SCREENING WITH PERSUASION

By

Dirk Bergemann, Tibor Heumann, Stephen Morris

July 2022

COWLES FOUNDATION DISCUSSION PAPER NO. 2338

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

Screening with Persuasion*

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

July 22, 2022

Abstract

We consider a general nonlinear pricing environment with private information. We characterize the information structure that maximizes the seller's profits. The seller who cannot observe the buyer's willingness to pay can control both the signal that a buyer receives about his value and the selling mechanism. The optimal screening mechanism has finitely many items even with a continuum of types. We identify sufficient conditions under which the optimal mechanism has a single item. Thus, the socially efficient variety of items is decreased drastically at the expense of higher revenue and lower information rents.

JEL Classification: D44, D47, D83, D84.

Keywords: Nonlinear Pricing, Finite Menu, Second-degree Price Discrimination, Recommender System

---

*We acknowledge financial support from NSF grants SES-2001208 and SES-2049744, and ANID Fondecyt Iniciacion 11200365. We thank seminar audiences at MIT IDSS, Boston College, Brown University, Emory University, 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, Stefano DellaVigna, Matt Gentzkow, Leon Musolff, Philipp Strack, and Alex Wolitzky. We thank Jack Hirsch, Shawn Thacker, 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 a world with a large variety of products and hence feasible matches between buyers and products, information can fundamentally affect the match between buyers and products. A notable feature of the digital economy is that sellers, or platforms and intermediaries that sellers use to place their products, commonly have information about the value of the match between any specific product and any specific buyer. In particular, by choosing how much information to disclose to the buyer about the value of the match between buyer and product, a seller can affect both the variety and the prices of the products offered.

We analyze the interaction between information and choice in a classic nonlinear pricing environment where the seller can offer a variety of products that are differentiated by their quality. We characterize the information structure and menu of choices that maximizes the expected profits of the seller. The buyer has a continuum of possible types–her willingness to pay for the quality. In the absence of any information design, the optimal menu offers a continuum of qualities to the buyer who then selects given her type as in Mussa and Rosen (1978). In this setting, we consider a seller who can control the selling mechanism and the information structure, but cannot observe the value or signal realization of the buyer. The selling mechanism could be any (possibly stochastic) menu.

We derive the basic structure of the optimal information and mechanism (in Section 3). The seller provides information in the form of a finite and monotone partition and consequently offers the buyer a finite menu (Theorem 1). The optimal menu is thus *short* relative to the menu with a continuum of choices in the absence of information design. The finiteness of the information structure has a straightforward intuition. Screening some open set of types requires maximizing the virtual surplus. By construction, distortions in the qualities from the profit-maximizing qualities will only cause second-order distortions to the total virtual surplus. But bundling a small interval of types into a single expected type causes a first-order decrease in the information rents. Hence, screening an open set of types is never optimal because pooling the types, and consequently the allocation, causes a first-order reduction in the information rents and only second-order distortions on profits. Thus the optimal menu with information design is small relative to the (continuum) menu in the absence of information design. But how small can the menu become?

We provide (weak) sufficient conditions on the cost function and value distribution under which a single-item menu is optimal (Theorem 2 in Section 4). The marginal cost must be (weakly) convex, and thus the cost function must be sufficiently convex and, in particular, more convex than quadratic costs. The value distribution must satisfy a *modest right tail* condition: The modest right tail condition holds if the buyer's expected valuation conditional on being above some threshold does not grow too fast in that threshold. The modest right tail condition is satisfied if the density is (i) uniform or linear, with any support; or, more generally, (ii) has a quasi-concave density which is concave on its decreasing component. Under the conditions of Theorem 2, the single-item menu may or may not involve some types being excluded, depending on the location of the support of the values. If the lower end of the support is far from the upper end and close to zero, then the single-item menu will lead to the exclusion of some types.

We report further results for the case where the cost function has a constant elasticity (in Section 5). These results offer insight into the structure of the optimal information and menu more generally, and, in particular, offer insight into the sufficient conditions for a single-item menu to be optimal as reported in Theorem 2. First, we provide a necessary and sufficient condition for the optimality of a single-item menu (in Proposition 10) depending only on the support of the distribution and the elasticity of the cost function. In particular, a single-item menu is optimal if the ratio of the upper bound of the support to the lower bound is less than the elasticity. Conversely, if this condition fails, there exists a distribution with that support where a single-item menu is not optimal. Note that this tight distribution-free condition will necessarily fail as the lower bound approaches zero, while the modest right tail condition assumed in Theorem 2 is easily satisfied even when the support includes 0. Second, we show that a single item menu will always be optimal for high enough elasticity and will never be optimal for sufficiently low elasticity, i.e., approach unit elasticity which corresponds to linear costs (Proposition 12). This explains why a lower bound on convexity of costs is required in Theorem 2. Note that a quadratic cost function has elasticity 2, and thus Theorem 2 requires a more convex cost than quadratic. Third, we establish that if we relax our maintained finite upper bound on values, the optimal menu is necessarily infinite (Proposition 11).

We provide a complete analysis of a model with only two values in Section 5. This analysis provides intuition for our general results. If the probability of a high-value buyer is too large relative to that of a low-value buyer, then the low-value buyer is excluded altogether and a single-item menu is optimal. This effect is already present without information design. However, even if the mass of low-value buyers is high enough, there will be no screening if the difference between values is not large enough. The reason is that the high-value buyer gains information rents from the quality provided to the low value buyer. These information rents are eliminated if the two types are pooled. The benefit of screening is that the high-value buyer receives the efficient allocation. The information rents grow linearly in the difference between high and low value whereas the efficiency gains are convex in the difference in values (for example, they are quadratic if the cost is quadratic). Hence, when the difference in values is too small, there will be pooling instead of screening. The critical ratio between the high value and low value below which pooling is optimal is increasing in the cost elasticity because the efficiency gains increase slower than the information rents. Our distribution-free result on the optimality of a single-item menu (Proposition 12) follows immediately from this observation. Conversely, if the difference between low and high values is large, there will be screening without exclusion. It is the existence of a high value-low probability type that gives rise to a failure of the modest right tail condition. If there is no upper bound on values in the general model (as in Proposition 11), the upper tail of the distribution plays the role of a high value-low probability type.

The binary type model also informs us about some of the welfare implications that come with the control of information. When the ratio of the high value to the low value is sufficiently large, the standard nonlinear pricing solution is to screen and exclude the low value buyer from the market. However, the optimal solution under information control is to pool the information and offer a single item to the entire market. In either case, the buyers receive zero information rent but the profits of the seller increases, and thus so will the social welfare. Thus, the ability to manage the information of the buyer can be social welfare increasing, but admittedly it is the seller who benefits rather than the consumers.

The above results are obtained in the setting first proposed by Mussa and Rosen (1978) where the product of willingness to pay and quality generate the gross utility of the buyer. In Section 6 we extend our analysis to general nonlinear utility functions of the buyer. Here we establish that all our previous results carry over entirely as long as willingness to pay and quality permit a multiplicative separable representation. We then further weaken the payoff environment to allow for general monotone and supermodular payoff functions. The final result, Theorem 3, establishes that the disclosure of any open set remains a suboptimal information policy. Theorem 3 thus generalizes Theorem 1, but with weaker implications. In particular, we show by means of two examples that a monotone partition is not always an optimal information structure under supermodularity conditions alone.

Our setting reflects three notable features of the digital economy. We already mentioned the fact that the sellers are well-informed about buyers' values. Our analysis considers the extreme case where the buyer only knows the prior and the seller has access to all feasible signals. A second notable feature is that the buyers have the ability to find which items are available at what price, due to search engines and price comparison sites. Thus, personalized prices (or more generally third-degree price discrimination) are not available, but menu pricing (or more generally second-degree price discrimination) can occur. Finally, particular items, that is quality-price pairs, are recommended to different buyers via recommendation and ranking services.

Our leading interpretation is that the seller can influence the information that the buyer has about his value but does not observe that value. Although we will not pursue it formally in the paper, an alternative interpretation of our model is that the seller does in fact observe the buyer's value but is unable, for regulatory or business model reasons, to offer prices for item that depend on the buyer's value. Thus, the seller cannot engage in perfect price discrimination (or third-degree price discrimination). In fact, the seller is constrained to offer a common menu of items. However, as long as all buyers are offered the same menu, he is allowed to credibly convey information about buyers' values. Now the implementation of the optimal information structure and selling mechanism is that the seller posts a menu and sends a signal to the buyer that recommends one item on the menu. The resulting recommendation policy is one which we commonly observe on e-commerce platforms. Namely, the seller does not engage in third-degree price discrimination, but rather, among the range of possible choices, every buyer is steered to a specific alternative at a price that is common to all consumers. This implementation makes sense in our model if we impose the interim obedience constraint that recommendations are optimal for the buyer conditional on the recommendation received.

Consistent with this interpretation, eBay personalizes the search results for each buyer through a machine learning algorithm and determines a personalized default order of search results in a process referred to as "Best Match," see eBay (2022). DellaVigna and Gentzkow (2019) provide strong evidence that large chains price uniformly across stores despite wide variation in consumer demographics and competition. Further, Cavallo (2017), (2019) documents that online and offline prices are identical or very similar for large multi-channel retailers, thus confirming the adherence to a uniform price policy. Related, Amazon apologized publicly to its customers when a price testing program offered the same product at different prices to different consumers, and committed to never "price on consumer demographics," see Weiss (2000).

1.2 Related Literature

We consider a model of nonlinear pricing as first analyzed in Mussa and Rosen (1978). Similar to Mussa and Rosen (1978), we focus on offering a menu of different qualities which can be produced at an increasing and convex cost. Our analysis could similarly be applied to the setting of quantity differentiation as in the model of Maskin and Riley (1984) with constant marginal costs (as we will discuss in Section 6).

In our analysis, the seller can control the information and the mechanism. It therefore combines Bayesian persuasion (Kamenica and Gentzkow (2011)) or information design more generally with mechanism design tools. We thus offer a solution to an integrated mechanism and information design problem in a classic economic environment. Perhaps surprisingly given the proximity of the tools as highlighted in the recent work by Kleiner, Moldovanu, and Strack (2021), we are not aware of any related work in optimal pricing that combines mechanism and information design. The closest work in this sense is Bergemann and Pesendorfer (2007), who consider a seller with many unit-demand buyers. There the seller can both choose the selling mechanism and the information each buyer receives about their own value. Their main result, showing that sellers have each buyer observe a finite and monotone partition about their value, relates to the current result. Because unit demand allocation is a linear problem, they are able to establish the optimality of a coarse partition by construction. Our nonlinear pricing analysis has to determine not only whether to assign the object, but also at what level of quality. Thus, the current results on the cardinality of the menu and the sufficient condition for a single-item menu do not have a counterpart in Bergemann and Pesendorfer (2007). Brooks and Du (2021), like Bergemann and Pesendorfer (2007), consider a many-player, single-unit setting where both the selling mechanism and the buyers' information are endogenous in some sense. However, there the seller is choosing the mechanism to maximize revenue under the worst-case information structure, so it is as if an adversary picks the information structure. Thus, the seller does not jointly design the information structure and selling mechanisms.

Our analysis is an instance of second-degree price discrimination. But the seller also creates segments (or pools) within the market. In doing so, he makes items intended for other segments less attractive to each buyer. In this sense, the seller is inducing partial third-degree price discrimination. By contrast, Bergemann, Brooks, and Morris (2015) and Haghpanah and Siegel (2022) explicitly allow full third-degree price discrimination, while also allowing discrimination within each segment by offering quality-differentiated products. Roesler and Szentes (2017) consider the buyer-optimal information structure with single-unit demand. Thus, the demand structure and the objective differ from the current work, but they share the focus on creating segmentations with a single aggregate market. Anderson and Dana (2009) ask when second-degree price discrimination is profitable in an environment where there is a priori a finite upper bound on the quality provided. They provide conditions under which all types receive the same quality, namely the quality at the upper bound. The condition can be stated in terms of log submodular versus log supermodular preferences. By contrast, our environment is log supermodular everywhere, and hence there is no reason to restrict the menu and offer a bunching solution in the sense of Anderson and Dana (2009).

Another explanation for pooling is that the distribution of values is "irregular" in the sense of Myerson (1981). This would imply an interval over which buyers with different values were offered the same quality, as established already in Mussa and Rosen (1978).

Loertscher and Muir (2022) consider a seller who offers fixed quantities of products of different qualities. They show that bundling different qualities, or randomizing the quality assignment via lotteries, can increase the revenue in the presence of irregular type distributions. Our results hold for regular distributions as well as irregular ones.

Rayo (2013) considers a model of social status provision that shares some features with our model. The utility function of the agent before any transfer is a product of his type (or an increasing function of his type) and a social status which is equal to his *expected type* given some information structure. Rayo (2013) then asks what is the optimal information structure to provide to the agent by a revenue maximizing monopolist. Thus, the allocation in Rayo (2013) is an information structure rather than a quality allocation. In consequence, the allocations all have the same constant cost, namely zero. In addition, the information structure only affects the allocation and not the expectation of the buyer regarding his own type. The main result, Theorem 1, is that the optimal information structure–he restricts attention to deterministic information structures–has an interval structure.

In Rayo (2013) and Loertscher and Muir (2022), distinct products are pooled and the buyer is offered a correspondingly smaller choice set. In our model, the optimal information design pools the types of the buyers and creates coarse information for the buyers. In turn, the coarse information leads to an optimally short menu of choices. While the prediction of the optimal nonlinear pricing problem is a continuum of choices–thus a very long menu–in many real-world application, the cardinality of the menus is finite and the menu is often very short.

2 Model

A seller supplies goods of varying quality $q \in \mathbb{R}_+$ to a buyer. The buyer has a willingness-to-pay (or value or type) $v \in \mathbb{R}_+$ for quality $q \in \mathbb{R}_+$. The utility net of the payment $p \in \mathbb{R}_+$ is:

$$u(v, q, p) = vq - p. \tag{1}$$

The value $v \in \mathbb{R}_+$ is distributed according to $F \in \Delta([\underline{v}, \bar{v}])$, with support $0 \leq \underline{v} < \bar{v} < \infty$, with strictly positive and non-vanishing density, except possibly at the upper bound (i.e., $F'(v) > 0$ for all $v < \bar{v}$). The seller's cost of providing quality q is $c(q)$, where $c : \mathbb{R}_+ \to \mathbb{R}_+$ is assumed to be a (weakly) increasing and (weakly) convex function.

The seller can choose the information the buyer has about their value and a menu of qualities and corresponding prices.

First, the seller chooses a signal (or information structure) $s : \mathbb{R}_+ \to \Delta\mathbb{R}_+$, where $s(v)$ is a signal realization observed by the buyer when the value is v. The buyer's expected value conditional on the signal realization s is denoted by:

$$w \triangleq \mathbb{E}[v \mid s]. \tag{2}$$

Since the utility is linear in v, w is a sufficient statistic for determining the buyer's preferences when they observe signal s. We denote by G the distribution of expected values; $\text{supp}(G)$ is the support of G. We refer to s, or more directly the induced distribution G of expected values, as the information structure. Second, the seller chooses a menu (or direct mechanism) with qualities $q(w)$ at prices $p(w)$:

$$M \triangleq \{(q(w), p(w))\}_{w \in \text{supp} G}$$

where the mechanism has to satisfy incentive compatibility and participation constraints. Thus for all $w, w' \in \text{supp}(G)$:

$$wq(w) - p(w) \geq wq(w') - p(w'); \tag{3}$$

$$wq(w) - p(w) \geq 0. \tag{4}$$

We refer to a mechanism as a pair (G, M) of information structure G and menu M. The seller's problem is to maximize expected profits, revenues minus cost:

$$\Pi \triangleq \max_{\substack{s:\mathbb{R} \to \Delta(S) \\ (q(w), p(w))}} \mathbb{E}[p(w) - c(q(w))].$$

The expected social surplus and the buyer's surplus are denoted by S and U, respectively:

$$S \triangleq \mathbb{E}[q(w)w - c(q(w))]; \quad U \triangleq \mathbb{E}[q(w)w - p(w)]. \tag{5}$$

We note that the direct mechanism M could also be expressed in terms of a simple indirect menu where the seller chooses (Q, p) consisting of a set of qualities $Q \subseteq \mathbb{R}_+$ and a pricing rule $p : Q \to \mathbb{R}_+$. Then the information structure can be expressed as a recommendation rule $s : \mathbb{R}_+ \to \Delta Q$.

3 Structure of the Optimal Mechanism

We now provide a complete characterization of the optimal mechanism for the general environment with a continuum of values. The main qualitative feature is that the optimal information structure partitions values into intervals. Subsequently, the buyer only learns to which element of the partition their realized value belongs.

3.1 Finite and Monotone Partition

We say that the information structure G has interval structure if there exists some countable collection of intervals $\{[\underline{x}_i, \bar{x}_i) \mid i \in I\}$, such that:

$$G^{-1}(q) = \begin{cases} F^{-1}(q), & \text{if } q \notin \cup_{i \in I}[\underline{x}_i, \bar{x}_i); \\ \frac{\int_{\underline{x}_i}^{\bar{x}_i} F^{-1}(t)dt}{\bar{x}_i - \underline{x}_i}, & \text{if } q \in [\underline{x}_i, \bar{x}_i). \end{cases} \tag{6}$$

We define the intervals in the quantile space rather than the value space. In other words, G has interval structure if it can be constructed from partitioning the space of quantiles and the associated values into intervals–possibly countably infinite. There is either *complete disclosure* and $q \notin \cup_{i \in I}[\underline{x}_i, \bar{x}_i)$, and then the buyer learns the value; or there is *pooling* and $q \in [\underline{x}_i, \bar{x}_i)$, and the buyer only learns that their value (and quantile) is somewhere in the interval $q \in [\underline{x}_i, \bar{x}_i)$. We say that G has a *finite pooling* interval structure if G has interval structure, I is finite and there is no interval of full disclosure, that is, $\bar{x}_i = \underline{x}_{i+1}$ for all i.

Our first main result establishes that the optimal information structure is a finite pooling interval structure. Here we have to add a qualifying remark. To the extent that some values may not receive a positive quality in the optimal mechanism, there may be some multiplicity in the information structure. For example, values which do not receive the good (they obtain quality zero) may or may not be pooled. But it is without loss of generality for the optimal revenue to always pool all values that receive zero quality. Hence, we consider mechanisms (G, M) such that, if $q(w) = q(w') = 0$ for any pair $w, w' \in \text{supp}(G)$, then $w = w'$. In other words, all values who are not served a positive quality are pooled in the same partition. Of course, this will not change the nature of the optimal mechanism beyond disciplining the information provided to values who do not buy a positive quality.

Theorem 1 (Structure of the Optimal Mechanism)
In every optimal mechanism the information structure is a finite pooling interval structure and the optimal quality $q^*(w)$ is increasing in w on the support of G.

This first main result therefore establishes that every optimal mechanism consists of a finite pooling information structure. In consequence, the optimal menu will contain only a finite number of items; it is a *short menu.* This contrasts with a long menu with a continuum of items which would be optimal in the absence of a choice regarding the information structure.

Corollary 1 (Finite Support)
The optimal information structure has finite support and the optimal menu offers a finite number of qualities.

We will prove Theorem 1 in several steps. In the first step we write the payment p as a function of the qualities q by applying the envelope theorem to the buyer's utility maximization problem (for example, as in Myerson (1981)).

The second step shows that there exists an optimal mechanism in which the optimal information structure has interval structure. This step will be proven by showing that, given a menu of qualities and prices, the seller's maximization problem is linear in the quantile function of the expected values. Hence, we can use recent results in Kleiner, Moldovanu, and Strack (2021) to characterize the optimal information structure in terms of the extreme points of the set of quantile functions that are a mean-preserving spread of the quantile function of values. We then proceed to show that in *every* optimal mechanism the information structure has interval structure. This follows from the fact that the policy and the information structure must be jointly designed. If the information structure does not have an interval structure, it is possible to write it as a linear combination of interval information structures, and each of one of these would have to be optimal. However, a given vector q^* cannot be optimal for more than one finite pooling information structure.

The third step shows that there is no interval of complete disclosure. This is the crucial step where we compute the trade-off between information rents and efficiency. We show that for small enough intervals, pooling information and allocation always increases the seller's profits.

The third step proves that the number of qualities offered is discrete, but leaves open the possibility that it is countably infinite. We then proceed in two additional steps to prove that the optimal mechanism is indeed finite. However, the crucial step in which we compute the trade-off between informational rents and efficiency is the aforementioned third step.

3.2 Restating the Seller's Problem

Before we can apply the envelope theorem as in the classic mechanism design literature, we need to address the following challenge. The distribution of expected values may not be absolutely continuous, so the resulting set of items of varying quality does not completely determine the incentive compatible payments. We therefore extend the assignment of qualities to cover the entire value space, $[\underline{v}, \bar{v}]$, the support of the distribution F.

Lemma 1 (Extension of Policies)
For all individually rational and incentive compatible mechanisms $(q(w), p(w))$ defined on $\operatorname{supp}(G)$, there exists a mechanism $(\hat{q}(w), \hat{p}(w))$ that is incentive compatible and individually rational for all $w \in \operatorname{supp}(F)$ with $\hat{q}(w) = q(w)$ and $\hat{p}(w) = p(w)$ for all $w \in \operatorname{supp}(G)$.

Proof of Lemma 1. To prove this lemma it is enough to consider the following mechanism. For every $w \notin \operatorname{supp}(G)$ we define:

$$(\hat{q}(w), \hat{p}(w)) \triangleq \underset{(q,p)\in\{(q(w),p(w))\}\cup\{(0,0)\}}{\arg\max} \{qw - p\}.$$

Clearly, incentive compatibility and individual rationality will be satisfied. Hence, we assume that q is defined for all $w \in \operatorname{supp}(F)$ (rather than $\operatorname{supp}(G)$). ■

We now provide the classic construction of the buyer's surplus U, defined earlier in (5) in terms of the qualities provided.

Lemma 2 (Buyer's Surplus)
In every incentive compatible mechanism (G, M), the buyer's surplus is:

$$U = \int_{\underline{v}}^{\overline{v}} q(w)(1 - G(w))dw.$$

This is the standard representation of the buyer's surplus, see Myerson (1981). As the seller's profits are equal to total surplus minus the buyer's surplus, we can write the profits in an incentive compatible mechanism as follows:

$$\Pi = \int_{\underline{v}}^{\overline{v}} (wq(w) - c(q(w)))\, dG(w) - \int_{\underline{v}}^{\overline{v}} q(w)(1 - G(w))dw, \tag{7}$$

which depends only on q (and not p).

The distribution G of expected values is generated by the prior distribution F and a signal s. By Blackwell (1951), Theorem 5, there exists a signal s that induces a distribution G of expected values if and only if F is a mean-preserving spread of G. F is defined to be a mean-preserving spread of G if

$$\int_v^\infty F(t)dt \le \int_v^\infty G(t)dt, \, \forall v \in \mathbb{R}_+,$$

with equality for $v = 0$. If F is a mean-preserving spread of G we write $F \prec G$. We will frequently use the fact that the distribution function $F \prec G$ if and only if $G^{-1} \prec F^{-1}$ (see Shaked and Shanthikumar (2007), Chapter 3).

Hence, the seller's problem can be restated in terms of G and q:

$$\max_{G,q} \left\{ \int_{\underline{v}}^{\overline{v}} (wq(w) - c(q(w)))\, dG(w) - \int_{\underline{v}}^{\overline{v}} q(w)(1 - G(w))dw \right\}, \tag{8}$$

subject to q being non-decreasing and $F \prec G$.

We denote by (G^*, q^*) a solution to this problem.

3.3 Interval Structure

We now reformulate the seller's problem in a way that it is linear in G^{-1}. For any mechanism (G, q) we can construct a random variable t uniformly distributed in the unit interval such that:

$$w = G^{-1}(t).$$

If G is continuously distributed, then $t = G^{-1}(w)$, otherwise, t is constructed introducing randomizations at every point of discontinuity of G. Similarly, we can write the quality q in terms of t, which we denote by:

$$q_t(t) \triangleq q(G^{-1}(t)). \tag{9}$$

Hence, we can describe the mechanism as (G, q_t).

We can write (7) as follows:

$$\Pi = \int_0^1 G^{-1}(t) q_t(t)dt - \int_0^1 c(q_t(t))dt - \int_0^1 q_t(t)(1 - t)dG^{-1}(t).$$

Integrating by parts the last term:

$$\Pi = \int G^{-1}(t)(1-t)dq_t(t) - \int c(q_t(t))dt.$$

Note that $q_t(t)$ is non-decreasing, so dq_t corresponds to the integral using q_t as a measure.

We can now fix an optimal mechanism (G^*, q^*). The optimal information structure must satisfy:

$$\Pi = \underset{\{G^{-1}:G^{-1}\prec F^{-1}\}}{\arg\max} \int_0^1 G^{-1}(t)(1-t)dq_t^*(t) - \int c(q_t^*(t))dt. \tag{10}$$

Note that if we change the information structure, but keep q_t^* constant, the expected production cost $\int c(q_t^*(t))dt$ will not change.

The optimization problem (10) is an upper semi-continuous linear functional of G^{-1}. The upper semi-continuity can be verified by noting that every $G^{-1} \prec F^{-1}$ is upper semi-continuous. Hence, if $\hat{G}^{-1} \to G^{-1}$ (taking the limit using the L^1 norm), we have that $\limsup \hat{G}^{-1}(t) \leq G(t)$ for all $t \in [0,1]$. Hence, $\limsup \int_0^1 \hat{G}^{-1}(t)(1-t)dq_t^*(t) \leq \int_0^1 G^{-1}(t)(1-t)dq_t^*(t)$.

Proposition 1 in Kleiner, Moldovanu, and Strack (2021) shows that the set $\{G^{-1} : G^{-1} \prec F^{-1}\}$ is a convex and compact set, and Theorem 1 shows that the extreme points of this set are given by (6). Following Bauer's maximum principle, the maximization problem attains its maximum at an extreme point of $\{G^{-1} : G^{-1} \prec F^{-1}\}$ and we can conclude:

Proposition 1

There exists an optimal mechanism (G^*, q^*) such that G^* has interval structure.

We now prove that an optimal mechanism must have an interval structure. Suppose that there exists an optimal mechanism (G^*, q^*) such that G^* does not have interval structure. Then there exists a collection of interval information structures G and a measure λ over these interval structure distributions such that:

$$G^{*-1} = \int G^{-1} d\lambda(G^{-1}).$$

Furthermore, since the functional (10) is linear in G^{-1}, each of these interval information structures must be optimal. Hence, there must exist two optimal mechanisms (G_1^*, q^*) and (G_2^*, q^*), with G_1^*, G_2^* having interval structure and $G_1^*(w) \neq G_2^*(w)$ for some w. However, G_1^* and G_2^* must have different supports, which we now show to be inconsistent with it being an optimal mechanism.

Lemma 3 (G^* and q^* are Increasing in the Support)
Let (G^*, q^*) be an optimal mechanism with $q^*(w) = 0$ if and only if $w \leq \hat{w}$. Then, G^* is constant in some interval (w_1, w_2) with $\hat{w} \leq w_1$ if and only if q^* is constant in (w_1, w_2).

Proof of Lemma 3. We first prove sufficiency. Suppose that in some interval (w_1, w_2), q^* is constant but G^* is not constant in this interval ($G^*(w_1) < G^*(w_2)$). Consider the following information structure:

$$\tilde{G}(w) = \begin{cases} \max\{G^*(w) - \varepsilon, 0\}, & \text{if } w \in (0, w_1]; \\ G^*(w_2), & \text{if } w \in (w_1, w_2]; \\ G^*(w), & \text{if } w > w_2; \end{cases}$$

where ε is chosen such that:

$$\int_0^{w_2} \tilde{G}(w)dw = \int_0^{w_2} G^*(w)dw.$$

Hence, this is a feasible information structure ($\tilde{G}$ is a mean-preserving contraction of G^*). Conditional on $w \leq w_1$, the distribution $\tilde{G}$ first order stochastically dominates G^* : for all $w \leq w_1$, $\tilde{G}(w) \leq G^*(w)$. Conditional on $w \geq w_2$, both distributions are the same. Since q^* is constant in (w_1, w_2), this clearly generates higher profits when $q^*(w_1) \neq 0$.

We now prove necessity. Suppose that in some interval (w_1, w_2), q^* is not constant ($q^*(w_1) < q^*(w_2)$) but G^* is constant in this interval. We can then consider the following variation:

$$\tilde{q}(w) = \begin{cases} q^*(v) & \text{if } w \notin (w_1, w_2); \\ q^*(w_1) & \text{if } w \in (w_1, w_2). \end{cases}$$

We then have that $(G^*, \tilde{q})$ generates the same total surplus but generates less bidder surplus. ■

Hence, following Lemma 3 there can only be one optimal finite pooling interval information structure. We thus conclude that there cannot be an optimal mechanism without a finite pooling interval structure, as this would require that there would be multiple optimal finite pooling interval structures. Thus we can conclude:

Proposition 2 (Necessity of Interval Structure)
In every optimal mechanism (G^*, q^*), the information structure G^* has interval structure.

The proposition leaves open the possibility that there are intervals of complete disclosure.

3.4 Optimality of Pooling

We next show that every interval is a pooling interval, and then in a second step that the number of intervals is finite. We first provide an intuition and then provide the proof.

Suppose the seller pools the allocation of all values in interval $[v_1, v_2]$, so that they all get the average quality in this interval: how much lower would the profits be? To make the notation more compact, we denote the virtual values by:

$$\phi(v) \triangleq v - \frac{1-F(v)}{f(v)}. \tag{11}$$

The revenue generated is the expectation of the product of the virtual values and the qualities:

$$R \triangleq \int_{\underline{v}}^{\bar{v}} q^*(v)\phi(v)f(v)dv.$$

We denote the mean and variance of the quality and virtual values in this interval by:

$$\mu_\phi \triangleq \frac{\int_{v_1}^{v_2} \phi(v)f(v)dv}{\int_{v_1}^{v_2} f(v)dv}; \qquad \mu_q \triangleq \frac{\int_{v_1}^{v_2} q^*(v)f(v)dv}{\int_{v_1}^{v_2} f(v)dv}; \tag{12}$$

$$\sigma_\phi^2 \triangleq \frac{\int_{v_1}^{v_2} (\phi(v)-\mu_\phi)^2 f(v)dv}{\int_{v_1}^{v_2} f(v)dv}; \qquad \sigma_q^2 \triangleq \frac{\int_{v_1}^{v_2} (q^*(v)-\mu_q)^2 f(v)dv}{\int_{v_1}^{v_2} f(v)dv}. \tag{13}$$

The first step of the proof consists of showing that the revenue losses due to pooling the allocation in the interval $[v_1, v_2]$ are bounded by:

$$\sigma_\phi \sigma_q (F(v_2) - F(v_1)).$$

The total cost will also (weakly) decrease if the allocation is pooled because the cost function is convex. Hence, pooling the allocation generates third-order profit losses when the interval is small (since each of the terms multiplied are small when the interval is small).

If in addition to pooling the allocation we pool the information of the values in this interval, we can reduce the buyer's information rent. When only the allocation is pooled–but not the information–then the quality increase that the pool gets relative to values just below the pool is the quality difference $\mu_q - q^*(v_1)$ priced at v_1. After pooling the information, the price of the quality increase is computed using the expected value conditional on being in this interval:

$$\mu_v \triangleq \mathbb{E}[v \mid v \in [v_1, v_2]]$$

Hence, pooling the information increases the transfers for every value higher than v_2 by an amount:

$$(\mathbb{E}[v \mid v \in [v_1, v_2]] - v_1)(\mu_q - q^*(v_1))(1 - F(v_1)).$$

Here the first two terms being multiplied are small when the interval is small. However, transfers are marginally increased for all values higher than v_2, which is a non-negligible mass of values (i.e., $(1 - F(v_1))$ is not small). In other words, pooling information increases the price of the quality increase $(\mu_q - q^*(v_1))$ for all values higher than v_2. Hence, pooling information generates a second-order benefit which always dominates the third-order distortions.

Proposition 3 (Pooling Intervals)
The optimal information structure consists of pooling intervals only.

Proof of Proposition 3. Following Proposition 1, the optimal information structure consists of intervals of pooling and intervals of full disclosure. We consider an optimal mechanism and an interval (v_1, v_2) such that the optimal information structure is full disclosure in this interval (i.e., such that $G^*(v) = F(v)$). We expose a contradiction by proving that there is an improvement. It is useful to write the interval (v_1, v_2) in terms of its mean and difference:

$$\hat{v} \triangleq \frac{v_1 + v_2}{2}; \quad \Delta \triangleq \frac{v_2 - v_1}{2}.$$

So, we have that $(v_1, v_2) = (\hat{v} - \Delta, \hat{v} + \Delta)$ and we will eventually take the limit $\Delta \to 0$.

Following Lemma 3, the qualities $q^*(v)$ must be strictly increasing in this interval. We consider qualities:

$$\tilde{q}(v) = \begin{cases} \mu_q & \text{if } v \in (v_1, v_2) \\ q^*(v) & \text{if } v \notin (v_1, v_2) \end{cases}$$

The difference between the optimal policy and the variation is given by:

$$\Pi^* - \tilde{\Pi} = \int_{v_1}^{v_2} (\phi(v)q^*(v) - c(q^*(v))) f(v)dv - (F(v_2) - F(v_1)) \left(\mu_\phi \mu_q - c(\mu_q)\right). \tag{14}$$

Note that we only need to consider the qualities in the interval $[v_1, v_2]$ to compute the difference. We can write this expression more conveniently as follows:

$$\Pi^* - \tilde{\Pi} = \int_{v_1}^{v_2} (\phi(v) - \mu_\phi)(q^*(v) - \mu_q) f(v)dv - \int_{v_1}^{v_2} \left(c(q^*(v)) - c(\mu_q)\right) f(v)dv.$$

The average production cost under μ_q is smaller than under q^* because the cost is convex, so the second integral is positive (which is being subtracted). Furthermore, using the Cauchy-Schwarz inequality, we can bound the first integral (and thus the whole expression) as follows:

$$\Pi^* - \tilde{\Pi} \leq \sigma_q \sigma_\phi (F(v_2) - F(v_1)). \tag{15}$$

Finally, using the Bhatia-Davis inequality, we can bound the variances as follows:

$$\sigma_q \leq \sqrt{(\mu_q - q^*(v_1))(q^*(v_2) - \mu_q)},$$

and similarly for ϕ. We can then conclude that:

$$\lim_{\Delta \to 0} \frac{\Pi^* - \tilde{\Pi}}{\Delta^3} \leq \frac{dq^*(\hat{v})}{dv} \frac{d\phi(\hat{v})}{dv} f(\hat{v}). \tag{16}$$

We thus have that the efficiency losses are of order Δ^3.

We now consider the following policy:

$$\hat{q}(v) = \begin{cases} q^*(v_1), & \text{if } v \in (v_1, \mu_v); \\ \mu_q, & \text{if } v \in [\mu_v, v_2); \\ q^*(v), & \text{if } v \notin (v_1, v_2). \end{cases} \tag{17}$$

Note that here μ_v is the mean value in the interval $[v_1, v_2]$. We additionally change the information structure so that all types in (v_1, v_2) are pooled. That is, the information structure is:

$$\hat{G}(v) = \begin{cases} F(v_1), & \text{if } v \in (v_1, \mu_v); \\ F(v_2), & \text{if } v \in [\mu_v, v_2); \\ F(v), & \text{if } v \notin [v_1, v_2). \end{cases} \tag{18}$$

Observe that the total surplus generated by $(\hat{G}, \hat{q})$ and by $(G, \tilde{q})$ is the same. Then, the difference in the generated profits is equal to the difference in the expected buyer's surplus:

$$\hat{\Pi} - \tilde{\Pi} = (\mu_q - q^*(v_1))(\mu_v - v_1)(1 - F(v_1)).$$

We conclude that:

$$\lim_{\Delta \to 0} \frac{\hat{\Pi} - \tilde{\Pi}}{\Delta^2} \geq \frac{dq^*(\hat{v})}{dv}(1 - F(v_1)). \tag{19}$$

Here we used that $(\mu_v - v_1)/\Delta \to 1$, as $\Delta \to 0$. The efficiency losses are of order Δ^2. We conclude that for Δ small enough, the new policy generates higher profits. ■

Given the optimality of pooling intervals we introduce the following notation. The distribution function G associated with any pooling interval structure is an increasing and piecewise constant step function G given by:

$$G_k \triangleq G(w_k), \quad g_k \triangleq G(w_k) - G(w_{k-1}). \tag{20}$$

Moreover, given the pooled interval structure, we can relate the probability of each interval to the underlying distribution of values, thus

$$g_k = F\left(v_k\right) - F\left(v_{k-1}\right), \quad w_k \in \left[v_{k-1}, v_k\right] .$$

Finally, the quality (and increment) allocated to value w_k is denoted by

$$q_k \triangleq q(w_k); \quad \Delta q_k \triangleq q_k - q_{k-1}. \tag{21}$$

In a finite-item menu, the profits are given by:

$$\Pi = \sum_{k=1}^{K} g_k\left(w_k q_k - c(q_k)\right) - \sum_{k=1}^{K} (w_{k+1} - w_k) q_k (1 - G_k), \tag{22}$$

where by convention $w_{K+1} \triangleq w_K$. This is the discrete counterpart of (7). We can also express the profits as follows:

$$\Pi = \sum_{k=1}^{K} w_k \Delta q_k (1 - G_{k-1}) - g_k c(q_k), \tag{23}$$

where $G_0 = 0$.

3.5 Menu Convexity

Before we can prove the finiteness of the menu, we establish a qualitative property of the menu that will support an argument for finiteness. Namely, we show that the menu will have increasing quality increments, thus

$$\Delta q_{k+1} = q_{k+1} - q_k \geq q_k - q_{k-1} = \Delta q_k,$$

for all k, whether k is finite or not.

Lemma 4 (Increasing Differences in Qualities)
In any optimal mechanism the quality increments Δq_k must be (weakly) increasing in k.

Proof of Lemma 4. We fix a mechanism (G, q) and assume that there exists $\tilde{k}$ such that $\Delta q_{\tilde{k}} > \Delta q_{\tilde{k}+1}$. We prove that the mechanism cannot be optimal.

We consider an alternative mechanism $(\tilde{G}, \tilde{q})$ that keeps the qualities the same:

$$\tilde{q}_k = q_k, \text{ for all } k \in \{1, ..., K\}.$$

The alternative information structure is defined as follows. First, for all $k \notin \{\tilde{k}, \tilde{k}+1\}$, the probabilities and expected values remain the same: $\tilde{w}_k = w_k$ and $\tilde{g}_k = g_k$. We modify the information structure as follows:

$$\tilde{g}_{\tilde{k}} = g_{\tilde{k}} + \varepsilon; \quad \tilde{g}_{\tilde{k}+1} = g_{\tilde{k}+1} - \varepsilon; \quad \tilde{w}_{\tilde{k}} = \frac{g_{\tilde{k}} w_{\tilde{k}} + \varepsilon w_{\tilde{k}+1}}{g_{\tilde{k}} + \varepsilon}; \quad \tilde{w}_{\tilde{k}+1} = w_{\tilde{k}+1}.$$

Note that:

$$\tilde{g}_{\tilde{k}} \tilde{w}_{\tilde{k}} + \tilde{g}_{\tilde{k}+1} \tilde{w}_{\tilde{k}+1} = g_{\tilde{k}} w_{\tilde{k}} + g_{\tilde{k}+1} w_{\tilde{k}+1},$$

so this is clearly a mean-preserving contraction of G, and thus it is clearly feasible. The new information structure will not be a partition, but for the purpose of the proof this is irrelevant because we will prove that the stated mechanism is suboptimal. To keep the expressions more compact, it is useful to introduce the following notation:

$$\kappa \triangleq \tilde{w}_{\tilde{k}} - w_{\tilde{k}} = \varepsilon \frac{w_{\tilde{k}+1} - w_{\tilde{k}}}{g_{\tilde{k}} + \varepsilon}.$$

Note that when $\varepsilon \approx 0$, $\kappa \approx \varepsilon (w_{\tilde{k}+1} - w_{\tilde{k}}) / g_{\tilde{k}}$.

The difference in the profits generated by the original mechanism and the new mechanism are given by:

$$\begin{aligned}
\Pi - \widetilde{\Pi} =& w_{\tilde{k}} \Delta q_{\tilde{k}} (1 - G_{\tilde{k}-1}) + w_{\tilde{k}+1} \Delta q_{\tilde{k}+1} (1 - G_{\tilde{k}}) - \big(\tilde{w}_{\tilde{k}} \Delta q_{\tilde{k}} (1 - \tilde{G}_{\tilde{k}-1}) + \tilde{w}_{k+1} \Delta q_{\tilde{k}+1} (1 - \tilde{G}_{\tilde{k}}) \big) \\
& - (g_{\tilde{k}} - \tilde{g}_{\tilde{k}}) c(q_{\tilde{k}}) - (g_{\tilde{k}+1} - \tilde{g}_{\tilde{k}+1}) c(q_{\tilde{k}+1}) \\
=& - \kappa (1 - G_{\tilde{k}-1}) \Delta q_{\tilde{k}} + w_{\tilde{k}+1} \Delta q_{\tilde{k}+1} \varepsilon + \varepsilon c(q_{\tilde{k}}) - \varepsilon c(q_{\tilde{k}+1}) \\
\leq& - \kappa (1 - G_{\tilde{k}-1}) \Delta q_{\tilde{k}} + w_{\tilde{k}+1} \Delta q_{\tilde{k}+1} \varepsilon - \varepsilon c'(q_{\tilde{k}}) \Delta q_{\tilde{k}+1} \\
<& \Delta q_{\tilde{k}+1} \bigg(w_{\tilde{k}+1} \varepsilon - \kappa (1 - G_{\tilde{k}-1}) - \varepsilon c'(q_{\tilde{k}}) \bigg)
\end{aligned}$$

The first inequality follows from the fact that c is convex, so $c(q_{\tilde{k}}) - c(q_{\tilde{k}+1}) \leq c'(q_{\tilde{k}}) \Delta q_{\tilde{k}+1}$; the second inequality follows from the fact that we are assuming (to reach a contradiction) that $\Delta q_{\tilde{k}+1} \leq \Delta q_{\tilde{k}}$. Taking the derivative and evaluating at 0, we get:

$$\frac{d(\Pi - \widetilde{\Pi})}{d\varepsilon} \bigg|_{\varepsilon = 0} < \frac{\Delta q_{\tilde{k}+1}}{g_{\tilde{k}}} \bigg(- (w_{\tilde{k}+1} - w_{\tilde{k}})(1 - G_{\tilde{k}-1}) + g_{\tilde{k}} w_{\tilde{k}+1} - g_{\tilde{k}} c'(q_{\tilde{k}}) \bigg).$$

The quality provided to expected valuation w_k satisfies the following first-order condition:

$$w_k - c'(q_k^*) = (w_{k+1} - w_k)\frac{1 - G_k}{g_k}.$$

Using this first-order condition we have that:

$$\left.\frac{d(\Pi - \widetilde{\Pi})}{d\varepsilon}\right|_{\varepsilon=0} < \frac{\Delta q_{\tilde{k}+1}}{g_{\tilde{k}}}\Big(- (w_{\tilde{k}+1} - w_{\tilde{k}})(1 - G_{\tilde{k}-1}) + g_{\tilde{k}} w_{\tilde{k}+1} + (w_{\tilde{k}+1} - w_{\tilde{k}})(1 - G_{\tilde{k}}) - g_{\tilde{k}} w_{\tilde{k}}\Big) = 0.$$

Hence, (G, q) is not optimal. ■

3.6 Optimality of Short Menu

We now show that the optimal information structure consists of finitely many intervals.

Proposition 4 (Finite Pooling Interval)
The optimal information structure has a finite pooling interval structure.

Proof of Proposition 4. Since the space of values is compact, this is equivalent to showing that there are no accumulation points of intervals. We consider three consecutive pooling intervals that generate expected values $w_{k-1} < w_k < w_{k+1}$.

Lemma 4 implies that there cannot be any accumulation points, except possibly at some $\hat{v}$ satisfying $q^*(\hat{v}) = 0$. Hence, it is a decreasing accumulation point (that is, the limit of expected valuations converges to $\hat{v}$ from the right). We denote by $\bar{f}$ and $\underline{f}$ the maximum and minimum density in $[\underline{v}, (\bar{v} + \hat{v})/2]$:

$$\underline{f} \triangleq \min_{v \in [\underline{v}, \frac{(\bar{v}+\hat{v})}{2}]} F'(v); \quad \bar{f} \triangleq \max_{v \in [\underline{v}, \frac{(\bar{v}+\hat{v})}{2}]} F'(v).$$

If such an accumulation point exists, we can find two consecutive pooling intervals, $(v_{k-1}, v_k]$ and $(v_k, v_{k+1}]$, generating expected values $w_k < w_{k+1}$, satisfying $g_k < g_{k+1}$, and:

$$c'(q_{k+1}^*) - c'(q_k^*) < \frac{\sqrt{\underline{f}}(1 - F(v_{k-1}))}{\sqrt{\bar{f}}(\sqrt{\bar{f}} + \sqrt{\underline{f}})}. \tag{24}$$

Note that the qualities q_k^* are monotonic, and so we must have that $(q_{k+1} - q_k)$ converge to 0 as we take intervals close enough to $\hat{v}$. So, we can take intervals close enough to $\hat{v}$ such that (24) is satisfied. Hence, we consider two intervals satisfying this inequality and reach a contradiction. We recall that the density is not vanishing except at the upper bound $\bar{v}$, so we must have that $\underline{f} > 0$.

Analogous to (11), we define:

$$\phi_k \triangleq w_k - (w_{k+1} - w_k)\frac{1 - G_k}{g_k},$$

and extend (12)-(13) in the natural way:

$$\mu_w \triangleq \frac{g_k w_k + g_{k+1} w_{k+1}}{g_k + g_{k+1}}; \quad \sigma_w^2 \triangleq \frac{g_k(w_k - \mu_w)^2 + g_{k+1}(w_{k+1} - \mu_w)^2}{g_k + g_{k+1}}$$

and analogously for μ_q, σ_q, μ_ϕ, σ_ϕ. Finally, $\hat{\Pi}$ and $\tilde{\Pi}$ are defined in the same way as before. Following the same steps as before, we have that:

$$\frac{\Pi^* - \tilde{\Pi}}{\hat{\Pi} - \tilde{\Pi}} \leq \frac{\sqrt{(\mu_q - q^*(w_k))(q^*(w_{k+1}) - \mu_q)}\sqrt{(\mu_\phi - \phi_k)(\phi_{k+1} - \mu_\phi)}(g_k + g_{k+1})}{(\mu_q - q^*(w_k))(\mu_w - w_k)(1 - F(v_{k-1}))}.$$

We note that:

$$\mu_\phi - \phi_k = \frac{g_{k+1}(\phi_{k+1} - \phi_k)}{g_k + g_{k+1}}, \quad \phi_{k+1} - \mu_\phi = \frac{g_k(\phi_{k+1} - \phi_k)}{g_k + g_{k+1}};$$

and the difference between other quantities can be written in an analogous way. We thus get that:

$$\frac{\Pi^* - \tilde{\Pi}}{\hat{\Pi} - \tilde{\Pi}} \leq \frac{g_k(\phi_{k+1} - \phi_k)(g_k + g_{k+1})}{g_{k+1}(w_{k+1} - w_k)(1 - F(v_{k-1}))}.$$

We then note that:

$$g_k + g_{k+1} \leq \bar{f}(v_{k+1} - v_{k-1}); \quad w_k \leq \frac{\sqrt{\bar{f}} v_k + \sqrt{\underline{f}} v_{k-1}}{\sqrt{\bar{f}} + \sqrt{\underline{f}}}; \quad w_{k+1} \geq \frac{\sqrt{\bar{f}} v_k + \sqrt{\underline{f}} v_{k+1}}{\sqrt{\bar{f}} + \sqrt{\underline{f}}}.$$

We also recall that $g_k < g_{k+1}$ and note that the optimal quality is given by $\phi_k = c'(q_k^*)$ and $\phi_{k+1} = c'(q_{k+1}^*)$. We thus get that:

$$\frac{\Pi^* - \tilde{\Pi}}{\hat{\Pi} - \tilde{\Pi}} \leq \frac{\sqrt{\bar{f}}(c'(q_{k+1}^*) - c'(q_k^*))(\sqrt{\bar{f}} + \sqrt{\underline{f}})}{\sqrt{\underline{f}}(1 - F(v_{k-1}))} < 1,$$

where the second inequality corresponds to (24). Thus, we reach a contradiction with this being an optimal mechanism. ■

4 Single-item Menu

So far, we have shown that the optimal mechanism generates a coarse menu that only offers a finite number of items. By contrast, with complete disclosure of information to the buyer, the optimal menu would typically offer a continuum of items. This leaves open the question of how coarse the menu can become. In this section, we introduce sufficient conditions that will guarantee that the optimal menu consists of a single item.

4.1 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:

$$(q^*, v^*) \in \arg\max_{q,v} \mathbb{P}[v' \geq v](\mathbb{E}[v' \mid v' \geq v]q - c(q)). \tag{25}$$

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.

The first-order conditions for the optimal single-item menu are given by:

$$\mu^* = c'(q^*) \quad \text{and} \quad c(q^*) = v^* q^*, \tag{26}$$

The first condition states that the quality is efficiently supplied given that the (expected) valuation of the buyer who buys the good is μ^*. The second condition states that the threshold v^* is also efficiently chosen: given that q^* units are going to be supplied, it is efficient to sell to a buyer with valuation 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 $v^* < \underline{v}$, which means there is no exclusion.

The reason there are no distortions in the quality supplied and the threshold is that 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.

4.2 Single-item Menu Optimality

We say a distribution has *modest tails* if for all $t \in [0, 1]$:

$$\mathbb{E}[v' \mid F(v') \geq t] \leq \mu^* + 2\left(1 - \sqrt{\frac{1-t}{1-F(v^*)}}\right)(\mu^* - v^*). \tag{MT}$$

The condition imposes an upper-bound on the conditional expected value of the tail of the distribution (i.e., values that are above some quantile t) based only on the threshold v^* and the conditional mean μ^* of the optimal single-item mechanism. Observe that the right hand side is equal to μ^* when $t = F(v^*)$, and thus the inequality holds with equality. The left hand side and right hand side are both increasing in t. Thus the condition requires that the conditional expected value of the tail does not increase too fast. The right hand side equals $3\mu^* - 2v^*$ when $t = 1$, and so the condition puts an upper bound on the support: $\bar{v} \leq 3\mu^* - 2v^*$. From a technical perspective, the condition guarantees that the distribution is a mean-preserving contraction of an appropriately constructed distribution that has linear density. We give our second main result and then provide additional interpretation of (MT).

Theorem 2 (Optimality of Single-item Mechanisms)
If the distribution satisfies (MT) *and* $c'''(q) \geq 0$*, then the optimal mechanism is a single-item menu.*

Theorem 2 shows that in a wide range of settings the optimal mechanism is a single-item menu. Condition (MT) is stated in terms of v^*, which cannot be easily inferred from F without making the explicit calculation. However, we can give stronger conditions that are easier to evaluate. Distribution F satisfies (MT) if the density f is quasi-concave and it satisfies:[1]

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

This condition states that f must be concave when it is decreasing. For example, any distribution with (weakly) increasing density satisfies (QC).

We prove Theorem 2 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 (MT) 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.

4.3 Linearly-decreasing Density and Quadratic Cost

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

$$L(v; \underline{v}, \bar{v}) \triangleq \frac{(v - \underline{v})(2\bar{v} - \underline{v} - v)}{(\bar{v} - \underline{v})^2}. \tag{27}$$

[1] In fact, (QC) alone implies quasi-concavity. We explicitly require quasi-concavity only to make the condition more transparent.

The density of this distribution, which we denote by $l(v; \underline{v}, \bar{v})$, is linearly-decreasing with zero density at the top of the support:

$$l(\bar{v}; \underline{v}, \bar{v}) = 0.$$

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

$$c(q) \triangleq \alpha q + \frac{\beta q^2}{2} + \gamma. \tag{28}$$

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.

Proposition 5 (Linear Density and Quadratic Cost Environment)
The optimal menu with linearly-decreasing density and a linear-quadratic cost function is always a single-item menu.

Proof of Proposition 5. Recall that in a finite-item menu, the profits can be written as in (23). We consider the optimality conditions of the highest two intervals of an optimal mechanism. For this, we define the profits from the highest two items:

$$\begin{aligned} \Pi_{K-1,K}(v_{K-1}, \Delta q_{K-1}, \Delta q_K) \triangleq \; & \big((g_{K-1} + g_K)\Delta q_{K-1} w_{K-1} - g_{K-1} c(q_{K-2} + \Delta q_{K-1}) \\ & + g_K \left(\Delta q_K w_K - c(q_{K-2} + \Delta q_{K-1} + \Delta q_K)\right)\big), \end{aligned} \tag{29}$$

which are the last two terms of the summations in (23). If the optimal mechanism is a multi-item mechanism, the solution to the following problem:

$$\Pi^*_{K-1,K} = \max_{\substack{v_{K-1} \in [v_{K-2}, \bar{v}], \\ \Delta q_{K-1}, \Delta q_K \geq 0}} \Pi_{K-1,K}(v_{K-1}, \Delta q_{K-1}, \Delta q_K) \tag{30}$$

must satisfy $\Delta q_{K-1},\ \Delta q_K > 0$ and $v_{K-2} < v_{K-1} < \bar{v}$, where q_{K-2} and v_{K-2} are parameters that are kept fixed in the optimization problem.

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

$$\Delta q_{K-1} = \max\{\frac{w_{K-1} - \alpha - \beta q_{K-2}}{\beta} - \frac{(w_K - w_{K-1})g_K}{\beta g_{K-1}}, 0\} \text{ and } \Delta q_K = \frac{w_K - \alpha - \beta q_{K-1}}{\beta}. \tag{31}$$

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

$$\frac{w_K - \alpha - \beta q_{K-2}}{w_{K-1} - \alpha - \beta q_{K-2}} < \frac{g_{K-1} + g_K}{g_K}. \tag{32}$$

To write expressions that are compact, we define:

$$z \triangleq \frac{\bar{v} - v_{K-1}}{\bar{v} - v_{K-2}}; \quad \kappa \triangleq 3\frac{v_{K-2} - \alpha - \beta q_{K-2}}{\bar{v} - v_{K-2}}.$$

Note that in an optimal mechanism we must have that $\kappa \geq 0$, as otherwise the mechanism would be offering a quality q_{K-2} whose marginal cost is higher than the value for all values $v \in [v_{K-3}, v_{K-2}]$, which is clearly suboptimal. Using these definitions, we have that (32) is satisfied if and only if:

$$\frac{(3 - 2z + \kappa)}{(1 - \frac{2z^2}{1+z} + \kappa)} < \frac{1}{z^2}. \tag{33}$$

And, for every z satisfying (33), (29) can be written as follows:

$$\Pi_z \triangleq \frac{(g_{K-1} + g_K)(\bar{v} - v_{K-2})^2}{18\beta\,(1 - z^2)} \left((3 - 2z + \kappa)z^2 \left(1 - 2z + \frac{4z^2}{1+z} - \kappa \right) + \left(1 + \kappa - \frac{2z^2}{1+z} \right)^2 \right). \tag{34}$$

Hence, Π_z is equal to $\Pi_{K-1,K}$ when (32) is satisfied. If the optimal mechanism is a multi-item mechanism, there must exist $z \in [0, 1]$ satisfying (33) that maximizes (34).

If z^* maximizes (34) and satisfies (33) 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:

$$(v^*(\alpha, \beta), \Delta q_K^*(\alpha, \beta)) \triangleq \underset{0 \leq \Delta q, v_{K-2} \leq v \leq \bar{v}}{\arg\max} \Pi_{K-1,K}(v, 0, \Delta q), \tag{35}$$

where we define the optimal quantity 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 (35) is then to find an interval $(v_{K-1}, \bar{v}] = (v^*(\alpha, \beta), \bar{v}]$ and an allocation $q_K^*(\alpha, \beta)$ so as to maximize the profits from all types in the given interval $(v_{K-2}, \bar{v}]$. This restricted maximization problem allows the interval $(v_{K-1}, \bar{v}]$ to be a strict inclusion of $(v_{K-2}, \bar{v}]$: that is, $(v_{K-1}, \bar{v}] \subsetneq (v_{K-2}, \bar{v}]$. In this case, all the types in the interval $(v_{K-2}, v_{K-1}]$ will receive the allocation q_{K-2}. Now, from Proposition 5, we know that:

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

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.

4.4 Linearly-decreasing Density and Convex Cost

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 (28)). We note that $c(q)$ and $\widehat{c}(q)$ intersect three times at most. 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 5 (Dominating Cost Function)
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 of Lemma 5. We begin by considering α, β, γ chosen such that:

$$\widehat{c}(q_{K-2}) = c(q_{K-2});\ \widehat{c}(q_{K-1}) = c(q_{K-1});\ \widehat{c}(q_K) = c(q_K). \tag{36}$$

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 $q \geq q_{K-2}$:

$$c(q) \leq \widehat{c}(q) \iff q \in [q_{K-1}, q_K]. \tag{37}$$

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 5. 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 5.

Suppose that:

$$q^*(\alpha, \beta) < q_{K-1}, \tag{38}$$

where (α, β) satisfy (36). 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 (38) and the second inequality follows from the fact that q^* by definition is larger than q_{K-2} (see (35)) . Thus, following the intermediate value theorem, there exists a $\hat{q} \in [q_{K-2}, q_K]$ such that:

$$q^*(\alpha(\hat{q}), \beta(\hat{q})) = \hat{q}. \tag{39}$$

Furthermore, note that $q_K > \max\{\hat{q}, q_{K-1}\}$, so we have that $\widehat{c}(q_K) < c(q_K)$. Thus, we are in Case 2 of Lemma 5.

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 (38) 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 $\hat{q}$ such that $q^*(\alpha(\hat{q}), \beta(\hat{q})) = \hat{q}$ 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 this Lemma we can now extend the optimality result to convex cost functions.

Proposition 6 (Optimality of Single-item Menu with Linear Density and Convex Cost) *The optimal menu with linear decreasing density and $c''' \geq 0$ is always a single-item menu.*

Proof of Proposition 6. 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 (29), we define:

$$\begin{aligned}&\hat{\Pi}_{K-1,K} \triangleq \\ &(g_{K-1} + g_K)\Delta q_{K-1} w_{K-1} - g_{K-1}\widehat{c}(q_{K-2} + \Delta q_{K-1}) + g_K\left(\Delta q_K w_K - \widehat{c}(q_{K-2} + \Delta q_{K-1} + \Delta q_K)\right).\end{aligned}$$

Now $c(\cdot)$ is the true cost function, which satisfied $c'''(\cdot) \geq 0$, and $\widehat{c}(\cdot)$ is a linear-quadratic cost. So $\hat{\Pi}_{K-1,K}$ 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:

$$\hat{\Pi}_{K-1,K}(v_{K-1}, \Delta q_{K-1}, \Delta q_K) < \hat{\Pi}_{K-1,K}(v^*_{K-1}(\alpha, \beta), 0, \Delta q^*_K(\alpha, \beta))$$

We now consider the three cases in Lemma 5.

If we take (α, β) so that the first case in Lemma 5 holds, then we have:

$$\Pi_{K-1,K}(v_{K-1}, \Delta q_{K-1}, \Delta q_K) = \hat{\Pi}_{K-1,K}(v_{K-1}, \Delta q_{K-1}, \Delta q_K); \quad (40)$$

$$\Pi_{K-1,K}(v^*_{K-1}(\alpha, \beta), 0, \Delta q^*_K(\alpha, \beta)) > \hat{\Pi}_{K-1,K}(v^*_{K-1}(\alpha, \beta), 0, \Delta q^*_K(\alpha, \beta)); \quad (41)$$

We thus have that:

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

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 5, then the argument is analogous but (41) will hold with equality and (40) will hold with strict inequality. ■

4.5 Distributions with Modest Tails

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 (26) are necessary and sufficient conditions for optimality when $c'''(\cdot) \geq 0$.

Proposition 7 (Sufficient Conditions for Optimality)
If $c'''(q) \geq 0$, the distribution is $L(v; \underline{v}, \bar{v})$, and $(\hat{q}, \hat{v})$ satisfy the first-order condition (26)*, then $(\hat{q}, \hat{v})$ solves* (25)*, i.e. $(\hat{q}, \hat{v}) = (q^*, v^*)$.*

Proof of Proposition 7. When the distribution is linearly decreasing, we have that:

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

Hence, if $(\hat{q}, \hat{v})$ satisfy the first-order condition (26) we have that:

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

We then have that:

$$\bar{v} = 3c'(\hat{q}) - 2\frac{c(\hat{q})}{\hat{q}}.$$

We now note that,

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

If $c'''(q) \geq 0$ we have that $c''(q)q \geq c'(q)$. Hence, we have that:

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

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

We now 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 $\bar{q} \in \{0, \infty\}$ is clearly never optimal. It is also easy to see that $v = \bar{v}$ cannot be an optimum as then the objective function of (25) 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 profits 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 (25):

$$L(v^*; \underline{z}, \overline{z}) = F(v^*) \text{ and } \mathbb{E}_L[v \mid v \geq v^*] = \mathbb{E}_F[v \mid v \geq v^*], \tag{42}$$

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 $\hat{F}(v)$:

$$\hat{F}(v) = \begin{cases} L(\hat{v}; \underline{z}, \overline{z}), & \text{if } v \in [0, \hat{v}]; \\ L(v; \underline{z}, \overline{z}), & \text{if } v \in [\hat{v}, \overline{z}]; \end{cases} \tag{43}$$

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

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

In the proof of Lemma 6 we will show that indeed such a $\hat{v}$ exists. Thus, $\hat{F}(v)$ 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, $\hat{F}$ is equal to $L(v; \underline{v}, \overline{v})$ for $v \geq \hat{v}$, and $\hat{F}$ has an atom of size $L(\hat{v}; \underline{z}, \overline{z})$ at $v = 0$.

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

Lemma 6 (Distribution Comparison)
Distribution $\hat{F}$ is a mean-preserving spread of F and $\hat{F}$ is first-order stochastically dominated by $L(v; \underline{z}, \overline{z})$.

Proof of Lemma 6. We first compare $L(v, \bar{z}, \underline{z})$ with F. Note that:

$$\begin{aligned}
\frac{\int_q^1 L^{-1}(v; \overline{z}, \underline{v}) dv}{1-q} =& \mu^* + 2\left(1 - \sqrt{\frac{1-q}{1-F(v^*)}}\right)(\mu^* - v^*);\\
\frac{\int_q^1 F^{-1}(v) dv}{1-q} =& \mathbb{E}[v \mid F(v) \geq q].
\end{aligned}$$

Hence, (MT) implies that for all $v' \in [0, \infty)$:

$$\int_{F(v')}^1 F^{-1}(v) dv \leq \int_{F(v')}^1 L^{-1}(v; \overline{z}, \underline{v}) dv. \tag{44}$$

If the inequality is satisfied with equality for $v' = 0$, we have that $\hat{v} = \underline{z}$ and, otherwise, $\hat{v} > \underline{z}$ (where $\hat{v}$ is used to construct $\hat{F}$ in (43)). Otherwise, we will have that $\hat{v} > \underline{z}$.

Since $\hat{F}$ 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 $\hat{F}$ is first-order stochastically dominated by $L(v; \underline{z}, \overline{z})$. We have that (44) implies that for all $v' \geq \hat{v}$:

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

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

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

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

We can now conclude the proof by establishing the main result of this section, Theorem 2.

Final Step of the Proof of Theorem 2. 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 7, for the linearly decreasing density the first-order condition is sufficient for optimality, and thus (v^*, q^*) given by (25) 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 $\hat{F}$, and $\hat{F}$ 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. ■

5 Additional Results

In this section we first analyze the optimal mechanism when the distribution has binary support and the cost is iso-elastic. Under these assumptions we can fully characterize the optimal mechanism, and the binary model will serve as a stepping stone to give more results about the general model. Second, we show that if the distribution has narrow support the optimal mechanism is pooling. We then show that, if the distribution has unbounded support, the optimal mechanism consists of infinitely many items. Finally, we offer the comparative statics of how the cost elasticity changes the nature of the optimal mechanism.

5.1 Optimal Mechanism with Binary Values

Throughout this section, the values v of the buyer have binary support $0 < v_L < v_H$, with probabilities $f_L$and f_H respectively (of course, $f_L + f_H = 1$). We also assume that the cost function is a power function:

$$c(q) = q^\eta/\eta,$$

with $\eta > 1$.

The disclosure policy of the seller always contains two extremal policies: (i) the seller discloses all information to the buyer and subsequently screens the different types through different qualities the *screening solution*; or, (ii) the seller does not disclose any information and subsequently pools all types and offers a single item for sale the *pooling solution*.

In between these two extremal disclosure policies, there is a large number of intermediate policies that would generate different optimal selling policies. The seller could combine low and high values in arbitrary proportions to create many intermediate values, and thus additional values to screen and to match with suitable qualities. Our first result is that the stochastic combination of low and high values will never be optimal. Either the values will be completely disclosed, thus yielding the screening solution, or completely pooled, thus yielding the pooling solution.

Proposition 8 (Complete Screening or Complete Pooling)
The optimal mechanism either exhibits complete screening *or* complete pooling.

This result with binary values generalizes appropriately to a case of continuum values as stated in Theorem 1. Namely, the optimal information is a monotone partition that pools adjacent values. The set of pooled values may be small or large but is never formed through stochastic combinations.

With these two possible forms of optimal disclosure policies, it is only necessary to consider three selling strategies: (a) with zero disclosure, the values are pooled and a single item is sold to the expected value; (b) with complete disclosure sell only to the high value; and, (c) with complete disclosure offer a menu that screens and serves distinct values. We notice that under either (a) or (b), the optimal menu consists of a *single item*.

If the seller pools the values, the profit is:

$$\Pi_P \triangleq \max_q \ \{(f_L v_L + f_H v_H)q - c(q)\}.$$

The buyer's expected value is $w = f_L v_L + f_H v_H$ and the seller provides the efficient quantity q_w given the expected value and extracts the expected surplus.

If the seller serves the high value only, the profit is:

$$\Pi_H \triangleq \max_q \{f_H(v_H q - c(q))\},$$

and the seller offers a single-item q_H at the efficient level to the high value buyer. Relative to the pooling solution, the profits are higher when there is a sale, but the probability of a sale is lower.

If the seller offers a menu (q_L, q_H), the profit is:

$$\Pi_M \triangleq \max_{q_L, q_H} \{(f_H(v_H q_H - c(q_H)) + f_L(q_L v_L - c(q_L))) - f_H q_L(v_H - v_L)\}.$$

Here the seller maximizes the difference between the total surplus (first term) and the information rents (second term). The high value buyer is offered the efficient quality q_H while the low value is offered a quantity q_L below the efficient level to reduce the information rents.

Using the fact that we restrict attention to cost functions that are power functions, we can determine the optimal quantities explicitly in the screening solution:

$$q_L = v_L^{\frac{1}{\eta-1}}\left(1 - \frac{(\frac{v_H}{v_L} - 1)f_H}{f_L}\right)^{\frac{1}{\eta-1}}, \text{ and } \ q_H = v_H^{\frac{1}{\eta-1}}. \tag{45}$$

If the optimal allocation is given by a single item, then the quality level is given by the efficient solution, thus either

$$q_w = (f_L v_L + f_H v_H)^{\frac{1}{\eta-1}}\ , \text{ or } \ q_H = v_H^{\frac{1}{\eta-1}}.$$

As a consequence, the profit function under either policy can be expressed explicitly as follows:

$$\Pi_P = v_L^{\frac{\eta}{\eta-1}} \frac{\eta-1}{\eta} (f_L + f_H \frac{v_H}{v_L})^{\frac{\eta}{\eta-1}}; \tag{46}$$

$$\Pi_H = v_L^{\frac{\eta}{\eta-1}} \frac{\eta-1}{\eta} f_H (\frac{v_H}{v_L})^{\frac{\eta}{\eta-1}}\ ; \tag{47}$$

$$\Pi_M = v_L^{\frac{\eta}{\eta-1}} \frac{\eta-1}{\eta} \left(\left(\frac{(1-f_H \frac{v_H}{v_L})^{\frac{\eta}{\eta-1}}}{(1-f_H)^{\frac{1}{\eta-1}}} \right) + f_H (\frac{v_H}{v_L})^{\frac{\eta}{\eta-1}} \right). \tag{48}$$

We can now compare the revenue from the two disclosure policies: zero disclosure, and thus Π_P, or complete disclosure, or max $\{\Pi_H, \Pi_M\}$. Toward a complete description of the optimal policy we first observe that under complete disclosure the choice of the optimal menu"one or two items"depends only on the values and frequencies, but not on the cost function for quality.

Lemma 7 (Screening with a One-item vs Two-item Menu)
With complete disclosure, a single-item menu yields higher profits than a two-item menu if and only if

$$v_L < v_H f_H. \tag{49}$$

Lemma 7 provides a well-known trade-off in screening problems. Serving the low value increases efficiency but it also increases the buyer's information rents. Hence, the quality offered to the low value is distorted downwards. The distortion is increasing in the probability that the buyer has a high value, and if this probability is too high, then the low value buyer is excluded completely and offered zero quality. Thus, we can express the optimality of the one-item versus two-item menus in terms of the ratio of the values v_H/v_L and the probability f_H (of the high value) alone. The exclusion condition (49) in fact holds for all convex cost functions provided with a marginal cost of zero at quality $q = 0$ (that is, $c'(0) = 0$).

In contrast to the classic screening problem, we allow the seller to disclose less information and in the limit pool the values of the buyers. The benefit of pooling is that the seller can extract all of the buyer's expected surplus. The cost of pooling is that there are efficiency losses associated to providing the same quality to low and high values.

We now characterize when pooling is used optimally. We define a threshold ratio r that determines when pooling is optimal:

$$r(f_H, \eta) \triangleq \{r \in \mathbb{R}_+ \mid \Pi_P \geq \max\{\Pi_H, \Pi_M\} \iff v_H/v_L \leq r\}. \tag{50}$$

In other words, $r(f_H, \eta)$ is such that the optimal mechanism is pooling if and only if the ratio between the high value and the low value, $v_H/v_L \leq r(f_H, \eta)$, is below the threshold. The threshold ratio r is a function of the primitives of the binary model, namely the prior probability f_H of a high value and the curvature of the cost function η:

$$r : [0, 1] \times \mathbb{R}_+ \to \mathbb{R}_+.$$

We now characterize r.

Proposition 9 (Optimality of Pooling)
The correspondence $r(f_H, \eta)$ as defined in (50) is not empty, it is single valued, and is increasing in η and decreasing in f_H, with $r(1, \eta) = \eta$.

Propositions 7 and 9 jointly give us a complete description of the optimal mechanism. Figure 1 illustrates the qualitative properties of the optimal mechanism for different values of (f_H, v_H). Figure 1 also illustrates how the optimal mechanism changes with the power η of the cost function. When $\eta = 2$, an explicit expression for the threshold function r can be given:

$$r(f_H, 2) = \begin{cases} 1 + \frac{1}{\sqrt{f_H}}, & \text{if } f_H > \frac{3-\sqrt{5}}{2}; \\ \frac{f_H^2 - 3f_H + 3}{f_H^2 - f_H + 1}, & \text{if } f_H \leq \frac{3-\sqrt{5}}{2}. \end{cases} \tag{51}$$

The threshold level for f_H is given by $(3-\sqrt{5})/2 \approx 0.38$. Pooling is therefore optimal only when the ratio between high and low value is sufficiently small. When the distribution gives more weight to the high value, pooling becomes less beneficial. We then see that offering a multi-item mechanism is optimal only when the probability of the high value is relatively small.

Corollary 2 (Single-item Menu)
If $\eta \geq 2$ and $f_H \geq (3 - \sqrt{5})/2$, then the optimal mechanism is a single-item menu.

In this case, we can provide predictions about when it is optimal to offer a single-item menu based on the distribution of the values and not the difference between the values. This is because excluding the low value is optimal when the difference between values is high enough, or the probability of the high value is high enough. We already solved analytically for the quadratic cost function, and then we can extend the result to $\eta > 2$ by observing the $r(f_H, \eta)$ is increasing in η.

5.2 Narrow Support Distributions

We can now show that, if the support of the distribution is narrow enough, the optimal mechanism will necessarily be pooling. When the distribution has binary values, this follows as a direct corollary of Proposition 9.

Corollary 3 (Small Differences Between Valuations)
If the ratio between the high and low value is lower than η :

$$\frac{v_H}{v_L} < \eta,$$

then the optimal mechanism pools both values.

We can extend the insight from the model with a binary value distribution to the general environment with a continuum of values. We then replace the low and high value, v_L and v_H by the lower and upper bound of the support of the distribution, thus $\underline{v}$ and $\overline{v}$.

Proposition 10
Pooling is optimal for every distribution F with support in $[\underline{v}, \overline{v}]$ if and only if

$$\frac{\overline{v}}{\underline{v}} < \eta. \tag{52}$$

Proof. Consider a distribution H with binary support in $\{\underline{v}, \bar{v}\}$ such that the probability of $\bar{v}$ is given by $(\mu_v - \underline{v})/(\bar{v} - \underline{v})$ (where μ_v is the mean of F). We then have that H is a mean-preserving spread of F. But, following Corollary 3, the optimal mechanism when the distribution is H pools all values. Hence, this must also be the optimal mechanism when the distribution is F. This proves the sufficiency part of the statement. The necessity part of the statement is straightforward because if (52) is not satisfied, then there is a binary value distribution with support in $\{\underline{v}, \bar{v}\}$ for which pooling is not the optimal mechanism. ■

To gain some intuition, we consider the case in which $v_H/v_L \approx 1$ and analyze whether pooling or offering a two-item menu generates higher profits (note that when $v_H/v_L \approx 1$ we will have that $v_H/v_L \leq f_H^{-1}$ so excluding the low value is not optimal). Of course, when $v_H/v_L = 1$, both selling strategies are equivalent. With nearby values, it is sufficient to consider a first-order approximation.

Let S_M and S_P be the total surplus generated when offering a two-item menu and single-item menu that pools, and let ΔS be the difference:

$$\Delta S \triangleq S_M - S_P.$$

It is possible to verify that:

$$\left.\frac{\partial \Delta S}{\partial v_H}\right|_{v_H=v_L} = 0.$$

In other words, at a first-order approximation the gains are 0 when the difference in values is small. As the cost is convex, small distortions around the socially optimal quantity generate only second-order losses. Hence, offering the optimal quantity for the low-value type to the high value type as well generates second-order losses when the ratio between both values is close enough to 1.

On the other hand, when the values are pooled, the information rent of the buyer is $U_P = 0$. By contrast, the information rents when offering a menu are:

$$U_M = f_H q_L (v_H - v_L).$$

The information rent is the surplus that the high value buyer obtains when reporting to be of low value. The information rent is proportional to the difference between values, and thus the information rents increase linearly with v_H.

We thus conclude that it is optimal to screen only when the ratio between the values is large enough. The reason is that the efficiency gains from a menu are convex in the difference of the values while the information rents are linear. Hence, it is profitable to separate the high value only when the high value is sufficiently high enough relative to the low value.

We note that the optimality of pooling has implications for social welfare as well. When pooling is optimal and when the optimal mechanism with complete disclosure would exclude the low value, then the profits of the seller are higher under pooling. After all, the seller controls the amount of information disclosed. Since the buyer receives zero information rent in either solution, it follows that the social welfare also increases with pooling. With binary values, it further follows that a buyer can never gain from the information policy of the seller, as pooling always removes any residual information rent of the buyer. However, this result is special for the binary model, and with more than two values, information control may increase the surplus of the buyer and the seller simultaneously, and a fortiori increase the social surplus.[2]

[2] For example, with three values, $v_L = 1/3,\ v_M = 1/2,\ v_L = 1$, and probabilities $f_L = 2/3,\ f_M = 1/6,\ f_H = 1/6$, it can be shown that the optimal mechanism pools type v_L and v_M. In this mechanism the buyer's surplus is larger than under complete disclosure.

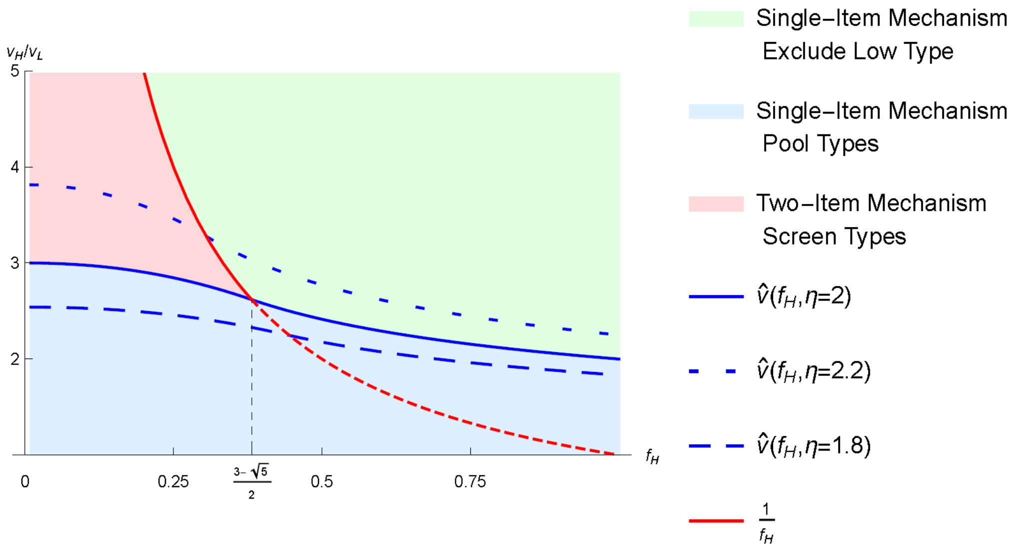

Figure 1: Illustration of the Optimal Mechanism for Binary Distributions

5.3 Unbounded Support

So far we have assumed that the support of values is bounded. We have shown that the optimal information structure is a finite pooling information structure. We now examine what happens when the distribution of values is unbounded.

Proposition 11 (Separation)
Suppose the support of values is unbounded ($\bar{v} = \infty$) and the cost is quadratic. Then, the optimal mechanism consists of infinite pooling intervals.

To prove this proposition, we use the same arguments as before to show that the optimal information structure will consists of pooling intervals and that there are no accumulation points in the intervals. When the distribution of values has finite support this is enough to prove there are finite intervals. Instead, when the distribution of values is unbounded, we now show that there must be infinite intervals.

We assume that the optimal information structure consists of finite pooling information structure, and reach a contradiction. To stay consistent with the notation previously introduced, the last interval is $[v_{K-1}, \infty)$. We consider an alternative mechanism in which we separate the last interval in $[v_{K-1}, \infty) = [v_{K-1}, v') \cup [v', \infty)$. The following definitions are useful:

$$g_H \triangleq \frac{1 - F(v')}{1 - F(v_{K-1})}; \qquad g_L \triangleq \frac{F(v') - F(v_{K-1})}{1 - F(v_{K-1})}; \tag{53}$$

$$v_H \triangleq \mathbb{E}[v \mid v \in [v', \infty)]; \qquad v_L \triangleq \mathbb{E}[v \mid v \in [v_{K-1}, v']]; \tag{54}$$

$$q_L \triangleq \Delta q_{K-1}; \qquad q_H \triangleq \Delta q_{K-1} + \Delta q_K. \tag{55}$$

We recall that the last two terms of the summation in (22) are written in (29), which corresponds to the maximization of v_{K-1}, Δq_{K-1}, Δq_K (while omitting the terms that do not depend on these variables). When the cost is $c(q) = q^2$, we can write the profit generated by the last two intervals $[v_{K-1}, v')$; $[v', \infty)$ as follows:

$$\Pi_{K-1,K} \triangleq \underset{\substack{v' \in [v_{K-1}, v_K] \\ q_L, q_H \geq 0}}{\arg\max} \quad (g_K + g_{K-1})\left(g_H v_H (q_H - q_L) + q_L v_L - g_H c(q_H + q_{K-2}) - g_L c(q_L + q_{K-2})\right).$$

Hence we obtain a binary distribution model. We conclude with the following lemma.

Lemma 8 (Optimality Condition for the Two Highest Intervals)
Let v_H, v_L, and g_H be defined as in (53)- (55) *as a function of $v' \in (v_{K-1}, \bar{v})$. An optimal mechanism that consists of a finite number of pooling intervals must satisfy that for all v', $v_H/v_L \leq r(g_H)$.*

However, when the distribution has unbounded support, one can separate a small enough tail of the distribution, and thus it is better to separate the last two intervals than to pool them. This is possible because by separating a small enough tail of the distribution one can generate a final interval that has low enough probability, but whose expected value is high enough compared to the previous intervals. For example, in Figure 2 we illustrate all the pair values of g_H, g_L, v_H, v_L that can be generated when a Pareto distribution is partitioned into only two intervals. We can clearly see that if the partition is done at a high enough value, the distribution of values generated is such that a two-item mechanism is better than pooling.

5.4 Comparative Statics with Respect to the Cost Elasticity

Consider some fixed information structure s that generates values $w_1, ..., w_K$ each with probability $g_1, ..., g_K$. We recall that the virtual values are:

$$\phi_k = w_k - (w_{k+1} - w_k)\frac{1 - G_k}{g_k},$$

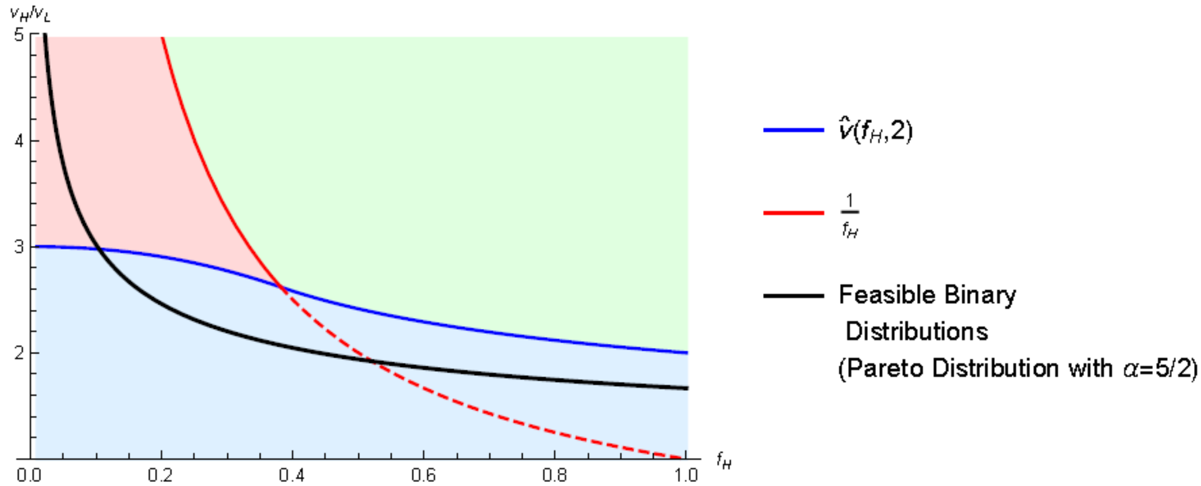

Figure 2: Feasible Binary Distributions given Pareto Distribution of Values

where $G_k = \sum_{i=1}^{k} g_i$ and $w_{K+1} = w_K$. Without loss of generality we assume that the virtual values ϕ_k are strictly increasing (since any optimal information structure will satisfy this) and $\phi_2 > 0$ (if $\phi_1 \leq 0$, there is exclusion on the first interval).

We define:

$$u(x) \triangleq \begin{cases} \frac{\eta-1}{\eta} x^{\frac{\eta}{\eta-1}} & \text{if } x \geq 0; \\ 0 & \text{otherwise} \end{cases} \tag{56}$$

If the cost is a power function $c(q) = q^\eta/\eta$, the profits generated by this information structure are:

$$\Pi_M = \sum_{k=1}^{K} g_k u(\phi_k).$$

The profits correspond to the expected utility that a risk-loving agent obtains when facing a lottery that has payments equal to the virtual values $\{\phi_k\}_{k \in K}$. The relative risk aversion for $x \geq 0$ is given by:

$$\frac{u''(x)x}{u'(x)} = \frac{1}{\eta - 1},$$

so the hypothetical risk-loving agent is more risk loving as η is closer to 1. For comparison, the pooling information structure generates profits:

$$\Pi_P = u(\mu_v).$$

We can now compare the profits generates by different information structures.

Proposition 12 (Profit Comparisons as a Function of Cost Elasticity)
Consider some information structure s:

1. *There exists η_s such that information structure s generates less profit than pooling if and only if $\eta \geq \eta_s$.*

2. *There exists η_s such that information structure s generates less profit than complete disclosure if $\eta \leq \eta_s$.*

To prove this result, we first compare the profits generated by some menu and by the pooling information structure. For this, we note that:

$$\sum_{k \in K} g_k \phi_k = w_1 \text{ and } \phi_K = w_K.$$

We thus have that:

$$\sum_{k \in K} g_k \phi_k < \mu_v \text{ and } \phi_K > \mu_v.$$

That is, the expected value of the virtual values is strictly less than the expected value, and the highest realization of the virtual values is higher than the expected value of the true values. Following the Arrow-Pratt characterization of risk aversion: a more risk-loving agent (lower η) always demands a lower certainty equivalent. Furthermore, in the limit $\eta \to \infty$ the agent becomes risk-neutral, so pooling generates higher profit than s. We then conclude that there exists a unique η_s such that:

$$\Pi_M \geq \Pi_P \iff \eta \geq \eta_s.$$

This proves the first statement.

We denote by $\hat{\Pi}$ the profits generated by complete disclosure:

$$\hat{\Pi} = \int u(\max\{\phi(v), 0\}) f(v) dv,$$

where ϕ is defined in (11). We bound the ratio between the profits generated by s and complete disclosure as follows:

$$\frac{\Pi_M}{\widehat{\Pi}_M} = \frac{\sum_{k=1}^{K} g_k u(\phi_k)}{\int u(\phi(v)) f(v) dv} < \frac{1}{\int_{\phi_K}^{\infty} \frac{u(\phi(v))}{u(\phi_K)} f(v) dv}.$$

We note that $\phi_K < \bar{v}$, so we have that:

$$\int_{\phi_K}^{\infty} f(v) dv > 0.$$

We thus have that:

$$\lim_{\eta \to 1} \frac{\Pi_M}{\widehat{\Pi}_M} < \lim_{\eta \to 1} \frac{1}{\int_{\phi_K}^{\infty} \frac{u(\phi(v))}{u(\phi_K)} f(v) dv} = 0.$$

The limit is obtained from observing that when $\eta \to 1$, the exponent in (56) converges to infinity, so the integrand diverges to infinity.

6 Discussion and Extensions

We stated our main results, Theorem 1 and 2, in the environment of nonlinear pricing first proposed by Mussa and Rosen (1978). There, the buyer's value is given by a multiplicatively separable function of willingness-to-pay and quality and the seller's cost of providing the quality is a general increasing and convex cost of quality. The objective of this section is to describe how our two main results extend to more general environments and determine which properties of the characterization are specific to the environment described in Section 2.

Towards this end, we now consider a payoff environment where the buyer's utility is nonlinear in quality q and type v. In this nonlinear environment the variable v does not directly present the value or (marginal) willingness-to-pay, and so we refer to v in this section as type v. We maintain an increasing and convex cost function $c(q)$ throughout:

$$u(v, q, p) \triangleq h(v, q) - p, \tag{57}$$

where h is a strictly increasing function and supermodular in both arguments v and q.

In the general nonlinear environment we show that the optimal menu will remain short. Theorem 3 establishes that there will be no intervals with complete disclosure about the type, small or larger. However, Theorem 3 does not provide a comprehensive characterization of the optimal information structure. In particular, in the general nonlinear environment, the optimal information structure is not guaranteed to be a monotone partition in the type v anymore.[3] In Example 1 we provide an instance where the optimal information structure is a non-monotone partition. Finally, we show that in the absence of supermodularity, the short menu result may disappear. Indeed, Example 2 gives an instance where complete disclosure is optimal.

[3] In a sender-receiver setting different from ours, Candogan and Strack (2021) also note that the linearity is critical in establishing a complete characterization of the optimal information structure.

6.1 Nonlinear Marginal Utility

Suppose then that the utility function of the buyer is given by

$$u(v,q,p) \triangleq h(v,q) - p,$$

where the function $h\left(v,q\right)$ permits a multiplicative separable representation

$$h\left(v,q\right) \triangleq h_v\left(v\right) h_q\left(q\right)$$

where $h_v\left(v\right)$ and $h_q\left(q\right)$ are continuous and strictly increasing functions: $h_v : [\underline{v}, \overline{v}] \to \mathbb{R}_+$ and $h_q : \mathbb{R}_+ \to \mathbb{R}_+$. The function $h_v\left(v\right)$ immediately suggests that we define the value in terms of $h_v\left(v\right)$ rather than v, and thus as long as h_v is a monotone function, the function h_v simply offers a new label for the types v without any further complication. Thus suppose we have

$$h(v,q) = vh_q(q), \tag{58}$$

where $h_q\left(q\right)$ is a strictly increasing concave function. In this case, we can recover our original model by defining:

$$\widehat{q} \triangleq h_q(q),$$

and writing the cost in terms of $\widehat{q}$:

$$\widehat{c}(\widehat{q}) \triangleq c(h_q^{-1}(\widehat{q})). \tag{59}$$

It is simple to verify that (given the assumptions we have made) $\widehat{c}$ will also be convex. Hence, we recover our original payoff environment by considering an appropriate change of variables. The above transformations generalize the multiplicatively separable environment with a continuum of types v and qualities q.

Maskin and Riley (1984) analyzed a model of nonlinear pricing with concave utility and constant marginal cost. We can then ask how the sufficient conditions of Theorem 2 can be adapted to their setting. Thus, we consider a model with non-linear utility

$$u(v,q,p) = vh(q) - p,$$

and constant marginal cost, $c\left(q\right) = c \cdot q$ for some constant $c \geq 0$.

We recover the original model be defining:

$$\widehat{q} = h_q(q)$$

and cost $c(\widehat{q}) = h^{-1}(\widehat{q})$. The conditions of Theorem 2 then translate to this model as follows. First, the modest-tail condition remains the same. This is because we just relabeled the quantities, which of course, does not affect how we define values. Second, the condition requiring convex marginal cost translates to:

$$\frac{\partial h(q)}{\partial q}\frac{\partial^3 h(q)}{\partial q^3} \leq 3\left(\frac{\partial^2 h(q)}{\partial q^2}\right)^2$$

A sufficient condition for this is that the marginal utility is concave (i.e., h has negative third derivative).

6.2 Nonlinear Utility in Type and Quality

We now return to the nonlinear model introduced earlier in (57). We shall assume that $h\,(v, q)$ is twice differentiable and given the supermodularity displays a positive cross-derivative. We assume that the cross derivative is bounded away from 0. That is, for all v, q:

$$\varepsilon < \frac{\partial^2 h(v,q)}{\partial v \partial q} < \frac{1}{\varepsilon}, \tag{60}$$

for some $\varepsilon > 0$. For this nonlinear model, we can now establish that for any interval $[v_1, v_2] \subset [\underline{v}, \overline{v}]$ in the support of the type distribution F, complete disclosure of the type to the buyer cannot be optimal. In consequence, the optimal information structure will always bundle types rather than disclose them.

Theorem 3 (Complete Disclosure is Never Optimal)
In every optimal mechanism, there is no interval $[v_1, v_2]$ *such that* $G^*(v) = F(v)$ *for every* $v \in [v_1, v_2]$.

Proof of Theorem 3. We assume that $G^*(v) = F(v)$ for every $v \in [v_1, v_2]$, for some interval $[v_1, v_2]$, and reach a contradiction. Following the same arguments as in Lemma 3, q^* must be strictly increasing in this interval. To reduce the amount of notation we assume that the cost function is identical to zero $c(q) = 0$ and that the buyer's utility function is strictly concave in q; if the cost is strictly increasing we can repeat the arguments by simply relabeling the buyer's utility function $\hat{h}(q, v) = h(q, v) - c(q)$.

The virtual surplus is given by:

$$\Phi(q, v) = h(v, q) - \frac{\partial h(v, q)}{\partial v} \frac{1 - F(v)}{f(v)}.$$

For any increasing quality function q, the seller's profits are:

$$\int \Phi(q(v), v) f(v) dv.$$

Analogous to the analysis in Section 3, we consider the profits generated by a mechanism that pools the allocation. The difference between the optimal policy and the variation is given by:

$$\Pi^* - \tilde{\Pi} = \int_{v_1}^{v_2} \Phi(q^*(v), v) f(v) dv - \int_{v_1}^{v_2} \frac{\int_{v_1}^{v_2} \Phi(q^*(w), v) f(w) dw}{(F(v_2) - F(v_1))} f(v) dv. \tag{61}$$

In other words, the values in $[v_1, v_2]$ are randomly assigned some of the qualities that were allocated in the optimal mechanism $q^*(v)$.

We now prove that:

$$\Pi^* - \tilde{\Pi} \leq \frac{1}{\varepsilon} \left(\int_{v_1}^{v_2} q^*(v) v f(v) dv - (F(v_2) - F(v_1)) \mu_q \mu_v \right), \tag{62}$$

where μ_v and μ_q are defined as in (12) (so here μ_v is the mean value in the interval $[v_1, v_2]$). For this, we note that the function:

$$\Psi(v, q) \triangleq \frac{1}{\varepsilon} q v - \Phi(q, v),$$

is supermodular. This follows from calculating the cross-derivative and using that the cross derivative of Φ is less than $1/\varepsilon$ (see (60)). Since $q^*(v)$ is increasing, when (w, v) are independently distributed according to F we have that $(q^*(w), v) \leq_{PDQ} (q^*(v), v)$, where the subscript PQD denotes the Positive Quadrant Dependence order (see Shaked and Shanthikumar (2007) Chapter 9). We then have that:

$$\frac{1}{F(v_2) - F(v_1)} \int_{v_1}^{v_2} \int_{v_1}^{v_2} \Psi(v, q^*(w)) f(v) f(w) dv dw \leq \int_{v_1}^{v_2} \Psi(v, q^*(v)) f(v) dv.$$

which corresponds to (9.A.17) in Shaked and Shanthikumar (2007). Replacing the definition of $\Psi(v, q)$ and replacing (61), we obtain (62).

Using the Cauchy-Schwarz inequality, as in (15), and following the same steps as before, we obtain that:

$$\lim_{\Delta \to 0} \frac{\Pi^* - \tilde{\Pi}}{\Delta^3} \leq \frac{1}{\varepsilon} \frac{dq^*(\hat{v})}{dv} f(\hat{v}). \tag{63}$$

Note that the right-hand-side of (62) is the same as the right-hand-side of (14) except for the fact that the cost does not appear (which is not necessary for the argument) and instead of ϕ we have v. This is the reason the derivative of ϕ does not appear in (63), while it does appear in (16). As before, we conclude that the efficiency losses are of order Δ^3.

We now consider the following policy (17). We additionally change the information structure so that all types in (v_1, v_2) are pooled, that is, the information structure is (17). Observe that the total surplus generated by $(\hat{q}, \hat{G})$ and by $(\tilde{q}, G)$ is the same. Then, the difference in the generated profits is equal to the difference in the expected buyer surplus. We then have that:

$$\begin{aligned}
\hat{\Pi} - \tilde{\Pi} =& (1 - F(v_1)) \left(\frac{\int_{v_1}^{v_2} \left(h(v, \mu_q) - h(v, q^*(v_1)) \right) f(v) dv}{(F(v_2) - F(v_1))} - \left(h(v_1, \mu_q) - h(v_1, q^*(v_1)) \right) \right) \\
=& \frac{(1 - F(v_1))}{(F(v_2) - F(v_1))} \left(\int_{v_1}^{v_2} \int_{v_1}^{v} \int_{q^*(v_1)}^{\mu_q} \frac{\partial^2 h(y, q^*(z))}{\partial v \partial q} dz dy f(v) dv \right) \\
\geq & \varepsilon (\mu_q - q^*(v_1))(\mu_v - v_1)(1 - F(v_1)).
\end{aligned}$$

As in Section 3, we can then conclude that:

$$\lim_{\Delta \to 0} \frac{\hat{\Pi} - \tilde{\Pi}}{\Delta^2} \geq \varepsilon \frac{dq^*(\hat{v})}{dv} (1 - F(v_1)).$$

Here we used that $(\mu_v - v_1)/\Delta \to 1$, as $\Delta \to 0$. We thus have that the efficiency losses are of order Δ^2. We thus conclude that for Δ small enough, the new policy generates higher profits. ■

The implications of Theorem 3 for the optimal information structure are notably weaker than the characterization offered in Theorem 1. In particular, we do not claim that the optimal information structure is given by a partition that is monotone and finite. The main implication of Theorem 1 is that with the above assumptions on supermodularity and differentiability there will be no complete disclosure of types anywhere. However, the pooling of types may be stochastic or not monotone in the type v. Thus the optimal menu will be coarse in the sense that some values in the distribution F will not appear in the support of the distribution G.

With weaker conditions on the differentiability and supermodularity than given by (60), the result of Theorem 1 may fail as the next example illustrates.

Example 1 (Weak Supermodularity) *Suppose the buyer has the following preferences:*

$$u(v,q,p) = \min\{q,v\} - p, \tag{64}$$

and the marginal cost of production of quality is normalized to zero. The buyer therefore has a marginal value of quality equal to 1 *until the quality reaches the level of their type* v*. Thus, higher types have higher demand for quality. We claim that in this setting the seller can extract the full surplus by offering a menu to the buyer that offers any level of quality at a per unit price of 1.*

In this example, the optimal mechanism offers complete disclosure of values and a continuum of qualities. Here, the marginal gains from higher quality or higher types are not bounded away from zero (and the utility function is not differentiable everywhere) in contrast to the above condition (60). In fact we have that

$$\frac{\partial^2 h(v,q)}{\partial q \partial v} = 0,$$

for almost every value-quality pair, which allows for full surplus extraction. Thus, we might consider the above example as a knife-edge case.

Indeed, the next example strengthens the supermodularity condition. It will lead to an optimal information structure that is a partition. However, the partition is not monotone anymore and the optimal partition combines non-adjacent types.

Example 2 (Variation of Weak Supermodularity) *Suppose there are three values* $v_L,\ v_M,\ v_H$*. The low and medium type have demand* $k < 1$ *with marginal values of 1 and 2, while the high type has unit demand with marginal value of 3:*

$$\begin{aligned} u(v_L, q) &= \ \min\{q, k\}; \\ u(v_M, q) &= 2\min\{q, k\}; \\ u(v_H, q) &= 3\min\{q, 1\}. \end{aligned}$$

The production cost is:

$$c(q) = \begin{cases} 0 & \textit{if } q \leq k; \\ \frac{1}{2}(q-k) & \textit{if } q \geq k. \end{cases}$$

We denote by f_L, f_M, f_H *the probability of each type and for the computations assume that* $f_L = f_H$ *and* $f_L, f_H < f_M$ *with* $f_L < 1/6$*.*

We claim that the optimal mechanism is to pool the low and the high types while separating the *medium type. First we observe that if the three types were pooled, then only k units are sold at a price of 2 per unit (i.e., thus a payment of $p = 2k$). Second, if the high type and the low type are pooled, then k units are sold at price a price of $2k$ to the medium type and 1 unit at a price $2k + 3(1 - k)/2$ to the pooled type. These two observations allow us to conclude that pooling the three types is dominated by pooling the high and low type and separating the medium type. By contrast, separating the three types is dominated by pooling the low and medium type. Moreover, if k is sufficiently close to 1, then pooling the three types is better than pooling the low and medium and separating the high type. We can therefore conclude that the above non-monotone pooling can be optimal.*[4]

In this second example, the supermodularity was strengthened as higher types have a higher marginal value for quality. Yet, without strictly increasing preferences for more quality everywhere, we showed that a non-monotone partitions arises as an optimal information structure. Thus, monotonicity and supermodularity appear as important conditions for the optimality of monotone partitions.

7 Conclusion

In the digital economy, the sellers and the digital intermediaries working on their behalf frequently have a substantial amount of information about the quality of the match between their products, the taste of the buyers, and ultimately the buyers' preferences. 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 buyer for the products offered.

We showed that in the presence of information and mechanism design, the seller offers a menu with only a small variety of items thus, a coarse menu. 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 buyer by narrowing the choice to a few items on the menu. We provided sufficient conditions for a broad class of distributions where this logic led the seller to offer only a single item on the menu. While

[4] This non-monotone partition is an example of a laminar partitional information structure as defined in Candogan and Strack (2021). Namely, the convex hulls of any two partition elements are such that either one contains the other or they have an empty intersection.

we obtained our results in the model of nonlinear pricing pioneered by Mussa and Rosen (1978), we showed that the coarse menu result remained a robust property in a larger class of nonlinear payoff environments.

In our analysis, the seller chooses the level of quality endogenously to match the expected taste of the buyer. In related work, McAfee (2002) matches two given distributions of, say, consumer demand and electricity supply, and shows how coarse 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 buyer, the product offered is therefore *opaque*, as its exact properties are not known to the buyer who is only guaranteed certain distributional properties of the product. This practice is sometimes referred to as *opaque pricing,* see Jiang (2007) and Shapiro and Shi (2008) for applications to services and transportation and Bergemann, Heumann, Morris, Sorokin, and Winter (2022) for auctions. Our analysis regarding the optimality of coarse 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 buyer opaque to find the optimal trade-off between efficient matching of quality and taste against the revenues from surplus extraction.

8 Appendix

Proof of Proposition 8. We can write (7) as follows:

$$\Pi = \int_{\underline{v}}^{\overline{v}} \left(wq(w) - c(q(w)) - \int_{\underline{v}}^{w} q(t)dt \right) dG(w).$$

The maximization over G (fixing q) corresponds to a classic Bayesian persuasion problem. When F has binary support, there exists an optimal information structure with binary support (see Kamenica and Gentzkow (2011)).

We then have that the distribution of expected values will have support in $\{w_L, w_H\}$ and the probabilities will be g_L, g_H satisfying:

$$g_L w_L + g_H w_H = f_L v_L + f_H v_H.$$

Thus, the optimal mechanism will be either pooling, a two-item mechanism, or a single-item mechanism in which the low type is excluded. The profits are given by (**??**)-(46), but replacing f_H, v_H, v_L with g_H, w_H, w_L. To make the notation more compact, we define:

$$\alpha \triangleq \frac{\eta}{\eta - 1}. \tag{65}$$

Clearly, if $\eta \in (1, \infty)$, $\alpha \in (1, \infty)$. Hence, we analyze this range of parameters.

We now compute the derivatives respect to g_H keeping w_L fixed (hence, adjust w_H so that the mean of expected valuations remains constant). For the second derivative, we have that:

$$\begin{aligned}
\frac{\partial^2 \Pi_M}{\partial g_H^2} =& \frac{(\alpha-1)\alpha(w_H - w_L)^2 \left(g_H^3 (1-g_H)^{-\alpha-1} (w_L - g_H w_H)^{\alpha-2} + w_H^{\alpha-2} \right)}{g_H} > 0; \\
\frac{\partial^2 \Pi_H}{\partial g_H^2} =& \frac{(\alpha-1)\alpha g_H^{-\alpha-1} (v_H - v_L)^2 (g_H v_H)^\alpha}{v_H^2} > 0
\end{aligned}$$

Hence, we will always have that in the optimum $g_H = f_H$ or $g_H \in \{0, 1\}$. We can then also conclude that, if the mechanism offers one item with exclusion, then we must have that $w_L = v_L$. So, if the mechanism offers one-item with exclusion we must have full disclosure. We now show that, if the optimal mechanism offers two items, we must have that $w_L = v_L$.

We now compute the derivative of Π_M with respect to g_H keeping $w_H = v_H$ fixed (hence, adjust w_L so that the mean of expected valuations remains constant). The first and second derivative are given by:

$$\begin{aligned}\frac{\partial \Pi_P}{\partial g_H} =& v_H^\alpha - \left(\frac{w_L - g_H v_H}{1-g_H}\right)^{\alpha-1} \frac{(2\alpha-1)(v_H - w_L) + (1-g_H)v_H}{1-g_H};\\ \frac{\partial^2 \Pi_P}{\partial g_H^2} =& \frac{2\alpha(v_H - w_L)\left(w_L - g_H v_H\right)^{\alpha-2}(1-g_H)\left((2\alpha-1)(v_H-w_L) - (1-g_H)v_H\right)}{(1-g_H)^{\alpha+1}}.\end{aligned}$$

We first observe that

$$\frac{\partial \Pi_P}{\partial g_H}\mid_{w_L = v_H} = 0 \quad \text{and} \quad \frac{\partial \Pi_P}{\partial g_H}\mid_{w_L = g_H v_H} = v_H^\alpha.$$

Furthermore, $\partial \Pi_P / \partial g_H = 0$ is convex in w_L and

$$\frac{\partial}{\partial w_L}\frac{\partial \Pi_P}{\partial g_H}\mid_{w_L = \frac{v_H(2(\alpha-1)+g_H)}{2\alpha-1}} = 0.$$

So, the first-order condition must be satisfied by some $w_L < \frac{v_H(2(\alpha-1)+g_H)}{2\alpha-1}$. From the second derivate, we have that

$$\frac{\partial^2 \Pi_P}{\partial g_H^2} \leq 0 \iff w_L \geq \frac{v_H(2(\alpha-1)+g_H)}{(2\alpha-1)}.$$

Thus, the first- and second-order condition are never satisfied. We thus reach a contradiction, so $w_L > v_L$ is never optimal. ∎

Proof of Proposition 9. Recall that the profits of the three strategies are given by (**??**)-(46). Hence, the comparisons between selling strategies will only depend on v_H/v_L. Thus, for the calculations, we can simply normalize $v_L = 1$.

We first note that, for any (f_H, v_H) such that $f_H v_H \geq 1$,

$$\Pi_H \geq \Pi_P \iff v_H \geq \frac{1-f_H}{f_H^{\frac{\eta-1}{\eta}} - f_H}.$$

We thus get the following lemma.

Lemma 9 (Pooling vs Serving Only High Type)
For all (f_H, v_H) such that $g_H v_H \geq 1$, it is optimal to pool types if and only if

$$v_H \geq \frac{1-f_H}{f_H^{\frac{\eta-1}{\eta}} - f_H}.$$

We denote by $\tilde{f}_H$ the solution to:

$$\frac{1-\tilde{f}_H}{\tilde{f}_H^{\frac{\eta-1}{\eta}} - \tilde{f}_H} \triangleq \frac{1}{\tilde{f}_H}. \tag{66}$$

We thus get that:

$$\hat{v}(f_H,\eta) = \frac{1-f_H}{f_H^{\frac{\eta-1}{\eta}} - f_H}, \text{ for all } f_H \geq \tilde{f}_H.$$

In this segment $\hat{v}(f_H,\eta)$ is decreasing in f_H and increasing in η. Note that when $v_H = 1/f_H$, we have $\Pi_M = \Pi_H$. Hence, $\hat{v}(f_H,\eta)$ will be continuous at $\tilde{f}_H$.

We now continue to analyze $\hat{v}(f_H,\eta)$ in the segment $v_H < 1/\tilde{f}_H$. We define:

$$\tilde{\Pi} \triangleq \Pi_M - \Pi_P.$$

Calculating the derivatives, we get:

$$\tilde{\Pi} \mid_{v_H=1} = 0; \tag{67}$$

$$\frac{\partial \tilde{\Pi}}{\partial v_H} \mid_{v_H=1} = -f_H\alpha < 0; \tag{68}$$

$$\frac{\partial^2 \tilde{\Pi}}{\partial v_H^2} = (\alpha-1)\alpha f_H \left(f_H(1-f_H)^{1-\alpha}(1-f_Hv_H)^{\alpha-2} - f_H(f_H(v_H-1)+1)^{\alpha-2} + v_H^{\alpha-2}\right) > 0, \tag{69}$$

where α is defined in (65). The last inequality can be verified as follows. If $\alpha < 2$, then:

$$(1-f_H)^{1-\alpha}(1-f_Hv_H)^{\alpha-2} > 1 \text{ and } (f_H(v_H-1)+1)^{\alpha-2} < 1.$$

On the other hand, if $\alpha > 2$, then

$$(f_H(v_H-1)+1) < v_H.$$

Hence, in either case $\tilde{\Pi}$ is convex with respect to v_H. We have that (67)-(69) implies that: (a) there exists a unique threshold v_H at which Π_M generates higher profits than Π_P (hence confirming that $\hat{v}(f_H,\eta)$ exists also for $f_H \leq \tilde{f}_H$), and (b) we have that, for v_H such that $\tilde{\Pi} = 0$, we must have that $\partial\tilde{\Pi}/\partial v_H > 0$.

We now show that $\hat{v}(f_H,\eta)$ is decreasing in η (in the segment $f_H \leq \tilde{f}_H$). The implicit function theorem states that:

$$\frac{\partial \hat{v}(f_H,\eta)}{\partial \eta} = \frac{\frac{\partial \tilde{\Pi}}{\partial \eta}}{-\frac{\partial \tilde{\Pi}}{\partial v_H}}.$$

We already proved the denominator will be negative, so we now prove that the numerator is negative. Hence, we prove that, for any (f_H, v_H) such that $\tilde{\Pi} = 0$, the numerator is also negative.

We define:

$$\tilde{\Pi}' \triangleq \log\left(\frac{\left(\frac{(1-f_Hv_H)^\alpha}{(1-f_H)^{\alpha-1}} + f_Hv_H^\alpha\right)}{(f_L+f_Hv_H)^\alpha}\right).$$

This is a monotonic transformation of $\tilde{\Pi}$, which will help get more compact expressions. We now check that:

$$\frac{\partial^2 \tilde{\Pi}'}{\partial \alpha^2} = \frac{f_H(1-f_H)^{\alpha+1} v_H^\alpha (1-f_H v_H)^\alpha (-\log(1-f_H v_H) + \log(1-f_H) + \log(v_H))^2}{(f_H(1-f_H)^\alpha v_H^\alpha - (f_H - 1)(1-f_H v_H)^\alpha)^2} > 0$$

That is, $\tilde{\Pi}'$ is convex in α, and we note that:

$$\tilde{\Pi}' \mid_{\alpha=1} = -\log(1 - f_H + f_H v_H) < 0;$$

$$\lim_{\alpha \to \infty} \tilde{\Pi}' = \infty > 0.$$

Thus, $\tilde{\Pi}' = 0$ for one $\alpha \in (1, \infty)$, and $\tilde{\Pi}'$ is increasing in α whenever $\tilde{\Pi}' = 0$. We can then conclude that there exists a unique $\alpha \in (1, \infty)$ such that $\tilde{\Pi} = 0$, and $\tilde{\Pi}$ is increasing in α whenever $\tilde{\Pi} = 0$, and thus decreasing in η. Hence, following the implicit function theorem, $\hat{v}(f_H, \eta)$ is increasing in η.

We now show that $\hat{v}(f_H, \eta)$ is decreasing in f_H (in the segment $f_H \leq \tilde{f}_H$). We prove this separately for the case $\eta \leq 2$ and $\eta > 2$. We prove the case $\eta \leq 2$ by appealing to the implicit function theorem. We note that:

$$\begin{aligned}
\tilde{\Pi} \mid_{f_H=0} &= 0 \\
\frac{\partial \tilde{\Pi}}{\partial f_H} \mid_{f_H=0} &= v_H^\alpha - 2\alpha(v_H - 1) - 1 \\
\frac{\partial^2 \tilde{\Pi}}{\partial f_H^2} &= (\alpha-1)\alpha(v_H-1)^2 \left(\frac{(1-f_H v_H)^{\alpha-2}}{(1-f_H)^{\alpha+1}} - (f_H(v_H-1)+1)^{\alpha-2} \right) > 0.
\end{aligned}$$

In this case, we can sign the last term when $\alpha < 2$, then:

$$\frac{(1-f_H v_H)^{\alpha-2}}{(1-f_H)^{\alpha+1}} > 1 \text{ and } (f_H(v_H-1)+1)^{\alpha-2} < 1.$$

Hence, for a fixed v_H, there exists f_H such that $\tilde{\Pi} = 0$ only if $\frac{\partial \tilde{\Pi}}{\partial f_H} \mid_{f_H=0} > 0$. And we then have that $\tilde{\Pi} = 0$ implies that $\frac{\partial \tilde{\Pi}}{\partial f_H} > 0$. Following the implicit function theorem, $\hat{v}(f_H, \eta)$ is decreasing in f_H.

We now prove that $\hat{v}(g_H, \eta)$ is decreasing also for $\eta > 2$. For this, we define:

$$q_P \triangleq (g_L v_L + g_H v_H)^{\frac{1}{\eta-1}},$$

which is the quality sold in the pooling mechanism. Using the envelope theorem, we have that:

$$\frac{\partial \tilde{\Pi}}{\partial g_H} = v_H(q_H - q_L) - (c(q_H) + c(q_L)) - (v_H - v_L) q_P,$$

where q_H and q_L are defined in (45). The objective function of Π_M and Π_P is linear in g_H. So, at any point such that $\tilde{\Pi} = 0$, we have that,

$$\frac{\partial \tilde{\Pi}}{\partial g_H} > 0$$

if and only if

$$q_L v_L - c(q_L) - (q_P v_L - c(q_P)) < 0. \tag{70}$$

The left-hand-side of the inequality is the difference between the intercept of Π_M and Π_P. We now prove that (70) is satisfied at any point such that $\tilde{\Pi} = 0$ when $\eta \geq 2$.

We begin by noting that:

$$1 = \arg\max_{q \in \mathbb{R}} qv_L - c(q).$$

We also note that $q_L < 1 < q_P$ and the objective function is concave. To make the notation more compact, we define:

$$\delta \triangleq f_H(v_H - v_L)$$

and note that $q_P = (1+\delta)^{1/(\eta-1)}$. We also note that:

$$q_L = (v_L - \frac{(v_H - v_L) f_H}{f_L})^{\frac{1}{\eta-1}} < (1-\delta)^{\frac{1}{\eta-1}} < 1.$$

So, we have that:

$$q_L v_L - c(q_L) - (q_P v_L - c(q_P)) < (1-\delta)^{\frac{1}{\eta-1}} v_L - \frac{(1-\delta)^{\frac{\eta}{\eta-1}}}{\eta} - ((1+\delta)^{\frac{1}{\eta-1}} v_L - \frac{(1+\delta)^{\frac{\eta}{\eta-1}}}{\eta}).$$

We now show that the right-hand-side is less than 0. For this, we write this term as a ratio and note that:

$$\frac{\eta(1-\delta)^{\frac{1}{\eta-1}} v_L - (1-\delta)^{\frac{\eta}{\eta-1}}}{\eta(1+\delta)^{\frac{1}{\eta-1}} v_L - (1+\delta)^{\frac{\eta}{\eta-1}}} |_{\delta=0} = 1$$

and

$$\frac{\partial}{\partial \delta}\left(\frac{(1-\delta)^{\frac{1}{\eta-1}} v_L - (1-\delta)^{\frac{\eta}{\eta-1}}}{(1+\delta)^{\frac{1}{\eta-1}} v_L - (1+\delta)^{\frac{\eta}{\eta-1}}}\right) = -\frac{2\delta^2(\eta-2)\eta(1-\delta)^{\frac{1}{\eta-1}-1}(\delta+1)^{-\frac{\eta}{\eta-1}}}{(\eta-1)(\delta-\eta+1)^2} < 0.$$

To check the inequality it is useful to recall that we are analyzing the range $f_H v_H < v_L$, and so $\delta < 1$. Hence, for all $\delta > 0$,

$$\frac{(1-\delta)^{\frac{1}{\eta-1}} v_L - (1-\delta)^{\frac{\eta}{\eta-1}}}{(1+\delta)^{\frac{1}{\eta-1}} v_L - (1+\delta)^{\frac{\eta}{\eta-1}}} < 1.$$

This proves that

$$q_L v_L - c(q_L) - (q_P v_L - c(q_P)) < 0,$$

which concludes the proof. ■

References

ANDERSON, E., AND J. DANA (2009): "When is Price Discrimination Profitable," *Management Science*, 55, 980–989.

BERGEMANN, D., B. BROOKS, AND S. MORRIS (2015): "The Limits of Price Discrimination," *American Economic Review*, 105, 921–957.

BERGEMANN, D., T. HEUMANN, S. MORRIS, C. SOROKIN, AND E. WINTER (2022): "Optimal Information Disclosure in Auctions," *American Economic Review: Insights*, forthcoming.

BERGEMANN, D., AND M. PESENDORFER (2007): "Information Structures in Optimal Auctions," *Journal of Economic Theory*, 137, 580–609.

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

BROOKS, B., AND S. DU (2021): "Optimal Auction Design with Common Values: An Informationally Robust Approach," *Econometrica*, 89, 1313–1360.

CANDOGAN, O., AND P. STRACK (2021): "Optimal Disclosure of Information to a Privately Informed Receiver," .

CAVALLO, A. (2017): "Are Online and Offline Prices Similar? Evidence from Large Multi-Channel Retailers," *American Economic Review*, 107(1), 283–303.

——— (2019): "More Amazon Effects: Online Competition and Pricing Behaviors," in *Jackson Hole Economic Symposium Proceedings*, ed. by F. R. B. of Kansas City, pp. 291–355.

DELLAVIGNA, S., AND M. GENTZKOW (2019): "Uniform Pricing in US Retail Chains," *Quarterly Journal of Economics*, 134, 2011–2084.

EBAY (2022): "Best Match," *https://tech.ebayinc.com/product/ebay-makes-search-more-efficient-through-personalization/.*

HAGHPANAH, N., AND R. SIEGEL (2022): "The Limits of Multi-Product Discrimination," *American Economic Review: Insights*, forthcoming.

Jiang, Y. (2007): "Price Discrimination with Opaque Products," *Journal of Revenue and Pricing Management*, 6, 118–134.

Kamenica, E., and M. Gentzkow (2011): "Bayesian Persuasion," *American Economic Review*, 101, 2590–2615.

Kleiner, A., B. Moldovanu, and P. Strack (2021): "Extreme Points and Majorization: Economic Applications," *Econometrica*, 89, 1557–1593.

Loertscher, S., and E. Muir (2022): "Monopoly Pricing, Optimal Randomization and Resale," *Journal of Political Economy*, 130, 566–635.

Maskin, E., and J. Riley (1984): "Monopoly with Incomplete Information," *RAND Journal of Economics*, 15, 171–196.

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

Mussa, M., and S. Rosen (1978): "Monopoly and Product Quality," *Journal of Economic Theory*, 18, 301–317.

Myerson, R. (1981): "Optimal Auction Design," *Mathematics of Operations Research*, 6, 58–73.

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

Roesler, A., and B. Szentes (2017): "Buyer-Optimal Learning and Monopoly Pricing," *American Economic Review*, 107, 2072–2080.

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

Shapiro, D., and X. Shi (2008): "Market Segmentation: The Role of Opaque Travel Agencies," *Journal of Economics and Management Strategy*, 17, 803–837.

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