Third-Party Data Providers Ruin Simple Mechanisms

Yang Cai* Federico Echenique† Hu Fu‡ Katrina Ligett§ Adam Wierman¶
Juba Ziani‖

February 22, 2018

Abstract

This paper studies the revenue of simple mechanisms in settings where a third-party data provider is present. When no data provider is present, it is known that simple mechanisms achieve a constant fraction of the revenue of optimal mechanisms. The results in this paper demonstrate that this is no longer true in the presence of a third party data provider who can provide the bidder with a signal that is correlated with the item type. Specifically, we show that even with a single seller, a single bidder, and a single item of uncertain type for sale, pricing each item-type separately (the analog of item pricing for multi-item auctions) and bundling all item-types under a single price (the analog of grand bundling) can both simultaneously be a logarithmic factor worse than the optimal revenue. Further, in the presence of a data provider, item-type partitioning mechanisms—a more general class of mechanisms which divide item-types into disjoint groups and offer prices for each group—still cannot achieve within a $\log\log$ factor of the optimal revenue.

1 Introduction

This paper investigates the effectiveness of simple mechanisms in settings where a bidder is uncertain about the type of good he is bidding on, and obtains information about the good from a third-party data provider. When no data provider is present, it is known that simple mechanisms can be used to provide a constant fraction of the revenue of optimal mechanisms; however, the results in this paper demonstrate that this is no longer possible when a data provider is present.

Mechanism design is the theory of informational asymmetries in economic environments. Indeed, information asymmetries are rampant in markets from ad auctions to art auctions; from acquiring a summer home to acquiring a startup. Naturally, whenever significant information asymmetries occur, agents have incentives to acquire information through outside channels. As a result, there is a proliferation of companies that seek to collect information that can be sold to participants in auctions with information asymmetries. Online adversing provides an extreme example. By tracking online behavior, *data providers* are able to sell valuable information about internet users (whose attention is the good for sale) to bidders in online advertising auctions. A recent FTC report 2014 details the scale and prevalence of such data providers – generating $426 million in annual revenue in 2012 and growing considerably in the years since.

*McGill University, cai@cs.mcgill.ca
†California Institute of Technology, fede@hss.caltech.edu
‡University of British Columbia, fu.hu.thu@gmail.com
§Hebrew University of Jerusalem, katrina.ligett@mail.huji.ac.il
¶California Institute of Technology, adamw@caltech.edu
‖California Institute of Technology, jziani@caltech.edu

Some of the most elegant results in mechanism design focus on providing simple characterizations of (near) optimal auctions. A particularly beautiful example is Myerson's characterization of optimal auctions [Myerson, 1981]. In the case of one bidder, Myerson [1981] characterizes the optimal pricing mechanism for a monopolist facing a bidder who has a one-dimensional private valuation, where the distribution over the bidder's valuation for the item is known to the seller, though the realization of that valuation is known only to the bidder. Our paper is part of an effort to understand optimal auctions when bidders have access to information from a third-party data provider, outside of the control of the seller.

Specifically, the goal of this paper is to investigate the impact of third-party data providers on the revenue of simple mechanisms. To do this, we consider a simple market—a single seller, a single bidder, and a single good—and a particular form of information asymmetry—the seller knows the type of the good she is selling, but the bidder has only partial information on the item type; the bidder knows her valuation for each of the n types, but the seller knows only distributional information about the valuations. The key to the model is that, in addition to a prior over the item type, the bidder obtains a signal about the item type from a third party data provider and, while seller knows the signaling scheme used by the data provider, the seller does not know the realization of signals.

Our model, though stylized, is already general enough to expose the difficulties created by third-party data providers. In particular, our results show that, even in this case, simple mechanisms cannot provide revenue within a constant factor of the revenue provided by an optimal mechanism. This result is perhaps surprising in the context of an elegant recent paper, where Daskalakis et al. [2016] study optimal auctions in a related model, in the *absence* of a data provider. Daskalakis et al. [2016] study the design of simple mechanisms in a setting where the only uncertainty about the item type is that it is drawn from a common prior. In this context, the question is whether it is valuable for the seller to share information with the bidder about the item type, or whether mechanisms that do not reveal information can be approximately optimal. Interestingly, Daskalakis et al. [2016] are able to characterize optimal auctions for this setting. Their insights—showing a direct correspondence between mechanisms for selling a single item of uncertain type and multi-item auctions—when combined with the work of Babaioff et al. [2014], allow them to further observe that the better mechanism of two simple approaches—setting a fixed price for the item (the parallel notion of grand bundling, which we term "item-type bundling"), or pricing each item type separately (the parallel notion of item pricing, which we term "item-type pricing")—is guaranteed to yield a constant factor of the revenue of the optimal mechanism. Thus, in the case where there is no third-party data provider, there is little need for the seller to reveal information to the bidders; simple mechanisms are sufficient.

Our results show, in contrast, that the presence of a third-party data provider complicates the mechanism design task dramatically. We first consider revenue-optimal mechanisms. While Daskalakis et al. [2016]'s characterization of optimal auctions extends naturally to our setting, these optimal mechanisms may be quite complex. Concretely, our setting satisfies a type of revelation principle stating that optimal mechanisms require only a single round of bidding, followed by a single round of information revelation (full revelation, in fact); however, the resulting menu of options presented to the bidder includes an option for each possible valuation vector, combined with each possible posterior of the bidder on the item type after receiving the information provider's signal, and requires the seller to condition the price charged on the realization of the item type.

Not only does the presence of a data provider complicate the design of the optimal mechanism, it also impacts the revenue achievable via simple mechanisms. Specifically, in the presence of a data provider, the better of item-type bundling and item-type pricing may achieve only a $\Omega(\log n)$ factor approximation of the revenue the seller could have achieved had she offered a richer menu to the buyer. In particular, a mechanism that divides the item types into disjoint groups and offers a price on each group can outperform both item-type pricing and item-type bundling by a logarithmic factor. Such mechanisms are known in the multi-item auction literature as partition mechanisms, and are seen as relatively simple mechanisms (see, e.g., Rubinstein [2016]). In our setting, we refer to such mechanisms as *item-type partition mechanisms*.

This separation between the revenue of item-type partitioning and that of item-type pricing and item-type bundling raises a natural question: if we expand our view of what constitutes "simple" mechanisms to include item-type partitioning, which generalizes both item-type pricing and item-type bundling, can we guarantee that simple mechanisms obtain a constant approximation of the optimal revenue in the presence of a data provider?

Our next result uses a more intricate argument to show that this is not the case. We demonstrate that, in the presence of a data provider, optimal mechanisms can outperform the best item-type partition mechanism by an $\Omega(\log\log n)$ factor. So, in the presence of a data provider, simple mechanisms truly are no longer optimal.

The results discussed so far focus on the seller, while not taking the data provider's incentives into account. They assume a fixed behavior by the data provider, without providing him with an objective. However, of course, the behavior of the data provider has an impact. To highlight the complexities of the data provider's participation in this market, we study the special case where the data provider is *adversarial*. This setting is of particular interest because it demonstrates behavior that is, perhaps, counter-intuitive. Specifically, a data provider that is attempting to negatively impact the revenue of the seller *may not* want to fully reveal her information about the item type to the bidder. Instead, there may be intermediate signals which, upon revelation, minimize the revenue of the seller. This serves to additionally highlight the complexity of mechanism design in the context of a third-party data provider.

To summarize, in this paper we make the following contributions. We propose a simple model of an auction in the presence of a third-party data provider, capturing information asymmetry regarding the type of the item for sale. Within this model, we first (Section 3) provide a characterization of the optimal auction based on that of Daskalakis et al. [2016], which may require a complex menu of options. Our main results study the potential for simple mechanisms to approximate the revenue of optimal mechanisms. In Section 4, we show that the item-type equivalents of item pricing and grand bundling cannot achieve within a $\Omega(\log n)$ factor of the revenue achievable by the optimal mechanism, nor even of the best item-type partition mechanism. Further, in Section 5, we show that there may be a $\Omega(\log\log n)$ gap between the revenue of the best item-type partitioning and that of the optimal mechanism. These results highlight that the presence of a data provider significantly reduces the ability of simple mechanisms to approximate the revenue of optimal mechanisms, even in the case of a single seller and a simple buyer. Finally, in Section 6, we turn to understanding the behavior of the data provider and show that *partial* information revelation may be more damaging to the seller than *full* information revelation.

Related work

There is a rich literature on information and signaling in auctions. One line of research focuses on designing a signaling scheme (on the part of the seller) given a certain auction format such as the second price auction [see, e.g., Emek et al., 2014, Miltersen and Sheffet, 2012, Cheng et al., 2015, Dughmi et al., 2015]; another line, closer to our setting, studies the design of *both* the auction and the signaling scheme (again, in this line of work, there is no data provider; any signal comes from the seller). Fu et al. [2012] showed that, if the auctioneer commits to a signaling scheme before choosing the form of the auction, full revelation followed by Myerson's auction for the revealed item type is the optimal design. Daskalakis et al. [2016] revealed the subtlety of this order of commitment and showed that, when the design of the auction and that of the signaling scheme are considered together (without having to commit to one before the other), the optimal strategy is to reveal no information at all, and the overall problem is in fact equivalent to the design of a multi-item auction. In particular, they show that, when the bidders have a publicly known common prior $\vec{\pi}$ on the type of the item, the optimal revenue for the seller is that of a multi-item auction with n items.

Furthermore, Theorem 2 of Daskalakis et al. [2016] shows a one-to-one correspondence between a selling a single item of uncertain type and classical multi-item auctions. In particular, item-type pricing, i.e., mechanisms in which the seller first reveals the item type and then charges a take-it-or-leave-it price, is equivalent to selling separately (i.e., item pricing) in the corresponding multi-item auction, and item-type bundling, i.e., mechanisms in which the seller does not reveal any information and offers a single take-it-or-leave-it price, is equivalent to grand bundling in the corresponding multi-item auction. When there is a single bidder, Daskalakis et al. [2016] further combine this correspondence with results of Babaioff et al. [2014] to show that the better of item-type pricing and item-type bundling gives at least 1/6 of the optimal revenue.

The results described above highlight the connection between our work and the study of simple mechanisms for multi-item auctions. Hart and Nisan [2012] pioneered this area. They showed that a seller, using item pricing, can extract a $\Omega\left(1/\log^2 n\right)$ fraction of the optimal revenue from an additive buyer whose values for n items are drawn independently, and selling these items as a bundle can achieve a $\Omega\left(1/\log n\right)$-fraction of the optimal revenue if the buyer's values are i.i.d. Li and Yao [2013] improved the approximation ratio for item pricing to $O\left(1/\log n\right)$, which is tight. Babaioff et al. [2014] showed that, surprisingly, the better of selling separately and grand bundling can achieve at least 1/6 of the optimal revenue. Subsequently there has been a surge of results generalizing the results of Babaioff et al. to broader settings [Cai and Huang, 2013, Yao, 2015, Rubinstein and Weinberg, 2015, Cai et al., 2016, Chawla and Miller, 2016, Cai and Zhao, 2017]. At this point, it is known that simple mechanisms such as sequential two-part tariffs can obtain a constant fraction of the optimal revenue for multiple buyers with combinatorial valuations that are, e.g., submodular, XOS [Cai and Zhao, 2017]. One might hope to extend these simple deterministic mechanisms to settings where the buyer has correlated values over the items; however, this is impossible. Hart and Nisan [2013] showed that even for a single additive buyer, when valuations are interdependent, the ratio between the revenue obtainable by a randomized mechanism and that of the best deterministic mechanism can be unbounded.

2 Model & Preliminaries

We consider a setting with a single, revenue-maximizing seller, a single item for sale, and a single bidder. The item for sale takes one of n possible types, and the bidder's valuation may depend on the item type. The bidder does not know the type i of the item, but has a publicly-known prior $\vec{\pi}$ over the item types. We let $\pi(i)$ denote the prior probability that the item is of type i.

The bidder's private value when the good is of type i is drawn from a publicly known distribution $D(i)$ over the space of non-negative real numbers $\mathbb{R}^+$. We denote by $V(i)$ the bidder's valuation for an item of type i, and denote by $\vec{V} = (V(1), \ldots, V(n))$ the bidder's valuation vector.

There is a third-party data provider who has (potentially imperfect) information on the type of the item, in the form of a random variable that can be arbitrarily correlated with the type of the item. Its distribution is publicly known, but its realization is only observed by the data provider. The data provider designs a signaling scheme in the form of a random variable S that is measurable with respect to the data-provider's information, and reveals the realization of S solely to the bidder. The seller does not observe the realization of S, but knows its distribution. After observing the realization s of S, the bidder updates his prior using Bayes' rule. We denote the resulting posterior by $\vec{\pi_s}$. While the bidder knows his posterior perfectly, the seller only knows the distribution over posteriors, conditional on the item type.

2.1 Simple mechanisms for a single bidder and no data provider

In the absence of a third-party data provider, in the single seller, single bidder, *multi-item* setting, *where each item has a single type*, Babaioff et al. [2014] show that, although the optimal mechanism may be complex, a simple mechanism achieves a constant factor of the optimal revenue. In particular, this mechanism is simply the better of either item pricing or grand bundling. This result was originally stated for multi-item auctions, but the results of Daskalakis et al. [2016] show that an equivalent result holds for the setting of a single item that can take on multiple possible types.

These simple mechanisms are important throughout our paper, so we formally define them here, in the context of selling a single item with multiple possible types.

Definition 1. An **item-type pricing** mechanism first reveals the type i of the item to the bidder, then offers to sell the item to the bidder at some price $P_{\text{it}}(i)$. We also refer to such mechanisms as "selling the types separately," in analogy to the concept of selling separately in the case of multi-item auctions.

Definition 2. An **item-type bundling** mechanism offers the item for sale at some price P_{gr} without revealing any information about the realized type of the item.

The following result from Babaioff et al. [2014], Daskalakis et al. [2016] summarizes the power of these simple mechanisms in the single-item, multi-type setting, without a data provider.

Proposition 1 (Babaioff et al. [2014], Daskalakis et al. [2016]). *In the absence of a data provider, the maximum of item-type pricing and item-type bundling yields at least a $\frac{1}{6}$-approximation to the optimal revenue when there is a single seller, a single bidder, a single item for sale, and the bidder has a publicly known prior over the type of the item.*

A generalization of both item-type bundling and item-type pricing mechanisms, inspired by Rubinstein [2016], is also important for the results in this paper.

Definition 3. An **item-type partition** mechanism first partitions the set of item types into non-empty groups $\mathcal{G}_1$ to $\mathcal{G}_g$, priced (resp.) P_1 to P_g. The mechanism then observes the type i of the item, and offers the item at price P_r, where r is uniquely chosen such that $i \in \mathcal{G}_r$.

Note that, after observing the offered price P_r, the bidder may infer that the realized item type must belong to group $\mathcal{G}_r$. Item-type pricing is an instantiation of item-type partitioning where the partition contains a separate group for each type; item-type bundling corresponds to item-type partitioning using the trivial partition. Item-type partitioning is, however, significantly more powerful than these other simple mechanisms, as it allows the seller to partition the item types into arbitrarily many groups of arbitrarily many sizes.

2.2 The equal revenue distribution

An important example in the derivation of lower bounds for mechanisms is the so-called equal revenue distribution. This distribution is crucial to a number of our examples in this paper and is defined as follows.

Definition 4. A random variable X with support $[1, +\infty)$ follows the **equal revenue (ER) distribution** if and only if $\mathbb{P}[X \leq x] = 1 - \frac{1}{x}$.

The equal revenue distribution gets its name from the fact that it has constant virtual value, and every price in the distribution's support offers the same expected revenue. The equal revenue distribution also has a number of other useful properties, proven in Hart and Nisan [2012], which we summarize in the following.

Lemma 1 (Hart and Nisan [2012])**.** *Let $n \geq 2$ be an integer, and let $I_1, \ldots, I_n$ be n i.i.d random variables that follow the ER distribution. Then:*

$$\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n} I_i \geq \frac{\log n}{2}\right] \geq \frac{1}{2},$$

and for any $P \geq 6 \log n$,

$$\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n} I_i \geq P\right] \leq \frac{9}{P}.$$

Unless otherwise specified, log is taken to be the natural logarithm.

3 Optimal mechanisms in the presence of a data provider

Here and throughout the rest of the paper, we assume the presence of a third-party data provider who knows (possibly imperfect) information about the item type, and who reveals some of this information to the bidder. The goal of this section is to characterize optimal mechanisms for a single bidder and a single item with several possible types, in the presence of such a data provider.

We consider a class of mechanisms that allow the seller to charge the bidder a price that is conditional on the type of the item. We observe that restricting attention to such item-contingent price mechanisms is without loss of generality. The characterization we present is a type of revelation principle, similar to that presented in Daskalakis et al. [2016], where the difference is the presence of a data provider. Note that the data provider is represented via a signaling scheme that, from the model perspective, is subsumed into a probability distribution over posteriors $\vec{\pi}$, representing beliefs of the bidder regarding the item's type. Therefore, the bidder has private information in the form of a valuation $\vec{v}$ and a posterior $\vec{\pi}$.

The characterization we present shows that the revenue achievable via any mechanism can be obtained with a conditional price menu, defined as follows.

Definition 5. A **menu with conditional prices** is a fixed collection of pairs $(\vec{Z}, \vec{P})$, where each $\vec{Z} \in [0, 1]^n$ is called an allocation rule, and each $\vec{P} \in \mathbb{R}^n_+$ is called a pricing rule. The bidder selects at most one pair $(\vec{Z}, \vec{P})$. After his choice has been made, the type i of the item is revealed. Given item type i, the bidder pays price $P(i)$, and receives the item with probability $Z(i)$.

A few comments about the mechanism are in order. Note that the allocation probability and price may both depend on the realized type of the item. So, one can think of the mechanism as requiring a single round of bidding, followed by a single round of information revelation (in fact, full information revelation), to determine which price $P(i)$ the bidder should pay. Additionally, note that the bidder pays regardless of whether he receives the item. Finally, note that conditional price mechanisms are strictly more general than item-type partition mechanisms. Item-type partitioning is, in fact, an instantiation of menus with conditional prices in which each $\vec{Z} \in \{0, 1\}^n$ (no fractional or probabilistic allocations are allowed), each item type is offered in exactly one option, and the conditional prices within an option are all identical. Each option then corresponds to a single subset of the partition.

Despite allowing prices and allocations to depend on the realization of the item type, the conditional price menus guarantee interim individual rationality, defined as follows.

Definition 6. A mechanism is **interim individually-rational** (interim IR) if and only if the bidder's expected utility from participating in the mechanism, conditional on a valuation $\vec{v}$ and posterior beliefs $\vec{\pi}$ over item types, is non-negative.

Interim IR can be seen as the bidder committing to an option from the menu offered by the mechanism. One justification for this notion is that a bidder might, in theory, be engaged in many auctions simultaneously. Therefore, the bidder might care only about his average payoff across multiple purchases. While for some type realizations such a bidder may lose, with high probability his overall utility is non-negative.

The following lemma shows that restricting design to conditional price menu mechanisms is without loss of generality.

Lemma 2. *For any equilibrium of any mechanism $\mathcal{M}$ in the presence of a data provider, such that the bidder, conditioned on the realization of her valuation vector and posterior beliefs over item types, obtains non-negative payoff in expectation, there is a conditional price menu that is incentive compatible, interim individually rational, and provides the same revenue.*

Lemma 2 implies the optimal revenue is given by the solution of a linear program whose size is proportional to the number of possible pairs of value vectors and posteriors (this number may be infinite). We make use of this linear program in Section 6. Additionally, note that Lemma 2 can easily be extended to the multi-bidder setting.

Proof. We treat the pair $(\vec{v}, \vec{\pi})$, where v is the valuation of the bidder, and π is his posterior, as the bidder's type. We follow the same steps as the proof of Theorem 1 and Appendix A of Daskalakis et al. [2016]. Consider a mechanism $\mathcal{M}$ with voluntary participation. $\mathcal{M}$ may use multiple rounds of communication and information revelation to the bidder. For each valuation vector $\vec{v}$ and posterior π_s, let $A(\vec{v}, \vec{\pi_s})$ be the (possibly randomized) equilibrium strategy of the bidder when his type is $(\vec{v}, \vec{\pi_s})$.

Let $Z(i, A)$ be an indicator random variable that indicates whether the bidder gets the item when he choses strategy A and the realized item type is i. Similarly, let $C(i, A)$ denote the price the bidder is asked to pay. The bidder's interim expected utility is then given as follows:

$$\mathbb{E}_{i \sim \pi_s}\left[\mathbb{E}\left[Z(i, A) \cdot V(i) - C(i, A)\right]\right]$$

where the first (outer) expectation is with respect to the randomness of the item type, while the second (inner) expectation is with respect to the randomness in the choices of the mechanism, the information revealed and the actions A of the bidder.

For all possible types $(\vec{v}, \vec{\pi_s})$, and for all possible misreports $\left(\vec{v'}, \vec{\pi_{s'}}\right)$ of the bidder, for A to be an equilibrium strategy it must be the case that

$$\begin{aligned} &\mathbb{E}_{i \sim \pi_s}\left[\mathbb{E}\left[Z\left(i, A\left(\vec{v}, \vec{\pi_s}\right)\right) V(i) - C\left(i, A\left(\vec{v}, \vec{\pi_s}\right)\right)\right]\right] \\ &\geq \mathbb{E}_{i \sim \pi_s}\left[\mathbb{E}\left[Z\left(i, A\left(\vec{v'}, \vec{\pi_{s'}}\right)\right) V(i) - C\left(i, A\left(\vec{v'}, \vec{\pi_{s'}}\right)\right)\right]\right]. \end{aligned}$$

Now define the variables $z_i(\vec{v}, \vec{\pi_s}) = \mathbb{E}[Z(i, A(\vec{v}, \vec{\pi_s}))]$ and $c_i(\vec{v}, \vec{\pi_s}) = \mathbb{E}[C(i, A(\vec{v}, \vec{\pi_s}))]$. The equation above can be rewritten as

$$\sum_i \pi_s(i)\left(z_i(\vec{v}, \vec{\pi_s}) V(i) - c_i(\vec{v}, \vec{\pi_s})\right) \geq \sum_i \pi_s(i)\left(z_i\left(\vec{v'}, \vec{\pi_{s'}}\right) V(i) - c_i\left(\vec{v'}, \vec{\pi_{s'}}\right)\right). \tag{IC}$$

Moreover, since the equilibrium A respects voluntary participation, the bidder's equilibrium payoff must be non-negative. As a consequence, we have

$$\sum_i \pi_s(i)\left(z_i(\vec{v}, \vec{\pi_s}) V(i) - c_i(\vec{v}, \vec{\pi_s})\right) \geq 0. \tag{IR}$$

Finally, we note that the revenue of the seller is given by

$$R = \sum_{\vec{\pi_s}, \vec{v}} \mathbb{P}[\vec{v}, \vec{\pi_s}] \sum_i \pi_s(i) \cdot c_i(\vec{v}, \vec{\pi_s}),$$

where $\mathbb{P}[\vec{v}, \vec{\pi_s}]$ is the probability the realized type of the bidder is $(\vec{v}, \vec{\pi_s})$.

A mechanism that satisfies constraints (IC) and (IR) and yields revenue R can clearly be implemented as an interim IR menu with conditional prices, in which the options are given by $(\vec{z}(\vec{\pi_s}, \vec{v}), \vec{c}(\vec{v}, \vec{\pi_s}))$ for each possible type $(\vec{\pi_s}, \vec{v})$. Hence, there exists an incentive compatible, individually rational, conditional price menu that provides the same revenue as mechanism $\mathcal{M}$. $\square$

4 Simple mechanisms achieve a logarithmic fraction of the optimal revenue: a warm-up example

Our main results focus on bounding the revenue achievable via simple mechanisms, in the presence of a third party data provider. In this section, we focus on "simple" mechanisms in which the seller runs the better of item-type pricing and item-type bundling. These are particularly interesting mechanisms to consider given Proposiation 1, where Daskalakis et al. [2016], using results of Babaioff et al. [2014], show that this style of mechanism obtains a constant fraction of the optimal revenue when a data provider is not present. To show that this is not the case when a data provider is present, we consider the following example.

Example 1. Let $n = m^2$ be the number of item types, for some integer m. The types are partitioned into m groups $I_1, \ldots, I_m$ such that each group contains exactly m types. The bidder's prior on the item type is uniform, i.e., the bidder initially believes that each item type is realized with probability $1/n = 1/m^2$, and that the probability that the realized type belongs to group I_k is therefore $1/m$. The bidder's valuation for type i in group I_k is $V(i)/k$, where $V(i)$ is a random variable drawn from the equal revenue distribution. The bidder's valuations for different item types are drawn independently.

We consider a data provider that releases a signal S revealing to the bidder the group to which the item belongs. The bidder's posterior probability on the item being of type i, upon observing signal s_k informing him that the group is I_k, is given by

$$\pi_{s_k}(i) = \begin{cases} 0 & i \notin I_k \\ \frac{1}{m} & i \in I_k \end{cases}.$$

Theorem 3. *There exists a single seller, single bidder, single item (taking one of n item types) setting where, in the presence of a data provider who signals information about the item type realization to the bidder, the expected revenue of the better of item-type pricing and item-type bundling is a $O\left(\frac{1}{\log n}\right)$ fraction of the expected revenue of the optimal mechanism.*

The proof of this theorem follows from Claims 1, 2, and 3, which we prove below.

Claim 1. *The expected revenue from item-type pricing in Example 1 is $O\left(\frac{\log n}{\sqrt{n}}\right)$.*

Claim 2. *The expected revenue from item-type bundling in Example 1 is $O\left(\frac{\log n}{\sqrt{n}}\right)$.*

Claim 3. *There exists an item-type partition mechanism that achieves expected revenue $\Omega\left(\frac{\log^2 n}{\sqrt{n}}\right)$ in Example 1. The optimal revenue in Example 1 is $\Theta\left(\frac{\log^2 n}{\sqrt{n}}\right)$.*

Proof of Claim 1. In item-type pricing, the seller announces the item type and then offers a price that is a function of the realized item type. The expected revenue of such a mechanism is simply given by

$$\frac{1}{n}\sum_{k=1}^{m}\frac{m}{k} = O\left(\frac{\log n}{\sqrt{n}}\right),$$

as the expected revenue from selling an item in the kth group is $\frac{1}{k}$. □

Proof of Claim 2. Let P^* be the optimal bundling price, and suppose the data provider announces signal s_k. There are two cases:

1. For k such that $P^* \geq \frac{6}{k}\log m$, by Lemma 1, the expected revenue is

$$P^* \cdot \mathbb{P}\left[\frac{1}{m}\sum_{i\in|I_k|}\frac{V(i)}{k} \geq P^*\right] = P^* \cdot \mathbb{P}\left[\frac{1}{m}\sum_{i\in|I_k|}V(i) \geq kP^*\right] \leq P^* \cdot \frac{9}{kP^*} = \frac{9}{k},$$

   as $|I_k| = m$.

2. Otherwise, we have k such that $P^* \leq \frac{6}{k}\log m$.

Letting $k^* = \min\{k : \; P^* > \frac{6}{k}\log m\}$, we see that the expected revenue of charging price P^* for the grand bundle is upper-bounded by

$$\begin{aligned}\frac{1}{m}\left(\sum_{k\geq k^*}\frac{9}{k} + \sum_{k<k^*}P^*\right) &\leq \frac{1}{\sqrt{n}}\left(9\cdot(1+\log m) + \sum_{k<k^*}\frac{6}{k^*-1}\log m\right)\\ &= \frac{1}{\sqrt{n}}\left(9\cdot(1+\log m) + 6\log m\right)\\ &= O\left(\frac{\log n}{\sqrt{n}}\right).\end{aligned}$$

□

Proof of Claim 3. Consider the following item-type partition mechanism: the seller first partition the item types into m groups in the same way as specified in Example 1. When the realized item type is in group I_k, she offers to sell the item to the bidder at price $P_k = \frac{\log m}{2k}$.

If the bidder receives signal s_k, then the price offered by the seller must be $\frac{\log m}{2k}$, and the bidder knows the item type is from group I_k. By Lemma 1, as $|I_k| = m$, we have:

$$\mathbb{P}\left[\frac{1}{m}\sum_{i\in I_k}\frac{V(i)}{k} \geq \frac{\log m}{2k}\right] \geq \frac{1}{2},$$

and hence with probability at least 1/2, conditional on $S = s_k$, he accepts the price, yielding expected revenue to the seller of at least $\frac{\log m}{4k}$. The total expected revenue for the seller is then given by

$$\frac{1}{m}\sum_{k=1}^{m}\frac{\log m}{4k} = \frac{\log n}{8\sqrt{n}}\sum_{k=1}^{m}\frac{1}{k} = \Omega\left(\frac{\log^2 n}{\sqrt{n}}\right).$$

A simple combination of Daskalakis et al. [2016], Li and Yao [2013] show that no truthful mechanism can achieve revenue higher than $\log m$ times the revenue of item-type pricing conditioned on receiving signal s_k, as selling separately achieves at least a $\Omega\left(\frac{1}{\log m}\right)$ fraction of the optimal revenue for selling m items. Thus, the optimal revenue is at most $O\left(\frac{\log^2 n}{\sqrt{n}}\right)$, and hence the item-type partition mechanism we just described yields a constant approximation to the optimal revenue. □

5 Item-type partitioning achieves a Log-log fraction of the optimal revenue

The previous section shows that neither item-type pricing nor item-type bundling can achieve a constant fraction of the optimal revenue in our setting. However, one may wonder if the result is due to the restrictive nature of the "simple" mechanisms considered. Here, we show that, in the presence of a data provider, even the more general class of item-type partition mechanisms is insufficient to guarantee a constant fraction of the optimal revenue. This is particularly tantalizing due to the fact that Example 1 admits a item-type partition mechanism that yields a constant approximation to the optimal revenue. However, in this section, we show an example where the best item-type partition mechanism only achieves a $O\left(\frac{1}{\log\log n}\right)$ fraction of the optimal revenue (and hence our "simpler" simple mechanisms, item-type bundling and item-type pricing, also do not yield a constant fraction of the optimal revenue, since they are special cases of item-type partitioning).

We consider the following class of examples.

Example 2. Given an integer m, let $n = 2^m$ be the number of item types. The bidder's prior on the item type is uniform, i.e., the bidder initially believes the item type takes each $i \in [n]$ with probability $1/n$. The bidder's valuation for each type is drawn i.i.d. from an equal revenue distribution.

We consider m possible partitions of the n items. Given a particular $k \in [m]$, we partition the set of all items into $n_k = 2^{m-k}$ subsets of size $2^k \geq 2$ each. Specifically, for $k \in [m]$ we partition the set of items into the subsets $I_{k,1}$ to I_{k,n_k}, where $I_{k,j} = \{(j-1)\cdot 2^k + 1, \dots j \cdot 2^k\}$ for all $j \in [n_k]$.

The signaling scheme used by the data provider is as follows. The data provider picks a value $k \in [m]$ according to the following distribution: for $k \leq m-1$, the provider picks k with probability $\frac{1}{k(k+1)}$; the provider picks m with the remaining probability $\frac{1}{m}$. The value k is drawn, importantly, independently of the type i of the item. The data provider's signal reveals to the bidder in which group (among $I_{k,1}$ to I_{k,n_k}) the item lies. We denote by $s_{k,j}$ the realization of the signal that indicates to the bidder that the item belongs to group $I_{k,j}$. We call k the size indicator.

Theorem 4. *There exists a single seller, single bidder, single item (taking one of n item types) setting where, in the presence of a data provider who signals information about the item type realization to the bidder, no item-type partition mechanism can achieve revenue higher than $O\left(\frac{1}{\log\log n}\right)$ of the optimal revenue.*

We prove the result via Claims 4 and 5 below.

Claim 4. *The expected revenue from the optimal item-type partition mechanism is $O(1)$ in Example 2.*

Claim 5. *There exists a mechanism that yields revenue $\Omega(\log\log n)$ in Example 2. The optimal revenue in Example 2 is $\Theta(\log\log n)$.*

Proof of Claim 4. Suppose the signal is $s_{k,j}$ for some $j \in [n_k]$. Suppose the item-type partition mechanism splits the item types into non-empty groups $\mathcal{G}_1$ to $\mathcal{G}_g$, where $g \leq 2^m$ is the number of such groups. If the seller offers to sell the item at price at P_r, then the item type is uniform over $\mathcal{G}_r \cap \mathcal{I}_{k,j}$ in the bidder's posterior. Note that $|\mathcal{G}_r \cap \mathcal{I}_{k,j}| \leq 2^k$. By Lemma 1, we have

$$P_r \cdot \mathbb{P}\left[\frac{1}{|\mathcal{G}_r \cap \mathcal{I}_{k,j}|}\sum_{t\in\mathcal{G}_r\cap\mathcal{I}_{k,j}} V(t) \geq P_r\right] \leq \begin{cases} 9 & \text{if } P_r \geq 6\log\left(2^k\right) = 6k\log 2 \\ P_r & \text{if } P_r < 6k\log 2, \end{cases},$$

following from $6\log\left(2^k\right) \geq 6\log|\mathcal{G}_r \cap \mathcal{I}_{k,j}|$.

Let $k^*(r) = \max\{k : \ P_r \geq 6k\log 2\}$. Further, let us denote by $\mathbb{P}[k]$ the probability that the data provider selects a partition of size 2^k. We have that the expected revenue from posting price P_r for every $r \in [g]$ is upper-bounded by

$$\begin{aligned} 9\sum_{k\leq k^*(r)} \mathbb{P}[k] + P_r \cdot \left(\sum_{k=k^*(r)+1}^{m} \mathbb{P}[k]\right) &= 9\sum_{k\leq k^*(r)} \frac{1}{k(k+1)} + P_r \cdot \left(\sum_{k=k^*(r)+1}^{m-1} \frac{1}{k(k+1)} + \frac{1}{m}\right) \\ &\leq 9\left(1 - \frac{1}{k^*(r)+1}\right) + 6\log 2 \cdot (k^*(r)+1)\left(\frac{1}{k^*(r)+1} - \frac{1}{m} + \frac{1}{m}\right) \\ &\leq 9 + 6\log 2, \end{aligned}$$

where the first step follows the fact that the probability of the data provider selecting a $k \leq m-1$ is $\frac{1}{k(k+1)}$, and the probability of him drawing $k = m$ is $\frac{1}{m}$. This concludes the proof. ☐

To prove Claim 5, we first construct a mechanism that achieves revenue $\Omega(\log\log n)$ in Example 2. In particular, we consider the following design.

Mechanism 1. *The seller offers a menu of $\sum_{k=1}^m n_k$ options. For every $\kappa \in [m]$, and every $\iota \in [n_\kappa]$, the menu contains the following option $L_{\kappa,\iota}$: the bidder first pays $P_\kappa = \frac{1}{8}\log 2^\kappa = \frac{\log 2}{8}\kappa$, then gets the item if and only if it is in group $I_{\kappa,\iota}$. Note that the price only depends on κ.*

To show that Mechanism 1 yields revenue $\Omega(\log\log n)$ in Example 2, we need the following claim, which characterizes the bidder's behavior in the mechanism. More specifically, we show in the following claim that if the bidder receives signal $s_{k,j}$, he purchases the corresponding option $L_{k,j}$ in Mechanism 1 with probability almost 1.

Claim 6. *In the the setting of Example 2, suppose the bidder receives signal $s_{k,j}$ (indicating that the item belongs to group $I_{k,j}$ of size 2^k) for $k \geq 2 \cdot 10^2 + 1$. Consider the menu of options proposed by the seller in Mechanism 1. With probability at least $1 - 10^{-3}$, no option $L_{\kappa,\iota}$ with either $\kappa \neq k$ or $\iota \neq j$ yields a higher utility for the bidder than option $L_{k,j}$, and $L_{k,j}$ yields positive utility to the bidder.*

Proof of Claim 6. The bidder's expected utility for $L_{k,j}$ when receiving signal $s_{k,j}$ is given by

$$U_{k,j} = \frac{1}{2^k} \sum_{i \in I_{k,j}} V(i) - \frac{1}{8} \log 2^k,$$

his expected utility for selecting option $L_{\kappa,\iota}$ for $\kappa > k$ is only less (his expected value for the item type is not more, but the price is higher), and his utility for selecting option $L_{\kappa,\iota}$ for $\kappa < k$ is

$$U_{\kappa,\iota} = \frac{1}{2^k} \sum_{i \in I_{\kappa,\iota} \cap I_{k,j}} V(i) - \frac{1}{8} \log 2^\kappa,$$

and his expected utility for selecting any option $L_{\kappa,\iota}$ such that $I_{\kappa,\iota} \cap I_{k,j} = \emptyset$ is negative, since he will pay but never be allocated the item.

Therefore, the bidder prefers $L_{\kappa,\iota}$ to $L_{k,j}$ with $\kappa \leq k$ and $I_{\kappa,\iota} \subset I_{k,j}$ only if

$$\frac{1}{2^k} \sum_{i \in I_{k,j} \setminus I_{\kappa,\iota}} V(i) \leq \frac{1}{8} \log 2^{k-\kappa}.$$

We want to upper bound the probability of the above event for all $\kappa < k$ and $I_{\kappa,\iota} \subset I_{k,j}$. Let us denote $V^M(i) = \min(V(i), M)$ for any M. We have immediately that $\mathbb{E}\left[V^M(i)\right] = \log M + 1$ and that its variance is upper-bounded by $2M$. Taking $M = 2^{k-1}$ and $X(i) = \log M + 1 - V^M(i)$ yields $|X(i)| \leq M$, $\mathbb{E}\left[X(i)\right] = 0$ and $\mathbb{E}\left[X(i)^2\right] \leq 2 \cdot 2^{k-1} = 2^k$. Recall Bernstein's inequality:

Lemma 5. *(Bernstein's Inequality): Suppose $X_1, ..., X_n$ are independent random variables with zero mean, and $|X_i| \leq B$ almost surely for all i. Then for any $t > 0$,*

$$\Pr\left[\sum_{i=1} X_i > t\right] \leq exp\left(-\frac{\frac{1}{2}t^2}{\sum_{i=1}^n E[X_i^2] + \frac{1}{3}Bt}\right)$$

We can then apply Bernstein's inequality to show that

$$\begin{aligned}
\mathbb{P}\left[\frac{1}{2^k} \sum_{i \in I_{k,j} \setminus I_{\kappa,\iota}} V(i) < \frac{1}{2^k} \cdot \left(\sum_{i \in I_{k,j} \setminus I_{\kappa,\iota}} (\log M + 1) - t\right)\right] &= \mathbb{P}\left[\sum_{i \in I_{k,j} \setminus I_{\kappa,\iota}} V(i) < \sum_{i \in I_{k,j} \setminus I_{\kappa,\iota}} (\log M + 1) - t\right] \\
&\leq \mathbb{P}\left[\sum_{i \in I_{k,j} \setminus I_{\kappa,\iota}} V^M(i) < \sum_{i \in I_{k,j} \setminus I_{\kappa,\iota}} (\log M + 1) - t\right] \\
&= \mathbb{P}\left[\sum_{i \in I_{k,j} \setminus I_{\kappa,\iota}} X(i) > t\right] \\
&\leq \exp\left(-\frac{1}{2} \cdot \frac{t^2}{2^k \cdot |I_{k,j} \setminus I_{\kappa,\iota}| + M \cdot t/3}\right) \\
&= \exp\left(-\frac{1}{2} \cdot \frac{t^2}{2^k\left(2^k - 2^\kappa\right) + M \cdot t/3}\right),
\end{aligned}$$

where the last inequality is due to Bernstein's inequality. Taking

$$t = \left(\frac{3}{4}\right)\left(2^k - 2^\kappa\right)(\log M + 1),$$

we have

$$\begin{aligned}
\frac{1}{2^k}\left(\sum_{i\in I_{k,j}\setminus I_{\kappa,\iota}} (\log M + 1) - t\right) &= \frac{1}{2^k}\cdot\frac{1}{4}\left(2^k - 2^\kappa\right)(\log M + 1)\\
&\geq \frac{1}{2^k}\cdot\frac{1}{4}\left(2^{k-1}\right)(\log M + 1)\\
&= \frac{1}{8}(\log M + 1),
\end{aligned}$$

and we thus obtain a bound on the probability of the event that a particular menu option $L_{\kappa,\iota}$ for $\kappa < k$ is better for the bidder than option $L_{k,j}$, given signal $s_{k,j}$:

$$\begin{aligned}
\mathbb{P}\left[\frac{1}{2^k}\sum_{i\in I_{k,j}\setminus I_{\kappa,\iota}} V(i) \leq \frac{1}{8}\log 2^{k-\kappa}\right] &< \mathbb{P}\left[\frac{1}{2^k}\sum_{i\in I_{k,j}\setminus I_{\kappa,\iota}} V(i) < \frac{1}{8}\left(\log 2^{k-1} + 1\right)\right]\\
&\leq \mathbb{P}\left[\frac{1}{2^k}\sum_{i\in I_{k,j}\setminus I_{\kappa,\iota}} V(i) < \frac{1}{2^k}\left(\sum_{i\in I_{k,j}\setminus I_{\kappa,\iota}}\left(\log 2^{k-1} + 1\right) - t\right)\right]\\
&\leq \exp\left(-\frac{k^2}{2}\cdot\frac{(3/4)^2\left(2^k - 2^\kappa\right)^2(\log 2)^2}{2^k\left(2^k - 2^\kappa\right) + \frac{1}{4}2^{k-1}\left(2^k - 2^\kappa\right)\left(\log 2^{k-1} + 1\right)}\right)\\
&\leq \exp\left(-\frac{(k-1)^2}{2}\cdot\frac{(3/4)^2\cdot 2^{k-1}\left(2^k - 2^\kappa\right)(\log 2)^2}{2^k\left(2^k - 2^\kappa\right) + \frac{1}{4}2^{k-1}\left(2^k - 2^\kappa\right)\left(\log 2^{k-1} + 1\right)}\right)\\
&\leq \exp\left(-\frac{k-1}{2}\cdot\frac{(3/4)^2(\log 2)^2}{\frac{2}{k-1} + \frac{1}{4}\left(\log 2 + \frac{1}{k-1}\right)}\right),
\end{aligned}$$

For $k \geq 2\cdot 10^2 + 1$, the above yields

$$\mathbb{P}\left[\frac{1}{2^k}\sum_{i\in I_{k,j}\setminus I_{\kappa,\iota}} V(i) < \frac{1}{8}\left(\log 2^{k-1} + 1\right)\right] \leq \exp\left(-(k-1)\cdot\frac{(3/4)^2(\log 2)^2}{\frac{4}{2\cdot 10^2} + \frac{1}{2}\left(\log 2 + \frac{1}{2\cdot 10^2}\right)}\right).$$

We now let

$$K = \exp\left(\frac{(3/4)^2(\log 2)^2}{\frac{4}{2\cdot 10^2} + \frac{1}{2}\left(\log 2 + \frac{1}{2\cdot 10^2}\right)}\right),$$

and note that we then have that for $k \geq 2\cdot 10^2 + 1$,

$$\mathbb{P}\left[\frac{1}{2^k}\sum_{i\in I_{k,j}\setminus I_{\kappa,\iota}} V(i) < \frac{1}{8}\left(\log 2^{k-1} + 1\right)\right] \leq \left(\frac{1}{K}\right)^{k-1}.$$

Since there are less than 2^k groups $I_{\kappa,\iota}$ such that $I_{\kappa,\iota} \subset I_{k,j}$, a union bound gives us that the probability that the bidder prefers a different option other than $L_{k,j}$ is upper bounded by $2\cdot\left(\frac{2}{K}\right)^{k-1}$. A direct calculation shows that $2\cdot\left(\frac{2}{K}\right)^{k-1} \leq 10^{-3}$. □

We are now ready to prove Claim 5.

Proof of Claim 5. The proof of Claim 6 directly implies that the revenue of the considered mechanism is lower-bounded by

$$\frac{(1-10^{-3})\log 2}{8}\left(\sum_{k\geq 2\cdot 10^2+1}^{m-1}\frac{k}{k(k+1)}+\frac{m}{m}\right)=\Omega\left(\log m\right)=\Omega\left(\log\log n\right),$$

as a bidder who receives signal $s_{k,j}$ picks option $L_{k,j}$ with price $\frac{\log 2}{8}k$ with probability at least $1-10^{-3}$.

The revenue of the best mechanism is upper-bounded by the optimal revenue the seller could obtain if she knew the realization of the signal. When facing signal $s_{k,j}$, the bidder's posterior is that the item type is taken uniformly at random from group $I_{k,j}$. By Babaioff et al. [2014], Daskalakis et al. [2016], the better of item-type pricing and item-type bundling (conditioning now on the realization of the signal) yields a constant approximation to the optimal revenue. The revenue from item-type pricing is clearly 1, and the revenue from item-type bundling is $O\left(\log 2^k\right)$ by Lemma 1 as setting $P>6\log 2^k$ yields constant revenue while setting $P\leq 6\log 2^k$ yields $O\left(\log 2^k\right)$. Therefore, the optimal revenue conditional on the signal being $s_{k,j}$ must be $O\left(\log 2^k\right)=O\left(k\right)$, and the optimal (unconditional) revenue is therefore

$$O\left(\sum_{k=1}^{m-1}\frac{k}{k(k+1)}+\frac{m}{m}\right)=O\left(\log m\right)=O\left(\log\log n\right).$$

□

We remark that Mechanism 1, although it has a concise description, is not "simple" in any of the usual senses, and is in fact carefully tailored to the incentives of the buyer. We do not know of "simpler" mechanisms that are approximately optimal in this setting.

6 The Behavior of an Adversarial Data Provider

Thus far, our focus has been on the impact of third-party information on the mechanism design problem of the *seller*, absent consideration or modeling of the incentives or behavior of the data provider. Here, we shift gears and consider the perspective of the data provider. Specifically, we consider the *strategic design of the signaling scheme of the data provider*.

To accomplish this, we need to consider the motivations of the data provider. One may consider the data provider to be profit maximizing, of course, but a particularly relevant behavior in the context of the results of the previous sections is to consider that the data provider may be *adversarial*, that is, he wishes to minimize the expected revenue of the seller. Understanding the behavior of an adversarial data provider is important for understanding how to design mechanisms in the presence of a data provider that are guaranteed, in the worst-case, to provide near-optimal revenue.

The question we seek to answer in this section is: what signaling structure would an adversarial data provider adopt?

To study this question, we consider a third-party data provider that has (potentially imperfect) information on the type of the item, in the form of some $X\in\mathcal{X}$. X is a random variable that can be arbitrarily correlated with the type of the item; its distribution is publicly known, but its realization is only observed by the data provider. The data provider publicly announces a signaling scheme, a random function $S : \mathcal{X} \to \mathcal{S}$, and reveals the realization of $S(X)$ solely to the bidder; the seller does not observe this realization, but knows the distribution of $S(.)$. The seller then sells the item so as to maximize her revenue, under the knowledge of the distribution of the signal X. The data provider's goal is to design the signaling scheme $S(.)$ so as to minimize the seller's revenue.

A first observation about this situation is that, when the data provider has perfect information about the item type, she minimizes the expected revenue of the seller by revealing the type.

Lemma 6. *If the data provider is adversarial and has full information about the type of the item (that is, if X is perfectly correlated with the item type), the optimal strategy for the data provider is to reveal X.*

Proof. Let $SREV$ be the optimal revenue that the seller can achieve when the type of the item is revealed, and $REV(S(X))$ be the optimal revenue achievable when the data provider's signal is the random variable $S(X)$. The seller can always guarantee a revenue of $SREV$ by revealing the type of the item then selling this type optimally, no matter what the distribution of $S(X)$ is, thus $REV(S(X)) \geq SREV$. The data provider can always set $S(.)$ to be the identity function I_d, which is equivalent to revealing X and hence the item type. Because $SREV$ is the optimal revenue when the item type is revealed, $REV(X) = SREV$, which concludes the proof. □

More interestingly, and perhaps counter-intuitively, if the data provider does not have full information, then only *partially* revealing information may minimize the revenue of the seller.

Lemma 7. *Let the number of item types be $n = 2$, and let $REV(S(X))$ be the optimal revenue for the seller when the data provider receives signal X and designs signaling scheme $S(.)$. There exists a distribution of valuations $\vec{V}$ and a prior $\vec{\pi}$ for the bidder, and a signaling scheme $S(.)$ such that $REV(S(X)) < REV(I_d(X))$ where I_d is the identity function, that is, fully revealing X is not the strategy for the provider that minimizes the expected revenue of the seller.*

The proof of this result uses the following example.

Example 3. Let the bidder's valuation for each item type be drawn i.i.d., taking value 1 with probability 1/2 and value 2.1 with probability 1/2. The bidder and the data provider share a common prior $\vec{\pi} = (3/4, 1/4)$. That is, they both initially believe the item is of type 1 with probability 3/4 and of type 2 with probability 1/4. The data provider receives a signal X on some support $\{x_1, x_2\}$. If the item is 1, the provider receives x_1 with probability 2/3 and x_2 with probability 1/3, and if the item is x_2, the provider receives x_2 with probability 1. Therefore, with probability 1/2, the data provider receives x_1 and has posterior $\vec{\pi_1} = (1, 0)$ (when receiving x_1, the provider knows the item must be 1); with probability 1/2, he receives x_2 and thus has posterior $\vec{\pi_2} = (1/2, 1/2)$.

Proof. We show that in Example 3, the data provider has a signaling scheme that is worse for the seller than fully revealing his information to the bidder.

If the data provider fully reveals his information X to the bidder, he induces the same posteriors $\vec{\pi_1}$ w.p. 1/2 and $\vec{\pi_2}$ w.p. 1/2 for the bidder. Computing the optimal revenue of the seller via linear programming, using the characterization of Section 3, yields revenue $R = 1.1062$.

Suppose the data provider instead reveals partial information to the bidder, in the following manner: when the provider receives a realization x of X that belongs to $\{x_1, x_2\}$, the provider outputs $S(X) = x$ w.p. $1 - \epsilon = 0.86$ and, for some distinct $x_3 \in \mathcal{S}$, $S(X) = x_3$ w.p. $\epsilon = 0.14$. Given this signaling scheme, when the bidder receives $S(X) = x_1$ (which occurs with probability $\frac{1}{2}(1 - \epsilon) = 0.43$), he infers that necessarily $X = x_1$, and so his posterior is $\vec{\pi_1}$. Similarly, when

$S(X) = x_2$ (which also occurs with probability 0.43), the bidder infers that it must be the case that $X = x_2$, hence he has posterior $\vec{\pi_2}$. Finally, when the bidder receives $S(X) = x_3$ (which occurs with probability $\epsilon = 0.14$), he infers that $X = x_1$ or $X = x_2$ with equal probability by symmetry, hence his posterior is $\frac{1}{2}(\vec{\pi_1} + \vec{\pi_2}) = (3/4, 1/4) = \vec{\pi}$. Computing the optimal revenue of the seller via linear programming, using the the results of Section 3, yields revenue $R = 1.0991 < 1.1062$, proving the result. □

It may seem counterintuitive that the information provider can harm the seller *more* by providing *less* information to the bidder. After all, one consequence of the characterization of Section 3 is that the seller can only lower her revenue by revealing more information to the bidder. However, information from the provider and information from the seller are not equivalent from the perspective of the seller, because the seller does not get to see the realization of the signal that the provider sends to the bidder. When the seller reveals information, she knows exactly what the bidder's posterior is, and can act as a function of the realized posterior; she is then faced with exactly the problem solved by Daskalakis et al. [2016] for that realized posterior. When the data provider reveals information, the seller, who only knows the signaling scheme but not the signal, faces a distribution of posteriors and doesn't know which of them is correct.

In particular, in Example 3, in the fully-revealing signaling scheme there are two posteriors $\vec{\pi_1}$ and $\vec{\pi_2}$. Each occurs with probability 1/2. The seller, intuitively, wishes to design a menu with one option for each of these two posteriors. In the partially revealing signaling scheme there is a third posterior, $\vec{\pi}$, which is an average of $\vec{\pi_1}$ and $\vec{\pi_2}$. In fact, the first signaling scheme is a mean-preserving spread of the second one. The seller, intuitively, wishes to design a menu with three options, one for each posterior.

This third posterior induces a trade-off in the linear program the seller solves to find the optimal mechanism. The second linear program has more IC constraints for the two posteriors than the linear program given the fully revealing signaling scheme. This makes the revenue the seller gets from buyers with posteriors $\vec{\pi_1}$ and $\vec{\pi_2}$ lower than before. The trade-off is that there is now a new posterior $\vec{\pi_3}$, from which the seller can make additional revenue. Example 3 is constructed so that the harm from the additional posterior exceeds the benefit.

Acknowledgements Cai thanks the NSERC for its support through the Discovery grant RGPIN-2015-06127 and FRQNT for its support through the grant 2017-NC-198956. Echenique thanks the National Science Foundation for its support through grants SES-1558757 and CNS-1518941. Fu thanks the NSERC for its support through Discovery grant RGPAS-2017-507934 and Accelerator grant RGPAS-2017-507934. Ligett's work was supported in part by NSF grants CNS-1254169 and CNS-1518941, US-Israel Binational Science Foundation grant 2012348, Israeli Science Foundation (ISF) grant 1044/16, a subcontract on the DARPA Brandeis Project, and the HUJI Cyber Security Research Center in conjunction with the Israel National Cyber Directorate (INCD) in the Prime Ministers Office. Wierman thanks the National Science Foundation for its support through grants NSF AitF-1637598, CNS-1518941, as well as the Linde Institute of Economic and Management Science at Caltech. Ziani thanks the National Science Foundation for its support through grants CNS-1331343 and CNS-1518941, and the US-Israel Binational Science Foundation through grant 20122348. We thank Noam Nisan for extremely useful comments and discussions.

References

Moshe Babaioff, Nicole Immorlica, Brendan Lucier, and S. Matthew Weinberg. A simple and approximately optimal mechanism for an additive buyer. *CoRR*, abs/1405.6146, 2014. URL http://arxiv.org/abs/1405.6146.

Yang Cai and Zhiyi Huang. Simple and Nearly Optimal Multi-Item Auctions. In *the 24th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2013.

Yang Cai and Mingfei Zhao. Simple mechanisms for subadditive buyers via duality. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*, pages 170–183, 2017. doi: 10.1145/3055399.3055465. URL http://doi.acm.org/10.1145/3055399.3055465.

Yang Cai, Nikhil R. Devanur, and S. Matthew Weinberg. A duality based unified approach to bayesian mechanism design. In *the 48th Annual ACM Symposium on Theory of Computing (STOC)*, 2016.

Shuchi Chawla and J. Benjamin Miller. Mechanism design for subadditive agents via an ex-ante relaxation. In *Proceedings of the 2016 ACM Conference on Economics and Computation, EC '16, Maastricht, The Netherlands, July 24-28, 2016*, pages 579–596, 2016. doi: 10.1145/2940716.2940756. URL http://arxiv.org/abs/1603.03806.

Yu Cheng, Ho Yee Cheung, Shaddin Dughmi, Ehsan Emamjomeh-Zadeh, Li Han, and Shang-Hua Teng. Mixture selection, mechanism design, and signaling. In *IEEE 56th Annual Symposium on Foundations of Computer Science, FOCS 2015, Berkeley, CA, USA, 17-20 October, 2015*, pages 1426–1445, 2015.

Constantinos Daskalakis, Christos Papadimitriou, and Christos Tzamos. Does information revelation improve revenue? In *Proceedings of the 2016 ACM Conference on Economics and Computation*, EC '16, pages 233–250, New York, NY, USA, 2016. ACM. ISBN 978-1-4503-3936-0. doi: 10.1145/2940716.2940789. URL http://doi.acm.org/10.1145/2940716.2940789.

Shaddin Dughmi, Nicole Immorlica, Ryan O'Donnell, and Li-Yang Tan. Algorithmic signaling of features in auction design. In *Algorithmic Game Theory - 8th International Symposium, SAGT 2015, Saarbrücken, Germany, September 28-30, 2015, Proceedings*, pages 150–162, 2015.

Yuval Emek, Michal Feldman, Iftah Gamzu, Renato Paes Leme, and Moshe Tennenholtz. Signaling schemes for revenue maximization. *ACM Trans. Economics and Comput.*, 2(2):5:1–5:19, 2014. doi: 10.1145/2594564. URL http://doi.acm.org/10.1145/2594564.

FTC. Ftc recommends congress require the data broker industry to be more transparent and give consumers greater control over their personal information, 2014. URL https://www.ftc.gov/news-events/press-releases/2014/05/ftc-recommends-congress-require-data-

Hu Fu, Patrick R. Jordan, Mohammad Mahdian, Uri Nadav, Inbal Talgam-Cohen, and Sergei Vassilvitskii. Ad auctions with data. In *Algorithmic Game Theory - 5th International Symposium, SAGT 2012, Barcelona, Spain, October 22-23, 2012. Proceedings*, pages 168–179, 2012.

Sergiu Hart and Noam Nisan. Approximate revenue maximization with multiple items. *CoRR*, abs/1204.1846, 2012. URL http://arxiv.org/abs/1204.1846.

Sergiu Hart and Noam Nisan. The menu-size complexity of auctions. In *the 14th ACM Conference on Electronic Commerce (EC)*, 2013.

Xinye Li and Andrew Chi-Chih Yao. On revenue maximization for selling multiple independently distributed items. *Proceedings of the National Academy of Sciences*, 110(28):11232–11237, 2013.

Peter Bro Miltersen and Or Sheffet. Send mixed signals: earn more, work less. In *ACM Conference on Electronic Commerce, EC '12, Valencia, Spain, June 4-8, 2012*, pages 234–247, 2012.

Roger Myerson. Optimal auction design. *Mathematics of Operations Research*, 6(1):pp. 58–73, 1981.

Aviad Rubinstein. On the computational complexity of optimal simple mechanisms. In *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science*, ITCS '16, pages 21–28. ACM, 2016. ISBN 978-1-4503-4057-1.

Aviad Rubinstein and S. Matthew Weinberg. Simple mechanisms for a subadditive buyer and applications to revenue monotonicity. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation, EC '15, Portland, OR, USA, June 15-19, 2015*, pages 377–394, 2015. doi: 10.1145/2764468.2764510. URL `http://doi.acm.org/10.1145/2764468.2764510`.

Andrew Chi-Chih Yao. An n-to-1 bidder reduction for multi-item auctions and its applications. In *SODA*, 2015. URL `http://arxiv.org/abs/1406.3278`.