Screening with Persuasion*

Dirk Bergemann[†] Tibor Heumann[‡] Stephen Morris[§]

March 6, 2025

Abstract

We analyze a nonlinear pricing model where the seller controls both product pricing (screening) and buyer information about their own values (persuasion).

We prove that the optimal mechanism always consists of finitely many signals and items, even with a continuum of buyer values. The seller optimally pools buyer values and reduces product variety to minimize informational rents.

We show that value pooling is optimal even for finite value distributions if their entropy exceeds a critical threshold. We also provide sufficient conditions under which the optimal menu restricts offering to a single item.

Jel Classification: D44, D47, D83, D84.

Keywords: Nonlinear Pricing, Screening, Bayesian Persuasion, Finite Menu, Second-Degree Price Discrimination, Recommender System.

---

*We acknowledge financial support from NSF grants SES-2001208 and SES-2049744, and ANID Fondecyt 1241302. We thank seminar audiences at Boston College, Brown University, Emory University, MIT, NBER Market Design, Pontificia Universidad Católica de Chile, University of Chile, the Warwick Economic Theory Workshop and the UCL memorial conference for Konrad Mierendorff for valuable suggestions, and, in particular, Mark Armstrong for his discussion at the latter conference. We benefited from discussions with Ian Ball, Roberto Corrao, Stefano DellaVigna, Piotr Dworczak, Fedor Sandomirskiy, Matt Gentzkow, Anton Kolotolin, Ellen Muir, Leon Musolff, Christopher Sandmann, Philipp Strack, and Alex Wolitzky. We thank Pedro Feijo de Moraes, Jack Hirsch, Hongcheng Li, Shawn Thacker, David Wambach, Michael Wang and Nick Wu for excellent research assistance.

[†]Yale University, dirk.bergemann@yale.edu.

[‡]Pontificia Universidad Católica de Chile, tibor.heumann@uc.cl.

[§]Massachusetts Institute of Technology, semorris@mit.edu

1 Introduction

1.1 Motivation

In the digital age, sellers often possess superior information about the match between products and buyers. This informational advantage, coupled with the ability to offer a variety of products, presents a novel challenge in the design of optimal pricing strategies. How should a seller leverage their informational advantage while determining the optimal variety of products to offer? With fewer products offered the seller can reveal less information to support the choice behavior of the buyer; conversely, more variety requires additional information be provided to all the buyers to distinguish between the product choices.

We analyze the interaction between the menu of items offered and the information provided to the buyers in the classic nonlinear pricing environment of Mussa & Rosen (1978). In our framework, the seller chooses the menu of products and prices *and* controls the information that buyers receive about their own willingness-to-pay. This setting reflects key features of the digital economy: (*i*) sellers' informational advantage, (*ii*) the use of recommendation systems, and (*iii*) the prevalence of menu pricing over personalized pricing.

We go beyond the setting of Mussa & Rosen (1978) by assuming that buyers initially only know the prior distribution of their values. The seller chooses how much information the buyers receive about their willingness-to-pay through the choice of an information structure. We refer to the prior distribution of values therefore as the *latent distribution* as the buyers only observe the expected values through the signals. The seller's problem consists of designing a public menu of products, prices and the information that buyers receive about their value. Thus the seller screens by offering quality-differentiated products while engaging in Bayesian persuasion to influence buyers' decisions–*screening with persuasion*. We maintain the assumption that the seller cannot engage in personalized price discrimination based on the information revealed about their values. We expand on the interpretation of this assumption after we introduce the model.

Our analysis yields several striking results. First, we prove that the optimal mechanism always consists of a finite menu, even when the underlying distribution of buyer values is continuous. This finding contrasts sharply with the continuous menus often found in classical screening models. The finiteness of the optimal menu arises from the seller's ability to pool buyer values, which reduces information rents more effectively than the traditional approach of distorting allocations. By pooling, we refer to any information structure that provides less than complete information to the buyer about their value. This encompasses any deterministic or stochastic information structure that sends the same signal for at least two distinct values.

Second, we establish sufficient conditions for the optimality of pooling. Specifically, we show that pooling is beneficial whenever the entropy of the value distribution exceeds $\log_2(9)$ bits. This result provides a clear, quantifiable threshold for when information design becomes crucial in pricing strategies.

Third, we offer sufficient conditions when a single-item menu is optimal and hence a binary signal that supports the binary choice of the buyer–to buy or not buy the single item on the menu. The sufficient conditions require modest tails of the value distribution and convexity of the marginal cost of quality.

These results together advance our theoretical understanding of nonlinear pricing and also offer practical insights for firms operating in information-rich environments. They suggest that sellers can often improve profits by strategically determining the information available to buyers and offering short, that is finite menus.

The seller reduces the granularity of the buyer's information by pooling the buyers' values into a small set of signals in the optimal mechanism, thereby compressing the information buyers receive. The central trade-off is between a suppression of the information rent that is enabled by a menu with a small variety of products and efficiency gains from a menu with a large variety. This trade-off is established by a variational argument that ultimately leads us to establish the optimality of a finite menu even when the underlying values form a continuum. The basic argument proceeds as follows. Suppose that the optimal mechanism were to display an interval of values in the support of the information structure. We ask what happens to profit if we pool the allocations associated with a small interval of values in that larger interval. By construction, distortions in the allocation from the candidate profit-maximizing allocation only cause second-order distortions to the virtual surplus. But if we additionally pool a small interval of values into a single *expected value,* then this causes a first-order decrease in the information rents. Thus, locally a suppression of variety that is supported by a reduction in information leads to an increase in profit. Hence, screening across a small open set of values is never optimal. This leads to the first main result, Theorem 1, which states that the optimal mechanism only generates a finite number of signals and a finite number of items. In particular, the number of signals received exceeds the number of items by at most one, to account for a signal to those buyers who optimally do not purchase any item.

We first illustrate the trade-off between the variety of the menu and the suppression of private information in an environment with finitely many values and finitely many given qualities. Proposition 1 provides a sufficient condition when pooling the lowest two values generates a profit improvement relative to complete disclosure. The local argument contains the critical ingredients for the general argument and thus provides a preview of the logic of Theorem 1. The sufficient condition also enables us to produce an upper bound on the entropy under which complete disclosure can be optimal (Corollary 1), which previews the more general result in Theorem 2.

We make no assumption about the support of the distribution of buyers' willingness-to-pay. In particular, it could be finite, countably infinite or - as in the classic treatment of Mussa & Rosen (1978) - a continuum. We first establish that the optimal mechanism is discrete (Proposition 3), thus has at most a countable number of items and signals. We then establish that the quality increments in an optimal discrete menu must be increasing, so that item qualities offered in the menu increase in a convex manner (Proposition 4). This help us to establish that the optimal menu always consists of a finite set of items. In the optimal information structure, the cardinality of items sold is always equal to the cardinality of buyers' expected values (or one less if there is exclusion). This result establishes that substantial pooling arises: whatever the true variety of willingness-to-pay in the population, it is optimal to offer a finite menu and offer only coarse information to buyers about their willingness-to-pay.

The results so far have not addressed the nature of the pooling of buyers' values: we have argued that there will be a finite set of possible expected values under the optimal information structure, but have not established which values get pooled together. A familiar argument establishes that information is optimally provided by a monotone partition (Corollary 3). We take a given, and possibly pooled distribution of qualities and observe that the information design problem is linear in the values. Thus, a maximum is attained at an extreme point of the set of feasible distributions as established by Kleiner et al. (2021). This characterization follows after our main results and highlights that the monotone partitional structure does not play a critical role for our main results, Theorem 1 and 2.

We further provide sufficient conditions for pooling to be optimal even if there are only a finite set of values to begin with. We consider a setting where buyers' willingness-to-pay are distributed on a finite grid (with equal distances between values). One measure of the variation in the distribution of values in this setting is the entropy of the distribution. We show that if the entropy of the distribution is above $\log_2 9$ bits, then the pooling of values must be optimal (Theorem 2). Thus if the distribution of values has more variation than a uniform distribution on nine values, then there must be some pooling in the optimal mechanism. This result is obtained by constructing a best case distribution where no pooling is optimal and establishing that its entropy has to be below $\log_2 9$.

Our results so far are obtained for a large class of cost functions for producing quality-differentiated products. In particular, we only assume that the cost of producing a distribution of qualities is increasing with respect to the increasing convex order. Thus the cost of a distribution of qualities increases if the distribution of qualities increases or becomes more variable. This class of cost functions includes standard separable cost functions as in Mussa & Rosen (1978), where there is a convex cost of producing a given quality and the cost of a distribution of qualities is just the sum of the costs of individual qualities, thus "variable inventory". But it also includes the "fixed inventory" case of Loertscher & Muir (2022), where the seller has an inventory of products that he can sell at zero cost but where any distribution not covered by the fixed inventory is infeasible (e.g., has infinite cost). Thus the results are insensitive to the nature of the cost function and do not rely on any cost savings from changing the inventory of products sold.

However, to obtain further results on the number of items in the optimal menu, we must appeal to properties of the cost function. We report a sufficient condition for when a single item menu is optimal with variable inventory (Theorem 3). In this case, the seller sells only one quality and divides buyers only into those who buy and those who are excluded. The sufficient condition is the convexity of the marginal cost function as well as a modest tails condition on the distribution of values.

Our setting reflects three notable features of the digital economy: (i) the sellers are better-informed than buyers about the match value; (ii) particular items on the menu, that is, specific quality-price pairs, are recommended to different buyers via recommendation services; (iii) personalized prices (or more generally third-degree price discrimination) are usually not available for the seller, but menu pricing (or second-degree price discrimination) can occur. The lack of third degree price discrimination occurs due to the presence of search engines and comparison sites, and business models that discourage third-degree price discrimination (see, for example, DellaVigna & Gentzkow (2019) for the uniform price policy of national chain store and Weiss (2000) for a commitment by Amazon to not price on demographic characteristics).

These three features are reflected in our analysis as follows: (i) the buyers only know the prior, or latent distribution; (ii) the seller chooses how much information to provide to the buyers (i.e., the information that is conveyed as a recommendation); (iii) the seller makes the products available to all buyers with a public menu of products and prices. Thus, our analysis predicts *personalized* product recommendations from a *non-personalized* common menu. We provide a more complete interpretation and discussion after we introduce the model formally.

1.2 Related Literature

We analyze a model of nonlinear pricing with a variable or a fixed distributions of qualities, as in Mussa & Rosen (1978) and Loertscher & Muir (2022), respectively. In either setting, these papers show that bundling different qualities, or randomizing the quality assignment via lotteries, can increase the revenue in the presence of irregular value distributions. By contrast, in our setting pooling is optimal for *all*– regular or irregular– distributions.

In our analysis, the seller controls the selling mechanism and the information that the buyers receive. The seller thus has access to the tools of Bayesian persuasion (Kamenica & Gentzkow (2011)) and mechanism design and we provide a solution to an integrated mechanism and information design problem in a classic economic environment. Rayo & Segal (2010) offer a sender-receiver model where the sender offers "prospects" to the receiver. Each prospect offers a pair of distinct individual rewards for sender and receiver, and those are private information to the seller. The receiver has a random outside option that is private information to the receiver. By bundling the prospects, the sender can increase the number and the probability of prospects that are accepted by the receiver (and profitable for the seller.) Thus, similar to the current environment, the sender controls how much information the receiver has about the offered items. But distinct from the current setting, the receiver only has unit demand. Building on Rayo & Segal (2010), Rayo (2013) considers a model of social status provision that also shares some features with our model. The utility function of an agent before any transfer is a product of their value (or an increasing function of their value) and a social status which is equal to their *expected value* given some information structure. Thus, the allocation in Rayo (2013) is an information structure rather than a quality allocation.

There are a number of related papers that investigate how information may influence the menu offered by multi-product sellers. Mensch & Ravid (2022) and Thereze (2022) consider a Mussa & Rosen (1978) style model and allow the buyer to acquire information about their willingness-to-pay. Each paper considers a different cost of information and then derives the optimal menu that the seller offers in anticipation of the endogenous response of the buyers. The resulting menu offers a continuum of choices in which the buyer is implicitly compensated for the cost of information acquisition. By contrast, Bergemann & Pesendorfer (2007) considers a seller with many unit-demand buyers. They establish the optimality of a finite and asymmetric allocation rule across the bidders. They analyze finite mechanisms with an exogenous bound on the number of items, and then show that as the bound grows, the number of items used in the optimal mechanism eventually stays constant. The present argument identifies the bilinear optimization problem of jointly determining information and allocation. This allows us to obtain results regarding entropy (Theorem 2) and the optimality of single-item mechanisms (Theorem 3), which has no counterparts in Bergemann & Pesendorfer (2007).

The arguments that lead to the monotone pooling result are related to those introduced by Wilson (1989). While exogenously limiting the number N of items in a nonlinear pricing environment, Wilson (1989) shows that in a *social surplus maximizing* mechanism this restriction only causes surplus losses of order $1/N^2$. Bergemann et al. (2021) extends this rate result to multi-dimensional types and allocations. Here, we complement this argument with the fact that the gains that come from reducing informational rents always have a larger order of magnitude than the efficiency losses. We thus conclude that there is always some amount of pooling in the *profit-maximizing* mechanism.

More distant, Johnson & Myatt (2003) and Ellison (2005) also ask what determines the variety of a menu. They provide sufficient conditions under which the seller may offer fewer or more products as a function of cost structure and the distribution of tastes and advertising costs, respectively. The resulting conditions are quite distinct from ours as the seller cannot control the information in these models. Given the control of the information by the seller, one could allow the seller to offer the information directly for the sale to the agents, see Eső & Szentes (2003), (2007), and then possibly extract an even larger surplus. We deliberately refrain from giving the seller this additional instrument. In the digital economy, the leading application of this paper, the platforms are selling differentiated products and typically bundle the information with the sale of products (through recommendations) rather than offer the information as a separate service at a separate price.

There is some earlier work that asks when second-degree price discrimination may optimally resolve in a single-item menu. Anderson & Dana (2009) impose an *a priori* finite upper bound on the quality in the setting of Mussa & Rosen (1978). They state conditions under which all values receive the same quality, namely the quality at the upper bound. Sandmann (2024) shows that a necessary condition for a single-item menu to be profit maximizing is that the single-item menu constitutes the socially optimal allocation. By contrast, in the current environment a continuum of qualities is socially optimal. Hence there would be no reason to restrict the menu and offer a bunching solution in the absence of information design.

2 Model

2.1 Payoffs

A seller supplies products of varying quality $q \in [q_l, q_h] \subset \mathbb{R}_+$ to a continuum of buyers with mass 1. Each buyer has unit demand and a willingness-to-pay, or value $v \in [v_l, v_h] \subset \mathbb{R}_+$, for quality q. The utility net of the payment $p \in \mathbb{R}_+$ is:

$$v{\cdot}q - p. \tag{1}$$

The buyers' values are distributed according to a common prior distribution $F \in \Delta([v_l, v_h])$, which we refer to as the *latent* distribution of values. We assume that space of values and qualities are bounded, that is, $q_h, v_h < \infty$.

The seller produces a set qualities Q and incurs a cost that depends on the distribution of qualities. A distribution of qualities $G \in \Delta([q_l, q_h])$ has a cost for the seller denoted by:

$$C(G) : \Delta[q_l, q_h] \to \mathbb{R}_+, \tag{2}$$

where $Q = \operatorname{supp} G$. We assume that C is monotone increasing with respect to the *increasing convex order*. Intuitively, G is greater in the increasing convex order – and thus more costly – than another distribution $\overline{G}$ if either G represents higher qualities than $\overline{G}$, or if G represents more dispersed qualities than $\overline{G}$, or both. The details of the cost function will not play a central role in our analysis and a formal definition of the order is given in Section 4 (see condition (18)). Importantly, the ordering of the cost function is flexible enough to accommodate the classic model of Mussa & Rosen (1978): if there is a convex function $c : [q_l, q_h] \to \mathbb{R}_+$ such that

$$C(G) = \int c(q) dG(q), \tag{3}$$

then C is monotone increasing in the increasing convex order. The model is also flexible enough to accommodate a fixed inventory of goods (see Loertscher & Muir (2022)), which is when the seller has a fixed mass of goods of various qualities that can be pooled and discarded (at zero cost); a formal expression is given in Section 4.6. We will be using the fixed inventory model as a leading example, as it highlights the independence of many of our main results from the exact nature of the cost function.

2.2 Mechanism

The selling mechanism (P, Q, S) consists of a menu Q of products, their prices P and an information structure S.

Menu Pricing The seller produces a set of qualities for sale $Q \subset [q_l, q_h]$ and prices $P : Q \cup \{0\} \to \mathbb{R}_+$, with the interpretation that $P(q)$ is the price at which every buyer can purchase quality $q \in Q$. The pricing function must satisfy $P(0) = 0$, meaning that buyers can choose to not purchase any good (or equivalently, acquire a zero quality good). We refer to (P, Q) as the menu. By the taxation principle, (Proposition 1, Guesnerie & Laffont (1984)), any truthful direct mechanism can be represented by an indirect mechanism in the form of a menu (P, Q). For our arguments, the use of the indirect mechanism is convenient.

Information Structure The seller also chooses an information structure $S : [v_l, v_h] \to \Delta\mathbb{R}_+$, with the interpretation that $S(v)$ is a distribution of real-valued signals observed by buyers with latent value v. A buyer's expected value conditional on the signal realization s is denoted by:

$$\overline{v}(s) \triangleq \mathbb{E}[v \mid s]. \tag{4}$$

Frequently, we can omit the dependence of the expected value $\overline{v}(s)$ on the signal s and simply write $\overline{v}$ for a generic expected value.

2.3 Profit Maximizing Mechanism

Now the seller chooses the mechanism (P, Q, S) to maximize profits. To calculate profits for any given mechanism, we must first identify buyers' optimal choice rule $q : [v_l, v_h] \to Q$ given the menu (P, Q), where item $q(\overline{v})$ bought by a buyer with expected value $\overline{v}$ is:

$$q(\overline{v}) \in \arg\max_{q \in Q} \{\overline{v}q - P(q)\}, \tag{5}$$

where (as usual) ties are broken in favor of the seller.

The information structure S induces a distribution of expected values $\overline{F}$, where

$$\overline{F}(\overline{v}) = \int_{v} \int_{s:\mathbb{E}[v|s] \leq \overline{v}} dS(s|v)\, dF(v).$$

Now the distribution of qualities $\overline{G}$ needed to supply the buyers' demand is given by:

$$\overline{G}(\overline{q}) = \int_{\{\overline{v}:q(\overline{v})\leq\overline{q}\}} d\overline{F}(\overline{v}), \tag{6}$$

and the seller's profit is

$$\mathbb{E}[P(q(\overline{v}))] - C(\overline{G}), \tag{7}$$

where the expectation is computed using distribution $\overline{F}$. The seller's problem is then to choose the mechanism (P, Q, S) to maximize profits. We conclude this section with a discussion of the interpretation of our model.

2.4 Interpretation

Justifying the Common Menu Restriction Our model assumes that the seller has the ability to endow the buyers with any information structure about their values. However, we did not specify whether the seller did or did not observe the signal realizations of an individual buyer. But this does not matter given our maintained assumption that the seller offers a common menu of qualities and prices available to all buyers, and that there is no personalized pricing.

However, the interpretation of our model is a little different depending on whether seller observes the buyers' signal realizations. In particular, suppose the seller could not observe the realizations of the signals. In this "information-constrained" case, the seller cannot offer different prices to buyers with different latent values or signals. A standard argument in nonlinear pricing, the taxation principle (Guesnerie & Laffont (1984)), establishes that even if the seller could elicit information from buyers, he would not be able to improve the profit compared to a public pricing schedule. Thus in this "information-constrained" scenario, there is no loss in assuming that the seller is restricted to offer a common menu.

But if the seller did observe the realization of the buyers' signals, there may be other reasons - outside our model - why the seller does not offer personalized prices. We call this the "instrument-constrained" case because we have imposed a constraint on the selling mechanism. As we discussed in the introduction, in the digital economy, the "information-constrained" argument against personalized pricing may be less relevant but there are a variety of forces pushing against personalized pricing: (i) platform sellers may fear that personalized pricing will scare consumers away from the platform; (ii) there may be legal restrictions on personalized pricing; (iii) buyers may have the ability to find others' personalized pricing by searching under alternative identities.

Recommendation as Information By a standard argument, it would be without loss of generality to assume that instead of sending abstract signals, the signals are restricted to items in the menu and can be interpreted as recommendations of what items to buy. Thus the information structure could be restricted to take the form $S : [v_l, v_h] \to \Delta(Q)$ and, subject to obedience constraints, buyers could be assumed to always select the item that was "recommended" to them.

This representation of the information structure is not particularly helpful to us analytically. However, we highlight this representation because we think that recommender systems are a relevant mechanism used in the digital economy to send information. In particular, when a search menu generates a list of buying options, we know that the first or top item is more likely to be chosen. One interpretation is that the top items represent recommendations of the seller.

Direct Allocations By a standard argument, it would be without loss of generality to assume a direct allocation mechanism, i.e., letting the buyers report their expected values to the mechanism and the mechanism assigns a quality-price pair to each buyer in such a way that it is incentive compatible for buyers to truthfully report their expected values. We use the indirect representation in part because it fits better with our interpretations of the common menu assumption. However, we emphasize that optimal choices in our mechanism $q(\overline{v})$ and $P(q(\overline{v}))$ correspond to incentive compatible and individually rational direct allocation rules (and some readers may find it helpful reading our arguments that way).

3 A First Example with Finitely Many Values

We first analyze the environment with a finite number N of values and qualities. In this section, we do not yet seek to derive the exact nature of the optimal mechanism. Rather, we compare the revenue of the seller under two specific information structures. The first is *complete disclosure* where each buyer learns their value as private information. The resulting optimal revenue arises from the standard optimal screening mechanism. By contrast, the second information structure pools the two lowest values of the buyer and then derives the optimal mechanism when these adjacent values are pooled, and we refer to this as *lower pooling*. We provide sufficient conditions for when this specific pooling of the lowest adjacent values improves the revenue relative to complete disclosure. This will provide key intuition for results that follow.

Throughout this section, we assume that the seller has a fixed inventory of goods, that is, a fixed distribution of qualities G (as in Loertscher & Muir (2022)). The assumption of an exogenous inventory simplifies the algebra as now revenue equals profit, but the conclusions go through for variable inventory (as in Mussa & Rosen (1978)).

The support of values and qualities are denoted by $\{v_1, ..., v_N\}$ and $\{q_1, ..., q_N\}$, with $0 < v_1 < \cdots < v_N$ and $0 < q_1 < \cdots < q_N$; both distributions have atoms of size $\{f_1, ..., f_N\}$. So there is an assortative allocation where buyers of value v_k are matched with allocation q_k. The discrete version of the virtual values are denoted by:

$$\phi_i \triangleq v_i - (v_{i+1} - v_i)\frac{\sum_{j=i+1}^{N} f_j}{f_i}. \tag{8}$$

The following lemma provides a way in which the virtual values shape the optimal mechanism in screening problems without persuasion.

Lemma 1 (Screening without Persuasion)
The profit-maximizing allocation is efficient if and only if the virtual values ϕ_i are increasing in i and greater than 0 for every i.

For the proof see Theorem 1 in Loertscher & Muir (2022), who characterize the optimal screening mechanism (without persuasion) when the seller has a fixed inventory of goods of various qualities. Throughout this section, we assume that the virtual values are increasing in i and greater than 0 for every i. Thus we focus on an environment that is most favorable towards complete disclosure as bundling due to ironing does not occur under complete disclosure. If the goods are sold efficiently, the optimal pricing policy is given by:

$$P(q_i) = v_1 q_1 + \sum_{j=2}^{i} (q_j - q_{j-1}) v_j, \tag{9}$$

by a standard argument using the incentive compatibility constraints. The seller is thus able to extract the value v_i only for the marginal quality, $q_i - q_{i-1}$, offered to each buyer v_i. The price for the product q_i is then the sum of the quality increments $(q_j - q_{j-1})$ multiplied by the value v_j for all values below v_i. The profits generated are given by:

$$\Pi = \sum_{i=1}^{N} P(q_i) f_i = v_1 q_1 + \sum_{i=2}^{N} (q_i - q_{i-1}) v_i (1 - \sum_{j=1}^{i-1} f_j), \tag{10}$$

where the second expression is obtained by writing $P(q_i)$ explicitly and changing the order of summation.

We can interpret the profit function from the perspective of a multi-product monopolist who sells N distinct products of qualities $(q_i - q_{i-1})$ and has a mass $(1 - \sum_{j=1}^{i-1} f_j)$ of each of these products. In this interpretation, the buyers have multi-unit demand so they can buy all products (or any ordered subset), which amounts to the same utility as buying a single good of quality q_N in the original interpretation of the model.

Let us now consider a mechanism where the *lowest two values* are pooled so the expected value and expected quality generated are given by:

$$\overline{q}_1 = \frac{f_1 q_1 + f_2 q_2}{f_1 + f_2}, \quad \overline{v}_1 = \frac{f_1 v_1 + f_2 v_2}{f_1 + f_2}. \tag{11}$$

After this pooling, there are only $N-1$ distinct qualities for sale among $N-1$ values, which we assume are sold assortatively (as before). The profits generated by pooling, say $\overline{\Pi}$, is computed as in (10) but accounting for the fact that the first two quality levels and the first two value levels are pooled. The difference between the profits after pooling and the profits under complete disclosure is:

$$\overline{\Pi} - \Pi = (\overline{v}_1 \overline{q}_1 - (v_1 q_1 + (q_2 - q_1) v_2 (1 - f_1))) + (q_2 - \overline{q}_1)\, v_3 (1 - f_1 - f_2). \tag{12}$$

Here we wrote the difference as the sum of two terms. The first term corresponds to the difference in the profits generated by values $\{v_1, v_2\}$ and qualities $\{q_1, q_2\}$: the seller now sells a pooled quality level $\overline{q}_1$ at price $\overline{v}_1$ to every buyer instead of selling quality q_1 at price v_1 to every buyer and a quality increment $(q_2 - q_1)$ at price v_2 to a mass of buyers $(1 - f_1)$. The second term accounts for the fact that pooling increases the quality difference between q_3 and the immediate quality predecessor– which is q_2 under complete disclosure and $\overline{q}_1$ when pooling– so we get a term proportional to $(q_2 - \overline{q}_1)$ and proportional to v_3 which is the price of this quality increment. Note that pooling does not change cost (which is zero) because we assume that there is a fixed inventory of goods, so changes in profits are the same as changes in revenue. If there was a non-trivial production cost, then the benefits of pooling would be even larger because pooling would also reduce cost.

In contrast to screening without persuasion, we now have that the distribution of qualities plays an important role to determine when pooling is optimal. This will lead to properties that any optimal distribution of qualities must have, and these properties will not have any counterpart in classic screening problems. Towards these properties, we observe that the payoff difference given by (12) is linear in q_1 and strictly positive when $q_1 = q_2$. Thus we have that:

$$(\overline{\Pi} - \Pi)\, |_{q_1 = 0} > 0 \Longrightarrow (\overline{\Pi} - \Pi) > 0. \tag{13}$$

Now, by evaluating the condition at $q_1 = 0$ we get a weaker condition for when pooling generates more profits than complete disclosure. Intuitively, pooling always increases the price at which q_1 is sold, so by evaluating at $q_1 = 0$ we obtain a weaker condition. We provide this condition explicitly.

Proposition 1 (Lower Pooling Improves Profits)
Lower pooling revenue exceeds complete disclosure revenue if:

$$\frac{f_2}{(1-(f_1+f_2))(f_1+f_2)} < \frac{v_3 - v_2}{v_2 - v_1}. \tag{14}$$

We have that (14) corresponds to the left-hand-side of (13) after re-arranging terms (the left-hand-side of (13) is linear in q_2, so (14) also does not depend on q_2). The trade-off is that pooling leads to q_2 being sold at price $\overline{v}_1$ (instead of v_2) but pooling also increases the quality increment q_3 minus its predecessor (which is $\overline{q}_1$ instead of q_2 under pooling), which is priced at v_3.

We can leverage (14) to provide a simpler condition for when complete disclosure is not optimal. We say that the values are *uniformly distanced* when

$$v_{i+1} - v_i = v_i - v_{i-1}, \text{ for all } i.$$

Assuming the values are uniformly distanced will allow us to significantly simplify the algebra.

Corollary 1 (Lower Pooling Improves Profit for Uniformly Distanced Values)
With uniformly distanced values*, lower pooling revenue exceeds complete disclosure revenue if:*

$$f_2 < \sqrt{f_1} - f_1. \tag{15}$$

With uniformly distanced and *uniformly distributed values ($f_i = f_j$ for all i, j), lower pooling revenue exceeds complete disclosure revenue if $N \geq 5$.*

The sufficient condition (15) is obtained by applying the *uniformly distanced* property to the improvability condition (14) of Proposition 1. This corollary then provides simple conditions for lower pooling to be optimal.

Pooling Fine Distributions If we approximate a continuous distribution with a sequence of finite distributions on a uniformly distanced grid, eventually (15) will be satisfied (as $\sqrt{f_1}$ will be larger than $f_2 + f_1$), and so complete disclosure will not be optimal. We will generalize this idea by showing that in any optimal mechanism the induced distribution of values is always finite (Theorem

1). Hence, the optimal mechanism is coarse even when the underlying distribution of values is very fine.

A striking feature of the analysis is that (15) is satisfied even when the distribution is relatively coarse. For example, with uniformly distanced and uniformly distributed values, lower pooling improves upon complete disclosure as soon as $N \geq 5$. We will generalize this idea by showing that there is an upper bound on the entropy of the distribution such that complete disclosure is optimal. This result will show that even when the latent distribution of values is finite, pooling will be optimal unless "most" of the distribution is located in "few large atoms"as measured by the entropy (Theorem 2).

There are no counterparts to these conditions in classic screening problems; we now provide some intuition for the stark difference with standard screening problems.

Screening without Persuasion To contrast our analysis with screening in the absence of persuasion, we note that (14) can be written as follows:

$$(\overline{\Pi} - \Pi) \mid_{q_1=0} = \frac{f_1 f_2 q_2}{f_1 + f_2}(\phi_1 - \phi_2) + \frac{f_2 q_2}{f_1 + f_2}(\overline{v}_1 - v_1) > 0. \tag{16}$$

The first term is the change in profits due to the pooling of the allocation only (without changing the information). Then, the first term is positive if and only if the virtual values are strictly decreasing (as in Lemma 1). The second term is the gain in profits from pooling the values after we have already pooled the allocation. In essence, we will show that the second term in (16) always dominates the first one when the atoms are small enough. A heuristic argument goes as follows. Suppose the atoms are small and the distance between values and between virtual values are small. Both terms are proportional to $f_2/(f_1 + f_2)$, but the first term is the product of two small factors $f_1(\phi_1 - \phi_2)$ while the second term consists of one small factor $(\overline{v}_1 - v_1)$. The intuition is that the profit losses due to pooling the allocation (without pooling information) are limited to the locally affected values. Yet the revenue benefits from pooling values increases the profits generated by all higher values. Due to the differences in the extensive margin, pooling is always beneficial when atoms are small enough. This is the reason we get distinct conditions for the optimality of pooling in screening with persuasion relative to screening without persuasion.

4 The Optimality of a Finite Menu

In Section 4.1 we state our first main result, Theorem 1, which asserts that the optimal mechanism is finite. The main steps of the proof are provided in Sections 4.2-4.4. In Section 4.2, we develop the properties of incentive compatible and feasible mechanisms and state the profit maximization problem (7) as a joint optimization over two distributions, the distribution of values $\overline{F}$ and the distribution of qualities $\overline{G}$. Section 4.3 establishes discreteness of the optimal mechanism, which is the main step in establishing finiteness of the optimal mechanism. Here we argue that pooling small intervals of values is always beneficial for the seller. Section 4.4 establishes an additional property of independent interest, the convexity of the menu. We show that the quality differences between any two adjacent intervals increase as the values increase. Incidentally, this allows us to go from discreteness to finiteness by ruling out accumulation points. In Section 4.5 we provide an additional property of the optimal mechanism, a monotone partitional property, that is not used to prove Theorem 1, but will be helpful to characterize the optimal mechanism in applications. Finally, in Section 4.6 we illustrate the optimal mechanism in two specific examples.

4.1 Optimality of a Finite Mechanism

We say a *mechanism is finite* if the support of the value distribution $\overline{F}$ is finite. Our first main result shows that every optimal mechanism offers a finite set of distinct expected values and expected qualities. Throughout the paper we consider mechanisms such that, if a buyer does not purchase any positive quality, then they are provided no additional information, thus,

$$q(\bar{v}) = q(\bar{v}') = 0 \Rightarrow \bar{v} = \bar{v}'.$$

This only disciplines the information that the optimally excluded buyers receive and it is without loss of generality to consider mechanisms with this property.

Theorem 1 (Optimality of Finite Mechanism)
Every optimal mechanism is finite.

The intuition for the result is that pooling small intervals generates losses that are “local”, meaning that they affect only the values being involved in the pooling. On the other hand, the reduction in informational rents affects all values higher than the interval being pooled. Hence, by considering an interval that is small enough we get that the benefit is an order of magnitude larger than the losses. This argument is formally presented in Section 4.3, where we show that any optimal mechanism is discrete.

Since the space of values and qualities is bounded, to conclude the proof of Theorem 1 we need to show that there are no accumulation points. Showing that there are no accumulation points in any interior part of the distribution can be proved in a similar way to how we prove that an optimal mechanism is discrete. However, proving that there are no accumulation points at the top of the distribution requires a different line of argument. We prove this by showing that the qualities offered in any optimal mechanism satisfy increasing differences. We provide this result after we prove that the optimal mechanism is discrete, and we argue in due course that this result is of independent interest. We note that if we were to relax the assumption that the space of values and qualities is bounded, we would get that the optimal mechanism is discrete without the property of finiteness.

Finally, we show that the endogenous distribution of values in an optimal mechanism can be obtained from monotone partitions of the exogenous distribution of values. This last result is not used to prove Theorem 1, but it will be useful when we characterize the optimal mechanism (with additional assumptions on values and qualities) in Section 6.

Before we provide the substantive and novel part of our analysis, we apply standard tools from information and mechanism design to write the seller's problem as a maximization problem over two distributions.

4.2 Optimization Over Two Distributions

The seller's design problem has two components: (i) the seller chooses the information that buyers have about their own value, and (ii) the menu. We now provide a more convenient representation of these two elements by writing the seller's problem as the maximization over two distributions. One of the distributions will be the distribution of expected values $\overline{F}$ generated by the information structure; the other distribution will be the mass of qualities that the seller supplies in the optimal mechanism $\overline{G}$.

The buyers' information structure is summarized by the distribution of expected values $\overline{F}$. By Blackwell (1951), Theorem 5, there exists an information structure that induces a distribution $\overline{F}$ of expected values if and only if $\overline{F}$ is a mean-preserving contraction of F, i.e.,

$$\int_v^{v_h} F(x)dx \leq \int_v^{v_h} \overline{F}(x)dx, \ \forall v \in [v_l, v_h], \tag{17}$$

with equality for $v = v_l$. If $\overline{F}$ is a mean-preserving contraction of F (or $\overline{F}$ majorizes F), we write

$F \prec \overline{F}$.

We assumed that the cost of providing a distribution G of qualities is $C(G)$ (see (2)). For any distributions $G, \overline{G} \in \Delta\mathbb{R}_+$, we say G is greater than $\overline{G}$ in the increasing convex order (or $\overline{G}$ weakly majorizes G), and denote this by $G \succ_c \overline{G}$, if

$$\int_q^{q_h} G(x)dx \leq \int_q^{q_h} \overline{G}(x)dx, \ \forall q \in [q_l, q_h], \tag{18}$$

see Theorem 4.A.2 in Shaked & Shanthikumar (2007). Unlike the definition of mean-preserving contractions given by (17), we now do not require (18) to be satisfied with equality when $q = q_l$. Theorem 4.A.6 in Shaked & Shanthikumar (2007) states that $\overline{G} \prec_c G$ if and only if there exists $\widehat{G}$ such that G is a mean preserving spread of $\widehat{G}$ and $\widehat{G}$ first-order stochastically dominates $\overline{G}$. Thus if $\overline{G} \prec_c G$, then $\overline{G}$ is both "lower" and "less variable" than G in the sense that $\overline{G}$ corresponds to lower qualities than $\widehat{G}$, which in turn corresponds to less variable qualities than G. Assuming that the cost function is monotone increasing in the increasing convex order means that for any pair of distributions $G, \overline{G} \in \Delta\mathbb{R}_+$, $\overline{G} \prec_c G$ implies that $C(\overline{G}) \leq C(G)$. So producing higher qualities and more variable qualities is more expensive.[1]

We recall that for any pricing function P, $q(\overline{v})$ denotes the optimal choice of a buyer with expected value $\overline{v}$. Analogously, we define the respective payment:

$$p(\overline{v}) = P(q(\overline{v})).$$

We refer to a pair $(q(\overline{v}), p(\overline{v}))$ as a *direct menu*. We depart from the classic terminology (that is "direct mechanism") because in our model a mechanism also includes the information structure.

Optimality of the choice function (5) implies that

$$\overline{v}q(\overline{v}) - p(\overline{v}) \geq \overline{v}q(\overline{v}') - p(\overline{v}'), \quad \forall \overline{v}, \overline{v}' \in [v_l, v_h]; \tag{19}$$

and

$$\overline{v}q(\overline{v}) - p(\overline{v}) \geq 0, \quad \forall \overline{v} \in [v_l, v_h]. \tag{20}$$

Thus, the direct menu $\{(q(\overline{v}), p(\overline{v}))\}$ induced by P satisfies the classic incentive compatibility and participation constraints. Following standard techniques, the incentive compatibility requires that

[1] While the majorization order ($\prec$) and the increasing convex order ($\prec_c$) have nearly identical definitions in terms of the integrals, (17) and (18) respectively, we adopt the precedence symbol in the opposite direction for each. This choice aligns with the conventions in the literature for majorization order and for increasing convex order (as used in the stochastic order literature), ensuring consistency within their respective contexts.

the allocation $q(\overline{v})$ is increasing and the payment $p(\overline{v})$ is determined by the allocation rule using the Envelope condition:

$$p(\overline{v}) = \overline{v}q(\overline{v}) - \int_{v_l}^{\overline{v}} q(s)ds. \tag{21}$$

Note that while $\overline{F}$ may have gaps, it is without loss of generality to assume that the allocation $q(\cdot)$ is defined on the entire domain $[v_l, v_h]$. We are thus left with determining the optimal distribution of values and the allocation function.

We thus have that for any mechanism (P, Q, S) there is an induced direct menu and a distribution of expected values $(p(\overline{v}), q(\overline{v}), \overline{F})$. Similarly, for any $(p(\overline{v})), q(\overline{v}), \overline{F})$ satisfying the above properties we have a mechanism (P, Q, S) that induces it.

However, rather than optimizing over the allocation function $q(\bar{v})$ we will optimize over the distribution of qualities $\overline{G}$ offered by the mechanism (as defined in (6)), and then back out what would be the implied allocation q using $\overline{F}$ and $\overline{G}$. More precisely, we will solve the problem:

$$\max_{\substack{\overline{F}^{-1} \prec F^{-1} \\ \overline{G}^{-1} \in \Delta\mathbb{R}_+}} \int_0^1 \overline{F}^{-1}(t)(1-t)\, d\overline{G}^{-1}(t) + \overline{G}^{-1}(0)\,\overline{F}^{-1}(0) - C\left(\overline{G}\right). \tag{22}$$

As in Section 3, the seller's revenue can be interpreted as the sum (integral) over all quality increments multiplied by the revenue generated per-unit-of-quality, (22) is the integral version of (10). By optimizing over distributions of qualities rather than the allocation function, the production cost $C(\overline{G})$ is independent of the information $\overline{F}$. Of course, varying $\overline{F}$ without varying $\overline{G}$ will change the implied allocation function $q(\overline{v})$. We write the expressions in terms of the quantile function (the inverse of $\overline{F}$ and $\overline{G}$) because this way the revenue is bilinear in the maximization variables.

Proposition 2 (Optimizing Over Two Distributions)
A mechanism (P, Q, S) is optimal if and only if there exists a solution of (22) of the form $(\overline{F}, \overline{G})$, such that

$$q(\overline{v}) = \overline{G}^{-1}(\overline{F}(\overline{v})),$$

and $p(\overline{v})$ is the direct mechanism induced by (P, Q, S).

We thus analyze a maximization problem over two distributions $(\overline{F}, \overline{G})$ given by (22), and we refer to a pair $(\overline{F}, \overline{G})$ as a mechanism. When solving the problem (22) we are only be interested in functions $(\overline{F}, \overline{G})$ such that at any t:

$$\overline{G}^{-1}(t) \text{ is increasing in } t \text{ only if } \overline{F}^{-1}(t) \text{ is increasing in } t. \tag{23}$$

This is without loss of generality: if $\overline{F}^{-1}(t)$ is constant in some interval (t_1, t_2) and $\overline{G}^{-1}(t)$ is not constant in this interval, then we can pool qualities in this interval and thus weakly increase revenue and weakly decrease cost. In a direct menu we construct $q(\overline{v})$ from $\overline{G}$, so this measurability condition is necessary and sufficient for the solution to be implementable with a direct menu.

4.3 Optimality of Discrete Mechanism

We say that a given distribution H has discrete support if there is no open interval $(t_1, t_2) \subset [0, 1]$ such that H^{-1} is strictly increasing in (t_1, t_2). We say a mechanism $(\overline{F}, \overline{G})$ is discrete if it consists of distributions with discrete support. As a first step towards proving Theorem 1 we prove that any optimal mechanism is discrete. We can describe a discrete mechanism by a countable collection of quantiles $\{t_k\}_{k \in K}$, where at each quantile t_k there is a discontinuous jump in expected values and expected qualities and both distributions are constant everywhere else. As shorthand notation, we denote by $\overline{v}_k$ the expected value and by $\overline{q}_k$ the quality at quantile t_k:

$$\overline{v}_k \triangleq \overline{F}^{-1}(t_k); \qquad \overline{q}_k \triangleq \overline{G}^{-1}(t_k). \tag{24}$$

Proposition 3 (Optimality of Discrete Mechanism)
Every optimal mechanism $(\overline{F}, \overline{G})$ is discrete.

Proof. The measurability condition (23) implies that, if $\overline{F}$ is discrete, then $\overline{G}$ (and thus the mechanism) is discrete. We consider a candidate mechanism (F, G) and correspondingly (F^{-1}, G^{-1}) such that F^{-1} is strictly increasing on some interval $[t_1, t_2)$ and show that this mechanism is not optimal. In writing the proof, the notation suggests that F and G are the exogenous distributions of values and qualities; this is only to make the notation more compact and is of no substantive difference for the argument.

We first consider the case in which (F^{-1}, G^{-1}) are both strictly increasing on some interval $[t_1, t_2)$ and show this mechanism is not optimal. It is useful to write the interval $[t_1, t_2)$ in terms of its mid-point and width:

$$\widehat{t} \triangleq \frac{t_1 + t_2}{2}; \quad \Delta \triangleq \frac{t_2 - t_1}{2}. \tag{25}$$

So, we have that

$$[t_1, t_2) = [\widehat{t} - \Delta, \widehat{t} + \Delta) \tag{26}$$

and we will eventually take the limit $\Delta \to 0$. Since both $F^{-1}(t)$ and $G^{-1}(t)$ are strictly increasing in this interval, they are differentiable at almost every t in this interval. So, without loss of generality we assume that $F^{-1}(t)$ and $G^{-1}(t)$ are differentiable at $\widehat{t}$ and the derivative is strictly positive.

We contrast the profits generated by (F, G) with ones generated by a mechanism in which values and qualities are pooled in this interval $[t_1, t_2)$. Before we compute the changes in profit we introduce some notation. We denote by v_1 and q_1 the value and the quality at the lower limit of the interval $[t_1, t_2)$ and denote by q_2 the quality at the upper limit of the interval $[t_1, t_2)$:

$$v_1 \triangleq F^{-1}(t_1);\ q_1 \triangleq G^{-1}(t_1);\ q_2 \triangleq G^{-1}(t_2).$$

We also denote by μ_v and μ_q the average value and average quality provided in this interval:

$$\mu_v \triangleq \frac{\int_{t_1}^{t_2} F^{-1}(t)dt}{t_2 - t_1} \text{ and } \mu_q \triangleq \frac{\int_{t_1}^{t_2} G^{-1}(t)dt}{t_2 - t_1}.$$

We only compute these quantities in the interval $[t_1, t_2]$, and we can therefore safely omit an index referring to the interval $[t_1, t_2]$ in the expressions.

We first consider the impact from pooling the allocation without changing the information structure, and thus pooling the values. The distribution of qualities can be written as follows:

$$G_q^{-1}(t) = \begin{cases} \mu_q & \text{if } t \in [t_1, t_2); \\ G^{-1}(t), & \text{if } t \notin [t_1, t_2). \end{cases} \tag{27}$$

As we only change the allocation of qualities in this first step, we use the subscript q to refer to this variation in $[t_1, t_2)$. We denote the profits generated by (F, G) by Π and the profits generated by (F, G_q) are denoted by Π_q. The change in profit can be expressed as follows:

$$\Pi - \Pi_q = \int_{t_1}^{t_2} F^{-1}(t)(1-t)dG^{-1}(t) - \int_{t_1}^{t_2} F^{-1}(t)(1-t)dG_q^{-1}(t) - (C(G) - C(G_q)).$$

We note that $F^{-1}(t)$ is differentiable at $\widehat{t}$ and so $F^{-1}(t)(1-t)$ is also differentiable at this point. Thus, there exists $r(t)$ such that:

$$F^{-1}(t)(1-t) = F^{-1}(\widehat{t})(1-\widehat{t}) + \frac{d\left[F^{-1}(t)(1-t)\right]}{dt}\Big|_{t=\widehat{t}}(t-\widehat{t}) + r(t)(t-\widehat{t}),$$

with $r(t) \to 0$ as $t \to \widehat{t}$. We thus get that:

$$\Pi - \Pi_q = \int_{t_1}^{t_2} r(t)(t-\widehat{t})dG^{-1}(t) - \int_{t_1}^{t_2} r(t)(t-\widehat{t})dG_q^{-1}(t) - (C(G) - C(G_q)).$$

Here we used that:

$$\int_{t_1}^{t_2} dG^{-1}(t) = \int_{t_1}^{t_2} dG_q^{-1}(t) = (q_2 - q_1) \text{ and } \frac{\int_{t_1}^{t_2} G^{-1}(t)dt}{t_2 - t_1} = \frac{\int_{t_1}^{t_2} G_q^{-1}(t)dt}{t_2 - t_1} = \mu_q,$$

so when we compute the difference in revenue the first two terms of the Taylor expansion of $F^{-1}(t)(1-t)$ cancel out. To obtain an upper bound on the difference we first note that $G_q \prec_c G$ so $C(G_q) \leq C(G)$ and we can also bound the value of each integral by:

$$\int_{t_1}^{t_2} r(t)(t-\widehat{t})dG^{-1}(t) \leq (q_2-q_1)\Delta \max_{t\in[\widehat{t}-\Delta,\widehat{t}+\Delta]} |r(t)|,$$
$$\int_{t_1}^{t_2} r(t)(t-\widehat{t})dG_q^{-1}(t) \geq -(q_2-q_1)\Delta \max_{t\in[\widehat{t}-\Delta,\widehat{t}+\Delta]} |r(t)|.$$

We thus have that:

$$\Pi - \Pi_q \leq 2(q_2-q_1)\Delta \max_{t\in[\widehat{t}-\Delta,\widehat{t}+\Delta]} |r(t)|.$$

We now note that:

$$\lim_{\Delta\to 0} \frac{(q_2-q_1)}{\Delta} = 2\frac{dG^{-1}(\widehat{t})}{dt}.$$

Thus:

$$\lim_{\Delta\to 0} \frac{\Pi-\Pi_q}{\Delta^2} \leq 4\frac{dG^{-1}(\widehat{t})}{dt} \lim_{\Delta\to 0}\left\{\max_{t\in[\widehat{t}-\Delta,\widehat{t}+\Delta]} |r(t)|\right\} = 0.$$

The limit is because $r(t) \to 0$ as $t \to \widehat{t}$.

We now compute the profit when the allocation is G_q and we now additionally pool the values in this interval, which we denote by F_v:

$$F_v^{-1}(t) = \begin{cases} \mu_v & \text{if } t \in [t_1,t_2); \\ F^{-1}(t), & \text{if } t \notin [t_1,t_2). \end{cases}$$

The corresponding profits are denoted by Π_{vq}. The change in profits is:

$$\Pi_{vq} - \Pi_q = \int_{t_1}^{t_2} F_v^{-1}(t)(1-t)dG_q^{-1}(t) - \int_{t_1}^{t_2} F^{-1}(t)(1-t)dG_q^{-1}(t) \tag{28}$$
$$= (\mu_q - q_1)(\mu_v - v_1)(1-t_1). \tag{29}$$

The production cost is the same in both mechanisms, so the cost does not appear in the difference. We obtain the second equality by noting that the only quantile in which F^{-1} and F_v^{-1} differ and $dG_q^{-1}(t)$ is not zero, is the discrete quality increment $\Delta q = (\mu_q - q_1)$ at $t = t_1$. Intuitively, if information about the values is pooled then the whole quality increment is priced at μ_v instead of v_1. We now note that:

$$\lim_{\Delta\to 0} \frac{(\mu_q-q_1)}{\Delta} = \frac{dG^{-1}(\widehat{t})}{dt}; \quad \lim_{\Delta\to 0} \frac{(\mu_v-v_1)}{\Delta} = \frac{dF^{-1}(\widehat{t})}{dt}$$

Hence,

$$\lim_{\Delta \to 0} \frac{\Pi_{vq} - \Pi_q}{\Delta^2} = \frac{dG^{-1}(\widehat{t})}{dt} \frac{dF^{-1}(\widehat{t})}{dt}(1 - \widehat{t}) > 0, \tag{30}$$

Hence, for a small enough Δ:

$$(\Pi_{vq} - \Pi_q) + (\Pi_q - \Pi) > 0.$$

Hence, the mechanism (F, G) is not optimal.

Finally, we consider the case in which $F^{-1}(t)$ is strictly increasing in $[t_1, t_2]$ and G^{-1}(t) is not strictly increasing in $[t_1, t_2]$. Without loss of generality we consider interval $[t_1, t_2]$ to be such that $G^{-1}(t)$ is constant in (t_1, t_2) and either $t_1 = 0$ or $G^{-1}(t)$ is increasing at t_1. If $G^{-1}(t)$ is continuously increasing in some interval $[t_1 - \epsilon, t_1]$ the first part of the proof applies to this interval so, without loss of generality, if $t_1 > 0$ we assume that $G^{-1}(t)$ is discontinuously increasing at t_1. Recall that whenever two values are excluded they have the same expected valuation so we must also have that $G^{-1}(t_1) \neq 0$. Since G^{-1} is constant, pooling information does not change the allocation. Pooling information, analogously to (29), generate revenue gains equal to:

$$\Pi_{vq} - \Pi_q = (\mu_q - q_1^-)(\mu_v - v_1)(1 - t_1) > 0,$$

where now $q_1^- \triangleq 0$ if $t_1 = 0$ and $q_1^- \triangleq \lim_{t \uparrow t_1} G^{-1}(t)$ if $t_1 > 0$. Hence, the mechanism is not optimal. This concludes the proof. ■

Before we proceed, we discuss the possibility of accumulation points in the optimal mechanism. Accumulation points in the interior of the distribution can be ruled out in an analogous way to the argument we just presented. However, when the accumulation point is at the top of the distribution the arguments no longer go through. Technically, we can see in (30) that it was necessary to take an interval bounded away from the top of the distribution (that is, $\widehat{t} < 1$). Conceptually, we need that the interval being pooled is small relative to the survival function. In the next subsection we use a different type of argument to rule out accumulation points at the top of the distribution, which will also provide novel insights into the optimal mechanism.

4.4 Optimality of Convex Menus

We previously defined a discrete menu by the expected values $\overline{v}_k$ and the quality $\overline{q}_k$ at quantile t_k (see 24):

$$\overline{v}_k \triangleq \overline{F}^{-1}(t_k); \qquad \overline{q}_k \triangleq \overline{G}^{-1}(t_k).$$

We say that the quality increments display increasing differences if

$$\overline{q}_{k+1} - \overline{q}_k > \overline{q}_k - \overline{q}_{k-1}, \tag{31}$$

for all $k < K$ and where $\overline{q}_0 = 0$. If the menu of offered qualities display increasing quality increments then we say that the *menu is convex.*

Proposition 4 (Optimality of Convex Menus)
Every optimal mechanism generates a convex menu.

Thus, the quality increments between menu items increase as we move up the quality ladder. This finding suggests implications for how sellers should structure their product lines. The convexity of the menu arises with some frequency in nonlinear pricing. In mobile phone pricing, the data packages offered are frequently convex. Similarly, mobile phones themselves are offered with a variety of memory chips with convex structure.[2]

The source of the convexity can be explained as follows. Suppose we have only two values and two quality levels and we compare the revenue generated by separating the values and qualities relative to the revenue generated when the values and the qualities are pooled (this trade-off is computed in (12), with $f_1 + f_2 = 1$ so the last term is 0). Before pooling, quality q_1 is sold at price v_1 and the quality increment $(q_2 - q_1)$ is sold at price v_2. After pooling, the average quality is sold at the average value. Hence, pooling increases the price at which the quality q_1 is being sold and decreases the price at which the quality increment $(q_2 - q_1)$ is being sold. One can then show that if $q_1 > (q_2 - q_1)$ we have that pooling will be optimal. When pooling any two values v_k, v_{k+1} the same trade-off appears: pooling decreases the price at which the quality increment $(\overline{q}_{k+1} - \overline{q}_k)$ is being sold and increases the price at which the quality increment $(\overline{q}_k - \overline{q}_{k-1})$ is being sold (when $k = 1$, $q_{k-1} = 0$). Hence, a necessary condition for pooling not to be optimal is that $(\overline{q}_{k+1} - \overline{q}_k) > (\overline{q}_k - \overline{q}_{k-1})$, which is the convex menu condition. Note that there is no counterpart to this result in classic screening models without persuasion. In particular, when there is no persuasion the optimality of pooling only depends on the distribution of values and not on the distribution of qualities (see Lemma 1).

The property of increasing differences allows us further to exclude the possibility of an accumulation point in the menu, not only at the top of the distribution, but at any point of the distribution

[2]For example, ATT offers three options : 3 GB, 15GB, 50GB for hotspot data packages. (https://www.att.com/plans/wireless/) and apple offers memory at levels of, 128, 256, 5124GB (https://www.apple.com/shop).

except possibly at the bottom of the distribution; we rule out accumulation points at the bottom of the distribution using arguments similar to the ones presented in Section 4.3. Since there are no accumulation points and the space of values and qualities is compact, every optimal mechanism is finite. When the seller has a fixed inventory, we can go further and bound the number of items offered in any mechanism.

Corollary 2 (Finite Upper Bound on the Number of Items)
With a fixed inventory on support $[q_l, q_h]$, the number of items K offered by an optimal mechanism is bounded above by

$$K < \frac{q_h}{q_l}.$$

The lowest quality item will be at least $\overline{q}_1 \geq q_l$. Using the convexity result of Proposition 4, the k-th item must have quality strictly exceeding kq_l. However, the last item, the K-th item, cannot have a quality higher than q_h, so we have that $Kq_l < q_h$, which proves the result.

The property that the menu satisfies increasing quality increments is of interest as it informs us about the structure of the menu independent of the distribution of values. The result predicts that in any multi-item menu the distance between any item and its next lower ranked item is increasing as one moves up the quality ladder, thus establishing that the menu offers qualities that are increasing in a convex manner.

4.5 Optimality of Monotone Partition

So far we have shown that the optimal mechanism consists of a finite menu. We now use the linear structure of the preferences to provide a sharper characterization of the optimal information structure. More precisely, we show that the discrete jumps of the distributions are obtained by pooling intervals of values. The arguments employed in this section are orthogonal to those used in the previous subsections.

A distribution of values $\overline{F}$ is said to be *monotone partitional* if $[0, 1]$ is partitioned into countable intervals $[t_i, t_{i+1})_{i \in I}$ and each interval either has complete disclosure, i.e., all buyers with values corresponding to quantiles in that interval know their value; or pooling, i.e., buyers know only that their value corresponds to a quantile in the interval $[t_i, t_{i+1})$, and so their expected value is

$$\overline{v}_i \triangleq \mathbb{E}[v \mid F(v) \in [t_i, t_{i+1})].$$

The expectation can be written explicitly in terms of the quantile function as follows:

$$\overline{v}_i = \frac{\int_{t_i}^{t_{i+1}} F^{-1}(t)\, dt}{t_{i+1} - t_i}.$$

Thus writing J for the labels of intervals with complete disclosure, we have

$$\overline{F}^{-1}(t) \triangleq \begin{cases} F^{-1}(t), & \text{if } t \in [t_i, t_{i+1}) \text{ for some } i \in J; \\ \overline{v}_i, & \text{if } t \in [t_i, t_{i+1}) \text{ for some } i \notin J. \end{cases} \tag{32}$$

Proposition 1 in Kleiner et al. (2021) shows that the set $\{\overline{F}^{-1} : \overline{F}^{-1} \prec F^{-1}\}$ is a convex and compact set, and their Theorem 1 shows that the extreme points of this set are given by (32).

Corollary 3 (Monotone Partition of Values)
There exists an optimal mechanism $(\overline{F}, \overline{G})$ in which $\overline{F}$ is monotone partitional.

The corollary shows that the optimal information structure can be constructed in a straightforward manner: the value space is partitioned into intervals, and buyers are only told to which interval their value belongs. We note that the results in the previous subsections, in particular Theorem 1 and Proposition 2-4 do not make use of the monotone partitional structure of the optimal information structure. The result of a monotone partitional information structure relies on the linear utility in v, while versions of our main results continue to hold in nonlinear utility environment, see Theorem 5 in the working paper version (Bergemann et al. (2024)).

4.6 Two Examples of Optimal Mechanisms

We now illustrate the optimal solution in a mechanism with exogenous inventory. A model with exogenous inventory (as in Section 3) corresponds to a cost function of the form:

$$C(\overline{G}) = \begin{cases} 0, & \text{if } \overline{G} \prec_c G; \\ \infty, & \text{otherwise.} \end{cases} \tag{33}$$

In this case, the cost in the optimal solution of (22) is always 0 but there is a constraint on the set of feasible qualities. That is, the feasible distribution of qualities are $\overline{G} \prec_c G$. The constraints on qualities $\overline{G}$ is almost the same as the constraint on values $\overline{F}$, except that the seller can discard goods but cannot discard values. In Figures 1 we illustrate the optimal mechanism in an example. The top panel illustrates the exogenous distributions of values and qualities (F, G) that the seller

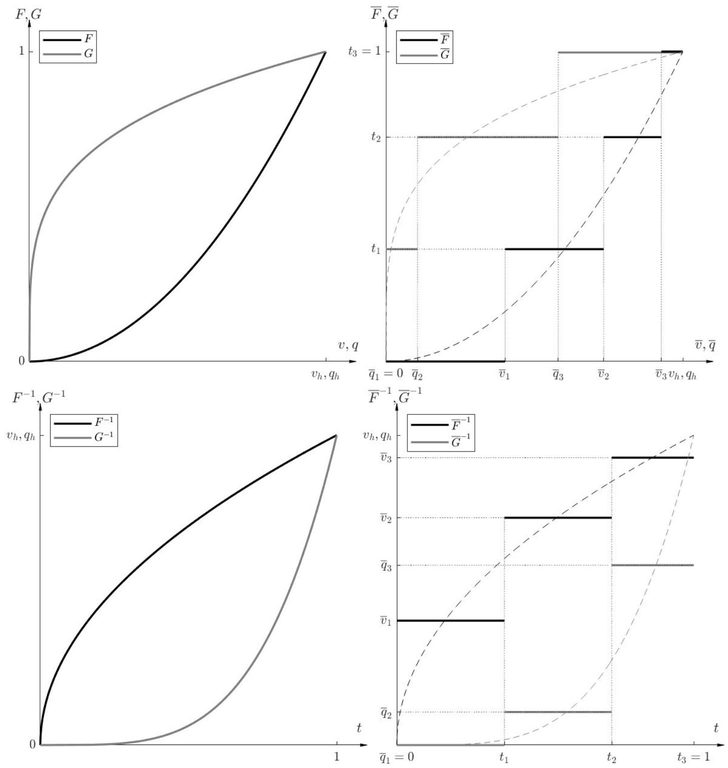

Figure 1: The given distributions of values $F(v) = v^2$ and qualities $G(q) = q^{1/4}$ are depicted on the top left. The associated optimal distributions $\overline{F}$ and $\overline{G}$, which are monotone partitional distributions are depicted on the top right.The corresponding quantile distributions $F^{-1}(t) = t^{1/2}$ and $G^{-1}(t) = t^4$ are at the bottom left. The optimal quantile distributions $\overline{F}^{-1}$ display jumps at the same quantiles at the bottom right.

can pool in the optimal mechanism $(\overline{F}, \overline{G})$ and the endogenous distributions of expected values and expected qualities generated by the optimal mechanism $(\overline{F}, \overline{G})$; the bottom panel illustrates the respective inverses.

The determination of the optimal menu with a variable supply follows the same logic. Below we display the optimal solution when the latent distribution is again given by a quadratic distribution of values, $F(v) = v^2$, but the supply of qualities is determined endogenously. We compute the solution under a quadratic cost function $c(q) = (1/2)\, q^2$. In Figures 2 we illustrate the optimal information structure and allocations. In Section 6, Theorem 3 establishes the optimality of a single item to be offered in the menu under sufficient conditions on distribution and cost. The current example with a quadratic distribution function and a quadratic cost function satisfies these sufficient condition. As might have been expected, the solution with a variable supply offers fewer items as it is costly to

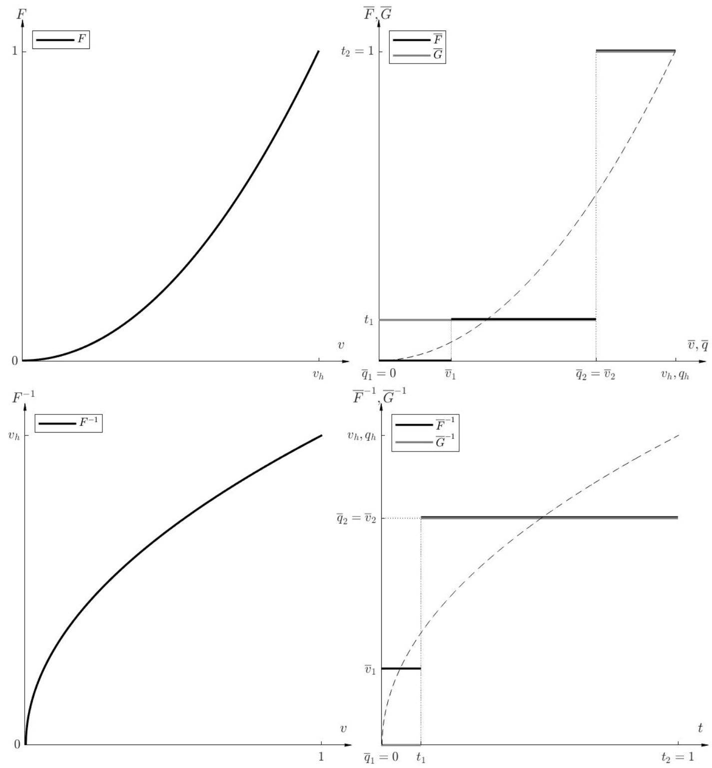

Figure 2: The given distribution of values $F(v) = v^2$ and the optimal distributions $\overline{F}$ and $\overline{G}$ given the cost function $c(q) = (1/2)q^2$ in the upper panel. The quantile distribution $F^{-1}(t) = t^{1/2}$ and optimal quantile distributions $\overline{F}^{-1}$ and $\overline{G}^{-1}$ given the cost function $c(q) = (1/2)q^2$ in the lower panel.

produce the supply. Perhaps less intuitively, the solution with variable quality supply can exclude fewer buyers than the solution with a fixed supply. But as we might expect, conditional on being offered a positive quality level, the quality offered is lower in the variable supply than in the fixed supply case for all latent values.

5 Pooling and Entropy

In Section 4 we provided a variety of results concerning the optimal mechanism, most notably, Theorem 1 that establishes that every optimal mechanism is finite. These results hold independent of the latent distribution of values F. An immediate corollary is that, if the latent distribution of values has infinite support, then the optimal mechanism will by necessity pool information.

Of course, if the latent distribution of values had finite support, we wouldn't be able to reach the same conclusion. So a natural question is: can we still assert that, in an optimal mechanism, the seller pools information regarding the value of the buyer? We now show that the answer is positive as long as the original distribution is not "too coarse". More precisely, we show that pooling is optimal as long as the entropy of the latent distribution of values is larger than $\log_2(9)$ bits. Note that, like Section 3, but unlike Section 4, we will not pursue general properties about the optimal mechanism, but simply explore when pooling some information is optimal. And thus, the answer will inevitably depend on the latent distribution of values.

Throughout this subsection we analyze the situation in which F has a distribution with support on a uniformly distanced (or equidistant) grid $\{v_1, ..., v_k, ...\}$, with atoms denoted by $\{f_1, ..., f_k, ...\}$, which we assume are strictly positive. The entropy of a discrete distribution F is defined by:

$$-\sum_k f_k \log_2(f_k). \tag{34}$$

The entropy is a measure of the amount of information that is contained in a random variable; we expand on the interpretation after we provide the result. The entropy of a random variable is independent of the domain of the random variable, so our assumption that the distribution is on a equidistant grid does not affect how entropy is measured. We define entropy using logarithms with base 2 so that it is measured in bits. We find an upper bound on the entropy generated by any distribution of values (on a equidistant grid) that satisfies a necessary condition for complete disclosure to be optimal, which we provide next.

Proposition 5 (Sufficient Condition for Optimality of Pooling)
Pooling improves revenue over complete disclosure if for some $k \in \{1, 2, ...\}$:

$$\frac{f_{k+1}}{\sum_{i \geq k} f_i} < \sqrt{\frac{f_k}{\sum_{i \geq k} f_i}} - \frac{f_k}{\sum_{i \geq k} f_i}. \tag{35}$$

This condition is a generalization of Corollary 1 and the earlier condition (15). In particular, the condition now covers the value of pooling of *any two adjacent* atoms, not only the lowest two values. We maximize (34) subject to:

$$\frac{f_{k+1}}{\sum_{i \geq k} f_i} \geq \sqrt{\frac{f_k}{\sum_{i \geq k} f_i}} - \frac{f_k}{\sum_{i \geq k} f_i}. \tag{36}$$

which is a necessary conditions for complete disclosure to be optimal. We show that the entropy-maximizing distribution is defined on an infinite grid and every constraint (36) is binding. Hence, we consider the family of distributions in which the first element f_1 is arbitrarily chosen, and each successive probability is found using that the constraints (36) are satisfied with equality.

Lemma 2 (Distribution with Binding Constraints)
For any f_1, there exists a unique distribution on an infinite support $\{\widehat{f}_1, \widehat{f}_2, ...\}$ such that every constraint (36) is satisfied with equality and $\widehat{f}_1 = f_1$.

We write the atoms of such distribution as $\widehat{f}_k(f_1)$, to emphasize that the atoms depend on the initial condition f_1. To prove this lemma, we write the inequality (36) in terms of the hazard rate h_i :

$$h_i \triangleq \frac{f_i}{\sum_{j \geq i} f_j},$$

where h_i is the hazard rate of i-th atom. With this the inequality(36) can be written as follows:

$$h_{i+1} \geq \frac{\sqrt{h_i} - h_i}{1 - h_i}. \tag{37}$$

The advantage of this expression is that the set of feasible hazard rates h_{i+1} depends only on the value of h_i, and we have that $f_1 = h_1$. Starting from the initial atom f_1, we find each successive atom using that the hazard rate is given by (37) (satisfied with equality). Since the hazard rates are always in $[0, 1]$ and the hazard rates converge to a unique fixed point at $h = \left(3 - \sqrt{5}\right)/2$, such a distribution can obviously be constructed for any initial probability f_1. In Figure 3 we illustrate (37): the shaded area represents the set of feasible hazard rates (h_i, h_{i+1}); the arrows show how the hazard rates evolve when (37) is binding. The hazard rates give an alternative definition of the critical distribution that maximizes entropy.

The entropy generated by the distribution that maintains all constraints (36) binding– parametrized by f_1– is given by:

$$E(f_1) \triangleq -\sum_{k \geq 1} \widehat{f}_k(f_1) \log_2(\widehat{f}_k(f_1)).$$

In the proof of Theorem 2 we show that the $E(f_1)$ is maximized by taking the limit as $f_1 \to 0$. We also show that:

$$E^* \triangleq \lim_{f_1 \to 0} E(f_1)$$

is bounded between:

$$\log_2(8) < E^* < \log_2(9) < \infty.$$

As reference, the entropy of a uniform distribution with support on N different values is given by $\log_2(N)$. Hence, complete disclosure is optimal only if F has an entropy below the entropy of a uniform distribution with support on 9 different values.

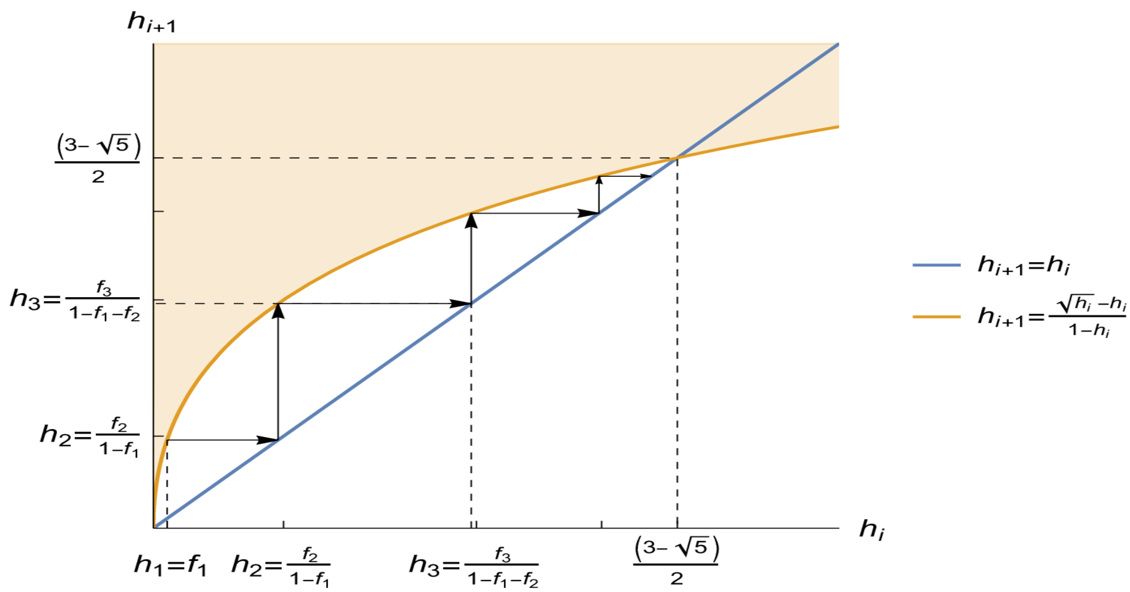

Figure 3: Behavior of the hazard rate h_i for the critical distribution $\widehat{f}$

Theorem 2 (Entropy and Optimality of Pooling Values)
The optimal mechanism has to pool values if the entropy of F is larger than $E^* \in (\log_2 8, \log_2 9)$.

The intuition for the result is that the atoms of the individual values in the distribution cannot have too small hazard rate. Otherwise, pooling the allocation of adjacent atoms generates small distortions to revenue while pooling information generates an order of magnitude larger revenue gains. With a uniform distribution the result is particularly stark, as illustrated in Section 3, and the same intuition goes through as long as both atoms being pooled are of similar size. However, when the hazard rate of the first atom is much smaller than the hazard rate of the second atom, then pooling can decrease revenue, even if both hazard rates are small. Hence, it may be possible to have "many small" atoms and still have complete disclosure being optimal. However, the sum of all these small atoms will end up having a negligible amount of mass, which is formalized by the fact that the entropy must be small.

To gain a more quantitative intuition for the result, consider a distribution with constant hazard rate h, that is, a Geometric distribution. The entropy of the Geometric distribution is:

$$E = -\frac{h\log_2(h) + (1-h)\log_2(1-h)}{h}.$$

As h converges to 0, the entropy diverges to infinity. Let's now contrast this with the dynamics in Figure 3. While the hazard rate might be initially small, the constraints guarantee that the hazard rates quickly increase. This leads to a bounded and relatively small entropy.

Theorem 2 states that there is an upper bound on the entropy distributions that induce complete disclosure to be optimal. The proof can be directly extended to show that in the optimal mechanism, there cannot be complete disclosure of any adjacent subset of values that have a *conditional entropy* larger than E^*. While we have assumed that the grid is uniform, the results would remain unchanged if the grid had increasing differences:

$$v_{i+1} - v_i \geq v_i - v_{i-1}.$$

Intuitively, we can see from (14) that in this case the conditions for pooling continue to hold (where one obviously needs to do the appropriate adjustments to account for all atoms and not only the first two).

6 Optimality of a Single-Item Menu

So far we have established that the optimal menu only invokes a finite number of items. We further proved that the pooling of values and thus the compression of information is optimal when the original distribution of values is not too coarse (as expressed by the entropy bound). A natural follow-up question is, does the optimal mechanism offer "few" items? (Naturally, the qualifier "few" would have different meanings according to the application.) An ideal answer would provide a sharp characterization of the number of items offered by the optimal mechanism as a function of the cost function and distribution of values. Such a general result is too ambitious, so we will instead pursue a more modest goal. In this section, we provide sufficient conditions under which the optimal menu only contains a single item. The conditions include many natural parametrizations (for example, uniform distribution and separable quadratic cost function) so this is a plausible outcome in many applications. Besides being a plausible outcome, the result will suggest that the optimal mechanism offers few items even when the conditions are not satisfied, and the intuition for the result will be informative of how properties of the cost function and distribution of values can lead to more or less pooling.

The results in previous sections were derived independently of the cost function, but a sharper characterization of the optimal mechanism necessarily depends on the cost function. Throughout this section, we assume that the seller has a separable cost function (see (3)). There are two reasons to focus on this case. First, it is the most widely studied model of second-degree price discrimination (see Mussa & Rosen (1978)). Hence, studying how properties of the cost function determine the number of items offered by a mechanism is of particular importance. Second, it allow us to build a connection between the values and the qualities (as these are endogenously determined), which provides enough structure to characterize when the optimal mechanism offers just one item.

To provide our single-item result we assume that the distribution of values is absolutely continuous and its density f is quasi-concave, and strictly concave on the decreasing part of the density. Formally, we say a distribution has *modest tails* if

$$f'(v) < 0 \Rightarrow f''(v) \leq 0. \tag{38}$$

This condition states that f must be concave when it is decreasing. For example, any distribution with (weakly) increasing density satisfies (38). It is also satisfied if the density is linearly decreasing. In contrast, the condition cannot be satisfied by any distribution with unbounded support. We say a mechanism offers a single-item menu if the range of $q(\overline{v})$ has at most two values and only one strictly positive value. Hence, every buyer is either offered a standard quality $q(\overline{v}) > 0$ or is excluded altogether $q(\overline{v}) = 0$.

Theorem 3 (Optimality of Single-Item with Modest Tails)
If the distribution F satisfies modest tails and $c'''(q) \geq 0$, then the optimal mechanism is a single-item menu.

Theorem 3 shows that in large class of environments the optimal mechanism is a single-item menu. The condition of the Theorem states that only a single-item is offered, but it does not imply that the seller finds it optimal to sell to all values, i.e. to pool all values. In other word, for some distribution and cost function the optimal mechanism consists of a single-item under which some buyers (with some values) are excluded (and hence do not buy any positive quality). Theorem 3 can be restated in the value space: It is optimal to provide a binary information structure, where the higher signal (higher expected value) suggests to buy the single item on offer, and the lower signal suggest to not buy. To gain some intuition we explain how the two assumptions of Theorem 3 lead to substantial pooling of values.

We begin by explaining why imposing a bound on the convexity of the marginal cost is expected to guarantee that a single-item mechanism is optimal. Consider some pooling of F into finitely many atoms. The quality offered to $\overline{v}_i$ is:

$$\overline{q}_i = c'^{-1}(\phi_i),$$

where ϕ_i is the discrete virtual value associated to $\overline{v}_i$ (as defined earlier in (8)). A necessary condition for a mechanism to be optimal is that the qualities generated by the mechanism satisfy the increasing difference property (see Proposition 4); this reflects that pooling is more likely to be optimal when the qualities offered to different buyers are more homogeneous. Hence, we can see how a "more convex" marginal cost pushes the trade-off to more pooling because the qualities generated by the mechanism (for any fixed information structure) are more homogeneous (since c'^{-1} becomes more concave as c' becomes more convex).

The second condition of the Theorem is the modest tails condition, which guarantees that the distribution of expected values generated by the mechanism cannot be too spread out. To gain intuition we provide a necessary condition for full pooling to be optimal. More precisely, we examine under what circumstances it is optimal to separate an infinitesimal sliver of values at the top of the distribution, starting from a mechanism that pools all values.

We denote by $\overline{q}_\mu$ the quality offered to the buyers when they are pooled (that is, when $\overline{v} = \mu_v$) and q_h the efficient quality offered to the highest value (that is, when $\overline{v} = v_h$):

$$\overline{q}_\mu \triangleq c'^{-1}(\mu_v) \text{ and } \overline{q}_h \triangleq c'^{-1}(v_h).$$

When all values are pooled the seller can leave the buyers with no surplus, but separating a small interval of values will generate some (infinitesimal) informational rents. The informational rents gained by the buyers in the interval that is being separated are:

$$U \triangleq \overline{q}_\mu(v_h - \mu_v).$$

That is, buyers being separated have a value of v_h and can buy $\overline{q}_\mu$ at price $p = \overline{q}_\mu \mu_v$, which gives U. Separating a sliver of high values also (infinitesimally) increases total surplus because the seller can offer them the efficient quality. The gains in total surplus from the separation are:

$$\Delta S \triangleq v_h(\overline{q}_h - \overline{q}_\mu) - (c(\overline{q}_h) - c(\overline{q}_\mu)).$$

Here we assume that the separation is small enough so that it does not affect the quality offered to the remaining pooled values.[3] The proposed separation increases profits if and only if $U < \Delta S$.

It is easy to check that U is linear in v_h while ΔS is convex in v_h with:

$$\frac{d\Delta S}{dv_h} \big|_{v_h = \mu_v} = 0.$$

[3]Separating a non-infinitesimal sliver would decrease the quality offered to the rest of the values, hence making ΔS small. This additional effect would tilt the balance towards more pooling, which would relax the necessary conditions for pooling to be optimal.

This condition is essentially saying that the surplus loss from pooling is of second order in the size of the distortions. Thus, when the highest value is sufficiently separated from the mean, then separating the infinitesimal sliver around the top increases profits, and a fortiori, full pooling will not be optimal. The analysis suggests (although it does not prove) that when the highest value is sufficiently close to the mean, then pooling will be optimal.

We can now relate the above intuition to the condition on the distribution of values. If f is non-decreasing, we have that $v_h/\mu_v \leq 2$ (where the inequality is tight when the distribution is uniform on $[0, v_h]$); if the distribution f is linearly decreasing, we then have that $v_h/\mu_v \leq 3$. Hence, the modest tails condition imposes a bound on the difference between the mean and the highest value of the distribution. Theorem 3 implies that $v_h/\mu_v \leq 3$ is not a sufficiently large difference for separating a sliver of high values to be optimal. For the separation to be optimal, we require f to be decreasing and convex, which is when the distribution does not have modest tails. If f is decreasing and sufficiently convex v_h/μ_v may be arbitrarily large, and some separation will always be optimal.

Theorem 3 provides general conditions under which a single-item mechanism is optimal. We know of distinct sufficient conditions under which a single-item mechanism is optimal and have provided additional results on how the cost function determines the optimal mechanism in Bergemann et al. (2024). For example, if the cost function is a power function $c(q) = q^\eta$ and the distribution of values has narrow support, more precisely $v_h/v_l < \eta$, then the optimal mechanism will pool all values (see Proposition 10 in Bergemann et al. (2024)). When the cost is a power function we were also able to show that full pooling is approximately optimal if η is large enough and complete disclosure is approximately optimal if η is close enough to 1. All these results support the intuitions we have provided in this section. Further, in an environment with a fixed inventory of qualities (see (33)), pooling all values is optimal if the distribution of qualities has increasing density (see Theorem 3 in Bergemann et al. (2024)).

7 Conclusion

In the digital economy, the sellers and intermediaries working on their behalf frequently have a substantial amount of information about the quality of the match between their products and the preferences of the buyers. Motivated by this, we considered a canonical nonlinear pricing problem that gave the seller control over the disclosure of information regarding the value of the buyers for the products offered.

We showed that in the presence of information and mechanism design, the seller offers a menu with only a small number of items. In considering the optimal size of the menu, the seller balances conflicting considerations of efficiency and surplus extraction. The socially optimal menu would provide a menu with a continuum of items to perfectly match quality and taste. By contrast, the profit-maximizing seller seeks to limit the information rent of the buyers by narrowing the choice to a few items on the menu. We provided sufficient conditions for a broad class of distributions under which this logic led the seller to offer only a single item on the menu. While we obtained our results in the model of nonlinear pricing pioneered by Mussa & Rosen (1978), we showed that the discrete menu result remained a robust property in a larger class of nonlinear payoff environments.

We analyzed a canonical model of second-degree price discrimination as in Mussa & Rosen (1978) or Maskin & Riley (1984). These models largely consider (pure) vertical differentiation among the buyers and in consequence in the choice and price of products. While vertical differentiation captures an important economic aspect, other specifications, in particular horizontal differentiation, might be of interest as well. Towards this end, we briefly discuss why horizontal differentiation is likely to lead to very different implications regarding the optimal information policy. Thus, consider a model of pure horizontal differentiation where there are many varieties of the product, and for each value of buyer there is some variety that attains the maximum value and all other varieties generate a lower surplus. Thus for example a utility function $u(v,q) = u - (v-q)^2$ would represent such a model of pure horizontal differentiation where the quadratic loss function expresses the fact that for every value v, there is an optimal variety, namely $q(v) = v$, and any deviation leads to a lower utility. In this setting of pure horizontal differentiation, the optimal information policy would be to completely disclose the information about the preferences, and then provide the optimal variety $q^*(v) = v$ at a constant price $p^* = u$ that would indeed extract the efficient social surplus from all values of buyers. This admittedly stark model of pure horizontal differentiation thus leads to a very different information policy than the model of vertical differentiation that we analyzed. For example, movie and TV series recommendations on Netflix and similar streaming services would seem to mirror the implications that a model of horizontal differentiation would predict. By contrast, service level agreements for utilities and telecommunications or tiered memberships for services would seem to be more directly related to the predictions from the vertical model we analyzed.

In related work, McAfee (2002) matches two given distributions of, say, consumer demand and electricity supply, and shows how discrete matching by pooling adjacent levels of demand and supply can approximate the socially optimal allocation. In this analysis, a range of different products are offered in the same class and with the same price. From the perspective of the buyers, the product offered is therefore *opaque*, as its exact properties are not known to the buyers who is only guaranteed certain distributional properties of the product. This practice is sometimes referred to as *opaque pricing,* see Jiang (2007) and Shapiro & Shi (2008) for applications to services and transportation and Bergemann et al. (2022) for auctions, in particular for digital advertising. Our analysis regarding the optimality of discrete menus would equally apply if we were to take the distribution of qualities as given and merely determine the partition of the distribution of the qualities. The novelty in our analysis is that the seller renders the preferences of the buyers opaque to find the optimal trade-off between efficient matching of quality and taste against the revenues from surplus extraction.

8 Appendix

The appendix contain all auxiliary results and the proofs omitted in the main body of the text.

Proof of Proposition 2. Using the change of variables

$$t = \overline{F}(\overline{v}) \Leftrightarrow \overline{F}^{-1}(t) = \overline{v},$$

and, given that $q(\overline{v})$ is non-decreasing, we can write the distribution of qualities in (6) in terms of the quantile $t \in [0,1]$:

$$\overline{G}^{-1}(t) \triangleq q(\overline{F}^{-1}(t)).$$

Using the expression for the payments (21) we have that:

$$\begin{aligned}\mathbb{E}[p(\overline{v})] &= \int_{v_l}^{v_h} \left(\overline{v} q(\overline{v}) - \int_{v_l}^{\overline{v}} q(s) ds \right) d\overline{F}(t) \\ &= \int_0^1 \left(\overline{F}^{-1}(t) \overline{G}^{-1}(t) - \int_0^t \overline{G}^{-1}(s) d\overline{F}^{-1}(s) \right) dt.\end{aligned}$$

Integrating by parts twice, the revenue generated by any incentive compatible mechanism is given by:

$$\mathbb{E}[p(\overline{v})] = \int_0^1 \overline{F}^{-1}(t)(1-t) d\overline{G}^{-1}(t) + \overline{F}^{-1}(0)\overline{G}^{-1}(0).$$

Of course, the above is only the revenue, and to write the profit we need to include the cost. Since F must satisfy the majorization constraint, we have that (22) is an upper bound for the profits that the seller can attain.

Consider $(\overline{F}, \overline{G})$ that solve (22). If $\overline{G}^{-1}$ is measurable with respect to $\overline{F}(\overline{v})$, then the mechanism in the proposition attains the upper bound. Suppose the measurability condition is not satisfied, then there exists $\{t_1, t_2\}$ such that $\overline{F}^{-1}(t)$ is constant in $[t_1, t_2]$ and $\overline{G}^{-1}(t)$ is not constant in $[t_1, t_2)$. We now define:

$$\widehat{G}^{-1}(t) = \begin{cases} \overline{G}^{-1}(t), & \text{if } t \notin [t_1, t_2); \\ \frac{\int_{t_1}^{t_2} \overline{G}^{-1}(t) dt}{t_2 - t_1}, & \text{if } t \in [t_1, t_2). \end{cases}$$

By construction we have that $\widehat{G} \prec_c \overline{G}$. We thus have that by construction $\widehat{G}$ is less costly to produce and it is easy to check that it generates (weakly) higher revenue. Hence, we can find a solution to (22) that satisfies the measurability condition and thus generates an optimal mechanism. ■

Proof of Proposition 4. In Proposition 3 we showed that the optimal mechanism is discrete. Hence, the mechanism consists of a countable collection of values, qualities and atom sizes $(\overline{v}_i, \overline{q}_i, \overline{f}_i)$

such that value $\overline{v}_i$ is allocated quality $\overline{q}_i$. We now provide a lemma that provides conditions for when a discrete mechanism is improvable.

Lemma 3 (Improvable Discrete Mechanism)
A discrete mechanism $(\overline{F}, \overline{G})$ *is not optimal if for any two consecutive atoms* $\{k, k+1\}$*:*

$$\frac{\overline{f}_{k+1}}{(\overline{f}_k + \overline{f}_{k+1})\left(1 - \frac{(\overline{f}_k + \overline{f}_{k+1})}{\sum_{j \geq k} \overline{f}_j}\right)} \left(1 - \frac{\overline{f}_k + \overline{f}_{k+1}}{\overline{f}_k} \frac{\overline{q}_k - \overline{q}_{k-1}}{\overline{q}_{k+1} - \overline{q}_k}\right) < \frac{\overline{v}_{k+2} - \overline{v}_{k+1}}{\overline{v}_{k+1} - \overline{v}_k}. \tag{39}$$

Proof. We introduce the language of a discrete mechanism with expected values $\overline{v}_k$ and qualities $\overline{q}_k$ at quantile t_k used earlier in (24):

$$\overline{v}_k \triangleq \overline{F}^{-1}(t_k); \qquad \overline{q}_k \triangleq \overline{G}^{-1}(t_k).$$

Incentive compatibility implies that $\overline{q}_k \leq \overline{q}_{k+1}$. We denote by $\overline{F}_k$ the probability that a value is at most v_k:

$$\overline{F}_k = \sum_{\{j: \overline{v}_j \leq \overline{v}_k\}} \overline{f}_j.$$

We can write the profits as follows:

$$\overline{\Pi} = \overline{v}_1 \overline{q}_1 + \sum_{k \geq 2} (\overline{q}_k - \overline{q}_{k-1}) \overline{v}_k (1 - \overline{F}_{k-1}).$$

This is the same expression as in (10). If atoms $\{k, k+1\}$ are pooled the difference in profits is given by:

$$\begin{aligned}
\Delta\Pi = &\left(\frac{\overline{f}_k \overline{q}_k + \overline{f}_{k+1} \overline{q}_{k+1}}{\overline{f}_k + \overline{f}_{k+1}} - \overline{q}_{k-1}\right) \left(\frac{\overline{f}_k \overline{v}_k + \overline{f}_{k+1} \overline{v}_{k+1}}{\overline{f}_k + \overline{f}_{k+1}}\right) (1 - \overline{F}_{k-1}) \\
&- \left(\overline{v}_k (\overline{q}_k - \overline{q}_{k-1})(1 - \overline{F}_{k-1}) + \overline{v}_{k+1} (\overline{q}_{k+1} - \overline{q}_k)(1 - \overline{F}_k)\right) \\
&+ \left(\overline{q}_{k+1} - \frac{\overline{f}_k \overline{q}_k + \overline{f}_{k+1} \overline{q}_{k+1}}{\overline{f}_k + \overline{f}_{k+1}}\right) \overline{v}_{k+2} (1 - \overline{F}_{k+1}).
\end{aligned}$$

We have that pooling these atoms is strictly optimal only if $\Delta\Pi > 0$. Re-arranging terms, we get the above inequality (39). ■

To conclude the proof of Proposition 4, we note that if $\overline{q}_k - \overline{q}_{k-1} > \overline{q}_{k+1} - \overline{q}_k$, then the left-hand-side of (39) is negative. Thus, in this case, the mechanism is not optimal. ■

Final Step of the Proof of Theorem 1. In Proposition 3 we established that the optimal mechanism is discrete and in Proposition 4 we proved that any optimal mechanism generates a convex menu of qualities. We now show that there are no accumulation points. Proposition 4 implies that there cannot be any accumulation points, except possibly at some quantile $\hat{t}$ satisfying $G^{*-1}(\hat{t}) = 0$. Hence, it is a decreasing accumulation point (that is, the limit of expected values converges to $\hat{t}$ from the right) and we recall ϕ_i is the discrete virtual value defined earlier in (8). Before we proceed, we provide a sufficient condition for a mechanism to be suboptimal.

Lemma 4 (Improvable Discrete Mechanism II)
A discrete mechanism $\left(\overline{F}, \overline{G}\right)$ is not optimal if for any two consecutive atoms $\{k, k+1\}$:

$$\frac{\overline{f}_k(\phi_{k+1} - \phi_k)(\overline{f}_k + \overline{f}_{k+1})}{\overline{f}_{k+1}(\overline{v}_{k+1} - \overline{v}_k)\sum_{j \geq k} \overline{f}_j} < 1. \tag{40}$$

Proof. If we omit the term in the parenthesis in (39):

$$\left(1 - \frac{\overline{f}_k + \overline{f}_{k+1}}{\overline{f}_k} \frac{\overline{q}_k - \overline{q}_{k-1}}{\overline{q}_{k+1} - \overline{q}_k}\right),$$

which is smaller than 1, then we obtain the following weaker bound:

$$\frac{\overline{f}_{k+1}}{(\overline{f}_k + \overline{f}_{k+1})(1 - \frac{(\overline{f}_k + \overline{f}_{k+1})}{\sum_{j \geq k} \overline{f}_j})} < \frac{\overline{v}_{k+1} - \overline{v}_k}{\overline{v}_k - \overline{v}_{k-1}}.$$

That is, if this condition is satisfied then the mechanism is not optimal. Re-arranging terms, we obtain (40). ■

We now consider any *infinite* optimal mechanism and show that it cannot be optimal by proving that (40) is satisfied. To make the notation more compact, we denote the left-hand-side of the inequality as follows:

$$B_k \triangleq \frac{\overline{f}_k(\phi_{k+1} - \phi_k)(\overline{f}_k + \overline{f}_{k+1})}{\overline{f}_{k+1}(\overline{v}_{k+1} - \overline{v}_k)\sum_{j \geq k} \overline{f}_j}.$$

We denote by $\{t_k\}_{k \leq K}$ the discontinuity quantiles, with $t_k < t_{k+1}$. The index runs from $-\infty$ to K. We denote the expected value in the limit by $\widehat{v} \triangleq \lim_{k \to -\infty} \overline{v}_k$.That is, the accumulation point is at $\widehat{v}$. We now define:

$$h_k \triangleq \frac{\overline{f}_k}{\overline{v}_{k+1} - \overline{v}_k},$$

and note that we can write ϕ_k as follows:

$$\phi_k = \overline{v}_k - \frac{(1 - t_{k+1})}{h_k}.$$

Since ϕ_k is increasing in k, we must have that the limit of ϕ_k and h_k exists and denote the respective limits as follows:

$$\widehat{\phi} \triangleq \lim_{k \to -\infty} \phi_k; \quad \widehat{h} \triangleq \lim_{k \to -\infty} h_k.$$

Since ϕ_k is positive we must have that $\widehat{h} > 0$. We must also have that $\widehat{h} < \infty$. To prove this, we note that for every k:

$$(\overline{v}_k - \widehat{v})(1 - t_k) = \sum_{\ell \leq k} g_l(\widehat{v} - \phi_l) \leq (\widehat{v} - \widehat{\phi}) \sum_{\ell \leq k} g_k.$$

The equality is obtained by simple algebraic manipulation of the terms while the inequality follows from the fact that the virtual values ϕ_k must be non-decreasing. Since $(\overline{v}_k - \widehat{v})(1 - t_k)$ is strictly positive for every k, we must have that $\widehat{v} > \widehat{\phi}$, which implies that $\widehat{h} < \infty$.

We now prove the result using that $\widehat{h} < \infty$. For this, we write the bound as follows:

$$B_k \leq \frac{h_k(\phi_{k+1} - \phi_k)(\overline{v}_{k+2} - \overline{v}_k) \max\{h_k, h_{k+1}\}}{(\overline{v}_{k+2} - \overline{v}_{k+1}) h_{k+1}(1 - t_{k-1})}.$$

Since $(\overline{v}_{k+1} - \overline{v}_k)$ converges to 0 in the limit $k \to -\infty$, we can consider intervals such that $(\overline{v}_{k+1} - \overline{v}_k) < (\overline{v}_{k+2} - \overline{v}_{k+1})$. We thus have that:

$$B_k \leq \frac{h_k(\phi_{k+1} - \phi_k) 2 \max\{h_k, h_{k+1}\}}{h_{k+1}(1 - t_{k-1})}.$$

Since ϕ_k is monotonic in k and positive, we must have that $(\phi_{k+1} - \phi_k)$ converges to 0 in the limit $k \to -\infty$. Taking the limit $k \to -\infty$, and using that h_k converges to $\widehat{h}$, we obtain:

$$\lim_{k \to -\infty} B_k \leq \lim_{k \to -\infty} \frac{h_k(\phi_{k+1} - \phi_k) 2 \max\{h_k, h_{k+1}\}}{h_{k+1}(1 - t_{k-1})} = 0.$$

Thus, in the limit $k \to -\infty$, we can find two consecutive intervals that can be pooled and increase revenue. ■

Proof of Corollary 3. The choice of information structure $\overline{F}$ must be optimal if we hold fixed a distribution G^* of qualities. So we consider the problem of choosing $\overline{F}$ to maximize

$$\Pi = \max_{\{\overline{F}^{-1} : F^{-1} \prec \overline{F}^{-1}\}} \int_0^1 \overline{F}^{-1}(t)(1 - t) dG^{*-1}(t) + G^{*-1}(0) F^{*-1}(0).$$

This optimization problem is an upper semi-continuous linear functional of $\overline{F}^{-1}$. Upper semi-continuity can be verified by noting that the quantile function $\overline{F}^{-1}$ is (by definition) upper semi-continuous. Hence, if $\widehat{F}^{-1} \to \overline{F}^{-1}$ (taking the limit using the L^1 norm), we have that $\limsup \widehat{F}^{-1}(t) \leq \overline{F}^{-1}(t)$ for all $t \in [0, 1]$. Hence, $\limsup \int_0^1 \widehat{F}^{-1}(t)(1 - t) dG^{*-1}(t) \leq \int_0^1 \overline{F}^{-1}(t)(1 - t) dG^{*-1}(t)$.

Following Bauer's maximum principle (Bauer (1958)), the maximization problem attains its maximum at an extreme point of $\{\overline{F}^{-1} : F^{-1} \prec \overline{F}^{-1}\}$. By Theorem 1, we excluded the possibility of intervals of complete disclosure. ■

Proof of Proposition 5. To obtain this result we omit the parenthesis in (39) and re-arrange terms. ■

Proof of Theorem 2. An upper bound on the entropy generated by any discrete distribution that induces complete disclosure to be optimal can be found by finding the distribution that maximizes entropy subject to (36):

$$E_K^* = \max_{\{f_1,\dots,f_K\}} -\sum_{k\geq 1} f_k \log_2(f_k), \text{ subject to: (36).} \tag{41}$$

Since (36) is a necessary condition for complete disclosure to be optimal, we have that this is an upper bound on the total entropy of any distribution that induces complete disclosure to be optimal among all distributions with a grid size of K. Throughout the analysis we assume that $5 \leq K < \infty$.[4]

We characterize the supremum over all grid sizes

$$E^* = \sup_{K\in\mathbb{N}} E_K^*.$$

This is an upper bound on entropy across all discrete-distributions that induce complete disclosure to be optimal. We will prove that the upper bound E^* is finite and also provide a way to compute the upper bound.

We begin the proof by analyzing the solution to (41) when $K < \infty$. We recall that the constraints are given by (36), which we can write as follows:

$$\frac{f_{k+1}}{\sum_{i\geq k} f_i} + \frac{f_k}{\sum_{i\geq k} f_i} - \sqrt{\frac{f_k}{\sum_{i\geq k} f_i}} \geq 0. \tag{42}$$

We first show that "most" of the constraints must bind.

Lemma 5 (Binding Constraints)
If $\{f_i^*\}_{i\in\{1,\dots,K\}}$ solves (41), then there exists $\iota \in \{K-5, K-4, K-3, K-2, K-1\}$ such that all the constraints with index $i \leq \iota$ in (36) bind.

[4] We disregard the cases with grid sizes less than 5 because the solution in these cases have a different structure; when $K \leq 4$ the entropy-maximizing distribution is the uniform distribution, so the constraints do not bind.

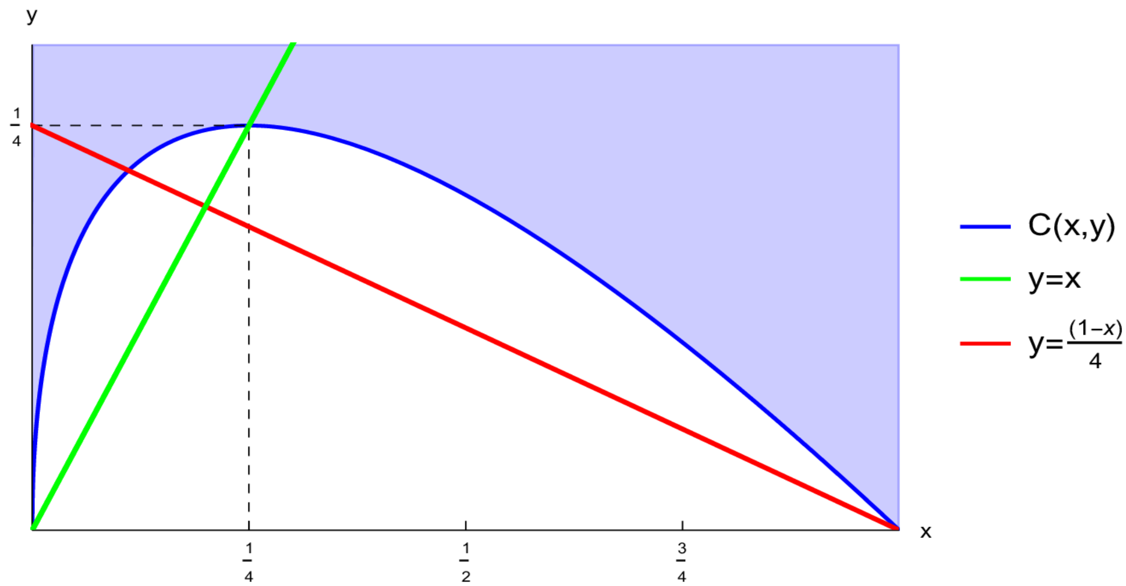

Figure 4: Geometric representation of the constraints via $\mathcal{C}(x, y)$.

Proof. We begin with some useful definitions. We define:

$$\mathcal{C}(x,y) \triangleq (x+y) - \sqrt{x},$$

as a compact representation of the constraints given by (42). For any distribution of values $\{f_1, ..., f_K\}$ we define:

$$f_{i,1} \triangleq \frac{f_i}{\sum_{j=i-1}^{K} f_j} \text{ and } f_{i,2} \triangleq \frac{f_i}{\sum_{j=i}^{K} f_j}.$$

Note that $f_{i,2}$ is the hazard rate, and $f_{i,1}$ is a modified hazard rate in which the denominator is evaluated one event before i. We can then write (42) as follows:

$$\mathcal{C}(f_{i,2}, f_{i+1,1}) \geq 0 \text{ for every } i \in \{1, ..., K-1\}.$$

We show the feasible set in Figure 4, where we also illustrate other curves that will be useful in the proof. We begin by providing some results about the feasible set.

Lemma 6 (Geometry of the Constraints)

1. *If* $\mathcal{C}(x, y) \geq 0$*, then* $\mathcal{C}(x, y') \geq 0$ *for all* $y' \geq y$.

2. *If* $\mathcal{C}(x, y) \geq 0$ *and* $x \leq 1/4$*, then* $\mathcal{C}(x', y') > 0$ *for all* $x' \leq x$ *and* $y' \geq y$.

3. *If* $\mathcal{C}(x, y) \geq 0$ *and* $y < (1-x)/4$, *then* $x < y$.

4. *If* $\mathcal{C}(x, y) \geq 0$ *and* $x \geq 1/4$ *then* $y/(1-x) > 1/4$.

We now prove that some of the final constraints must bind.

Lemma 7 (Last Constraint is Binding)
There exists constraint $i \in \{K-5, K-4, K-3, K-2\}$ *such that*

$$\mathcal{C}(f^*_{i,2}, f^*_{i+1,1}) = 0.$$

Proof. We prove by contradiction. Suppose that all these constraints are satisfied with slack. We then must have that:

$$f^*_{K-4} = f^*_{K-3} = f^*_{K-2} = f^*_{K-1} = f^*_K.$$

We then must have that $f^*_{K-4,2} = f^*_{K-3,1} = 1/5$. But in this case the constraint is not satisfied:

$$C(f^*_{K-4,2}, f^*_{K-3,1}) < 0,$$

so we reach a contradiction. ■

We thus have that some constraint must bind, and we denote by $\iota \in \{K-5, K-4, K-3, K-2\}$ the last binding constraint. That is,

$$C(f^*_{i,2}, f^*_{i+1,1}) > 0, \text{ for all } i > \iota. \tag{43}$$

We now proceed to show that the constraints $i \leq \iota$ must also bind. We assume that there exists $i \leq \iota$ such that $\mathcal{C}(f^*_{i,2}, f^*_{i+1,1}) > 0$ and reach a contradiction.

Case 1. We first consider the case that

$$f^*_{i+1,1} \leq \frac{1-f^*_{i,2}}{4}. \tag{44}$$

We note that:

$$f^*_{i+1,2} = \frac{f^*_{i+1,1}}{1-f^*_{i,2}}, \tag{45}$$

which can be obtained by simply using the definitions and manipulating the expressions. Hence, (44) implies that $f^*_{i+1,2} \leq 1/4$. We then consider the following distribution:

$$\tilde{f}_j = \begin{cases} f^*_j + \epsilon, & \text{if } j = i; \\ f^*_j - \epsilon, & \text{if } j = i+1; \\ f^*_j, & \text{otherwise;} \end{cases}$$

where ϵ is small enough such that $f^*_{i+1} - \epsilon > 0$ and $\mathcal{C}(\tilde{f}_{i,2}, \tilde{f}_{i+1,1}) > 0$.

We first prove that $\{\tilde{f}\}$ is feasible. For this, we note that

$$\mathcal{C}(\tilde{f}_{j,2}, \tilde{f}_{j+1,1}) = \mathcal{C}(f^*_{j,2}, f^*_{j+1,1}), \text{ for all } j \notin \{i-1, i, i+1\},$$

since these constraints do not change with ϵ. We assumed that ϵ is small enough such that constraint with index $j = i$ continues to be satisfied with slack. Hence, need to prove that

$$\mathcal{C}(\tilde{f}_{j,2}, \tilde{f}_{j+1,1}) \geq 0, \text{ for all } j \in \{i-1, i+1\}. \tag{46}$$

We first note that $\tilde{f}_{i-1,2} = f^*_{i-1,2}$ and $\tilde{f}_{i,1} > f^*_{i,1}$, so $\mathcal{C}(\tilde{f}_{i-1,2}, \tilde{f}_{i,1}) > 0$ (see item 1 of Lemma 6). We now note that $\tilde{f}_{i+1,2} < f^*_{i+1,2}$ and $\tilde{f}_{i+2,1} > f^*_{i+2,1}$, so $\mathcal{C}(\tilde{f}_{i+1,2}, \tilde{f}_{i+2,1}) > 0$ (see item 2 of Lemma 6). Thus (46) is also satisfied.

We now prove that $\{\tilde{f}\}$ generates higher entropy. We have that (44) and $\mathcal{C}(f^*_{i,2}, f^*_{i+1,1}) > 0$ imply that $f^*_i < f^*_{i+1}$ (see item 3 of Lemma 6). Hence, we have that

$$|\tilde{f}_i - \tilde{f}_{i+1}| < |f^*_i - f^*_{i+1}|.$$

Hence $\{\tilde{f}_j\}$ is feasible and generates higher entropy, thus reaching a contradiction.

(Case 2) We now consider the case that

$$f^*_{i+1,1} > \frac{1 - f^*_{i,2}}{4}. \tag{47}$$

We have that (45) continues to be satisfied so in this case we have that $f^*_{i+1,2} > 1/4$. Furthermore, $f^*_{j,2} \geq 1/4$ for all $j > i+1$ (see item 4 in Lemma 6). Without loss of generality, we can assume f^*_{i+1} is such that:

$$f^*_{i+1} = \underset{j \in \{i+1, ..., K\}}{\arg\max} \ f_j. \tag{48}$$

If multiple maximizers exist, then we take the lowest index that that maximizes the atom's size. To show this is without loss of generality, let j^* be the maximizing index. We now show that we must also have that $\mathcal{C}(f_{j^*-1,2}, f_{j^*,1}) > 0$, so we can relabel i to be $i = j^* - 1$. To check this, note that $f_{j^*-1,2} > 1/4$ and $f_{j^*} > f_{j^*-1}$ so the constraint must slack.

We then consider the following distribution:

$$\tilde{f}_j = \begin{cases} f^*_j - \epsilon. & \text{if } j = i+1; \\ f^*_j(1 + \frac{\epsilon}{\sum_{l=i+2}^{K} f^*_l}). & \text{if } j \geq i+2; \\ f^*_j, & \text{otherwise;} \end{cases}$$

where ϵ is small enough such that $f_j^* - \epsilon > 0$ and $\mathcal{C}(\tilde{f}_{i,2}, \tilde{f}_{i+1,1}) > 0$. We prove this distribution is feasible and generates higher entropy than $\{f^*\}$.

We first prove that $\{\tilde{f}\}$ is feasible. For all $j \notin \{i, i+1\}$ we have that $C(\tilde{f}_{j,2}, \tilde{f}_{j+1,1})$ is satisfied since the constraint does not change relative to $\{f^*\}$. For $j = i$ we have that $\mathcal{C}(\tilde{f}_{i,2}, \tilde{f}_{i+1,1}) > 0$ because ϵ is small enough. We thus check that $\mathcal{C}(\tilde{f}_{i+1,2}, \tilde{f}_{i+2,1}) \geq 0$. If $\mathcal{C}(f^*_{i+1,2}, f^*_{i+2,1}) > 0$, then the constraint will be trivially satisfied when ϵ is close enough to 0. We now suppose that $\mathcal{C}(f^*_{i+1,2}, f^*_{i+2,1}) = 0$, and we prove that $\mathcal{C}(\tilde{f}_{i+1,2}, \tilde{f}_{i+2,1}) \geq 0$. For this, we note that:

$$-\frac{\frac{\partial C(f^*_{i+1,2}, f^*_{i+2,1})}{\partial x}}{\frac{\partial C(f^*_{i+1,2}, f^*_{i+2,1})}{\partial y}} = \frac{1}{2\sqrt{f^*_{i+1,2}}} - 1;$$

$$\frac{\frac{\partial \tilde{f}_{i+2}}{\partial \epsilon}}{\frac{\partial \tilde{f}_{i+1}}{\partial \epsilon}} = \frac{\frac{f^*_{i+2}}{\sum_{j=i+2}^{K} f^*_j}}{-1} = -f_{i+2,2} = -\frac{\sqrt{f^*_{i+1,2}} - f^*_{i+1,2}}{1 - f^*_{i+1,2}}.$$

We thus have that:

$$\frac{\frac{\partial \tilde{f}_{i+2}}{\partial \epsilon}}{\frac{\partial \tilde{f}_{i+1}}{\partial \epsilon}} < \frac{\frac{\partial C(f^*_{i+1,2}, f^*_{i+2,1})}{\partial x}}{\frac{\partial C(f^*_{i+1,2}, f^*_{i+2,1})}{\partial y}}.$$

We have that $\tilde{f}_{i+1,2} < f^*_{i+1,2}$ and $\tilde{f}_{i+2,1} > f^*_{i+2,1}$, so for a small enough ϵ the constraint is feasible.

We now prove that $\{\tilde{f}\}$ generates higher entropy. For this, we write the derivative of the entropy:

$$\frac{\partial}{\partial \epsilon} - \sum_{j=1}^{K} \tilde{f}_j \log_2(\tilde{f}_j) = (\log_2(f^*_{i+1}) + 1) - \sum_{j=i+2}^{K} \frac{f^*_j}{\sum_{l=i+2}^{K} f^*_l} (\log_2(f^*_j) + 1).$$

Recall that (48) is satisfied, and hence, we have that:

$$|\log_2(f^*_{i+1})| < |\log_2(f^*_j)|, \text{ for all } j \in \{i+2, ...\}.$$

Hence, $\{\tilde{f}\}$ generates higher entropy. ■

We now characterize the distribution that maximizes entropy for a fixed N. For this, we describe the probabilities $\{f^*_{1,2}, ..., f^*_{\iota+1,2}\}$. If we have these hazard rates we can recover the true probability inductively using that:

$$f^*_i = f^*_{i,2}(1 - \sum_{j=1}^{i-1} f^*_j).$$

Finally, $\{f_{\iota+2}, ..., f^*_K\}$ are found using that the total probability must add up to 1 and these probabilities must all be the same.

We can find the different hazard rates $\{f^*_{1,2}, ..., f^*_{K-1,2}\}$ inductively using that (36) is satisfied with equality. For this we define we first define:

$$\mathcal{S}(x) \triangleq \frac{\sqrt{x} - x}{1 - x}.$$

We note that $\mathcal{C}(f_{i,2}, f_{i+1,1}) = 0$ if and only if

$$\mathcal{C}(f_{i,2}, f_{i+1,2}(1 - f_{i,2})) = 0.$$

Solving for $f_{i+1,2}$, we get that the constraint is satisfied if and only if

$$f_{i+1,2} = \mathcal{S}(f_{i,2}).$$

We denote by $\mathcal{S}^i$ the function composed with itself i times and S^0 is the identity function.

Corollary 4 (Characterization of Probabilities)
For every K, $\{f^*_1, ..., f^*_{K-1}, f^*_K\}$ maximizes (41) if and only if, there exists $f_0 \leq \frac{1}{2}\left(3 - \sqrt{5}\right)$ and $\iota \in \{K - 5, ..., K - 2\}$ such that:

$$f^*_{i,2} = \mathcal{S}^{i-1}(f_0), \text{ for all } i \in \{1, ..., \iota + 1\},$$
$$f^*_{\iota+2} = ... = f^*_K, \text{ for all } i \in \{\iota + 2, ..., K\};$$

and $\{f^*_{\iota+2}, ..., f^*_K\}$ are determined by the constraint that the probabilities must add up to one.

The Corollary shows how to find the probabilities that maximize entropy up to two parameters (f_0, ι). The Corollary described how to find the hazard rates $\{f^*_{i,2}\}$, but we can recover the absolute probabilities as follows:

$$f_i = (1 - \sum_{j=1}^{i-1} f_j)\mathcal{S}\left(\frac{f_{i-1}}{(1 - \sum_{j=1}^{i-1} f_j)}\right). \tag{49}$$

For any K, most of the constraints will be binding (in fact, all of them except at most 6). Motivated by this fact, we examine the entropy generated by a distribution with infinite grid points in which every constraint binds:

$$\widehat{E}(f_1) = \sum_{n=1}^{\infty} f_i \log_2(f_i), \text{where } \{f_i\}_{i>1} \text{ are found inductively as in (49).}$$

We now prove this is an appropriate bound for the total entropy E^* that can be attained by distributions that induce complete disclosure to be optimal.

Lemma 8 (Bound on Entropy with Infinite Support Distributions)
As $K \to \infty$,

$$\lim_{K\to\infty} E_K^* = \sup_{f_0\in[0,1]} \widehat{E}(f_0),$$

where possibly both sides of the equality are infinite. Furthermore the supremum of $\widehat{E}(f)$ is attained at $f = 0$.

Proof. We prove separately the cases in which $E_K^* \to \infty$ as $K \to \infty$ and the case in which E_K^* converges to a finite number as $K \to \infty$.

(Case 1) We first prove that if $E_K^* \to \infty$ as $K \to \infty$, then $\widehat{E}$ also diverges. To prove this, we first note that:

$$E_K^* = -\sum_{i=1}^{K} f_i^* \log_2(f_i^*) \leq -\sum_{i=1}^{K-4} f_i^* \log_2(f_i^*) + \log_2(5).$$

Hence, if $K \to \infty$, we have that:

$$\lim_{K\to\infty} \frac{-\sum_{i=1}^{K-4} f_i^* \log_2(f_i^*)}{E_K^*} = 1.$$

We now note that:

$$-\sum_{i=1}^{K-4} f_i^* \log_2(f_i^*) \leq \widehat{E}(f_1^*).$$

Hence, if $E_K^* \to \infty$ as $K \to \infty$, then we must also have that:

$$\sup_{f\in[0,1]} \widehat{E}(f) = \infty.$$

Note that, if $\widehat{E}(f)$ diverges, then it must also diverge in the limit $f \to 0$. To verify this, note that if $\widehat{E}(f) = \infty$ for some f, then $\widehat{E}(\mathcal{S}^{-1}(f)) = \infty$. Iterating the inverse of $\mathcal{S}$, we would get that $\widehat{E}(f) = \infty$ in the limit $f \to 0$.

(Case 2) We now prove that, if

$$\lim_{k\to\infty} E_K^* < \infty, \tag{50}$$

then it converges to the supremum of $\widehat{E}(f)$ which is attained in the limit $f \to 0$. We first prove that $f_K^* \to 0$ as $K \to \infty$. To prove this, suppose the grid size increases by 1 grid point so it is labelled as $\{1, ..., K, K+1\}$. In this case, we can split f_K^*, into two atoms, say $\widetilde{f}_K^*$ and $\widetilde{f}_{K+1}^*$ with

$$\mathcal{C}\left(\frac{f_{K-1}^*}{f_K^* + f_{K-1}^*}, \frac{\widetilde{f}_K^*}{f_K^* + f_{K-1}^*}\right) = 0 \text{ and } \widetilde{f}_K^* + \widetilde{f}_{K+1}^* = f_K^*.$$

This would clearly be a feasible distribution. If f_K^* is bounded away from zero, the increase in entropy will be bounded away from zero. However, if (50) is satisfied we then must have that the entropy increments as the grid size increases must converge to 0. Hence, we must have that $f_K^* \to 0$ as $K \to \infty$.

The fact that $f_K^* \to 0$ as $K \to \infty$ implies that:

$$\lim_{K\to\infty} E_K^* = -\lim_{K\to\infty} \sum_{i=1}^{\iota+1} f_i^* \log_2(f_i^*),$$

where ι is the last binding constraint (as defined in (43)). That is, in the limit we can omit the contribution of the last atoms to the entropy. However, we also have have that:

$$\lim_{K\to\infty} \sum_{i=1}^{\iota} f_i^* \log_2(f_i^*) = \lim_{K\to\infty} \widehat{E}(f_1^*),$$

where the limit $K \to \infty$ also includes how the probability f_i^* change with K. Thus, if (50) is satisfied,

$$\lim_{K\to\infty} E_K^* = \lim_{K\to\infty} \widehat{E}(f_1^*).$$

We must also have that:

$$\lim_{K\to\infty} \widehat{E}(f_1^*) = \sup_{f\in[0,1]} \widehat{E}(f),$$

as the supremum of $\widehat{E}(f)$ can be attained by E_K^* in the limit $K \to \infty$ by the distribution $\{\widehat{f}_k(f)\}$. Hence, the supremum of $\widehat{E}(f)$ is attained by the limit of f_1^*, as $K \to \infty$.

Finally, we verify that $f_1^* \to 0$ as $K \to \infty$. To prove this, suppose the grid size increases by 1 grid point so it is labelled as $\{0, 1, ..., K\}$. We can then construct the following distribution on the new grid:

$$\widetilde{f}_i^* = (1 - \widetilde{f}_0^*) f_i^*, \text{ for all } i \in \{1, ..., K\}.$$

Finally, we explain how to determine $\widetilde{f}_0^*$. We define x, y as follows:

$$x \triangleq \max\{w \in [0,1] : \mathcal{C}(z, f_1^*) \geq 0 \text{ for all } z \leq w\};$$
$$y \triangleq \arg\max_{z\in[0,1]} \{z \log_2 z + (1-z)\log_2(1-z) + (1-z)E_K^*\},$$

and let

$$\widetilde{f}_0^* \triangleq \min\{x, y\}.$$

This generates a feasible distribution and if f_1^* is bounded away from zero, the entropy increment will be bounded away from zero. Hence, if (50) is satisfied we then have that $f_1^* \to 0$ as $K \to \infty$. So the supremum of $\widehat{E}(f)$ is attained at $f = 0$. We thus obtain the result. ■

While are ultimately interested only in the maximum of $\widehat{E}(f)$, we can find this maximum by characterizing the function $\widehat{E}(\cdot)$ for all f. We can write the function $\widehat{E}(\cdot)$ recursively as follows:

$$\widehat{E}(f) = -f \log_2(f) - (1-f)\log_2(1-f) + (1-f)\widehat{E}(\mathcal{S}(f)). \tag{51}$$

We can now prove that $\widehat{E}(0)$ is finite.

Lemma 9 (Properties of $\widehat{E}(f)$)
The supremum of $\widehat{E}(\cdot)$ is finite.

Proof. We already proved the supremum of $\widehat{E}(f)$ is attained at $f = 0$. We can write (49) alternatively as follows:

$$\frac{\widehat{E}(f) - (1-f)\widehat{E}(\mathcal{S}(f))}{f - \mathcal{S}(f)} = \frac{-f\log_2(f) - (1-f)\log_2(1-f)}{f - \mathcal{S}(f)}.$$

Taking limits $f \to 0$, we obtain:

$$\lim_{f\to 0} \frac{\widehat{E}(f) - (1-f)\widehat{E}(\mathcal{S}(f))}{f - \mathcal{S}(f)} = \widehat{E}'(0);$$
$$\lim_{f\to 0} \frac{-f\log_2(f) - (1-f)\log_2(1-f)}{f - \mathcal{S}(f)} = 0.$$

We thus get that $\widehat{E}'(0) = 0$, which implies that $\widehat{E}(0) < \infty$. ■

We can now conclude the proof of Theorem 2 by computing $\widehat{E}(0)$. We observe that (51) is a linear functional and it is thus easy to verify numerically that $\widehat{E}(0) \approx 3.04$ while $\log_2\ 9 \approx 3.17$ and $\log_2\ 8 = 3$. We thus obtain the result of Theorem 2. ■

Proof of Theorem 3. We establish Theorem 3 through a sequence of optimal menus for successively richer environments. We begin with the optimal single-item menu.

Suppose that the seller were constrained to only offer a single item, which item would he then offer and at what price? The optimal single-item menu is found by solving the following problem:

$$(q^*, v^*) \in \underset{q,v\in\mathbb{R}}{\arg\max}\ \mathbb{P}[v' \geq v](\mathbb{E}[v' \mid v' \geq v]q - c(q)). \tag{52}$$

We denote by μ^* the expectation of v conditional on v exceeding v^*:

$$\mu^* \triangleq \mathbb{E}[v' \mid v' \geq v^*].$$

The single-item mechanism consists of selling quality q^* at a price $p^* = \mu^* q^*$, which is sold to all values higher than v^*. The buyer is only informed whether he should buy the good. Note that the buyer is left with no surplus.

Lemma 10 (Single Item Menu)
The optimal single item menu satisfies the following first-order conditions:

$$\mu^* = c'(q^*) \quad \text{and} \quad c(q^*) = v^* q^*, \tag{53}$$

Proof. The first-order condition is obtained by taking the derivative of (52) with respect to v and q and equating these to zero. ■

The first condition states that the quality is efficiently supplied given that the (expected) value of the buyer who buys the good is μ^*. The second condition states that the threshold v^* is also efficiently chosen: given that q^* quality is going to be supplied, it is efficient to sell to a buyer with value v if and only if the utility he obtains from this quality is larger than the cost of producing it. We note that the second equality might eventually be satisfied by some $v^* < v_l$, which means there is no exclusion. There are no distortions in the quality supplied (given the threshold v^*) and there is no distortion in the threshold v (given the quality q^*) because in a single-item mechanism there is zero buyer surplus. So these quantities are not distorted to reduce consumer surplus. In general, when the optimal mechanism is a multi-item mechanism, both the thresholds and the qualities provided are distorted to reduce the consumer surplus.

The rest of the proof proceeds as follows. We first show that when the cost is quadratic and the density is linearly decreasing, the optimal mechanism is a single-item mechanism. We then show that the optimal mechanism is a single-item mechanism when the marginal cost is convex and the density is linearly decreasing. Finally, we show that the distributions (38) are mean-preserving contractions of an (appropriately constructed) linearly-decreasing density, which we use to prove that the optimal mechanism is a single-item mechanism.

Before we begin, we introduce some notation. An optimal mechanism is described by cutoffs $\{w_1, ..., w_{K-1}\}$, with $w_K \triangleq v_h$, such that:

$$\bar{f}_k = \int_{w_{k-1}}^{w_k} dF(v); \quad \bar{v}_k = \int_{w_{k-1}}^{w_k} v dF(v).$$

The qualities offered are denoted by $\{q_1, ..., q_K\}$. To make the notation more compact, we define: $\Delta q_k = q_k - q_{k-1}$. Recall that in a finite-item menu, the revenue can be written as in (10). We consider the optimality conditions of the highest two intervals of an optimal mechanism. For this, we define the profit from the highest two items:

$$\Pi_{K-1,K}(w_{K-1}, \Delta q_{K-1}, \Delta q_K) \triangleq \big((\overline{f}_{K-1} + \overline{f}_K)\Delta q_{K-1}\overline{v}_{K-1} - \overline{f}_{K-1}c(q_{K-2} + \Delta q_{K-1}) + \overline{f}_K\left(\Delta q_K \overline{v}_K - c(q_{K-2} + \Delta q_{K-1} + \Delta q_K)\right)\big), \tag{54}$$

which are the last two terms of the summations in (10) and subtracting the respective cost. If the optimal mechanism is a multi-item mechanism, the solution to the following problem:

$$\Pi^*_{K-1,K} = \max_{\substack{w_{K-1} \in [w_{K-2}, v_h], \\ \Delta q_{K-1}, \Delta q_K \geq 0}} \Pi_{K-1,K}(w_{K-1}, \Delta q_{K-1}, \Delta q_K)$$

must satisfy $\Delta q_{K-1},\ \Delta q_K > 0$ and $w_{K-2} < w_{K-1} < v_h$, where q_{K-2} and w_{K-2} are parameters that are kept fixed in the optimization problem. We show the optimal mechanism is a single-item mechanism in a specific parametrization of the model.

Proposition 6 (Linear Density and Quadratic Cost Environment)
The optimal menu is always a single-item menu when the density is linearly-decreasing and the cost is linear-quadratic.

■

Proof. We analyze the optimal mechanism when the distribution of values is given by:

$$L(v; v_l, v_h) \triangleq \frac{(v - v_l)(2v_h - v_l - v)}{(v_h - v_l)^2}.$$

The density of this distribution, which we denote by $l(v; v_l, v_h)$, is linearly-decreasing with zero density at the top of the support:

$$l(v_h; v_l, v_h) = 0.$$

We begin by proving that a single-item mechanism is optimal when the cost is linear-quadratic:

$$c(q) \triangleq \alpha q + \frac{\beta q^2}{2} + \gamma. \tag{55}$$

The fixed cost γ plays no role in the analysis and is added to the cost function only to simplify the exposition of some arguments.

Given the quadratic cost function, the optimality conditions for q_{K-1} and q_K are:

$$\Delta q_{K-1} = \max\left\{\frac{\overline{v}_{K-1} - \alpha - \beta q_{K-2}}{\beta} - \frac{(\overline{v}_K - \overline{v}_{K-1})\overline{f}_K}{\beta \overline{f}_{K-1}}, 0\right\} \text{ and } \Delta q_K = \frac{\overline{v}_K - \alpha - \beta q_{K-1}}{\beta}. \tag{56}$$

Hence, $\Delta q_{K-1} > 0$ only if

$$\frac{\overline{v}_K - \alpha - \beta q_{K-2}}{\overline{v}_{K-1} - \alpha - \beta q_{K-2}} < \frac{\overline{f}_{K-1} + \overline{f}_K}{\overline{f}_K}. \tag{57}$$

To write expressions that are compact, we define:

$$z \triangleq \frac{v_h - w_{K-1}}{v_h - w_{K-2}}; \quad \kappa \triangleq 3\frac{w_{K-2} - \alpha - \beta q_{K-2}}{v_h - w_{K-2}}.$$

Note that in an optimal mechanism $\kappa \geq 0$, as otherwise the mechanism would be offering a quality q_{K-2} whose marginal cost is higher than the value for all values $v \in [w_{K-3}, w_{K-2}]$, which is clearly suboptimal. Using these definitions, we have that (57) is satisfied if and only if:

$$\frac{3 - 2z + \kappa}{1 - \frac{2z^2}{1+z} + \kappa} < \frac{1}{z^2}. \tag{58}$$

And, for every z satisfying (58), (54) can be written as follows:

$$\Pi_z \triangleq \frac{(\overline{f}_{K-1} + \overline{f}_K)(v_h - w_{K-2})^2}{18\beta\,(1 - z^2)} \left((3 - 2z + \kappa) z^2 \left(1 - 2z + \frac{4z^2}{1+z} - \kappa \right) + \left(1 + \kappa - \frac{2z^2}{1+z} \right)^2 \right). \tag{59}$$

More precisely, we have that

$$\Pi_z = \Pi_{K-1,K} + (\overline{f}_{K-1} + \overline{f}_K) c(q_{K-2}),$$

when (57) is satisfied. The last term on the right-hand-side is constant from the optimization perspective, so we can just focus on optimizing Π_z over z. If the optimal mechanism is a multi-item mechanism, there must exist $z \in (0,1)$ satisfying (58) that maximizes (59).

If z^* maximizes (59) and satisfies (58) with strict inequality, then z^* must satisfy the first- and second-order conditions. However, there is no $z^* \in (0,1)$ that satisfies the first- and second-order conditions:

$$\left.\frac{\partial \Pi_z}{\partial z}\right|_{z=z^*} = 0 \quad \text{and} \quad \left.\frac{\partial^2 \Pi_z}{\partial z^2}\right|_{z=z^*} \leq 0.$$

Hence, there is no interior solution. This is a contradiction, so in the optimal mechanism $\Delta q_{K-1} = 0$. ■

We now deploy the argument for the optimality of a single-item menu beyond the quadratic model. Towards this end, we define the solution to a restricted optimization problem for a linear-quadratic cost function with parameters α and β:

$$(v^*(\alpha,\beta), \Delta q_K^*(\alpha,\beta)) \triangleq \underset{0 \leq \Delta q, w_{K-2} \leq w \leq v_h}{\arg\max} \; \Pi_{K-1,K}(w, 0, \Delta q), \tag{60}$$

where we define the optimal quality for the last interval:

$$q_K = q^*(\alpha, \beta) \triangleq q_{K-2} + \Delta q_K^*(\alpha, \beta).$$

Thus, we consider a restricted optimization problem where the seller takes as given the first $K-2$ intervals and allocations. The restricted problem (60) is then to find an interval $(w_{K-1}, v_h] = (v^*(\alpha, \beta), v_h]$ and an allocation $q_K^*(\alpha, \beta)$ so as to maximize the profit from all values in the given interval $(w_{K-2}, v_h]$. This restricted maximization problem allows the interval $(w_{K-1}, v_h]$ to be a strict inclusion of $(w_{K-2}, v_h]$: that is, $(w_{K-1}, v_h] \subsetneq (w_{K-2}, v_h]$. In this case, all the values in the interval (w_{K-2}, w_{K-1}) will receive the allocation q_{K-2}. Now, from Proposition 6, we know that when the cost is linear-quadratic:

$$\Pi_{K-1,K}(w_{K-1}, \Delta q_{K-1}, \Delta q_K) < \Pi_{K-1,K}(v^*(\alpha, \beta), 0, \Delta q_K^*(\alpha, \beta)),$$

for every $\Delta q_{K-1},\ \Delta q_K > 0$. We add (α, β) as an argument because we will eventually vary these parameters; we don't add γ because the solution $(v^*(\alpha, \beta), \Delta q_K^*(\alpha, \beta))$ evidently does not depend on the constant γ.

We now analyze the entire class of convex cost functions with $c'''(q) \geq 0$. We assume that the optimal mechanism consists of multiple items and reach a contradiction. We denote by $\widehat{c}(q)$ a linear-quadratic cost function (as in (55)). We note that $c(q)$ and $\widehat{c}(q)$ intersect at most three times. Furthermore, if $c(q)$ and $\widehat{c}(q)$ are equal at qualities $q_1,\ q_2,\ q_3$, then the difference $\widehat{c}(q) - c(q)$ satisfies:

$$\widehat{c}(q) - c(q) \geq 0 \iff q \in (-\infty, q_1] \cup [q_2, q_3].$$

We use this for the following result.

Lemma 11 (Convex Marginal Cost Functions)
For every convex cost function with $c'''(q) \geq 0$ and for every (q_{K-2}, q_{K-1}, q_K) with $q_{K-2} \leq q_{K-1} \leq q_K$, there exists (α, β, γ) satisfying $c(q_{K-2}) = \widehat{c}(q_{K-2})$ and one of the following three conditions:

1. $c(q_K) = \widehat{c}(q_K);\ c(q_{K-1}) = \widehat{c}(q_{K-1});\ c(q^*(\alpha, \beta)) < \widehat{c}(q^*(\alpha, \beta))$;
2. $c(q_K) > \widehat{c}(q_K);\ c(q_{K-1}) = \widehat{c}(q_{K-1});\ c(q^*(\alpha, \beta)) = \widehat{c}(q^*(\alpha, \beta))$;
3. $c(q_K) = \widehat{c}(q_K);\ c(q_{K-1}) > \widehat{c}(q_{K-1});\ c(q^*(\alpha, \beta)) = \widehat{c}(q^*(\alpha, \beta))$.

Proof. We begin by considering α, β, γ chosen such that:

$$\widehat{c}(q_{K-2}) = c(q_{K-2});\ \widehat{c}(q_{K-1}) = c(q_{K-1});\ \widehat{c}(q_K) = c(q_K). \tag{61}$$

For this, we need to set the parameters α, β, γ as follows:

$$\alpha = \frac{c(q_K)\left(q_{K-2}^2 - q_{K-1}^2\right) + c(q_{K-1})(q_K^2 - q_{K-2}^2) + c(q_{K-2})\left(q_{K-1}^2 - q_K^2\right)}{(q_K - q_{K-1})(q_K - q_{K-2})(q_{K-1} - q_{K-2})},$$
$$\beta = \frac{2(c(q_K)(q_{K-1} - q_{K-2}) + c(q_{K-1})(q_{K-2} - q_K) + c(q_{K-2})(q_K - q_{K-1}))}{(q_K - q_{K-1})(q_K - q_{K-2})(q_{K-1} - q_{K-2})},$$
$$\gamma = \frac{c(q_K)q_{K-1}q_{K-2}(q_{K-1} - q_{K-2}) + c(q_{K-1})q_K q_{K-2}(q_{K-2} - q_K) + c(q_{K-2})q_K q_{K-1}(q_K - q_{K-1})}{(q_K - q_{K-1})(q_K - q_{K-2})(q_{K-1} - q_{K-2})}.$$

These are the coefficients one obtains from the interpolation of a second-degree polynomial. Since $\widehat{c}$ is a linear-quadratic cost function and since $c''' \geq 0$, we have that for all $q \geq q_{K-2}$:

$$c(q) \leq \widehat{c}(q) \iff q \in [q_{K-1}, q_K].$$

In other words, $\widehat{c}$ is equal to c at the qualities implemented by the mechanism and exhibits higher costs at qualities that are in between these two qualities and lower cost outside this interval. If

$$q^*(\alpha, \beta) \in [q_{K-1}, q_K],$$

then we are in Case 1 of Lemma 11. We now show that, if $q^*(\alpha, \beta) \notin [q_{K-1}, q_K]$, then we can find different α, β, γ such that we are in Case 2 or 3 of Lemma 11.

Suppose that:

$$q^*(\alpha, \beta) < q_{K-1}, \tag{62}$$

where (α, β) satisfy (61). We then need to find different parameters α, β. We consider parameters α, β as a function of q implicitly defined as follows:

$$\widehat{c}(q_{K-2}) = c(q_{K-2});\ \widehat{c}(q_{K-1}) = c(q_{K-1});\ \widehat{c}(q) = c(q).$$

We can write α, β, γ explicitly as before but replacing $c(q_K)$ with $c(q)$ and q_K with q. Since $\alpha,\ \beta,\ \gamma$ are functions of q, we write $\alpha(q),\ \beta(q), \gamma(q)$ and observe that they are continuous functions of q (while some of the denominators converge to 0 as $q \to q_{K-1}$, the limits exist). We also have that:

$$q^*(\alpha(q_K), \beta(q_K)) - q_K < 0 \text{ and } q^*(\alpha(q_{K-2}), \beta(q_{K-2})) - q_{K-2} \geq 0,$$

where the first inequality follows from (62) and the second inequality follows from the fact that q^* by definition is larger than q_{K-2} (see (60)) . Thus, following the intermediate value theorem, there exists a $\widehat{q} \in [q_{K-2}, q_K]$ such that: $q^*(\alpha(\widehat{q}), \beta(\widehat{q})) = \widehat{q}$. Furthermore, note that $q_K > \max\{\widehat{q}, q_{K-1}\}$, so we have that $\widehat{c}(q_K) < c(q_K)$. Thus, we are in Case 2 of Lemma 11.

Finally, if $q^*(\alpha, \beta) > q_K$, we can find α, β, γ such that Case 3 is satisfied in an analogous way to the case when (62) was satisfied. In particular, we consider parameters α, β as functions of q implicitly defined as follows:

$$\widehat{c}(q_{K-2}) = c(q_{K-2});\ \widehat{c}(q) = c(q);\ \widehat{c}(q_K) = c(q_K).$$

And we can show there exists $\widehat{q}$ such that $q^*(\alpha(\widehat{q}), \beta(\widehat{q})) = \widehat{q}$ and:

$$c(q_K) = \widehat{c}(q_K);\ c(q_{K-1}) > \widehat{c}(q_{K-1});\ c(q^*(\alpha, \beta)) = \widehat{c}(q^*(\alpha, \beta)).$$

This concludes the proof. ■

With Lemma 11 we can extend the optimality result to convex cost functions.

Proposition 7 (Optimality of Single-item Menu with Linear Density and Convex Cost) *The optimal menu with linear decreasing density and $c'''(q) \geq 0$ is always a single-item menu.*

Proof. We now suppose that the optimal mechanism satisfies $\Delta q_{K-1}, \Delta q_K > 0$ and reach a contradiction. In the same manner as (54), we define:

$$\widehat{\Pi}_{K-1,K} \triangleq$$
$$(\overline{f}_{K-1} + \overline{f}_K)\Delta q_{K-1}\overline{v}_{K-1} - \overline{f}_{K-1}\widehat{c}(q_{K-2} + \Delta q_{K-1}) + \overline{f}_K\left(\Delta q_K \overline{v}_K - \widehat{c}(q_{K-2} + \Delta q_{K-1} + \Delta q_K)\right).$$

Now $c(\cdot)$ is the true cost function, which satisfied $c'''(\cdot) \geq 0$, and $\widehat{c}(\cdot)$ is a linear-quadratic cost. So $\widehat{\Pi}_{K-1,K}$ is computed as $\Pi_{K-1,K}$ but using the linear-quadratic cost instead of the true cost. With a linear-quadratic cost the optimal mechanism is a single-item menu and thus:

$$\widehat{\Pi}_{K-1,K}(w_{K-1}, \Delta q_{K-1}, \Delta q_K) < \widehat{\Pi}_{K-1,K}(w^*_{K-1}(\alpha, \beta), 0, \Delta q^*_K(\alpha, \beta)).$$

We now consider the three cases in Lemma 11.

If we take (α, β) so that the first case in Lemma 11 holds, then we have:

$$\Pi_{K-1,K}(w_{K-1}, \Delta q_{K-1}, \Delta q_K) = \widehat{\Pi}_{K-1,K}(w_{K-1}, \Delta q_{K-1}, \Delta q_K); \tag{63}$$
$$\Pi_{K-1,K}(w^*_{K-1}(\alpha, \beta), 0, \Delta q^*_K(\alpha, \beta)) > \widehat{\Pi}_{K-1,K}(w^*_{K-1}(\alpha, \beta), 0, \Delta q^*_K(\alpha, \beta)). \tag{64}$$

We thus have that:

$$\Pi_{K-1,K}(w_{K-1}, \Delta q_{K-1}, \Delta q_K) < \Pi_{K-1,K}(w^*_{K-1}(\alpha, \beta), 0, \Delta q^*_K(\alpha, \beta)),$$

which contradicts the assumption that the multi-item mechanism is optimal.

If we consider (α, β) that satisfy the cases 2 or 3 of Lemma 11, then the argument is analogous but (64) will hold with equality and (63) will hold with strict inequality. ■

We now analyze distributions with modest tails. We begin with an important property of the optimal single-item mechanism when the distribution has a linearly-decreasing density. For these distributions, the first-order conditions (53) are necessary and sufficient conditions for optimality when $c'''(\cdot) \geq 0$.

Proposition 8 (Sufficient Conditions for Optimality)
If $c'''(q) \geq 0$, the distribution is $L(v; v_l, v_h)$, and $(\widehat{q}, \widehat{v})$ satisfy the first-order condition (53), then $(\widehat{q}, \widehat{v})$ solves (52), i.e. $(\widehat{q}, \widehat{v}) = (q^*, v^*)$.

Proof. When the distribution is linearly decreasing, we have that:

$$\mathbb{E}[v \mid v \geq \widehat{v}] = \frac{2\widehat{v} + v_h}{3}.$$

Hence, if $(\widehat{q}, \widehat{v})$ satisfy the first-order condition (53) we have that:

$$\frac{2\widehat{v} + v_h}{3} = c'(\widehat{q}) \quad \text{and} \quad \widehat{v} = \frac{c(\widehat{q})}{\widehat{q}}.$$

We have that:

$$v_h = 3c'(\widehat{q}) - 2\frac{c(\widehat{q})}{\widehat{q}}.$$

We now note that:

$$\frac{d}{dq}\left(3c'(q) - 2\frac{c(q)}{q}\right) = \frac{2}{q}\left(c''(q)q - c'(q) + \frac{c(q)}{q}\right) + c''(q).$$

If $c'''(q) \geq 0$ we have that $c''(q)q \geq c'(q)$ and hence:

$$\frac{d}{dq}\left(3c'(q) - 2\frac{c(q)}{q}\right) > 0.$$

Thus, there is a unique pair $(\widehat{v}, \widehat{q})$ such that the first-order condition is satisfied.

We verify that the first-order condition is sufficient for optimality. For this, we check that the solution is always interior, and since there is only one point that satisfies the first-order condition, this must be the optimum. We first note that $\widehat{q} \in \{0, \infty\}$ is clearly never optimal. It is also easy to see that $v = v_h$ cannot be an optimum as then the objective function of (52) is 0. We finally note that $v^* = 0$ is never optimal, which can be checked by noting that the first-order condition with respect to the cutoff gives $c(q^*) \leq v^* q^*$. Hence, the solution is always interior and it must be the only point that satisfies the first-order conditions. ■

For a given distribution F, we now introduce two related distributions, one generated by a linear decreasing density, and the other by a truncated version of the former. These latter two distributions are constructed in such a way as to allow us to compare the profit from the optimal mechanism under F (which we do not know) to the optimal mechanism under these two related distributions. Jointly with a cost-dominating argument, we can then establish the optimality of a single-item menu in a large class of environments.

Towards this end, we consider a distribution $L(v; \underline{z}, \overline{z})$ with a linearly-decreasing density where the lower and upper bounds of the distribution L, namely $\underline{z}, \overline{z}$, are chosen to satisfy the following properties relative to the distribution F and the optimal single-item threshold v^* under F given by (52):

$$L(v^*; \underline{z}, \overline{z}) = F(v^*) \text{ and } \mathbb{E}_L[v \mid v \geq v^*] = \mathbb{E}_F[v \mid v \geq v^*], \tag{65}$$

where the subscripts in the expectation indicate the distribution used to compute the expectation. Namely, at the threshold v^*, L and F obtain the same quantile, and the conditional expectation above the threshold v^* are identical. To satisfy these conditions, it is necessary to set:

$$\begin{aligned} \underline{z} =& 3\mathbb{E}_F[v \mid v \geq v^*] - 2v^* - \frac{3(\mathbb{E}_F[v \mid v \geq v^*] - v^*)}{\sqrt{1 - F(v^*)}}; \\ \overline{z} =& 3\mathbb{E}_F[v \mid v \geq v^*] - 2v^*. \end{aligned}$$

We also consider the following distribution $\widehat{F}(v)$:

$$\widehat{F}(v) = \begin{cases} L(\widehat{v}; \underline{z}, \overline{z}), & \text{if } v \in [0, \widehat{v}]; \\ L(v; \underline{z}, \overline{z}), & \text{if } v \in [\widehat{v}, \overline{z}]; \end{cases} \tag{66}$$

where $\widehat{v}$ is chosen such that:

$$\int_0^\infty v dF(v) = \int_0^\infty v d\widehat{F}(v).$$

In the proof of Lemma 12 we will show that indeed such a $\widehat{v}$ exists. Thus, $\widehat{F}(v)$ is constructed by taking the mass of the lower tail of $L(v; \underline{z}, \overline{z})$ and moving it to 0. In other words, $\widehat{F}$ is equal to $L(v; v_l, v_h)$ for $v \geq \widehat{v}$, and $\widehat{F}$ has an atom of size $L(\widehat{v}; \underline{z}, \overline{z})$ at $v = 0$.

We can now relate these three distributions in terms of stochastic orders.

Lemma 12 (Distribution Comparison)
Distribution $\widehat{F}$ is a mean-preserving spread of F and $\widehat{F}$ is first-order stochastically dominated by $L(v;\underline{z},\overline{z})$.

Proof. We first compare $L(v,\bar{z},\underline{z})$ with F. Since f satisfies (38) and $L'(v,\bar{z},\underline{z})$ is linearly decreasing, we must have that f and $L'(v,\bar{z},\underline{z})$ intersect at most twice. However, by construction L is constructed to satisfy (65), so they intersect exactly twice at two values $v_1, v_2 \geq v^*$. Hence, (38) implies that for all $v' \in [0,\infty)$:

$$\int_{F(v')}^{1} F^{-1}(v)dv \leq \int_{F(v')}^{1} L^{-1}(v;\overline{z},v_l)dv. \tag{67}$$

If the inequality is satisfied with equality for $v'=0$, we have that $\widehat{v}=\underline{z}$ and, otherwise, $\widehat{v}>\underline{z}$ (where $\widehat{v}$ is used to construct $\widehat{F}$ in (66)). Since $\widehat{F}$ is constructed by taking the mass of the lower tail of $L(v;\underline{z},\overline{z})$ and moving it to 0, it is transparent that $\widehat{F}$ is first-order stochastically dominated by $L(v;\underline{z},\overline{z})$. We have that (67) implies that for all $v' \geq \widehat{v}$:

$$\int_{F(v')}^{1} F^{-1}(v)dv \leq \int_{F(v')}^{1} \widehat{F}^{-1}(v)dv.$$

We also have that by construction $\widehat{F}$ has the same mean as F. It then follows that for all v'

$$\int_{F(v')}^{1} F^{-1}(v)dv \leq \int_{F(v')}^{1} \widehat{F}^{-1}(v)dv,$$

with equality for $v'=0$. Hence, $\widehat{F}$ is a mean-preserving spread of F (see Theorem 3.A.5 in Shaked & Shanthikumar (2007)). ■

We can now conclude the proof by establishing Theorem 3.

Final Step of the Proof of Theorem 3. We first verify that the optimal single-item mechanism when the distribution is $L(v;\underline{z},\overline{z})$ is the same as when the distribution is F. By construction of $L(v;\underline{z},\overline{z})$, the first-order condition that is satisfied for F is also satisfied for $L(v;\underline{z},\overline{z})$. Following Proposition 8, for the linearly decreasing density the first-order condition is sufficient for optimality, and thus (v^*,q^*) given by (52) do in fact form the optimal mechanism for L. We have that $L(v;\underline{z},\overline{z})$ generates at least as much profit as $\widehat{F}$, and $\widehat{F}$ generates at least as much profit as F. Since the optimal mechanism for distribution $L(v;\underline{z},\overline{z})$ is a single-item mechanism, and this mechanisms generates the same profit (by construction) under distribution F, this must also be the optimal mechanism under distribution F. ■

References

Anderson, E. & Dana, J. (2009). When is price discrimination profitable, *Management Science* **55**: 980–989.

Bauer, H. (1958). Minimalstellen von funktionen und extremalpunkte, *Archiv der Mathematik* **9**: 389–393.

Bergemann, D., Heumann, T. & Morris, S. (2024). Screening with persuasion, *Technical Report 2338 R3*, Cowles Foundation for Research in Economics, Yale University.

Bergemann, D., Heumann, T., Morris, S., Sorokin, C. & Winter, E. (2022). Optimal information disclosure in classic auctions, *American Economic Review: Insights* **4**: 371–388.

Bergemann, D. & Pesendorfer, M. (2007). Information structures in optimal auctions, *Journal of Economic Theory* **137**: 580–609.

Bergemann, D., Yeh, E. & Zhang, J. (2021). Nonlinear pricing with finite information, *Games and Economic Behavior* **130**: 62–84.

Blackwell, D. (1951). Comparison of experiments, *Proceedings of the Second Berkeley Symposium in Mathematical Statistics and Probability*, University of California Press, Berkeley, pp. 93–102.

DellaVigna, S. & Gentzkow, M. (2019). Uniform pricing in us retail chains, *Quarterly Journal of Economics* **134**: 2011–2084.

Ellison, G. (2005). A Model of Add-On Pricing, *Quarterly Journal of Economics* **120**(2): 585–637.

Eső, P. & Szentes, B. (2003). The one who controls the information appropriates its rents, *Technical Report Discussion Paper 1369*, Kellog School of Management.

Eső, P. & Szentes, B. (2007). Optimal information disclosure in auctions, *Review of Economic Studies* **74**: 705–731.

Guesnerie, R. & Laffont, J. (1984). A complete solution to a class of principal-agent problems with an application to the control of a self-managed firm, *Journal of Public Economics* **25**: 329–369.

Jiang, Y. (2007). Price discrimination with opaque products, *Journal of Revenue and Pricing Management* **6**: 118–134.

Johnson, J. & Myatt, D. (2003). Multiproduct quality competition: Fighting brands and product line pruning, *American Economic Review* **93**: 748–774.

Kamenica, E. & Gentzkow, M. (2011). Bayesian persuasion, *American Economic Review* **101**: 2590–2615.

Kleiner, A., Moldovanu, B. & Strack, P. (2021). Extreme points and majorization: Economic applications, *Econometrica* **89**: 1557–1593.

Loertscher, S. & Muir, E. (2022). Monopoly pricing, optimal randomization and resale, *Journal of Political Economy* **130**: 566–635.

Maskin, E. & Riley, J. (1984). Monopoly with incomplete information, *RAND Journal of Economics* **15**: 171–196.

McAfee, P. (2002). Coarse matching, *Econometrica* **70**: 2025–2034.

Mensch, J. & Ravid, D. (2022). Monopoly, product quality and learning, *Technical report*.

Mussa, M. & Rosen, S. (1978). Monopoly and product quality, *Journal of Economic Theory* **18**: 301–317.

Rayo, L. (2013). Monopolistic signal provision, *B.E. Journal of Theoretical Economics* **13**: 27–58.

Rayo, L. & Segal, I. (2010). Optimal information disclosure, *Journal of Political Economy* **118**: 949–987.

Sandmann, C. (2024). When are sparse menus profit maximizing?, *Technical report*.

Shaked, M. & Shanthikumar, J. (2007). *Stochastic Orders*, Springer Verlag.

Shapiro, D. & Shi, X. (2008). Market segmentation: The role of opaque travel agencies, *Journal of Economics and Management Strategy* **17**: 803–837.

Thereze, J. (2022). Screening costly information, *Technical report*, Princeton University.

Weiss, T. (2000). Amazon apologizes for price-testing program that angered customers, *Computerworld* .
URL: *https://www.computerworld.com/article/2588337/amazon-apologizes-for-price-testing-program-that-angered-customers.html*

Wilson, R. (1989). Efficient and competitive rationing, *Econometrica* **57**: 1–40.