred to Appendix B.

Proposition 2. *If the set of possible experiments is a singleton, i.e., the agent cannot acquire any additional information, for convex and separable valuations, the optimal mechanism is to post a deterministic price for revealing full information.*

Proposition 2 shows that the optimal mechanism has a simple form when the agent cannot acquire endogenous information. This implies that, for example, when the data broker sells information to a firm that supplies products to consumers with quadratic cost for qualities (c.f., Example 2), the optimal mechanism is posted pricing. This is in contrast to Yang (2020), where the firm has a linear cost function. The optimal mechanism there is a φ-quasi-perfect mechanism, which is considerably more complicated than posted pricing.

4.2 Endogenous Information

When agents can acquire endogenous information, the incentive constraints in Lemma 1 cannot be simplified to the monotonicity constraint, and the individual rational constraint may bind for types higher than the lowest type $\underline{\theta}$. Thus the point-wise optimization method for classical auction design cannot be applied when there is endogenous information. In the following theorem, we provide a full characterization of the optimal mechanism under Assumption 1 and 2 by directly tackling the constraints on the integration of allocations. Note that the regularity assumption (Assumption 2) is only made to simplify the exposition. The same characterization holds for irregular distributions by adopting the ironing techniques in Toikka (2011). The detailed proof of Theorem 2 is provided in Appendix B.

---

[14]Sinander (2019) shows that under some regularity conditions, experiments that are monotone in Blackwell order can be implemented by incentive compatible mechanism. However, those conditions are violated in our paper and Blackwell monotone experiments may not be implementable. This issue of implementation with Blackwell monotone experiments for selling information has also been observed in the model of Yang (2020).

Theorem 2. *For convex and separable valuations, under Assumption 1 and 2, there exists an optimal mechanism $\widehat{\mathcal{M}}$ with allocation rule $\hat{\pi}$ such that,*[15]

- *for any type $\theta \geq \theta^*$, the data broker reveals full information, i.e., $\hat{\pi}(\theta) = \sigma^F$;*

- *for any type $\theta < \theta^*$, the data broker commits to information structure*

$$\hat{\pi}(\theta) = \arg\max_{\hat{\sigma} \in \hat{\Sigma}} \mathbf{E}_{\mu \sim \hat{\sigma}|D}[V(\mu, \theta)] - C^A(\hat{\sigma}, D)$$

  where ties are broken by maximizing the cost $C^A(\hat{\sigma}, D)$;

- $U(\underline{\theta}) = V(D, \hat{\Sigma}, \underline{\theta})$.

Theorem 2 implies that there is no distortion at the top in the optimal mechanism. Intuitively, when the agent has sufficiently high type, i.e., $\phi(\theta) > 0$, by fully revealing the information to the agent, the expected virtual value is maximized since $\mathbf{E}_{\mu \sim \sigma|D}\left[\mathbf{E}_{\hat{\mu} \sim \hat{\sigma}_{\theta,\mu}|\mu}[v(\hat{\mu})]\right]$ is maximized when the signal σ fully reveals the state. Moreover, with fully revealed state, the posterior belief of the agent is a singleton, and hence the cost of the endogenous information is zero since there is no additional information available. By Equation (1), this allocation maximizes the virtual surplus, and hence the expected revenue of the data broker.

According to the characterization in Theorem 2, for any type $\theta < \theta^*$, the utility of the agent in the optimal mechanism $\widehat{\mathcal{M}}$ is

$$U(\theta) = \int_{\underline{\theta}}^{\theta} \mathbf{E}_{\mu \sim \hat{\pi}(z)|D}\left[\mathbf{E}_{\hat{\mu} \sim \hat{\sigma}_{z,\mu}|\mu}[v(\hat{\mu})]\right] \, \mathrm{d}z + U(\underline{\theta})$$

$$= \int_{\underline{\theta}}^{\theta} \mathbf{E}_{\mu \sim \hat{\sigma}_{\theta,D}|D}[v(\mu)] \, \mathrm{d}z + V(D, \hat{\Sigma}, \underline{\theta}) = V(D, \hat{\Sigma}, \theta).$$

Thus the individual rational constraint does not only bind for the lowest type, but also for all types below the monopoly type θ^*. Note that this is different from the Myerson's auction design problem or selling information when the agent cannot acquire additional information. In those cases, the utility of the low type agents coincide with the outside option because the seller chooses not to sell to these agents. Here the data broker provide valuable information with positive payment to the agent such that the low type agents are exactly indifferent between participation and choosing the outside option.

Note that although the optimal mechanism may require a complex price discrimination scheme against different types when the agent can acquire additional costly information, in the remaining part of this section, we will provide sufficient conditions on the prior distribution, the cost function or the type distribution such that pricing for revealing full information is optimal or approximately optimal.

Proposition 3. *For any cost C^A and any prior D, if $\sigma^N \in \arg\max_{\hat{\sigma} \in \hat{\Sigma}} \mathbf{E}_{\hat{\mu} \sim \hat{\sigma}|D}[V(\hat{\mu}, \theta^*)] - C^A(\hat{\sigma}, D)$, the optimal mechanism is to post a price for revealing full information.*

In Appendix B, we will show that when the monopoly type θ^* find it optimal to not acquire costly information, all types below θ^* will have strictly incentives to not acquire costly information. Combining this observation with the characterization in Theorem 2, we directly obtain the result that the optimal mechanism is pricing for revealing full information.

---

[15]The characterization on allocation actually holds in any optimal mechanism except for a set of types with measure zero.

Note that the condition in Proposition 3 is also necessary for pricing for revealing full information to be revenue optimal when $\theta^* > \underline{\theta}$ and the type distribution is continuous with positive density everywhere. This is because if the monopoly type θ^* has strict incentive to acquire costly information given the prior, there exists a positive measure of types below θ^* that also have strict incentive to acquire costly information given the prior. Hence the data broker can have strictly revenue increase by price discriminating those low types.

There are two interpretations for Proposition 3. Fixing the cost function C^A, we say the prior D is *sufficiently informative* if $\sigma^N \in \arg\max_{\hat{\sigma} \in \hat{\Sigma}} \mathbf{E}_{\hat{\mu} \sim \hat{\sigma}|D}[V(\hat{\mu}, \theta^*)] - C^A(\hat{\sigma}, D)$. This is intuitive since when the prior is sufficiently close to the degenerate pointmass distribution, the marginal cost for additional information is sufficiently high while the marginal benefit of additional information is bounded. Thus it is not beneficial for the agent to not acquiring any additional information. We formalize the intuition in Appendix B.

An alternative interpretation for Proposition 3 is that by holding the prior D as fixed, when the cost C^A of acquiring additional information is sufficiently high, for agent with type θ^*, she has no incentive to acquire additional information given the prior. Note that this can be achieved by scaling any cost function up by a sufficiently large constant. Thus posting a deterministic price for revealing full information is also optimal when the information acquisition is sufficiently costly. This is a generalization for Proposition 2 where the cost of information is infinite for any experiment except σ^N.

So far we have shown the optimality of revealing full information with conditions on the prior or the cost of acquiring additional information. Without any such assumptions, the optimal mechanism may contain a continuum of menus, which discriminate different types of the agent by offering experiments with increasing level of informativeness. However, we show that the additional benefit of price discrimination is limited, as posting a deterministic price for revealing full information is approximately optimal for revenue maximization given the same set of assumptions as in Theorem 2. The proof of Theorem 3 is provided in Appendix B.

Theorem 3. *For convex and separable valuations, under Assumption 1 and 2, for any prior D and any cost function C^A, posting a deterministic price for revealing full information achieves at least half of the optimal revenue.*

Theorem 3 is shown by identifying the worst case type distribution and cost function that maximize the multiplicative gap between the optimal revenue and the revenue from posted pricing, and then directly proving that the gap in the worst case is 2. Note that Theorem 3 is a worst case analysis. The revenue from posted pricing can be substantially higher than half of the optimal revenue for broad classes of type distributions and cost functions, e.g., when the type distribution is uniform and the cost is the reduction in entropy.

In Proposition 1 we have shown that the ability of acquiring additional information distorts the incentives of the agent, and reduces the optimal revenue of the data broker. In the following proposition, we discuss the implication of endogenous information acquisition on the social welfare and the expected utility of the agent. Recall that $|\hat{\Sigma}| = 1$ is equivalent to the setting where the agent cannot acquire additional information.

Proposition 4. *For any valuation function V, any prior D, and any type distribution F, in the revenue optimal mechanism,*

- *the social welfare is minimized when $|\hat{\Sigma}| = 1$;[16]*

---

[16]Our result actually implies that the expected value for each type of the agent is minimized when $|\hat{\Sigma}| = 1$.

- *the utility of the agent is minimized when $|\hat{\Sigma}| = 1$.*

The proof of statement 1 in Proposition 4 is provided in Appendix B, and the second statement is implied by the first statement and Proposition 1. Intuitively, when the agent cannot acquire additional information, in the optimal mechanism, the data broker will not provide information to lower types to reduce the information rent from higher types, which minimizes the social welfare.

4.3 Applications

In this section, we apply the characterizations of the optimal mechanisms to several leading examples of selling information.

Error Minimization. Here we consider the model where the agent is a decision maker trying to minimize the square error of the chosen action. That is, let the state space and action space be $\Omega = A \subseteq \mathbb{R}$, and the agent minimizes the square error between the chosen action and the true state, i.e., the utility of the agent is $u(a, \omega; \theta) = -\theta \cdot (a - \omega)^2$. This is one of the models illustrated in Example 1.

Recall that in Example 1, we show that the valuation function of the agent is separable with the form $V(\mu, \theta) = \theta \cdot v(\mu)$, where $v(\mu) = -\text{Var}(\mu)$ is the variance of distribution μ. Let F be the distribution over the types and let θ^* be the monopoly type in distribution F. We assume that the prior distribution D over states is a Gaussian distribution $\mathcal{N}(0, \eta^2)$ with variance η^2. The agent can repeatedly pay a unit cost c to observe a Gaussian signal $s = \omega + \epsilon$ where $\epsilon \sim \mathcal{N}(0, 1)$.

Next we illustrate the optimal mechanism in this setting by applying Theorem 2.

- For any $\theta \geq \theta^*$, the data broker reveals the states to the firm with price $p = \theta^* \cdot \text{Var}(D)$.

- For any $\theta < \theta^*$, the optimal allocation solves the following Bayesian persuasion problem

$$\hat{\pi}(\theta) = \arg\max_{\hat{\sigma} \in \hat{\Sigma}} \mathbf{E}_{\mu \sim \hat{\sigma}|D}[V(\mu, \theta)] - C^A(\hat{\sigma}, D).$$

  Note that in this example, the agent can only decide the number of Gaussian signals to observe, and with k signals, the cost is k, and the variance of the posterior is $\frac{\eta^2}{1+k\eta^2}$ regardless of the realized sequence of the observed signals. Thus, letting

$$k_\theta = \text{argmax}_k -\theta \cdot \frac{\eta^2}{1 + k\eta^2} - kc,$$

  in the optimal mechanism, the data broker commits to a signal structure that is Blackwell equivalent to revealing k_θ Gaussian signals with unit variance, and charges the agent with price $k_\theta \cdot c$.

Note that for any $\theta < \theta^*$, the optimal number of signals revealed to the agent k_θ is weakly increasing in θ and η, and is weakly decreasing in c. Moreover, fixing distribution F and correspondingly the monopoly type θ^*, when η is sufficiently small or when c is sufficiently large, $k_\theta = 0$ for any $\theta < \theta^*$, and the optimal mechanism reduces to posted pricing mechanism.

Monopoly auction. Here we consider the monopoly auction model introduced in Mussa and Rosen (1978). This is introduced in Example 2. We consider a simple case that the state space
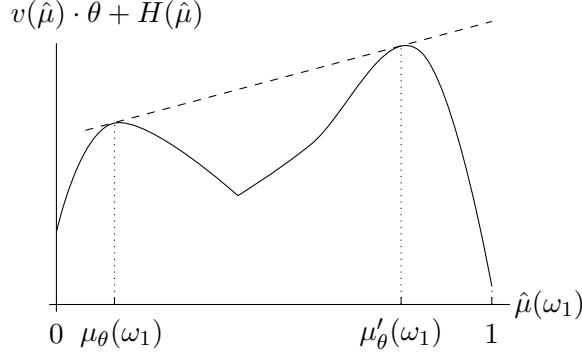
Figure 1: The figure is the value of $v(\hat{\mu}) \cdot \theta + H(\hat{\mu})$ as a function of $\hat{\mu}(\omega_1)$.

$\Omega = \{\omega_1, \omega_2\}$ is binary, where $\omega_i \in \mathbb{R}$ represents the value of the consumer and $0 < \omega_1 < \omega_2$. In this case, given posterior belief μ of the firm, the virtual value of the consumer simplifies to

$$\phi_\mu(\omega_1) = \omega_1 - \frac{\mu(\omega_2)(\omega_2 - \omega_1)}{\mu(\omega_1)};$$

$$\phi_\mu(\omega_2) = \omega_2.$$

According to Mussa and Rosen (1978), the optimal mechanism of the firm with cost c is to provide the product with quality $q(\omega_i) = \frac{\max\{0, \phi_\mu(\omega_i)\}}{2c}$ to the agent with value ω_i, and the expected profit of the firm is $\frac{1}{c} \cdot v(\mu)$ where

$$v(\mu) \triangleq \mu(\omega_1) \cdot \frac{\max\{0, \phi_\mu(\omega_1)\}^2}{4} + \mu(\omega_2) \cdot \frac{\omega_2^2}{4}.$$

Suppose the cost c is the private information of the firm, and let $\theta = 1/c$. Recall that F is the distribution over the types and D is the prior over states. Let θ^* be the monopoly type in distribution F. We assume that the firm can flexibly design any experiment, i.e., $\hat{\Sigma}$ contains all possible experiments. In addition, for any $\hat{\sigma} \in \hat{\Sigma}$, the cost is the reduction in entropy, i.e., $C^A(\hat{\sigma}, \mu) = H(\mu) - \mathbf{E}_{\hat{\mu} \sim \hat{\sigma}|\mu}[H(\hat{\mu})]$ for any posterior μ where $H(\mu) = -\sum_i \mu(\omega_i) \log \mu(\omega_i)$ is the entropy function.

Since the valuation function of the firm satisfies the separable condition and convexity, next we illustrate the optimal mechanism in this setting by applying Theorem 2.

- For any $\theta \geq \theta^*$, or equivalently for any $c \leq 1/\theta^*$, the data broker reveals full information to the firm with price

$$p = \theta^* \cdot (\mathbf{E}_{\omega \sim D}[v(\mu_\omega)] - v(D)),$$

where μ_ω is the pointmass distribution on state ω.

- For any $\theta < \theta^*$, or equivalently for any $c > 1/\theta^*$, the optimal allocation solves the following Bayesian persuasion problem

$$\hat{\pi}(\theta) = \arg\max_{\hat{\sigma} \in \hat{\Sigma}} \mathbf{E}_{\mu \sim \hat{\sigma}|D}[V(\mu, \theta)] - C^A(\hat{\sigma}, D)$$

$$= \arg\max_{\hat{\sigma} \in \hat{\Sigma}} \mathbf{E}_{\mu \sim \hat{\sigma}|D}[\theta \cdot v(\mu) + H(\mu)] - H(D).$$

14

By the concavification approach in Kamenica and Gentzkow (2011), the optimal signal structure has signal space of size 2. As illustrated in Fig. 1, if the prior satisfies $\mu_\theta(\omega_1) < D(\omega_1) < \mu'_\theta(\omega_1)$, the data broker induces posterior either μ_θ or μ'_θ for type θ. Otherwise, the data broker reveals no information to the firm.

5 Conclusions and Extensions

In this paper, we study the model of selling information to an agent with the ability to acquire costly information upon receiving the signals from the data broker. We show that in the optimal mechanism, there is no distortion at the top, and the agent has no incentive to acquire additional information under equilibrium. Our results apply to a broad class of setting including selling information to a firm providing products to consumers with different level of qualities, and selling information to a decision maker taking an action to maximize her payoff based on her posterior belief on the states. In addition, we show that if the agent cannot acquire additional information, or if the prior distribution is sufficiently informative, then posting a deterministic price is optimal among all possible mechanisms. In the remaining part of the section, we briefly discusses the potential extensions of our model.

Limited Experiments of the Data Broker. In this paper, the data broker can sell any experiment to the agent. In reality the data broker may only collect additional information about the states with certain formats. Specifically, let $\Sigma^S \subseteq \Sigma$ be the set of possible experiments of the data broker. In the case that $\hat{\Sigma} \subseteq \Sigma^S$, i.e., the ability of the data broker is stronger than the agent in learning the states, the optimal mechanism still guarantees that the agent does not have incentive to acquire any additional information under equilibrium. In addition, if there exists an experiment $\bar{\sigma} \in \Sigma^S$ that is Blackwell more informative than any other experiments in Σ^S, under the separable valuation assumption, there is no distortion at the top and the optimal mechanism sells experiment $\bar{\sigma}$ if the private type of the agent is sufficiently high. For example, if Σ^S only contains experiments that partition the state space, experiment $\bar{\sigma}$ is the one that provides the finest partition.

Cost of Information. Throughout the paper, we assumed that the data broker has free access to the information. The cost of the data broker can be easily incorporated into our model. If the cost of the data broker is always higher than the cost of the agent for acquiring any information, one can verify that in this case, the data broker cannot make any profit by selling information to the agent. If the cost of the data broker is always smaller, the data broker can make positive profit from selling information, and in the optimal mechanism, the agent still does not have incentive to acquire any additional information under equilibrium.

Dependence on the Action of the Agent. In many application, the agent purchases information to make better predictions, and takes an action based on the posterior belief after seeing the signals from the data broker. It is possible that the action chosen by the agent will also affect the utility of the data broker. For example, the data brokers in our model are users selling their personal characteristic information to online platforms in exchange of services, and online platforms may use these additional information to price discriminate the users in the future. This additional factor that affects the utility of the data broker can also be included in our model as the cost of information. Note that we can assume that the action of the agent only depends on the posterior belief of the agent,[17] and hence the utility of the data broker for the action of the agent is simply a

---

[17]This is satisfied in many applications (e.g., Example 1).

function of the posterior. This is mathematically equivalent to a cost function, and all techniques in our paper can be applied here equivalently.

Endogenous Timing of Additional Experiments. We have focused on the timing where the agent can only acquire information after receiving the signal from the seller. Note that this is without loss if the seller can commit to the mechanism before the agent acquires additional information.[18] It is an interesting open question for characterizing the optimal mechanism when the agent optimally acquires information before the seller announces the mechanism.

Non-separable Valuation. In Section 4, we have focused on the case when the agent has separable valuation. This assumption can be relaxed by considering the model when there is a commonly known monotone supermodular function g and a convex valuation function $v(\mu)$ such that for any type θ and any posterior μ, the value of the agent is $V(\mu, \theta) = g(v(\mu), \theta)$. Essentially, the crucial assumption here is that θ measures how much the agent values the information, and it should not affect how the agent processes the signals from the data broker or how the agent updates her posteriors. Under the relaxed assumption, the qualitative results in our paper extend naturally.

When different types of the agent have heterogeneous information about the states in ex ante (e.g., Bergemann, Bonatti, and Smolin, 2018), the agent with different prior information may have heterogeneous preference order over the information structures. This heterogeneity in preference creates room for increasing the revenue through price discrimination. Thus results such as posted pricing is optimal may fail even if the agent cannot acquire endogenous information or the prior is sufficiently informative. However, we can still show that in the optimal mechanism, the agent does not have incentive to acquire additional information under equilibrium.

References

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

Anat R Admati and Paul Pfleiderer. Direct and indirect sale of information. *Econometrica: Journal of the Econometric Society*, pages 901–928, 1990.

Moshe Babaioff, Nicole Immorlica, Brendan Lucier, and S Matthew Weinberg. A simple and approximately optimal mechanism for an additive buyer. *Journal of the ACM (JACM)*, 67(4): 1–40, 2020.

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

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

Dirk Bergemann, Benjamin Brooks, and Stephen Morris. The limits of price discrimination. *American Economic Review*, 105(3):921–57, 2015.

---

[18]For any mechanism such that the agent has incentives to acquire additional information before interacting with the seller, there exists an equivalent mechanism where the seller elicits the private type of the agent, and simulates the information acquisition process for the agent. Moreover, the agent has no incentives to acquire additional information in the new mechanism.

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

Dirk Bergemann, Yang Cai, Grigoris Velegkas, and Mingfei Zhao. Is selling complete information (approximately) optimal? *working paper*, 2021.

Alexander W Bloedel and Weijie Zhong. The cost of optimally-acquired information. Technical report, Technical report, Working paper, Stanford University, 2020.

Andrew Caplin, Mark Dean, and John Leahy. Rational inattention, optimal consideration sets, and stochastic choice. *The Review of Economic Studies*, 86(3):1061–1094, 2019.

Shuchi Chawla, Yifeng Teng, and Christos Tzamos. Menu-size complexity and revenue continuity of buy-many mechanisms. In *Proceedings of the 21st ACM Conference on Economics and Computation*, pages 475–476, 2020.

Jacques Crémer and Fahad Khalil. Gathering information before signing a contract. *The American Economic Review*, pages 566–578, 1992.

Nikhil R Devanur and S Matthew Weinberg. The optimal mechanism for selling to a budget constrained buyer: The general case. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, pages 39–40, 2017.

Péter Eső and Balazs Szentes. Optimal information disclosure in auctions and the handicap auction. *The Review of Economic Studies*, 74(3):705–731, 2007.

George Georgiadis and Balazs Szentes. Optimal monitoring design. *Econometrica*, 88(5):2075–2107, 2020.

Jason D Hartline. Approximation in mechanism design. *American Economic Review*, 102(3): 330–36, 2012.

Jason D Hartline and Tim Roughgarden. Simple versus optimal mechanisms. In *Proceedings of the 10th ACM conference on Electronic commerce*, pages 225–234, 2009.

Emir Kamenica and Matthew Gentzkow. Bayesian persuasion. *American Economic Review*, 101 (6):2590–2615, 2011.

Hao Li and Xianwen Shi. Discriminatory information disclosure. *American Economic Review*, 107 (11):3363–85, 2017.

Shuze Liu, Weiran Shen, and Haifeng Xu. Optimal pricing of information. *Proceedings of the 22nd ACM Conference on Economics and Computation*, 2021.

David Martimort and Lars Stole. The revelation and delegation principles in common agency games. *Econometrica*, 70(4):1659–1673, 2002.

Flavio M Menezes and Paulo K Monteiro. Auctions with endogenous participation. *Review of Economic Design*, 5(1):71–89, 2000.

Jeffrey Mensch. Screening inattentive agents. *Available at SSRN 3405884*, 2019.

Paul Milgrom and Ilya Segal. Envelope theorems for arbitrary choice sets. *Econometrica*, 70(2): 583–601, 2002.

Michael Mussa and Sherwin Rosen. Monopoly and product quality. *Journal of Economic theory*, 18(2):301–317, 1978.

Roger B Myerson. Optimal auction design. *Mathematics of operations research*, 6(1):58–73, 1981.

Luis Rayo and Ilya Segal. Optimal information disclosure. *Journal of political Economy*, 118(5): 949–987, 2010.

Xianwen Shi. Optimal auctions with information acquisition. *Games and Economic Behavior*, 74 (2):666–686, 2012.

Christopher A Sims. Implications of rational inattention. *Journal of Monetary Economics*, 50(3): 665–690, 2003.

Ludvig Sinander. The converse envelope theorem. *arXiv preprint arXiv:1909.11219*, 2019.

Alex Smolin. Disclosure and pricing of attributes. *Available at SSRN 3318957*, 2020.

Juuso Toikka. Ironing without control. *Journal of Economic Theory*, 146(6):2510–2526, 2011.

Dong Wei and Brett Green. (Reverse) price discrimination with information design. *Available at SSRN 3263898*, 2020.

Kai Hao Yang. Selling consumer data for profit: Optimal market-segmentation design and its consequences. *working paper*, 2020.

Weijie Zhong. Optimal dynamic information acquisition. *Available at SSRN 3024275*, 2017.

A Optimal Mechanism

Theorem 1. *In the revenue optimal mechanism, the following two properties hold.*

1. *The agent does not acquire costly information under equilibrium. That is,*

$$\mathbf{E}_{\theta \sim F}\big[\mathbf{E}_{\mu \sim \pi(\theta)|D}\big[C^A(\hat{\sigma}^*_{\mu,\theta}, \mu)\big]\big] = 0$$

   where $\hat{\sigma}^*_{\mu,\theta} \in \arg\max_{\hat{\sigma} \in \hat{\Sigma}} \mathbf{E}_{\hat{\mu} \sim \hat{\sigma}|\mu}[V(\hat{\mu}, \theta)] - C^A(\hat{\sigma}, \mu)$.[19]

2. *Revealing full information is in the menu of the optimal mechanism, i.e., there exists a type $\theta \in \Theta$ such that $\pi(\theta) = \sigma^F$.*

Proof of Statement 1 of Theorem 1. Let $\mathcal{M} = (\pi, p)$ be the optimal mechanism. Let $\kappa_{\theta,\sigma}$ be the optimal choice of experiments for agent with type θ when she receives the realized experiment σ from mechanism $\mathcal{M}$. By contradiction, let $\hat{\Theta}$ be the set of types with positive measure such that for any $\hat{\theta} \in \hat{\Theta}$, the cost for additional experiments given optimal best response strategy $\kappa_{\hat{\theta},\sigma}$ for agent with type $\hat{\theta}$ is positive, i.e.,

$$\int_{\Sigma} \int_{\Omega} \int_{S} C^A(\kappa_{\hat{\theta},\sigma}(s), \hat{\mu}_{\sigma,s,D}) \, \mathrm{d}\sigma(s|\omega) \, \mathrm{d}D(\omega) \, \mathrm{d}\pi(\sigma|\hat{\theta}) > 0,$$

where $\hat{\mu}_{\sigma,s,D}$ is the posterior given experiment σ and signal s, assuming the prior is D. Let $\hat{\pi}$ and $\hat{p}$ be the allocation and payment rule such that

- for any $\theta \notin \hat{\Theta}$, $\hat{\pi}(\theta) = \pi(\theta)$ and $\hat{p}(\theta) = p(\theta)$;

- for any $\hat{\theta} \in \hat{\Theta}$, for any $\sigma \in \Sigma$, $\hat{\pi}(\sigma|\hat{\theta}) = \pi(\sigma \circ \kappa_{\hat{\theta},\sigma}|\hat{\theta})$,[20] and

$$\hat{p}(\hat{\theta}) = p(\hat{\theta}) + \int_{\Sigma} \int_{\Omega} \int_{S} C^A(\kappa_{\hat{\theta},\sigma}(s), \hat{\mu}_{\sigma,s,D}) \, \mathrm{d}\sigma(s|\omega) \, \mathrm{d}D(\omega) \, \mathrm{d}\pi(\sigma|\hat{\theta}).$$

Let $\hat{\mathcal{M}} = (\hat{\pi}, \hat{p})$. It is easy to verify that

$$\text{Rev}(\hat{\mathcal{M}}) = \int_{\Theta} \hat{p}(\theta) \, \mathrm{d}F(\theta) = \int_{\Theta \setminus \hat{\Theta}} \hat{p}(\theta) \, \mathrm{d}F(\theta) + \int_{\hat{\Theta}} \hat{p}(\theta) \, \mathrm{d}F(\theta)$$

$$= \int_{\Theta \setminus \hat{\Theta}} p(\theta) \, \mathrm{d}F(\theta) + \int_{\hat{\Theta}} \left(p(\theta) + \int_{\Sigma} \int_{\Omega} \int_{S} C^A(\kappa_{\hat{\theta},\sigma}(s), \hat{\mu}_{\sigma,s,D}) \, \mathrm{d}\sigma(s|\omega) \, \mathrm{d}D(\omega) \, \mathrm{d}\pi(\sigma|\hat{\theta}) \right) \mathrm{d}F(\theta)$$

$$< \int_{\Theta \setminus \hat{\Theta}} p(\theta) \, \mathrm{d}F(\theta) + \int_{\hat{\Theta}} p(\theta) \, \mathrm{d}F(\theta) = \text{Rev}(\mathcal{M}).$$

The inequality holds because the types in set $\hat{\Theta}$ occur with positive measure. Thus the revenue of mechanism $\hat{\mathcal{M}}$ is strictly higher. Moreover, in mechanism $\hat{\mathcal{M}}$, the utility of the agent has at least the same expected utility compared to mechanism $\mathcal{M}$ by not acquiring any additional information upon receiving the signal. Therefore, mechanism $\hat{\mathcal{M}}$ is individual rational. Next it is sufficient to show that mechanism $\hat{\mathcal{M}}$ is incentive compatible. It is easy to verify that the agent with any type θ has no incentive to deviate to type $\hat{\theta} \notin \hat{\Theta}$ since her utility for reporting truthfully weakly increases,

---

[19]Since $C^A(\hat{\sigma}, \mu) \geq 0$ for any $\hat{\sigma}$ and μ, the agent only acquires costly information for a set with measure zero.

[20]Note that under this new sequential experiment $\hat{\pi}(\hat{\theta})$, the signals generated by experiments in all stages will be revealed to the agent.

while her utility for misreporting $\hat{\theta}$ remains the same. Finally, for any type θ, under mechanism $\hat{\mathcal{M}}$, the utility for deviating the report from type θ to type $\hat{\theta} \in \hat{\Theta}$ is

$$
\begin{aligned}
U(\theta; \hat{\theta}, \hat{\mathcal{M}}) &= \mathbf{E}_{\mu \sim \hat{\pi}(\hat{\theta})|D}\left[V(\mu, \hat{\Sigma}, \theta)\right] - \hat{p}(\hat{\theta}) \\
&= \mathbf{E}_{\mu \sim \hat{\pi}(\hat{\theta})|D}\left[V(\mu, \hat{\Sigma}, \theta)\right] - p(\hat{\theta}) - \int_{\Sigma}\int_{\Omega}\int_{S} C^{A}(\kappa_{\hat{\theta},\sigma}(s), \hat{\mu}_{\sigma,s,D})\, \mathrm{d}\sigma(s|\omega)\, \mathrm{d}D(\omega)\, \mathrm{d}\pi(\sigma|\hat{\theta}) \\
&= \int_{\Sigma}\int_{\Omega}\int_{S}\left(\mathbf{E}_{\hat{\mu} \sim \kappa_{\hat{\theta},\sigma}(s)|\hat{\mu}_{\sigma,s,D}}\left[V(\hat{\mu}, \hat{\Sigma}, \theta)\right] - C^{A}(\kappa_{\hat{\theta},\sigma}(s), \hat{\mu}_{\sigma,s,D})\right)\, \mathrm{d}\sigma(s|\omega)\, \mathrm{d}D(\omega)\, \mathrm{d}\pi(\sigma|\hat{\theta}) - p(\hat{\theta}) \\
&\leq \int_{\Sigma}\int_{\Omega}\int_{S} V(\hat{\mu}_{\sigma,s,D}, \hat{\Sigma}, \theta)\, \mathrm{d}\sigma(s|\omega)\, \mathrm{d}D(\omega)\, \mathrm{d}\pi(\sigma|\hat{\theta}) - \hat{p}(\hat{\theta}) \\
&= U(\theta; \hat{\theta}, \mathcal{M}).
\end{aligned}
$$

The inequality holds because by Assumption 1, given any posterior $\hat{\mu}_{\sigma,s,D}$, a feasible choice for the agent is to choose $\kappa_{\hat{\theta},\sigma}(s) \in \hat{\Sigma}$, pay cost $C^{A}(\kappa_{\hat{\theta},\sigma}(s), \hat{\mu}_{\sigma,s,D})$, and then choose additional experiments optimally given the realized signal. The utility of this choice is upper bounded by directly choosing the optimal experiment from $\hat{\Sigma}$, which induces value $V(\hat{\mu}_{\sigma,s,D}, \hat{\Sigma}, \theta)$ for the agent. Thus, we have

$$
U(\theta; \hat{\mathcal{M}}) - U(\theta; \hat{\theta}, \hat{\mathcal{M}}) \geq U(\theta; \mathcal{M}) - U(\theta; \hat{\theta}, \mathcal{M}) \geq 0,
$$

and mechanism $\hat{\mathcal{M}}$ is incentive compatible. $\qquad\square$

Proof of Statement 2 of Theorem 1. For any mechanism $\mathcal{M} = (\pi, p)$, let $\bar{p} = \sup_{\theta} p(\theta)$. By adding the choice $(\sigma^{F}, \bar{p})$ into the menu of mechanism $\mathcal{M}$, the revenue of the data broker only increases. $\quad\square$

Proposition 1. *For any set of experiments $\hat{\Sigma} \subseteq \hat{\Sigma}' \subseteq \Sigma$, any type distribution F, we have*

$$
\mathrm{OPT}(F, \hat{\Sigma}') \leq \mathrm{OPT}(F, \hat{\Sigma}).
$$

Proof. For any set of experiments $\hat{\Sigma} \subseteq \hat{\Sigma}' \subseteq \Sigma$, for any type distribution F, let $\mathcal{M}$ be the optimal mechanism when the type distribution is F, and the agent can conduct additional experiments in $\hat{\Sigma}'$. Next we show that mechanism $\mathcal{M}$ is incentive compatible and individual rational when the set of additional experiments for the agent is $\hat{\Sigma}$. By Theorem 1, for any type θ, when the set of experiments is $\hat{\Sigma}'$, the agent has no incentive to acquire additional information on equilibrium path. Thus when the set of additional experiments is $\hat{\Sigma}$, by reporting the type truthfully, the utility of the agent remains the same. Therefore, mechanism $\mathcal{M}$ is individual rational. In addition, since $\hat{\Sigma} \subseteq \hat{\Sigma}'$, in mechanism $\mathcal{M}$, the utility of deviating to any other type is weakly smaller when the set of additional experiments is $\hat{\Sigma}$. Thus mechanism $\mathcal{M}$ is incentive compatible as well given the set $\hat{\Sigma}$. Therefore, we have

$$
\mathrm{OPT}(F, \hat{\Sigma}') = \mathrm{Rev}(\mathcal{M}, \hat{\Sigma}') = \mathrm{Rev}(\mathcal{M}, \hat{\Sigma}) \leq \mathrm{OPT}(F, \hat{\Sigma}). \qquad\square
$$

B Separable Valuation

Before the proof of the theorems in Section 4, we first present the following lemma showing that experiment σ^{F} that reveals full information is the most valuable for the agent. Recall that $\hat{\sigma}_{\theta,\mu} \in \hat{\Sigma}$ is the optimal experiment the agent chooses when her type is θ and her posterior belief after receiving the signal from the data broker is μ.

Lemma 2. *Let σ^F be the experiment that reveals full information. For any experiment $\sigma \in \Sigma$, any prior D, and any type θ, we have*

$$\mathbf{E}_{\mu \sim \sigma^F | D}\Big[\mathbf{E}_{\hat{\mu} \sim \hat{\sigma}_{\theta,\mu}|\mu}[v(\hat{\mu})]\Big] \geq \mathbf{E}_{\mu \sim \sigma | D}\Big[\mathbf{E}_{\hat{\mu} \sim \hat{\sigma}_{\theta,\mu}|\mu}[v(\hat{\mu})]\Big].$$

Proof. For the fully informative experiment σ^F, for any experiment σ, any prior D, and any type θ, we have

$$\mathbf{E}_{\mu \sim \sigma^F | D}\Big[\mathbf{E}_{\hat{\mu} \sim \hat{\sigma}_{\theta,\mu}|\mu}[v(\hat{\mu})]\Big] = \mathbf{E}_{\omega \sim D}[v(\omega)] \geq \mathbf{E}_{\mu \sim \sigma | D}\Big[\mathbf{E}_{\hat{\mu} \sim \hat{\sigma}_{\theta,\mu}|\mu}[v(\hat{\mu})]\Big].$$

The inequality holds since v is convex in μ and σ^F fully reveals the states. $\qquad\square$

Lemma 1. *An allocation rule π can be implemented by an incentive compatible and individual rational mechanism if and only if for any $\theta, \theta' \in \Theta$,*[21]

$$\int_{\theta'}^{\theta} \mathbf{E}_{\mu \sim \pi(z)|D}\big[\mathbf{E}_{\hat{\mu} \sim \hat{\sigma}_{z,\mu}|\mu}[v(\hat{\mu})]\big] - \mathbf{E}_{\mu \sim \pi(\theta')|D}\big[\mathbf{E}_{\hat{\mu} \sim \hat{\sigma}_{z,\mu}|\mu}[v(\hat{\mu})]\big] \ \mathrm{d}z \geq 0, \tag{IC}$$

$$\int_{\underline{\theta}}^{\theta} \mathbf{E}_{\mu \sim \pi(z)|D}\big[\mathbf{E}_{\hat{\mu} \sim \hat{\sigma}_{z,\mu}|\mu}[v(\hat{\mu})]\big] \ \mathrm{d}z + U(\underline{\theta}) \geq V(D, \hat{\Sigma}, \theta). \tag{IR}$$

Proof of Lemma 1. Given allocation rule π, by the envelope theorem, for any incentive compatible mechanism $\mathcal{M}$, the interim utility $U(\theta)$ is convex in θ and

$$U(\theta) = \int_{\underline{\theta}}^{\theta} \mathbf{E}_{\mu \sim \pi(z)|D}\big[\mathbf{E}_{\hat{\mu} \sim \hat{\sigma}_{z,\mu}|\mu}[v(\hat{\mu})]\big] \ \mathrm{d}z + U(\underline{\theta}).$$

Note that the mechanism $\mathcal{M} = (\pi, p)$ is individual rational if and only if $U(\theta) \geq V(D, \hat{\Sigma}, \theta)$ for any type θ, i.e.,

$$\int_{\underline{\theta}}^{\theta} \mathbf{E}_{\mu \sim \pi(z)|D}\big[\mathbf{E}_{\hat{\mu} \sim \hat{\sigma}_{z,\mu}|\mu}[v(\hat{\mu})]\big] \ \mathrm{d}z + U(\underline{\theta}) \geq V(D, \hat{\Sigma}, \theta).$$

Moreover, the corresponding payment rule for mechanism $\mathcal{M}$ is

$$p(\theta) = \mathbf{E}_{\mu \sim \pi(\theta)|D}\Bigg[V(\mu, \hat{\Sigma}, \theta) - \int_{\underline{\theta}}^{\theta} \mathbf{E}_{\mu \sim \pi(z)|D}\big[\mathbf{E}_{\hat{\mu} \sim \hat{\sigma}_{z,\mu}|\mu}[v(\hat{\mu})]\big] \ \mathrm{d}z\Bigg] - U(\underline{\theta}).$$

Next we verify the incentive constraints of the given mechanism. Note that for any $\theta, \theta' \in \Theta$, letting $U(\theta; \theta')$ be the utility of the agent with type θ when she reports θ' in mechanism $\mathcal{M}$, we have

$$\begin{aligned}
&U(\theta) - U(\theta; \theta') \\
&= U(\theta) - U(\theta') - \mathbf{E}_{\mu \sim \pi(\theta')|D}\Big[V(\mu, \hat{\Sigma}, \theta)\Big] + \mathbf{E}_{\mu \sim \pi(\theta')|D}\Big[V(\mu, \hat{\Sigma}, \theta')\Big] \\
&= \int_{\theta'}^{\theta} \mathbf{E}_{\mu \sim \pi(z)|D}\big[\mathbf{E}_{\hat{\mu} \sim \hat{\sigma}_{z,\mu}|\mu}[v(\hat{\mu})]\big] \ \mathrm{d}z - \mathbf{E}_{\mu \sim \pi(\theta')|D}\Bigg[\int_{\theta'}^{\theta} V_3(\mu, \hat{\Sigma}, z) \ \mathrm{d}z\Bigg] \\
&= \int_{\theta'}^{\theta} \Big(\mathbf{E}_{\mu \sim \pi(z)|D}\big[\mathbf{E}_{\hat{\mu} \sim \hat{\sigma}_{z,\mu}|\mu}[v(\hat{\mu})]\big] - \mathbf{E}_{\mu \sim \pi(\theta')|D}\big[\mathbf{E}_{\hat{\mu} \sim \hat{\sigma}_{z,\mu}|\mu}[v(\hat{\mu})]\big]\Big) \ \mathrm{d}z.
\end{aligned}$$

Thus $U(\theta) - U(\theta; \theta') \geq 0$ if and only if the integral constraint in the statement of Lemma 1 is satisfied. $\qquad\square$

---

[21]If $\theta < \theta'$, we use $\int_{\theta'}^{\theta}$ to represent $-\int_{\theta}^{\theta'}$.

Proposition 2. *If the set of possible experiments is a singleton, i.e., the agent cannot acquire any additional information, for convex and separable valuations, the optimal mechanism is to post a deterministic price for revealing full information.*

Proof. Let $\mathcal{M}$ be the mechanism that reveals full information if $\phi(\theta) \geq 0$ and reveals no information if $\phi(\theta) < 0$. When $|\hat{\Sigma}| = 1$, for any type θ and any posterior μ, we have that $\mathbf{E}_{\hat{\mu} \sim \hat{\sigma}_{\theta,\mu}|\mu}[v(\hat{\mu})] = v(\mu)$, and hence the integral constraint simplifies to

$$\int_{\theta'}^{\theta} \mathbf{E}_{\mu \sim \pi(z)|D}\big[\mathbf{E}_{\hat{\mu} \sim \hat{\sigma}_{z,\mu}|\mu}[v(\hat{\mu})]\big] - \mathbf{E}_{\mu \sim \pi(\theta')|D}\big[\mathbf{E}_{\hat{\mu} \sim \hat{\sigma}_{z,\mu}|\mu}[v(\hat{\mu})]\big] \, \mathrm{d}z$$

$$= \int_{\theta'}^{\theta} \mathbf{E}_{\mu \sim \pi(z)|D}[v(\mu)] - \mathbf{E}_{\mu \sim \pi(\theta')|D}[v(\mu)] \, \mathrm{d}z \geq 0.$$

Note that this is equivalent to the condition that $\mathbf{E}_{\mu \sim \pi(\theta)|D}[v(\mu)]$ is non-decreasing in θ. Moreover, similar to Myerson (1981), the individual rational constraint simplifies to the case that the utility of the lowest type $\underline{\theta}$ is at least her outside option. Thus the problem of revenue maximization simplifies to

$$\max \quad \mathbf{E}_{\theta \sim F}\big[\phi(\theta) \cdot \mathbf{E}_{\mu \sim \pi(\theta)|D}[v(\mu)]\big] - U(\underline{\theta})$$

$$\text{s.t.} \quad \mathbf{E}_{\mu \sim \pi(\theta)|D}[v(\mu)] \text{ is non-decreasing in } \theta,$$

$$U(\underline{\theta}) \geq V(D, \hat{\Sigma}, \underline{\theta}).$$

Note that $\mathbf{E}_{\mu \sim \pi(\theta)|D}[v(\mu)]$ is maximized by revealing full information, and minimized by revealing no information. By Myerson (1981), it is easy to verify that the allocation rule of mechanism $\mathcal{M}$ maximizes the expected virtual surplus, and hence mechanism $\mathcal{M}$ is the revenue optimal mechanism. $\qquad \square$

Theorem 2. *For convex and separable valuations, under Assumption 1 and 2, there exists an optimal mechanism $\widehat{\mathcal{M}}$ with allocation rule $\hat{\pi}$ such that,*[22]

- *for any type $\theta \geq \theta^*$, the data broker reveals full information, i.e., $\hat{\pi}(\theta) = \sigma^F$;*

- *for any type $\theta < \theta^*$, the data broker commits to information structure*

$$\hat{\pi}(\theta) = \arg\max_{\hat{\sigma} \in \hat{\Sigma}} \mathbf{E}_{\mu \sim \hat{\sigma}|D}[V(\mu, \theta)] - C^A(\hat{\sigma}, D)$$

  where ties are broken by maximizing the cost $C^A(\hat{\sigma}, D)$;

- $U(\underline{\theta}) = V(D, \hat{\Sigma}, \underline{\theta})$.

Proof of Theorem 2. We first show that allocation rule $\hat{\pi}$ combined with $U(\underline{\theta}) = V(D, \hat{\Sigma}, \underline{\theta})$ can be implemented as an incentive compatible and individual rational mechanism. One way to prove this is to verify the constraints specified in Lemma 1. However, directly verifying the incentive constraints in Lemma 1 for allocation $\hat{\pi}$ might be challenging as we impose little structure on the information costs.[23] Thus we adopt an alternative approach by explicitly constructing an incentive

---

[22]The characterization on allocation actually holds in any optimal mechanism except for a set of types with measure zero.

[23]Without additional structures on the costs, it is hard to characterize the optimal strategy $\hat{\sigma}_{\theta,\mu}$ given any type θ and posterior μ.

compatible and individual rational mechanism $\widehat{\mathcal{M}}$. Then we show that the constructed mechanism has allocation $\hat{\pi}$ and utility for the lowest type $U(\underline{\theta}) = V(D, \hat{\Sigma}, \underline{\theta})$.

First consider a mechanism $\mathcal{M}'$ that post a deterministic price p for revealing full information. The price p is chosen such that the agent purchases information from the seller if and only if $\theta \geq \theta^*$. Note that given mechanism $\mathcal{M}'$, for agent with type $\theta < \theta^*$, she will choose not to participate the auction, and then subsequently conduct experiment

$$\hat{\sigma}_\theta = \arg\max_{\hat{\sigma} \in \hat{\Sigma}} \mathbf{E}_{\mu \sim \hat{\sigma}|D}[V(\mu, \theta)] - C^A(\hat{\sigma}, D).$$

We assume that the agent breaks tie by maximizing the cost $C^A(\hat{\sigma}, D)$. Now let $\widehat{\mathcal{M}}$ be the mechanism that reveals full information for types $\theta \geq \theta^*$ with price p, and commits to information structure $\hat{\sigma}_\theta$ for types $\theta < \theta^*$ with price $C^A(\hat{\sigma}_\theta, D)$. It is easy to verify that $\widehat{\mathcal{M}}$ has allocation rule $\hat{\pi}$ and the utility of the lowest type $\underline{\theta}$ in $\widehat{\mathcal{M}}$ is $V(D, \hat{\Sigma}, \underline{\theta})$. Moreover, by the proof of Theorem 1, mechanism $\widehat{\mathcal{M}}$ is incentive compatible and individual rational.

Note that when the posterior μ is in the support of $\sigma^F|D$, the agent will not acquire additional costly information since σ^F fully reveals the state. Moreover, when the posterior μ is in the support of $\hat{\sigma}_\theta|D$, by Assumption 1, the agent will not acquire additional costly information because otherwise $\hat{\sigma}_\theta$ is not the utility maximization information structure given prior D. Combining the observations, we have that $C^A(\hat{\sigma}_{\theta,\mu}, \mu) = 0$ for μ in the support of $\hat{\pi}(\theta)|D$, and hence by Equation (1), the revenue of mechanism $\widehat{\mathcal{M}}$ is

$$\begin{aligned}
\text{Rev}(\widehat{\mathcal{M}}) &= \mathbf{E}_{\theta \sim F}\left[\mathbf{E}_{\mu \sim \hat{\pi}(\theta)|D}\left[\phi(\theta) \cdot \mathbf{E}_{\hat{\mu} \sim \hat{\sigma}_{\theta,\mu}|\mu}[v(\hat{\mu})] - C^A(\hat{\sigma}_{\theta,\mu}, \mu)\right]\right] - U(\underline{\theta}) \\
&= \mathbf{E}_{\theta \sim F}\left[\phi(\theta) \cdot \mathbf{E}_{\mu \sim \hat{\pi}(\theta)|D}\left[\mathbf{E}_{\hat{\mu} \sim \hat{\sigma}_{\theta,\mu}|\mu}[v(\hat{\mu})]\right]\right] - V(D, \hat{\Sigma}, \underline{\theta}) \\
&= \int_{\theta^*}^{\bar{\theta}} \phi(\theta) \cdot \mathbf{E}_{\mu \sim \sigma^F|D}[v(\mu)] \, \mathrm{d}\theta + \int_{\underline{\theta}}^{\theta^*} \phi(\theta) \cdot \mathbf{E}_{\mu \sim \hat{\sigma}_{\theta,D}|D}[v(\mu)] \, \mathrm{d}\theta - V(D, \hat{\Sigma}, \underline{\theta}). \quad (2)
\end{aligned}$$

Now consider any incentive compatible and individual rational mechanism $\mathcal{M}$ with allocation π, again by Equation (1), the revenue of mechanism $\mathcal{M}$ is

$$\begin{aligned}
\text{Rev}(\mathcal{M}) &= \mathbf{E}_{\theta \sim F}\left[\mathbf{E}_{\mu \sim \pi(\theta)|D}\left[\phi(\theta) \cdot \mathbf{E}_{\hat{\mu} \sim \hat{\sigma}_{\theta,\mu}|\mu}[v(\hat{\mu})] - C^A(\hat{\sigma}_{\theta,\mu}, \mu)\right]\right] - U(\underline{\theta}) \\
&\leq \mathbf{E}_{\theta \sim F}\left[\phi(\theta) \cdot \mathbf{E}_{\mu \sim \pi(\theta)|D}\left[\mathbf{E}_{\hat{\mu} \sim \hat{\sigma}_{\theta,\mu}|\mu}[v(\hat{\mu})]\right]\right] - U(\underline{\theta}),
\end{aligned}$$

where the inequality holds since $C^A(\hat{\sigma}_{\theta,\mu}, \mu) \geq 0$ for any posterior μ. For any type $\theta \geq \theta^*$, i.e., $\phi(\theta) \geq 0$, by applying Lemma 2, the contribution of revenue from type θ is

$$\begin{aligned}
\text{Rev}(\mathcal{M}; \theta) &\triangleq \phi(\theta) \cdot \mathbf{E}_{\mu \sim \pi(\theta)|D}\left[\mathbf{E}_{\hat{\mu} \sim \hat{\sigma}_{\theta,\mu}|\mu}[v(\hat{\mu})]\right] \\
&\leq \phi(\theta) \cdot \mathbf{E}_{\mu \sim \sigma^F|D}\left[\mathbf{E}_{\hat{\mu} \sim \hat{\sigma}_{\theta,\mu}|\mu}[v(\hat{\mu})]\right] \\
&= \phi(\theta) \cdot \mathbf{E}_{\mu \sim \sigma^F|D}[v(\mu)]. \quad (3)
\end{aligned}$$

Next we bound the revenue contribution from types $\theta < \theta^*$, i.e., $\phi(\theta) < 0$.

$$\mathbf{E}_{\theta \sim F}[\text{Rev}(\mathcal{M}; \theta) \cdot \mathbf{1}\left[\theta < \theta^*\right]] - U(\underline{\theta})$$

$$= \int_{\underline{\theta}}^{\theta^*} f(\theta) \cdot \phi(\theta) \cdot \mathbf{E}_{\mu \sim \pi(\theta) | D}\Big[\mathbf{E}_{\hat{\mu} \sim \hat{\sigma}_{\theta, \mu} | \mu}[v(\hat{\mu})]\Big] \, \mathrm{d}\theta - U(\underline{\theta})$$

$$= -\int_{\underline{\theta}}^{\theta^*} (f(\theta) \cdot \phi(\theta))' \int_{\underline{\theta}}^{\theta} \mathbf{E}_{\mu \sim \pi(z) | D}\Big[\mathbf{E}_{\hat{\mu} \sim \hat{\sigma}_{z, \mu} | \mu}[v(\hat{\mu})]\Big] \, \mathrm{d}z \, \mathrm{d}\theta - U(\underline{\theta})$$

$$\leq -\int_{\underline{\theta}}^{\theta^*} (f(\theta) \cdot \phi(\theta))' \cdot (V(D, \hat{\Sigma}, \theta) - U(\underline{\theta})) \, \mathrm{d}\theta - U(\underline{\theta})$$

$$= -\int_{\underline{\theta}}^{\theta^*} (f(\theta) \cdot \phi(\theta))' \cdot V(D, \hat{\Sigma}, \theta) \, \mathrm{d}\theta - U(\underline{\theta})(f(\underline{\theta}) \cdot \phi(\underline{\theta}) + 1)$$

$$\leq -\int_{\underline{\theta}}^{\theta^*} (f(\theta) \cdot \phi(\theta))' \cdot V(D, \hat{\Sigma}, \theta) \, \mathrm{d}\theta - V(D, \hat{\Sigma}, \underline{\theta})(f(\underline{\theta}) \cdot \phi(\underline{\theta}) + 1)$$

$$= \int_{\underline{\theta}}^{\theta^*} \phi(\theta) \cdot \mathbf{E}_{\mu \sim \hat{\sigma}_{\theta, D} | D}[v(\mu)] \, \mathrm{d}\theta - V(D, \hat{\Sigma}, \underline{\theta}). \tag{4}$$

The second equality holds by integration by parts. The first inequality holds by (1) $(f(\theta) \cdot \phi(\theta))'$ is non-negative for $\theta \leq \theta^*$ under Assumption 2 (c.f., Devanur and Weinberg, 2017); and (2) $\int_{\underline{\theta}}^{\theta} \mathbf{E}_{\mu \sim \pi(z) | D}\big[\mathbf{E}_{\hat{\mu} \sim \hat{\sigma}_{z, \mu} | \mu}[v(\hat{\mu})]\big] \, \mathrm{d}z \geq V(D, \hat{\Sigma}, \theta) - U(\underline{\theta})$ according to the individual rational constraints in Lemma 1. The last inequality holds since $U(\underline{\theta}) \geq V(D, \hat{\Sigma}, \underline{\theta})$ and $f(\underline{\theta}) \cdot \phi(\underline{\theta}) + 1 = f(\underline{\theta}) \cdot \underline{\theta} + F(\underline{\theta}) \geq 0$. Finally, the last inequality holds by integration by parts and the facts that $\phi(\theta^*) = 0$ and

$$V(D, \hat{\Sigma}, \theta) = V(D, \hat{\Sigma}, \underline{\theta}) + \int_{\underline{\theta}}^{\theta} \mathbf{E}_{\mu \sim \hat{\sigma}_{z, D} | D}[v(\mu)] \, \mathrm{d}z.$$

Combining Equations (2) to (4), we have

$$\text{Rev}(\mathcal{M}) \leq \mathbf{E}_{\theta \sim F}[\text{Rev}(\mathcal{M}; \theta) \cdot \mathbf{1}\left[\theta \geq \theta^*\right]] + \mathbf{E}_{\theta \sim F}[\text{Rev}(\mathcal{M}; \theta) \cdot \mathbf{1}\left[\theta < \theta^*\right]] - U(\underline{\theta})$$

$$\leq \int_{\theta^*}^{\bar{\theta}} \phi(\theta) \cdot \mathbf{E}_{\mu \sim \sigma^F | D}[v(\mu)] \, \mathrm{d}\theta + \int_{\underline{\theta}}^{\theta^*} \phi(\theta) \cdot \mathbf{E}_{\mu \sim \hat{\sigma}_{\theta, D} | D}[v(\mu)] \, \mathrm{d}\theta - V(D, \hat{\Sigma}, \underline{\theta}) = \text{Rev}(\widehat{\mathcal{M}}).$$

Thus mechanism $\widehat{\mathcal{M}}$ is revenue optimal. $\square$

Proposition 3. *For any cost C^A and any prior D, if $\sigma^N \in \arg\max_{\hat{\sigma} \in \hat{\Sigma}} \mathbf{E}_{\hat{\mu} \sim \hat{\sigma} | D}[V(\hat{\mu}, \theta^*)] - C^A(\hat{\sigma}, D)$, the optimal mechanism is to post a price for revealing full information.*

Proof. By Theorem 2, it is sufficient to show that if $\sigma^N \in \arg\max_{\hat{\sigma} \in \hat{\Sigma}} \mathbf{E}_{\hat{\mu} \sim \hat{\sigma} | D}[V(\hat{\mu}, \theta^*)] - C^A(\hat{\sigma}, D)$, then $\hat{\sigma}_{\theta, D} = \sigma^N$ for any $\theta < \theta^*$. Suppose by contradiction that there exists $\theta < \theta^*$ such that $C^A(\hat{\sigma}_{\theta, D}, D) > 0$, i.e.,

$$\mathbf{E}_{\mu \sim \hat{\sigma}_{\theta, D} | D}[v(\mu)] \cdot \theta - C^A(\hat{\sigma}_{\theta, D}, D) \geq \mathbf{E}_{\mu \sim \sigma^N | D}[v(\mu)] \cdot \theta.$$

Since $\theta^* > \theta$, we have that

$$\mathbf{E}_{\mu \sim \hat{\sigma}_{\theta, D} | D}[v(\mu)] \cdot \theta^* - C^A(\hat{\sigma}_{\theta, D}, D) - \mathbf{E}_{\mu \sim \sigma^N | D}[v(\mu)] \cdot \theta^*$$

$$> \Big(\mathbf{E}_{\mu \sim \hat{\sigma}_{\theta, D} | D}[v(\mu)] - \mathbf{E}_{\mu \sim \sigma^N | D}[v(\mu)]\Big) \cdot \theta - C^A(\hat{\sigma}_{\theta, D}, D) \geq 0,$$

contradicting to the assumption that σ^N is one of the optimal choices for type θ^* given the prior D.

□

In the following lemma, in the case that the state space Ω is finite, we formalize the intuition that when the prior is sufficiently close to the degenerate pointmass distribution, the marginal cost for additional information exceeds the marginal benefit of additional information.

Definition 4. *The set of possible experiments $\hat{\Sigma}$ is finitely generated if it is generated by σ^N and a finite set $\hat{\Sigma}'$ through sequential learning, where σ^N is the one that always reveals no additional information with zero cost, and any $\hat{\sigma} \in |\hat{\Sigma}'|$ is an experiment that provides an informative signal about the state with fixed cost $c_{\hat{\sigma}} > 0$.*

Lemma 3. *Suppose Ω is finite and $\hat{\Sigma}$ is finitely generated. Suppose that there exists $\bar{v} < \infty$ such that $\max_{\omega \in \Omega} v(\omega) \leq \bar{v}$ and $v(\mu) \geq \min_{\omega \in \Omega} \mu(\omega) \cdot v(\omega)$ for any μ. Then there exists $\epsilon > 0$ such that any prior D satisfying $D(\omega) > 1 - \epsilon$ for some $\omega \in \Omega$ is sufficiently informative.*[24]

Proof. Let $c_m = \min_{\hat{\sigma} \in \hat{\Sigma}'} > 0$. By construction, for any experiment $\hat{\sigma} \in \hat{\Sigma}$, we have $C^A(\hat{\sigma}, \mu) \geq c_m$ for any μ. Let ω^* be the state such that $D(\omega^*) > 1 - \epsilon$. Given prior D, the utility increase of type θ^* for additional information is at most

$$\theta^* \cdot \left(\sum_{\omega \in \Omega} D(\omega) v(\omega) - v(D) \right) \leq \theta^* \cdot \left(\sum_{\omega \neq \omega^*} D(\omega) v(\omega) \right) < \theta^* \cdot \epsilon \cdot \bar{v}.$$

The first inequality holds since $v(D) \geq D(\omega^*) v(\omega^*)$, and the second inequality holds since $v(\omega) \leq \bar{v}$ and $\sum_{\omega \neq \omega^*} D(\omega) < \epsilon$. Thus, when $\epsilon = \frac{c_m}{\theta^* \cdot \bar{v}}$, the cost of information is always higher than the benefit of information, and the agent with type θ^* will never acquire any additional information given prior D. □

Before the proof of Theorem 3, we first introduce the definition of quantiles and revenue curves, which are helpful for bounding the approximation ratio. For any distribution F, let $q_F(\theta) \triangleq \Pr_{z \sim F}[z \geq \theta]$ be the quantile corresponding to type θ. Accordingly, we can define $\theta(q)$ as the type corresponds to quantile q. The revenue curve as a function of the quantile is defined as $R_F(q) \triangleq q \cdot \theta(q)$. Note that the regularity condition in Assumption 2 is equivalent to the concavity assumption for the revenue curve.

Lemma 4 (Myerson, 1981). *A distribution F is regular if and only if $R_F(q)$ is concave in q.*

Theorem 3. *For convex and separable valuations, under Assumption 1 and 2, for any prior D and any cost function C^A, posting a deterministic price for revealing full information achieves at least half of the optimal revenue.*

Proof of Theorem 3. We first normalize the primitives such that $\theta^* \cdot q(\theta^*) = 1$. For any type θ, let $c(\theta) \triangleq V(D, \hat{\Sigma}, \theta)$ be the outside option of the agent for not participating the auction. It is easy to verify that $c(\theta)$ is convex in θ. Let $\bar{x} \triangleq \mathbf{E}_{\omega \sim D}[v(\omega)]$ be the maximum possible allocation. According to Theorem 2, if the distribution F is regular, in the revenue optimal mechanism, the expected utility of the agent is $c(\theta)$ for any $\theta < \theta^*$ and is $(\theta - \theta^*) \cdot \bar{x} + c(\theta^*)$ for any $\theta \geq \theta^*$.

---

[24]We can have similar results when $\hat{\Sigma}$ is not finitely generated. For example, when the cost function is the reduction in entropy, by applying the techniques in Caplin, Dean, and Leahy (2019), for any valuation function v, there exists $\epsilon > 0$ such that any prior D satisfying $D(\omega) > 1 - \epsilon$ for some $\omega \in \Omega$ is sufficiently informative.
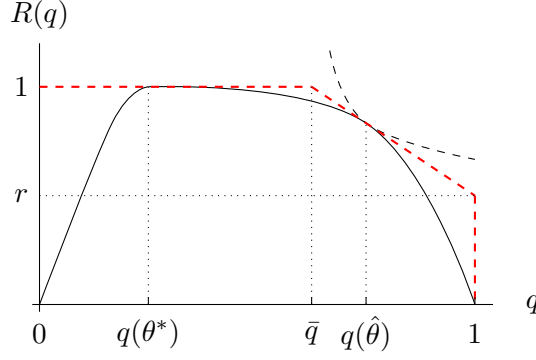
Figure 2: The figure illustrates the reduction on the type distribution that maximizes the approximation ratio between the optimal revenue and the price posting revenue. The black solid curve is the revenue curve for distribution F and the red dashed curve is the revenue curve for distribution $\hat{F}$. The black dashed curve is the revenue curve $\bar{F}$ such that the seller is indifferent at deterministically selling at any prices with negative virtual value.

Suppose $\hat{\theta}$ is the cutoff type that participates the auction in the optimal price posting mechanism for distribution F. It is easy to verify that $\hat{\theta} \leq \theta^*$ since revealing full information to any type above the monopoly type only increases the expected revenue. Moreover, for any type $\theta < \theta^*$, since the payment that inducing $\hat{\theta}$ to be the cutoff type is $\hat{\theta} \cdot \bar{x} - c(\hat{\theta})$, we have that

$$(\hat{\theta} \cdot \bar{x} - c(\hat{\theta})) \cdot q_F(\hat{\theta}) \geq (\theta \cdot \bar{x} - c(\theta)) \cdot q_F(\theta).$$

That is, any type $\theta < \theta^*$,

$$q_F(\theta) \leq \frac{(\hat{\theta} \cdot \bar{x} - c(\hat{\theta})) \cdot q_F(\hat{\theta})}{\theta \cdot \bar{x} - c(\theta)}.$$

Let $\bar{F}$ be the distribution such that

$$q_{\bar{F}}(\theta) = \frac{(\hat{\theta} \cdot \bar{x} - c(\hat{\theta})) \cdot q_F(\hat{\theta})}{\theta \cdot \bar{x} - c(\theta)}$$

for any type θ. Thus the virtual value function $\bar{\phi}(\theta)$ for distribution $\bar{F}$ is

$$\bar{\phi}(\theta) = \theta - \frac{\theta \cdot \bar{x} - c(\theta)}{\bar{x} - c'(\theta)} \leq 0.$$

Moreover,

$$\bar{\phi}'(\theta) = \frac{c''(\theta) \cdot (c(\theta) - \theta \cdot \bar{x})}{(\bar{x} - c'(\theta))^2} \leq 0.$$

Thus the revenue curve such that the seller is indifferent at selling at any price is convex. Let $\hat{F}$ be the distribution with piecewise linear revenue curve illustrated in Figure 2. Thus we have that

$$q_{\hat{F}}(\theta) = \begin{cases} \frac{1}{\theta} & \theta \geq \frac{1}{\bar{q}}, \\ \frac{1 - r\bar{q}}{\theta(1-\bar{q}) + 1 - r} & \theta < \frac{1}{\bar{q}}. \end{cases}$$

26

Let $p(\theta) \triangleq \theta \cdot \bar{x} - c(\theta) \geq 0$. First note that distribution $\hat{F}$ is first order stochastically dominated by $\bar{F}$, the optimal revenue from posted pricing is weakly smaller for distribution $\hat{F}$. Moreover, both distributions achieve the same price posting revenue by choosing the price $p(\hat{\theta})$ such that the cutoff type is $\hat{\theta}$. Thus the optimal price posting revenue for distribution $\hat{F}$ is attained by choosing price $p(\hat{\theta})$. This further indicates that optimal price posting revenue is the same for distribution F and $\hat{F}$, i.e., $\mathrm{PP}(F, c) = \mathrm{PP}(\hat{F}, c)$. Secondly, since distribution F is first order stochastically dominated by $\hat{F}$, it is easy to verify that $\mathrm{OPT}(F, c) \leq \mathrm{PP}(\hat{F}, c)$. Therefore, the ratio between the price posting revenue and the optimal revenue is minimized when the type distribution is $\hat{F}$.

For distribution $\hat{F}$, since the optimal price is $p(\hat{\theta})$, we have

$$(\hat{\theta} \cdot \bar{x} - c(\hat{\theta})) \cdot q_{\hat{F}}(\hat{\theta}) \geq (\theta \cdot \bar{x} - c(\theta)) \cdot q_{\hat{F}}(\theta).$$

Let $\zeta = (\hat{\theta} \cdot \bar{x} - c(\hat{\theta})) \cdot q_{\hat{F}}(\hat{\theta})$ and let

$$\hat{c}(\theta) = \theta \cdot \bar{x} - \frac{\zeta}{q_{\hat{F}}(\theta)}.$$

It is easy to verify that $\mathrm{PP}(\hat{F}, c) = \mathrm{PP}(\hat{F}, \hat{c}) = \zeta$. Moreover, $\hat{c}(\theta)$ is convex and $c(\theta) \geq \hat{c}(\theta)$ for any θ, which implies that any feasible mechanism for c is also feasible for $\hat{c}$, and hence $\mathrm{OPT}(\hat{F}, c) \leq \mathrm{OPT}(\hat{F}, \hat{c})$. Thus to prove Theorem 3, it is sufficient to bound $\frac{\mathrm{PP}(\hat{F}, \hat{c})}{\mathrm{OPT}(\hat{F}, \hat{c})}$. Note that by construction, the monopoly type for distribution $\hat{F}$ is $\frac{1}{\bar{q}}$. Hence the optimal revenue is

$$\mathrm{OPT}(\hat{F}, \hat{c}) = \int_r^{\frac{1}{\bar{q}}} \hat{f}(\theta)(\theta \cdot \hat{c}'(\theta) - \hat{c}(\theta)) \, \mathrm{d}\theta + \left(\frac{1}{\bar{q}} \cdot \bar{x} - \hat{c}(\frac{1}{\bar{q}}) \right) \cdot \bar{q}$$

$$= \int_r^{\frac{1}{\bar{q}}} \hat{f}(\theta) \cdot \zeta \cdot \frac{1 - r}{1 - r\bar{q}} \, \mathrm{d}\theta + \zeta$$

$$= \zeta \cdot \left(\frac{(1 - \bar{q})(1 - r)}{1 - r\bar{q}} + 1 \right) \leq 2\zeta,$$

where the inequality is tight if $\bar{q} = r = 0$. Combining the observations, for any distribution F and any outside option function c induced by the set of experiments $\hat{\Sigma}$, we have

$$\frac{\mathrm{PP}(F, c)}{\mathrm{OPT}(F, c)} \geq \frac{\mathrm{PP}(\hat{F}, \hat{c})}{\mathrm{OPT}(\hat{F}, \hat{c})} \geq \frac{1}{2}.$$

$\square$

Proposition 4. *For any valuation function V, any prior D, and any type distribution F, in the revenue optimal mechanism,*

- *the social welfare is minimized when $|\hat{\Sigma}| = 1$;*[25]

- *the utility of the agent is minimized when $|\hat{\Sigma}| = 1$.*

Proof for Statement 1 of Proposition 4. Recall that $\theta^* = \inf_\theta \{\phi(\theta) \geq 0\}$. Note that for any type $\theta \geq \theta^*$, the agent receives full information regardless of the set of possible experiments $\hat{\Sigma}$ for the agent. For any type $\theta < \theta^*$, by the proof of Proposition 2, the agent receives no information when $|\hat{\Sigma}'| = 1$. Since no information is the least preferred allocation for the agent, the social welfare is minimized when $|\hat{\Sigma}'| = 1$. $\square$

---

[25]Our result actually implies that the expected value for each type of the agent is minimized when $|\hat{\Sigma}| = 1$.

C Competing Sellers

In the previous section, we have provided the characterization of the optimal mechanism when the ability of acquiring additional information is given as exogenous. In reality, the agent's ability of acquiring additional information could arise when there are multiple data brokers providing different menus for selling experiments, and the agent can choose to purchase information sequentially from different data brokers. In this section, we will focus on the model of two data brokers competing for selling experiments to a single agent. The qualitative results generalize directly to the setting with multiple data brokers.

To simplify the analysis, we assume that the cost of information for data broker 1 is zero, while the cost function C_2 for data broker 2 is non-negative. Note that our result can be directly generalized to the setting where the the cost of information for data broker 1 is positive but at most the cost for data broker 2. In addition, we assume that the set of possible experiments $\hat{\Sigma}_1$ for data broker 1 is a superset of the set of possible experiments $\hat{\Sigma}_2$ for data broker 2.

Here we still impose the constraint that the set of possible experiments and the cost functions for both data brokers satisfy Assumption 1. The timeline of the model is formalized as follows.

1. Data brokers 1 and 2 commit to mechanisms $\mathcal{M}_1$ and $\mathcal{M}_2$ respectively.

2. The agent chooses entries from both mechanisms sequentially based on the past signals she received. Note that the agent can purchase the experiments from each data broker multiple times until it is not beneficial for her to do so.

3. The reward $V(\hat{\mu}, \theta)$ of the agent is realized based on her posterior belief $\hat{\mu}$ and her private type θ.

Note that an important part in the above model is that the agent can purchase from the same data broker multiple times given the committed mechanism. We will briefly discuss the implication on classical model where purchase occurs once at the end of the section. This setup of multiple purchases has been studied in Chawla et al. (2020) as the buy-many mechanism. To justify this, consider an alternative interpretation for the single-agent model where there is a continuum of agents with total measure 1 and the distribution over this continuum of agents is F. In reality, after the data broker committed to a selling mechanism, the data broker does not observe the identity of the agent in the continuum, and each agent can simply interact with the data broker multiple times to purchase multiple experiments from the same data broker. For example, Amazon turk provides services for assigning surveys and questionnaires to individual workers with a fixed 20% fee per assignment. Agents with higher valuations for more precise signals can always delegate their multiple surveys to Amazon turk with a linear rate on the commission fee. Note that this assumption serves as a limitation for the data brokers to price discriminate different types of the agent, and as will become clear in the later analysis, under equilibrium, it suffices for the agent to interact with each data broker for at most one time.[26]

Another restriction on the model is that we will only focus on the pure strategy subgame perfect equilibria of the data brokers. More specifically, we require each data broker to deterministically choose a menu for selling experiments. Note that we do not require the entries inside the menu to be deterministic. In addition, we do not require the agent to have pure strategy. It is an interesting

---

[26]In the model of the previous section where the agent has the exogenous ability to acquire additional information, we can also allow the agent to purchase from the data broker for multiple times given the offered menu. Applying the techniques we will develop in the following theorem, our main observation that the agent does not acquire costly information under equilibrium remains unchanged.

open question to pin down the behavior of the equilibrium when the data brokers can adopt mixed strategies.

Theorem 4. *In any pure strategy subgame perfect equilibrium, the expected revenue of data broker 2 is 0.*

Theorem 4 implies that under equilibrium, the agent will only purchase information from the data broker with lower cost for providing information. Note that in order for Theorem 4 to be meaningful, there should exist at least one pure subgame perfect equilibrium. It is easy to verify that the following strategies is a pure strategy subgame perfect equilibrium:[27] data broker 2 offers each experiment $\hat{\sigma} \in \hat{\Sigma}_2$ with price equals the cost, data broker 1 best responds to it, and the agent acquires information optimally from data broker 1. Under this equilibrium, data broker 1 will commit to the mechanism such that the agent has no incentive to purchase from data broker 2. Theorem 4 states that this uniquely pin down the revenue of data broker 2 in any pure strategy subgame perfect equilibrium.[28] This observation is similar to the Bertrand competition model, where the firms set price at the marginal cost under equilibrium. Finally, note that although in this section we have focused on the model where the buyer can purchase from each data broker repeatedly, it is easy to verify that in the classical model where the buyer can only purchase from each data broker once, the strategy profile described above is still an equilibrium strategy for both data brokers. In addition, under equilibrium, the agent never purchase from data broker 2.

Proof of Theorem 4. The idea for proving Theorem 4 is similar to Theorem 1. By taxation principle for the common agency game, it is without loss to focus on mechanisms that all data brokers offer menus to the agent (Martimort and Stole, 2002). Let $(\mathcal{M}_1, \mathcal{M}_2)$ be a pure Nash equilibrium where $\mathcal{M}_1, \mathcal{M}_2$ are mechanisms that offer menus. Suppose by contradiction under equilibrium the expected payment of the agent to data broker 2 is $\epsilon > 0$. For any type θ, let π_θ be the distribution over sequence of experiments that agent with type θ receives under equilibrium, and p_θ be the expected payment of the agent to both of the data brokers. Let $\bar{p} = \sup_\theta p_\theta$ Let $\hat{\mathcal{M}}_1 = (\hat{\pi}_1, \hat{p}_1)$ be the mechanism that offers option $(\pi_1(\theta), \hat{p}_1(\theta))$ for each type θ, where $\pi_1(\theta) = \pi_\theta$ and $\hat{p}_1(\theta) = p_\theta - \frac{\epsilon \cdot p_\theta}{2\bar{p}}$.[29] Since both $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$ satisfies Assumption 1 and $\hat{\Sigma}_2 \subseteq \hat{\Sigma}$, any sequence of experiments chosen by the agent is an element in $\hat{\Sigma}_1$, and hence $\hat{\mathcal{M}}_1$ is a feasible mechanism for data broker 1. It is easy to verify that when the agent interacts with data broker 1 by revealing the type truthfully, the utility of the agent is non-negative and the revenue of data broker 1 is

$$\int_\Theta \hat{p}_1(\theta) \, dF(\theta) \geq \int_\Theta \left(p_\theta - \frac{\epsilon}{2} \right) \, dF(\theta) > \text{Rev}_1(\mathcal{M}_1, \mathcal{M}_2)$$

where $\text{Rev}_1(\mathcal{M}_1; \mathcal{M}_2)$ is the equilibrium revenue for data broker 1 given mechanism profile $(\mathcal{M}_1, \mathcal{M}_2)$. Next it is sufficient to show that given mechanism profile $(\hat{\mathcal{M}}_1, \mathcal{M}_2)$, when the agent purchases information optimally from both data brokers, the equilibrium revenue of data broker 1 weakly increases, i.e.,

$$\text{Rev}_1(\hat{\mathcal{M}}_1, \mathcal{M}_2) \geq \int_\Theta \hat{p}_1(\theta) \, dF(\theta).$$

---

[27]This equilibrium is not unique. Another equilibrium is that both data brokers offer full information with zero payment and the agent acquires information optimally from data broker 1.

[28]This result is not a direct application of Theorem 1, since the agent can potentially purchase multiple entries from the menu posted by data broker 1.

[29]The additional discount $-\frac{\epsilon \cdot p_\theta}{2\bar{p}}$ is used to provide incentives for the agent to break ties in favor of experiments with higher payments when there are multiple best response choices given mechanism profile $(\mathcal{M}_1, \mathcal{M}_2)$.

Let π_θ^* be any optimal sequence of experiments chosen by the agent given mechanism profile $(\hat{\mathcal{M}}_1, \mathcal{M}_2)$, and let p_θ^* be the corresponding payment. Note that by Assumption 1, the sequence of experiments π_θ^* chosen by the agent is also feasible when the mechanism profile is $(\mathcal{M}_1, \mathcal{M}_2)$. Let p_θ' be the payment of the agent for choosing the sequence of experiments π_θ^* given mechanism profile $(\mathcal{M}_1, \mathcal{M}_2)$. By the definition of the payment rule $\hat{p}$, we have $p_\theta^* = p_\theta' - \frac{\epsilon \cdot p_\theta'}{2\bar{p}}$. Let $V(\pi_\theta^*, \theta)$ be the value of the agent with type θ and sequence of experiments π_θ^*. Since π_θ^* is the optimal choice in $(\hat{\mathcal{M}}_\infty, \mathcal{M}_2)$ and π_θ is the optimal choice in $(\mathcal{M}_1, \mathcal{M}_2)$, we have

$$V(\pi_\theta^*, \theta) - p_\theta^* \geq V(\pi_\theta, \theta) - \hat{p}_\theta = V(\pi_\theta, \theta) - p_\theta - \frac{\epsilon \cdot p_\theta}{2\bar{p}}$$

and

$$V(\pi_\theta^*, \theta) - p_\theta' \leq V(\pi_\theta, \theta) - p_\theta.$$

Combining the inequalities, we have $p_\theta^* \geq \hat{p}_\theta$ for any type θ, and hence $\mathrm{Rev}_1(\hat{\mathcal{M}}_1, \mathcal{M}_2) \geq \int_\Theta \hat{p}_1(\theta) \, dF(\theta)$. $\qquad\square$

D   Necessity of Assumption 1

In this section, we show that Assumption 1 is necessary for Theorem 1. We will prove this by constructing an instance where the set of possible experiments for the agent violates Assumption 1. One thing to keep in mind is that the goal for the following contrived construction is not to provide an instance that fits the economic applications. It only serves the purpose of showing that Assumption 1 is necessary for our observation in Theorem 1.

Suppose there is a binary state $\Omega = \{\omega_1, \omega_2\}$. Thus the posterior is uniquely determined by the probability of ω_1, and thus we will also use μ to represent $\mu(\omega_1)$. We assume that the prior $D = \frac{1}{2}$. The data broker can choose any experiment, and the set of possible experiments for the agent is $\{\sigma^N, \sigma^F, \hat{\sigma}_1\}$. Recall that σ^N reveals no information and σ^F reveals full information. The signal space S for $\hat{\sigma}_1$ is $\{s_1, s_2\}$, and $\hat{\sigma}_1(s_1|\omega_1) = \hat{\sigma}_1(s_2|\omega_2) = \frac{2}{3}$. The cost function satisfies $C(\sigma^N, \mu) = 0$ and $C(\hat{\sigma}_1, \mu) = 1$ for all μ, and

$$C(\sigma^F, \mu) = \begin{cases} 10 & \mu \in [\frac{1}{3}, \frac{2}{3}]; \\ 1 & \text{otherwise.} \end{cases}$$

We assume that the type space is also binary, i.e., $\Theta = \{\theta_1, \theta_2\}$. There is common prior over the types F where $F(\theta_1) = F(\theta_2) = \frac{1}{2}$. We assume that

$$V(\mu, \theta_1) = \begin{cases} 3 & \mu = \frac{4}{5} \text{ or } \frac{1}{5}; \\ 0 & \text{otherwise;} \end{cases}$$

$$V(\mu, \theta_2) = \begin{cases} 10 & \mu = 0 \text{ or } 1; \\ 0 & \text{otherwise.} \end{cases}$$

It is not difficult to verify that in the optimal mechanism $\mathcal{M} = (\pi, p)$, we have that $\pi(\theta_1) = \hat{\sigma}_1, p(\theta_1) = \frac{2}{3}$ and $\pi(\theta_2) = \sigma^F, p(\theta_2) = 10$. Under equilibrium, for type θ_1, the agent has incentive to acquire additional information by conducting experiment $\hat{\sigma}_1$ regardless of the signal realization. In the proof of Theorem 1, if Assumption 1 is satisfied, the data broker can simulate the behavior of θ_1 by offering experiment $\hat{\sigma}_1 \circ \hat{\sigma}_1$ to agent with type θ_1, raise the payment for type θ_1, and accordingly

30

increase the expected revenue. However, this option is not profitable in this constructed example. The main reason is that if the data broker offers type θ_1 with experiment $\hat{\sigma}_1 \circ \hat{\sigma}_1$, agent with type θ_2 will have strong incentive to deviate her report to θ_1, since the cost for conducting experiment σ^F is significantly reduced given the posterior distribution induced by experiment $\hat{\sigma}_1 \circ \hat{\sigma}_1$. Thus the payment extracted from type θ_2 will be significantly smaller and the expected revenue of the data broker is reduced in this case.