Screening with Persuasion*

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

October 15, 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, Christopher Sandmann, 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 buyers. In particular, by choosing how much information to disclose to the buyers about the value of the match between buyers 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. The seller can offer a variety of products that are differentiated by their quality and so screen the buyers for their willingness-to-pay. The seller can also control how much information to disclose to the buyers about their willingness-to-pay for the qualities, thus screens while engaging in Bayesian persuasion.

We characterize the information structure and menu of choices that maximizes the expected profits of the seller. The buyers have a continuum of possible values–their willingness to pay for the quality. In the absence of any information design, the optimal menu offers a continuum of qualities to the buyers who then selects given their type as in Mussa & Rosen (1978). In this setting, the seller controls the selling mechanism and the information structure. Yet, the seller cannot observe the value or signal realization of the buyers. The selling mechanism could be any (possibly stochastic) menu. Our main analysis considers the case where the distribution of ex post qualities to be sold by the seller is exogenously given (as in the recent work of Loertscher & Muir (2022)). The seller can create bundles of different qualities to sell to the buyers. We then extend our results to the case where the seller produces goods with different qualities with a convex cost function, as in the classic analysis of Mussa & Rosen (1978) or Maskin & Riley (1984).

Our main result is a characterization of the structure of the optimal information and menu, namely pooling of values and pooling of qualities. The seller provides information via monotone pooling, i.e., every value is pooled with a positive mass of nearby values and the pool takes the form of an interval. There is also monotone pooling of qualities, where each pool of buyers is allocated a specific pool of qualities. Using the convexity of the information rent, the buyers are indeed pooled into finitely many intervals and are offered a menu with a finitely many items. This contrasts with the menu with a continuum of options in the absence of information design.

The argument is as follows. Once one fixes the distribution of expected values induced by an information structures and a distribution of expected qualities, the revenue is pinned down by standard Myersonian arguments. The problem then reduces to maximizing a linear functional subject to *two* majorization constraints. As a first step, we establish that information is given by a monotone partition, where elements of the partition are either intervals (with pooling) or singletons (with full disclosure of the value). Theorem 1 follows from fixing the pooled distribution of qualities and solving the problem with one majorization constraint. The monotone partition argument then follows from Myerson (1981) and Kleiner et al. (2021). The second step uses an orthogonal variational argument. Suppose that an interval of values were perfect revealed and screened by the seller. We ask what happens to profits if we pool the allocation of a small interval of values. By construction, distortions in the allocation from the profit-maximizing allocation will only cause second-order distortions to the total virtual surplus. But if we additionally pool a small interval of values into a single expected type causes a first-order decrease in the information rents. Hence, screening an open set of values is never optimal because pooling the values, and consequently the allocation, causes a first-order reduction in the information rents and only second-order distortions on profits. Theorem 2 establishes that the optimal menu with persuasion is discrete and thus "small" relative to the (continuum) menu in the absence of persuasion. But how small will the optimal menu be?

A second set of results establishes sufficient conditions under which the partitions of values and qualities will be very coarse and thus the number of items in the menu will be small. If the distribution of qualities is convex (i.e., the density is weakly increasing), then a single item menu is optimal (Theorem 3). More generally, Proposition 2 shows that the number of items in the menu is no more than (the next integer lower than) the ratio of the upper bound to the lower bound of the exogenously quality distribution. By contrast, with complete disclosure of information there would typically be a continuum of items in the optimal menu.

Our main result goes through unchanged in the classic setting of Mussa & Rosen (1978) where the distribution of qualities for sale is endogenous and the seller has a convex cost of producing quality. The intuition is that the above argument shows that pooling will increase revenue, and with a convex production cost, pooling will additionally reduce cost and so pooling will increase profits. The number of items in the optimal menu will now be sensitive to the convexity of the cost function. In Section 6 we report results for the case where the cost function has a constant elasticity. In Theorem 5 we provide a necessary and sufficient condition for the optimality of a single-item which depends only on the support of the distribution and the elasticity of the cost function. In particular, a single-item menu is optimal for all distribution of values and qualities if and only if the ratio of upper bound and lower bound of the support of the quality distribution is less than the cost 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 and distribution-free condition will necessarily fail as the lower bound approaches zero. Second, any given information structure generates less profits than pooling all values (i.e., providing no information) when the cost elasticity is high enough, and will generate less profits than full disclosure if the cost elasticity is low enough (Proposition 3).

Our setting reflects three notable features of the digital economy. We already mentioned the fact that the sellers are well-informed about buyers' values and specifically their match value with the products of the seller. Our analysis considers the extreme case where the buyers 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. In Section 7, we show that the optimal mechanism can be implemented as an indirect mechanism in the form of a recommender system.

1.2 Related Literature

We analyze a model of nonlinear pricing and consider first the case where the distribution of qualities is exogenous as in the recent work of Loertscher & Muir (2022). Subsequently, we establish that our main result continues to hold in the setting of Mussa & Rosen (1978) and Maskin & Riley (1984) where the distribution of qualities is endogenous and an increasing function of quality. Loertscher & 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 on optimal pooling hold for all regular and irregular distributions.

In our analysis, the seller can control the information of buyers, as well as the pooling of qualities and the mechanism. It therefore combines Bayesian persuasion (Kamenica & 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 et al. (2021), we are not aware of any related work in optimal pricing that combines mechanism and information design. The closest work is that of Bergemann & Pesendorfer (2007), who consider a seller with many unit-demand buyers. We postpone a detailed discussion to Section 3.2 after we obtain a canonical statement of our problem.

We analyze a second-degree price discrimination problem. As the seller pools buyers with adjacent values, the seller creates segments within the single aggregate market. In doing so, the seller makes each items intended for one segment less attractive to the other segments. In this sense, the seller is inducing an element of third-degree price discrimination. By contrast, Bergemann et al. (2015) and Haghpanah & Siegel (2022) allow many segments and thus full third-degree price discrimination while offering quality-differentiated products. Roesler & Szentes (2017) consider the buyer-optimal information structure for a single-item demand and a single aggregate market. Thus, the demand structure and the objective differ from the present work, but they share the focus on creating segments within a single aggregate market.

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. But importantly, the information structure only affects the allocation and not the expectation of the buyers 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.

2 Model

A seller supplies goods of varying quality $q \in \mathbb{R}_+$ to a continuum of buyers with mass 1. The seller has a mass 1 of goods with qualities $q \in [\underline{q}, \bar{q}]$ distributed according to

$$Q \in \Delta\left([\underline{q}, \bar{q}]\right),$$

where $0 \leq \underline{q} \leq \bar{q} < \infty$. Thus, the seller can offer a fixed and exogenously given distribution Q of qualities. In Section 6 we extend the analysis to a setting with endogenously chosen quantities as in Mussa & Rosen (1978) and Maskin & Riley (1984). We introduce the additional notation at that point.

The buyers have unit demand and a willingness-to-pay, or short value $v \in \mathbb{R}_+$ for quality q. The utility net of the payment $p \in \mathbb{R}_+$ is:

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

The buyers' values $v \in [\underline{v}, \bar{v}]$ are distributed according to

$$F \in \Delta\left([\underline{v}, \bar{v}]\right),$$

where $0 \leq \underline{v} < \bar{v} < \infty$, with strictly positive and non-vanishing density (i.e., $F^{'}(v) > 0$).

The seller's choice has two components: (i) the seller chooses the information that buyers have about their own value, and (ii) the seller chooses a direct mechanism (or short menu) that specifies the (expected) quality and payments for any reported (expected) value. We now describe these elements in turn.

First, the seller chooses a signal (or information structure):

$$S : \mathbb{R}_+ \to \Delta\mathbb{R}_+, \tag{2}$$

where $s(v)$ denotes a signal realization when the value is v observed by a buyer. A buyer's expected value conditional on the signal realization s is denoted by:

$$w \triangleq \mathbb{E}[v \mid s]. \tag{3}$$

Since the utility is linear in v, w is a sufficient statistic for determining the buyers' preferences when they observe signal s. The information that the buyers receive about their expected value $w \in \mathbb{R}_+$ is represented by a distribution of expected values G:

$$G \in \Delta\left([\underline{v}, \overline{v}]\right),$$

and $\operatorname{supp} G$ denotes the support of the distribution G.

Second, the seller choose a menu of qualities and payments. The seller has the ability to shape the distribution of qualities sold by pooling goods of different quality. If the buyers are offered a distribution of qualities (or bundle) of qualities B, the expected quality offered is denoted by

$$q = \int t dB(t).$$

If $\int dB(t)dt < 1$, then the probability of receiving an object is less than 1. This is of course equivalent to receiving a quality zero with probability $1 - \int dB(t)$. The expected utility of a bundle B net of the payment $p \in \mathbb{R}_+$ is:

$$u(w, q, p) = w \int t dB(t) - p. \tag{4}$$

A menu (or direct mechanism) with bundles $B(\cdot|\, w)$ at prices $p(w)$ for every value w is given by:

$$M \triangleq \{(B(\cdot|\, w), p(w))\}_{w \in \text{supp}G}.$$

The menu has to satisfy incentive compatibility and participation constraints:

$$wq(w) - p(w) \geq wq(w') - p(w'), \quad \forall w, w' \in \text{supp}G; \tag{5}$$

$$wq(w) - p(w) \geq 0; \quad \forall w \in \text{supp}G; \tag{6}$$

as well as feasibility. Namely, the amount of goods offered must be weakly less than the amount of goods the seller owns:

$$\int_{\underline{v}}^{\bar{v}} \int_{q}^{q'} dB(t|\, w)\, dG(w) \leq \int_{q}^{q'} dQ(t), \quad \forall q, q' \in [\underline{q}, \bar{q}] \tag{7}$$

Hence, the total amount of goods of quality in any interval $[q, q']$ offered to the buyers must be weakly less than the goods of the same quality in the seller's endowment.

We refer to a mechanism as a pair (S, M) of information structure S and menu M. The seller's problem is to maximize expected profits subject to the above incentive, participation and feasibility constraints (5)-(7):

$$\Pi \triangleq \max_{\substack{S:\mathbb{R}_+ \to \Delta(\mathbb{R}_+) \\ (B(q;w), p(w))}} \mathbb{E}[p(w)]. \tag{8}$$

Since there is no cost, the profits are equal to revenue.

A first significant step in the analysis is to show that the above seller's problem can be stated entirely in terms of a choice of a pair of distributions over expected values and expected qualities respectively subject to the appropriate majorization constraints. The menu, namely allocation $B(\cdot|\, w)$ and transfer $p(w)$ can be entirely represented in terms of these distributions.

3 A Reformulation with Majorization Constraints

We start with a re-statement of the seller's problem exploiting standard properties of incentive compatible and feasible mechanisms. This will lead us to state the revenue maximization (8) as an optimization problem subject to two majorization constraints that is bilinear in the distributions of expected values and expected qualities. This reformulation uses familiar insights to obtain a maximization problem subject to majorization constraints. Yet, the specific solution is different since we are jointly optimizing in the space of values and qualities (allocations). Thus, we obtain two majorization constraints, one for each dimension of the design problem with significant implications for the nature of the optimal solution.

3.1 Two Majorization Constraints

The seller's choice has two components: (i) the seller chooses the information that buyers have about their own value, and (ii) the seller chooses a mechanism that specifies the (expected) quality and payments for any reported (expected) value.

Following standard techniques, the incentive compatibility requires that the allocation $q(w)$ is increasing and the payments $p(w)$ are determined by the allocation rule using the Envelope condition:

$$\mathbb{E}[p(w)] = \int_{\underline{v}}^{\bar{v}} \left(wq(w) - \int_{\underline{v}}^{w} q(w)dw \right) dG(w), \tag{9}$$

where G is the distribution of expected valuations. As it is standard in the literature, the second term inside the integral is the buyers' surplus. Initially, the supply of qualities is given and there are no production costs and so the expected profits are equal to the expected payments. The seller's objective is to maximize (9) subject to the constraint that G can be induced by some information structure and the quality assignments are feasible (of course, whether $q(w)$ is feasible depends on G). Note that G may have gaps but it is without loss of generality to assume that q is defined on the whole domain $[\underline{v}, \bar{v}]$ and hence we can pin down payments uniquely with the allocation rule.

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

$$\int_{v}^{\bar{v}} F(t)dt \leq \int_{v}^{\bar{v}} G(t)dt, \, \forall v \in [\underline{v}, \bar{v}],$$

with equality for $v = \underline{v}$. If G is a mean-preserving contraction of F (or G majorizes F), we write $G \succ F$. Following Shaked & Shanthikumar (2007) (Chapter 3), we have that $G \succ F$ if and only if $F^{-1} \succ G^{-1}$.

We describe the set of feasible allocations $q(w)$. Whether a given allocation rule $q(w)$ is feasible depends on G. However, we can eliminate this implicit dependence by describing the allocation rule in terms of quantiles. For this, we define a distribution of expected qualities

$$R \in \Delta\left([\underline{q}, \overline{q}]\right),$$

and an associated quantile distribution of qualities $R^{-1} : [0,1] \to [\underline{q}, \bar{q}]$:

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

That is, R is the distribution of expected qualities of the goods offered to the buyers (and hence the quantile allocation rule is R^{-1}). Following Kleiner et al. (2021) (see Proposition 4), the allocation rule $q(w)$ is feasible if and only if the distribution of qualities R^{-1} satisfies

$$\int_q^{\overline{q}} R^{-1}(t)dt \leq \int_q^{\overline{q}} Q^{-1}(t)dt,\ \forall q \in [\underline{q}, \overline{q}],$$

where we do not require that the inequality becomes an equality at $q = \underline{q}$. In this case, we say Q^{-1} *weakly majorizes* R^{-1} and we write $R^{-1} \prec_w Q^{-1}$. The seller has the ability to pool goods of different quality, which corresponds to choosing distributions R that are a mean preserving contraction of the distribution Q. However, the seller has the ability to not sell some goods, which is equivalent to setting the quality of some good to 0. This is equivalent to allow downward shifts in the distribution of quality in the sense of first-order stochastic dominance.

Replacing $q(w)$ in (9) and integrating by parts, the expected payments can be written as follows:

$$\mathbb{E}[p(w)] = \int_0^1 G^{-1}(t)(1-t)dR^{-1}(t).$$

The seller's problem is then given by:

$$\max_{\substack{G^{-1} \prec F^{-1} \\ R^{-1} \prec_w Q^{-1}}} \int_0^1 G^{-1}(t)(1-t)dR^{-1}(t), \tag{10}$$

and R^{-1} is measurable with respect to G. The additional measurability condition is to guarantee that we can implement the allocation rule using a direct mechanism $q(w) = R^{-1}(G(w))$. We thus obtain an optimization problem subject to two majorization constraints that is bilinear in R and G. We denote by (G^*, R^*) a solution to this problem.

3.2 One Vs Two Majorization Constraints

We can now relate our result more precisely to the existing literature. It is useful to focus on the re-formulation of the seller's problem as a linear function that is maximized subject to two majorization constraints, as in (10). While there is a rich literature studying related problems, typically the concern is with the optimization over one of the two distributions. For example, if we take (10) and impose the constraint that the seller cannot pool the values of the buyers (i.e. $F = G$), then we recover the recent work of Loertscher & Muir (2022) who characterize the optimal selling policy for a distribution of qualities to a continuum of buyers.

Both Loertscher & Muir (2022) and the present work consider a classic second degree price discrimination environment without competition among the buyers. But the analysis naturally extends to competing buyers, and hence auctions. We can then interpret q as the probability of receiving the object in the auction and have a model of quantity discrimination. With this interpretation, if we impose next to the condition of $F = G$ the constraint that the distribution of quantities is given by $Q(q) = q^{\frac{1}{N-1}}$, then we characterize the optimal symmetric auction of an indivisible good when there are N symmetric bidders, as in Myerson (1981).[1] If instead we impose that $R = Q = q^{\frac{1}{N-1}}$ and optimize only over G, we recover the setting of Bergemann, Heumann, Morris, Sorokin & Winter (2022). They study the problem of finding the revenue-maximizing information structure G in a symmetric second-price auction. As the allocation is efficient conditional on the information in a second price auction, the allocation of quantities must be $R = q^{\frac{1}{N-1}}$. If we relax the assumption that $Q = q^{\frac{1}{N-1}}$ (but maintain the assumption that $R = Q$), we recover the problem of finding the pooling of qualities that maximize buyers surplus in a two-sided matching market (see Proposition 4 in Kleiner et al. (2021)). All the above optimization problems can be solved using ironing techniques first developed by Myerson (1981) and later generalized by Kleiner et al. (2021). In all of the above work pooling may arise as a result of ironing (if the buyers' distribution of values is *regular*, then pooling will never arise). In contrast, in our work where we are maximizing over both G and R, pooling arises as an inevitable property of the optimal mechanism via a very distinct argument. Furthermore, the optimal mechanism is *very* coarse in the sense that it offers very few distinct qualities.

The closest work to our paper is Bergemann & Pesendorfer (2007). They study the problem of characterizing the optimal auction of an indivisible good when there are N bidders and the seller can control the bidders' information. The main result in Bergemann & Pesendorfer (2007) establishes that even in a symmetric environment, the optimal asymmetric auction is better than the

[1] When the seller chooses the optimal selling mechanism of an indivisible object with N bidders, the allocation rule is a function $q_i(v_i, v_{-i})$ that determines the probability that agent i wins the obejct given the report of the other bidders. The winning probability is $q_i(v_i) = \int q_i(v_i, v_{-i}) f(v_{-i}) dv_{-i}$. In a symmetric auction, a non-decreasing allocation rule can be implemented if and only if $q^{\frac{1}{N-1}} \prec_w Q$ as developed in detail in Kleiner et al. (2021).

optimal symmetric auction. This result does not have a direct counterpart in our model of second degree price discrimination (without competition among buyers). Bergemann & Pesendorfer (2007) develop ad hoc arguments for the asymmetry of the optimal auction that are based on the convexity of the information rents. By contrast, we derive the optimality of a monotone pooling distribution using majorization constraints and extreme points as in Kleiner et al. (2021). In particular, the majorization argument for pooling allows us to extend the results to the case where the qualities are endogenously determined by the seller as in Section 6. The argument that leads to the monotone pooling result are related to those used by Wilson (1989). While exogenously limiting the number N of items, he shows that in a *surplus-maximizing* mechanism this restriction only causes surplus losses of order $1/N^2$. We complement this argument with the fact that the gains that come from reducing informational rents are always an order of magnitude larger. We thus conclude that we always get some degree of pooling in the *profit-maximizing* mechanism.

4 Structure of the Optimal Mechanism

We now provide a characterization of the optimal mechanism. The mechanism specifies a distribution of expected values G as the choice of information and a distribution of (expected) qualities R as the choice of available qualities. It is often useful to work with the inverses of the value and quality distributions G^{-1} and R^{-1}. Now $G^{-1}(t)$ is the value of quantile t under G and $R^{-1}(t)$ is the expected quality of quantile t under R. We introduce some language for critical properties of the value and quality distributions.

A distribution of values G is said to be *monotone partitional* if the support is partitioned into intervals and buyers know only what element of the partition they are in. Formally, G is monotone partitional if $[0,1]$ is partitioned into countable intervals $[x_i, x_{i+1})_{i\in I}$ and each interval either has full disclosure, i.e., all buyers with values corresponding to quantiles in that interval know their value; or pooling, i.e., buyers know only that their value corresponds to a quantile in the interval $[x_i, x_{i+1})$, and so their expected value is

$$w_i \triangleq \mathbb{E}[v_i \mid F(v_i) \in [x_i, x_{i+1})].$$

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

$$w_i = \frac{\int_{x_i}^{x_{i+1}} F^{-1}(t)\, dt}{x_{i+1} - x_i}.$$

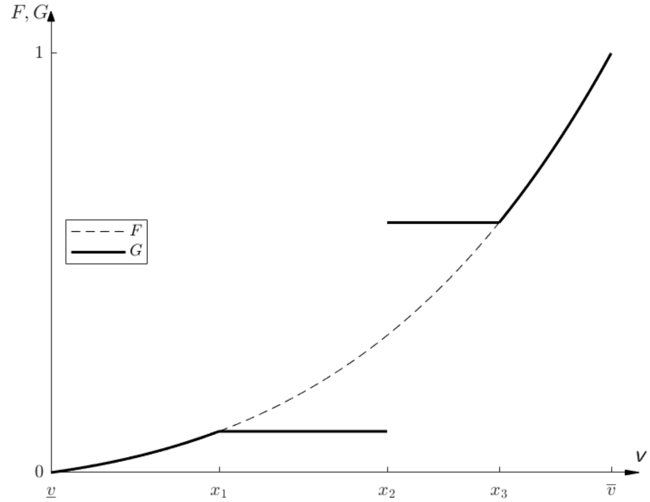

Figure 1: A monotone partitional distribution G which majorizes F and has intervals of complete disclosure and of pooled disclosure.

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

$$G^{-1}(t) \triangleq \begin{cases} F^{-1}(t), & \text{if } t \in [x_i, x_{i+1}) \text{ for some } i \in J; \\ w_i, & \text{if } t \in [x_i, x_{i+1}) \text{ for some } i \notin J. \end{cases} \tag{11}$$

The distribution G is said to be *monotone pooling* if all intervals are pooling. In Figure 1 and 2 we illustrate a monotone partitional and a monotone pooling distribution G of the same underlying distribution F (represented by the dashed curve).

We can similarly define monotone partitional and monotone pooling distributions R of qualities. However, the quality distribution R needs to be only *weakly* majorized by Q, and so we need to make a small change to the definition. We say R is a *weak* monotone partitional distribution if there exists a monotone partitional distribution $\hat{R}$ and an indicator function $\mathbb{I}_{t\geq\widehat{x}}$ such that:

$$R^{-1}(t) = \hat{R}^{-1}(t) \cdot \mathbb{I}_{t\geq\hat{x}}(t),$$

where $\mathbb{I}_{t\geq\widehat{x}}(t)$ is the indicator function, which we recall is defined as follows:

$$\mathbb{I}_{t\geq\widehat{x}}(t) \triangleq \begin{cases} 1 & \text{if } t \geq \widehat{x} \\ 0 & \text{if } t < \widehat{x} \end{cases}$$

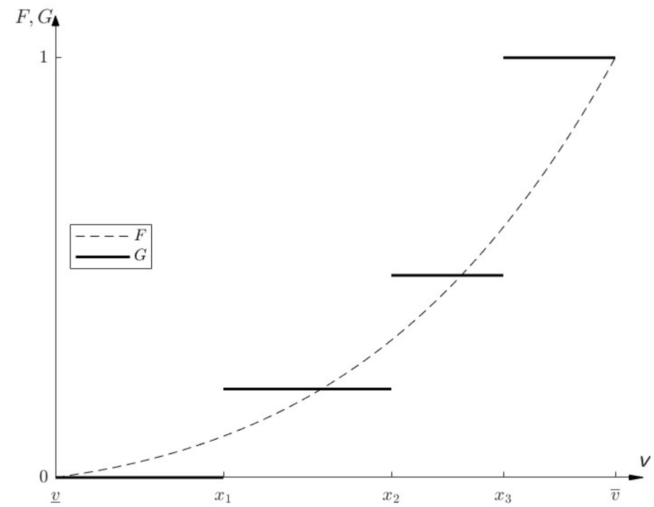

Figure 2: A monotone pooling distribution G which majorizes F and has only intervals of pooled disclosure.

In other words, a weak monotone partitional distribution is generated by first taking all qualities corresponding to quantiles below $\widehat{x}$ and reducing these qualities to zero, and then generating monotone partitional distribution (where some low qualities have been reduced to 0). A *weak* monotone pooling distribution is defined in the analogous sense, where all intervals are pooling.

To simplify the exposition, henceforth we omit the qualifier "weak". Thus, when we say a distribution of qualities R is a monotone partitional (or monotone pooling) distribution R, it is always in the weak sense.

We say that G and R have *common support* if the induced *partitions of quantiles* that generate the monotone partition distributions are the same. Our first main result establishes that the optimal distributions G and R are monotone pooling distributions.

Theorem 1 (Structure of the Optimal Mechanism)
In every optimal mechanism, G^* and R^* are monotone pooling distributions and they have common quantile support.

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 as represented by G^*. 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 $(G, R, (q,t))$ 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 interval of the 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.

This first result establishes that every optimal mechanism has monotone pooling information structure. In consequence, the optimal menu will contain only a countable number of items. This contrasts with the continuum of items which would be optimal in the absence of a choice regarding the information structure.

We will prove Theorem 1 in two steps. The first step shows that there exists an optimal mechanism in which the optimal value and quality distribution is monotone partitional. This step shows 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 et al. (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 distributions are partitional. This follows from the fact that the menu and the information structure are jointly optimized. If the information structure were not monotone partitional, then it would be possible to write it as a linear combination of partitional 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 monotone partitional information structure.

The second step shows that there is no interval of complete information disclosure, and thus the distribution are always monotone pooling. 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 allocations jointly always increases the seller's profits.

4.1 Monotone Partitional Distribution

The choice of information structure G must be optimal if we hold fixed a distribution R^* of qualities. So we consider the problem of choosing G to maximize

$$\Pi = \underset{\{G^{-1}: G^{-1} \prec F^{-1}\}}{\arg\max} \int_0^1 G^{-1}(t)(1-t) dR^{*-1}(t). \tag{12}$$

The optimization problem (12) is an upper semi-continuous linear functional of G^{-1}. 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, $\limsup \int_0^1 \hat{G}^{-1}(t)(1-t)dR^{*-1}(t) \leq \int_0^1 G^{-1}(t)(1-t)dR^{*-1}(t)$.

Proposition 1 in Kleiner et al. (2021) shows that the set $\{G^{-1} : G^{-1} \prec F^{-1}\}$ is a convex and compact set, and their Theorem 1 shows that the extreme points of this set are given by (11). Following Bauer's maximum principle, the maximization problem attains its maximum at an extreme point of $\{G^{-1} : G^{-1} \prec F^{-1}\}$.

The objective function is also linear R, so the same analysis applies. The only difference is that the set of extreme points of the set $\{R^{-1} : R^{-1} \prec_w Q^{-1}\}$ is the set of weak partitional structures (see Corollary 2 in Kleiner et al. (2021)). We thus obtain the following result.

Lemma 1 (Sufficiency of Monotone Partitional Distributions)
There exists an optimal mechanism (G^*, R^*) such that G^* and R^* are monotone partitional distributions.

We now prove that an optimal mechanism must have a monotone pooling structure. For this, we first prove that, if both distributions have partitional structure, then they must be on a common support (in the quantile space).

Lemma 2 (G^* and R^* Have Common Quantile Support)
If (G^*, R^*) form an optimal mechanism with monotone partitional distributions, then G^* and R^* must have common support.

Proof. We assume that (G, R) form an optimal mechanism with monotone partitional distributions, we show that G^{-1} is increasing if and only if R^{-1}, and so they must have a common support.

We first note that, whenever G^{-1} is constant, R^{-1} must also be constant. This follows from the measurability condition $q(w) = R^{-1}(G(w))$. In other words, if there is an atom at some w (so that G^{-1} is constant), all quantiles in the atom must receive the same allocation as these quantiles correspond to the same expected valuation.

Suppose that R is a monotone partition distribution and consider interval $[x_k, x)$ with $k \notin J$ (if such interval does not exists, then $R = Q$). We thus have that $R^{-1}(t)$ is constant in $[x_k, x)$, and and $R^{-1}(x_k) < R^{-1}(x_{k+1})$. We show that $G^{-1}(t)$ must also be constant in $[x_k, x)$.

Suppose G^{-1} is not constant in $[x_k, x_{k+1})$, and consider

$$\hat{G}^{-1}(x) \triangleq \begin{cases} G^{-1}(x) & \text{for all } x \notin [x_k, x_{k+1}); \\ \frac{\int_{x_k}^{x_{k+1}} G^{-1}(t)dt}{x_{k+1}-x_k} & \text{for all } x \in [x_k, x_{k+1}). \end{cases}$$

We obviously have that $G \prec \hat{G}$, so $\hat{G}$ is feasible. Let the profits generated by mechanism (G, R) and $(\hat{G}, R)$, denoted by Π and $\hat{\Pi}$ respectively, and note that:

$$\Pi - \hat{\Pi} = \Big((R^{-1}(x_k) - R^{-1}(x_k^-))(1 - x_k)G^{-1}(x_k)\Big) - \Big((R^{-1}(x_k) - R^{-1}(x_k^-))(1 - x_k)\hat{G}^{-1}(x_k)\Big).$$

As before, $R^{-1}(x_k^-)$ denotes the left limit of R^{-1} at x_k. Since G is not constant in $[x_k, x_{k+1})$ we have that $G^{-1}(x_k) < \hat{G}^{-1}(x_k)$. We thus conclude that $\Pi - \hat{\Pi} \leq 0$, and the inequality is strict if $R^{-1}(x_k) > 0$. ■

Suppose that there exists an optimal mechanism (G^*, R^*) such that either G^* or R^* do not have partitional structure. Then there exists a collection of partitional information structures G and a measure λ over these partitional structure distributions such that:

$$G^{*-1} = \int_{\{G^{-1}:G^{-1}\prec F\}} G^{-1} d\lambda(G^{-1}).$$

The same characterization applies for the set of distributions $\{R^{-1} : R^{-1} \prec_w Q\}$.

Since the functional (10) is linear in G^{-1} and R^{-1}, each of these partitional distributions must be optimal. Hence, there must exist two optimal mechanisms (G_1^*, R^*) and (G_2^*, R^*), with R^* having partitional structure and G_1^*, G_2^* having partitional structure and $G_1^*(w) \neq G_2^*(w)$ for some w. However, they cannot both be increasing at a quantile t if and only if R^* is increasing.

Lemma 3 (Necessity of Monotone Partitional Distribution)
Every optimal mechanism (G^*, R^*) has monotone partitional distributions with common support.

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

4.2 Monotone Pooling Distribution

We next show that indeed every interval is a pooling interval. 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? We begin the argument with the virtual values given by:

$$\phi(v) \triangleq v - \frac{1 - F(v)}{f(v)}. \tag{13}$$

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

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

We denote the mean and variance of the quality and virtual values in the interval $[v_1, v_2]$ by:

$$\mu_\phi \triangleq \frac{\int_{v_1}^{v_2} \phi(v) f(v) dv}{\int_{v_1}^{v_2} f(v) dv}; \qquad \mu_q \triangleq \frac{\int_{v_1}^{v_2} q^*(v) f(v) dv}{\int_{v_1}^{v_2} f(v) dv}; \tag{14}$$

$$\sigma_\phi^2 \triangleq \frac{\int_{v_1}^{v_2} (\phi(v) - \mu_\phi)^2 f(v) dv}{\int_{v_1}^{v_2} f(v) dv}; \qquad \sigma_q^2 \triangleq \frac{\int_{v_1}^{v_2} (q^*(v) - \mu_q)^2 f(v) dv}{\int_{v_1}^{v_2} f(v) dv}. \tag{15}$$

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

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

Hence, pooling the qualities 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 qualities we pool the values in this interval, we can reduce the buyers' information rent. When only the qualities are pooled–but not the values–then the quality increase that the values which are assigned the pooled quality get relative to values just below the pool is the quality difference $\mu_q - q^*(v_1)$ priced at v_1. After pooling the values, 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 values increases the payments 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, payments 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 values increases the price of the quality improvement $(\mu_q - q^*(v_1))$ for all values higher than v_2. Hence, pooling values generates a second-order benefit which always dominates the third-order distortions.

Lemma 4 (Monotone Pooling Distributions)
The optimal mechanism (G^*, R^*) has monotone pooling distributions.

Proof. 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 establish 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 2, 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:

$$\Pi^* - \tilde{\Pi} = \int_{v_1}^{v_2} \phi(v) q^*(v) f(v) dv - (F(v_2) - F(v_1)) \mu_\phi \mu_q. \tag{16}$$

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:

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

Using the Cauchy-Schwarz inequality, we can bound the first integral (and thus the whole expression) as follows:

$$\Pi^* - \tilde{\Pi} \leq \sigma_q \sigma_\phi (F(v_2) - F(v_1)). \tag{17}$$

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. So we have that:

$$\Pi^* - \tilde{\Pi} \leq \sqrt{(\mu_q - q^*(v_1))(q^*(v_2) - \mu_q)} \sqrt{(\mu_\phi - \phi_1)(\phi_2 - \mu_\phi)} (F(v_2) - F(v_1)). \tag{18}$$

We can then conclude that:

$$\lim_{\Delta \to 0} \frac{\Pi^* - \tilde{\Pi}}{\Delta^3} \leq \frac{dq^*(\hat{v})}{dv} \frac{d\phi(\hat{v})}{dv} f(\hat{v}). \tag{19}$$

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

We now consider the following allocation policy:

$$\hat{q}(v) = \begin{cases} q_1^-, & \text{if } v \in (v_1, \mu_v); \\ \mu_q, & \text{if } v \in [\mu_v, v_2); \\ q^*(v), & \text{if } v \notin (v_1, v_2). \end{cases} \tag{20}$$

where

$$q_1^- \triangleq \lim_{v \uparrow v_1} q^*(v),$$

so the limit is taken from below. 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 values in (v_1, v_2) are pooled. That is, the information structure is:

$$\hat{G}(v) = \begin{cases} F(v_1), & \text{if } v \in (v_1,\mu_v); \\ F(v_2), & \text{if } v \in [\mu_v, v_2); \\ F(v), & \text{if } v \notin [v_1, v_2). \end{cases} \tag{21}$$

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 buyers' surplus:

$$\hat{\Pi} - \tilde{\Pi} = (\mu_q - q_1^-)(\mu_v - v_1)(1 - F(v_1)). \tag{22}$$

Since, $q_1^- \leq q^*(v_1)$, we have that:

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

We conclude that:

$$\lim_{\Delta \to 0} \frac{\hat{\Pi} - \tilde{\Pi}}{\Delta^2} \geq \frac{dq^*(\hat{v})}{dv}(1 - F(v_1)). \tag{23}$$

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. ■

With Lemma 4 we have therefore completed the proof of Theorem 1. The optimal distributions of values and qualities, G^* and R^* are monotone pooling distribution. Furthermore, the corresponding distributions G^{*-1} and R^{*-1} have common support. Thus, the distributions G^* and R^* have the same value in their range as illustrated in Figure 3 below. More directly, when we plot the quantile distributions G^{*-1} and R^{*-1} as in Figure 4 then we find that the distributions share the same quantile at which the quantile functions jump upwards. Of course, they can jump to different levels in terms of values and qualitities but the jumps occur at the same quantiles which reflects that values and qualities are assorted in monotone manner.

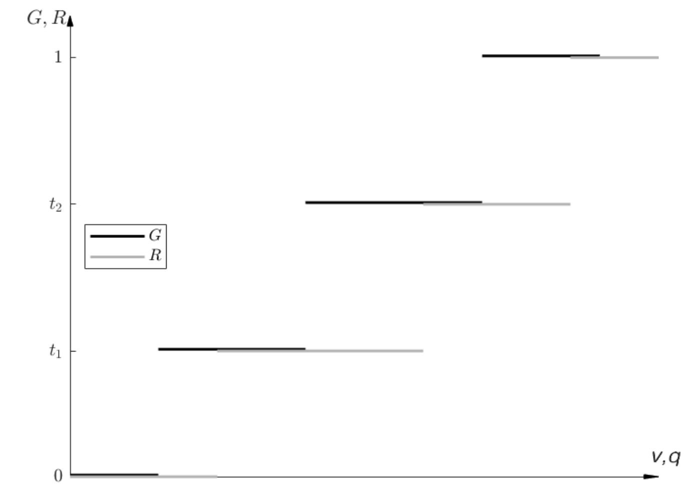

Figure 3: The optimal distributions G and R are monotone pooling distributions that have common support.

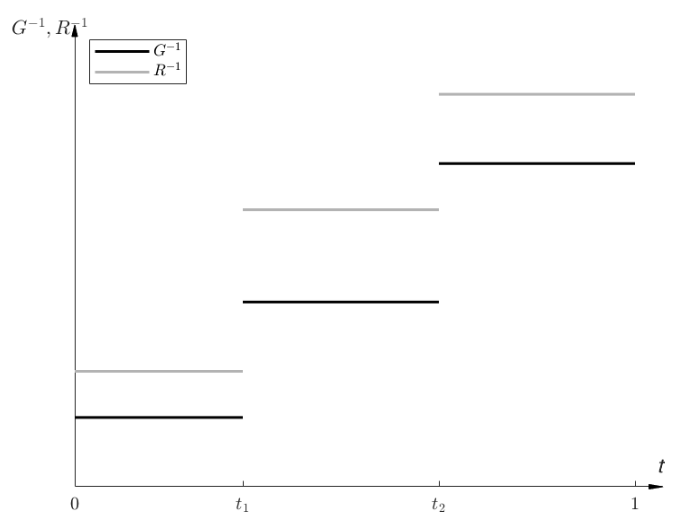

Figure 4: The optimal quantile distributions G^{-1} and R^{-1} have common support and therefore display jumps at the quantiles.

5 On the Number of Items in the Optimal Menu

5.1 The Convexity of Optimal Qualities

We next establish a property of the optimal mechanism that will be important for our results on the size of the menu. Namely, we show that the optimal menu will have increasing quality increments. By Theorem 1, the optimal distributions (G^*, R^*) are monotone pooling distributions. The distribution function G associated with any pooling structure is an increasing and piecewise constant step function G denoted by:

$$G_k \triangleq G(w_k), \quad g_k \triangleq G(w_k) - G(w_{k-1}). \tag{24}$$

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

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

Also, by Theorem 1, the optimal distributions (G^*, R^*) have a common support in the quantile space. Namely, the boundary points of the intervals in the quantile space are denoted by $\{x_k\}_{k\in I}$ and are related to the corresponding distribution G as follows:

$$G_k \triangleq x_k, \quad g_k \triangleq x_k - x_{k-1}. \tag{25}$$

The expected values and qualities are denoted by:

$$w_k \triangleq \frac{\int_{x_{k-1}}^{x_k} F^{-1}(t)dt}{x_k - x_{k-1}}; \quad r_k \triangleq \frac{\int_{x_{k-1}}^{x_k} Q^{-1}\left(t\right) \mathbb{I}_{t\geq\widehat{x}}(t)dt}{x_k - x_{k-1}},$$

where we recall that $\mathbb{I}_{t\geq\widehat{x}}(t)$ is the indicator function. In a finite-item menu, the profits are given by:

$$\Pi = \sum_{k=1}^{K} g_k w_k r_k - \sum_{k=1}^{K}(w_{k+1} - w_k)r_k(1 - G_k), \tag{26}$$

where by convention $w_{K+1} \triangleq w_K$. This is the discrete counterpart of (9). Finally, the quality increments are denoted by

$$\Delta r_k \triangleq r_k - r_{k-1}. \tag{27}$$

We then express the profits as follows:

$$\Pi = \sum_{k=I} w_k \Delta r_k (1 - G_{k-1}), \tag{28}$$

where $G_0 = 0$. This is the discrete counterpart of (10).

Proposition 1 (Increasing Differences in Qualities)
In any optimal mechanism the quality increments Δr_k must be (weakly) increasing in k.

Proof. We fix a mechanism (G, R) and assume that there exists l such that $\Delta r_l > \Delta r_{l+1}$. We prove that the mechanism cannot be optimal.

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

$$\tilde{g}_l = g_l + \varepsilon; \qquad \tilde{g}_{l+1} = g_{l+1} - \varepsilon;$$
$$\tilde{w}_l = \frac{g_l w_l + \varepsilon w_{l+1}}{g_l + \varepsilon}; \qquad \tilde{w}_{l+1} = w_{l+1};$$
$$r(\tilde{w}_l) = \frac{g_l r(w_l) + \varepsilon r(w_{l+1})}{g_l + \varepsilon}; \qquad \widetilde{r}(\tilde{w}_{l+1}) = r(w_{l+1}).$$

Note that:

$$\tilde{g}_l \tilde{w}_l + \tilde{g}_{l+1} \tilde{w}_{l+1} = g_l w_l + g_{l+1} w_{l+1};$$
$$\tilde{g}_l \widetilde{r}(\tilde{w}_l) + \tilde{g}_{l+1} \widetilde{r}(\tilde{w}_{l+1}) = g_l r(w_l) + g_{l+1} r(w_{l+1})$$

so this is clearly a feasible allocation. 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. The difference in the profits generated by the original mechanism and the new mechanism are given by:

$$\Pi - \widetilde{\Pi} = w_l \Delta r_l (1 - G_{l-1}) + w_{l+1} \Delta r_{l+1} (1 - G_l) - \left(\tilde{w}_l \Delta \widetilde{r}_l (1 - \tilde{G}_{l-1}) + \tilde{w}_{l+1} \Delta \widetilde{r}_{l+1} (1 - \tilde{G}_l) \right)$$

Since $g_k = \tilde{g}_k$ for all $g \geq l+2$, we can write the difference as follows:

$$\begin{aligned} \Pi - \widetilde{\Pi} = & w_l (r(w_l) - r(w_{l-1}))(1 - G_{l+1} + g_l + g_{l+1}) + w_{l+1}(r(w_{l+1}) - r(w_l))(1 - G_{l+1} + g_{l+1}) \\ & - \left(\tilde{w}_l (\widetilde{r}(\tilde{w}_l) - r(w_{l-1}))(1 - G_{l+1} + \tilde{g}_l + \tilde{g}_{l+1}) + \tilde{w}_{l+1}(\widetilde{r}(\tilde{w}_{l+1}) - \widetilde{r}(\tilde{w}_l))(1 - G_{l+1} + \tilde{g}_{l+1}) \right) \end{aligned}$$

Taking the derivative and evaluating at 0, we get:

$$\left. \frac{d(\Pi - \widetilde{\Pi})}{d\varepsilon} \right|_{\varepsilon = 0} = - \frac{(1 - G_{l-1})(\Delta r_{l+1} - \Delta r_l)}{g_l}.$$

The optimality condition requires this derivative is less than 0, so we require that $(\Delta r_{k+1} - \Delta r_k) \geq 0$, which concludes the proof. ■

5.2 Optimality of Finite Item Menu

We can now show that the optimal information structure consists of finitely many signals when $\bar{q} < \infty$.

Theorem 2 (Finite Item Menu)
The optimal mechanism (G^*, R^*) is given by finite and monotone pooling distributions.

Proof. 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 $w_{k-1} < w_k < w_{k+1}$.

Proposition 1 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 values converges to $\hat{v}$ from the right). We denote by $\underline{f}$ and $\bar{f}$ the minimum and maximum 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:

$$\phi_{k+1} - \phi_k < \frac{\sqrt{\underline{f}}(1 - F(v_{k-1}))}{\sqrt{\bar{f}}(\sqrt{\bar{f}} + \sqrt{\underline{f}})}. \tag{29}$$

Note that the virtual values ϕ_k are monotonic, and so we must have that $(\phi_{k+1} - \phi_k)$ converge to 0 as we take intervals close enough to $\hat{v}$. So, we can take intervals close enough to $\hat{v}$ such that (29) 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 (13), we define:

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

and extend (14)-(15) 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 observe 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 finally recall that $g_k < g_{k+1}$ so we get that:

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

where the second inequality corresponds to (29). Thus, we reach a contradiction with this being an optimal mechanism. ■

We assume throughout the analysis that the values and qualities have compact support, in particular that there are finite upper bounds $\bar{q}, \bar{v} < \infty$. However, we could relax these conditions. In fact, Theorem 1 remains valid with unbounded supports for values and qualities. For Theorem 2 to remain valid, we can likewise allow for unbounded support of values but do need a bounded support of qualities, thus $\bar{q} < \infty$. With unbounded support for values *and* qualities countably many items still remain optimal.So far, we have shown that the optimal mechanism generates a coarse menu with finitely many items. By contrast, with complete disclosure of information to the buyers, the optimal menu would typically offer a continuum of items.

5.3 Optimality of Single-Item Mechanism

We now provide results on the number of items in the optimal menu. We identify a sufficient condition under which it is optimal to offer only a single item. We then provide a general upper bound on the number of items independent of any distributional assumptions.

The first result offers a sufficient condition for the optimal menu to contain a single-item menu.

Theorem 3 (Optimality of Single-Item Mechanism)
If the distribution Q of qualities has increasing density, then the optimal mechanism is a single-item menu.

Proof. Recall that in a finite-item menu, the profits can be written as in (28). 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:

$$\Pi_{K-1,K} \triangleq \big((g_{K-1}+g_K)\Delta r_{K-1}w_{K-1}+g_K\Delta r_K w_K\big),$$

which are the last two terms of the summations in (28). If the last two intervals are pooled, then the profits generated would be:

$$\hat{\Pi}_{K-1,K} \triangleq \frac{1}{(g_{K-1}+g_K)}(g_{K-1}w_{K-1}+g_K w_K)(g_{K-1}r_{K-1}+g_K r_K-(g_{K-1}+g_K)r_{K-2}).$$

We now note that:

$$\Pi_{K-1,K}-\hat{\Pi}_{K-1,K}=\frac{(w_K-w_{K-1})g_K(g_{K-1}(r_K-r_{K-1})-(g_K+g_{K-1})(r_{K-1}-r_{K-2}))}{g_K+g_{K-1}}.$$

If the distribution Q has increasing density, then we have that:

$$r_{K-2}\le q_{K-2};\quad r_{K-1}\ge \frac{q_{K-2}+q_{K-1}}{2};\quad r_K\le \frac{\frac{\bar{q}+q_{K-1}}{2}\big(\frac{(\bar{q}-q_{K-1})g_{K-1}}{q_{K-1}-q_{K-2}}\big)+\bar{q}\big(g_K-\frac{(\bar{q}-q_{K-1})g_{K-1}}{q_{K-1}-q_{K-2}}\big)}{g_K}.$$

The first inequality follows from the fact that r_{K-2} must be smaller than the upper bound of the interval $[q_{k-3},q_{K-2}]$; the second equality follows from the fact that r_{K-1} must be greater than the midpoint of the interval $[q_{K-2},q_{K-1}]$ (since the density is increasing). The third inequality follows from the fact that the density in the last interval $[q_{K-1},\bar{q}]$ is at least $g_{K-1}/(q_{K-1}-q_{K-2})$ so the expected value in the last interval is bounded by the expected value generated by having an atom of size $(g_K-\frac{(\bar{q}-q_{K-1})g_{K-1}}{q_{K-1}-q_{K-2}})$ at the top of the support. We thus get that:

$$\Pi_{K-1,K}-\hat{\Pi}_{K-1,K}\le -\frac{(w_K-w_{K-1})(g_{K-1}(\bar{q}-q_{K-1})-g_K(q_{K-1}-q_{K-2}))^2}{2(g_K+g_{K-1})(q_{K-1}-q_{K-2})}<0.$$

Thus, there is an optimal solution in which the last two intervals are pooled. Inductively, we conclude there is always an optimal single-item mechanism. If $\underline{v}>0$ or $q_{K-2}>0$, then the inequality is strict. Note that, if the distribution is uniform, we have that $g_K(q_{K-1}-q_{K-2})+g_{K-1}(\bar{q}-q_{K-1})=0$. ■

This theorem states that for a large class of distributions Q of qualities the seller will optimally offer a single quality (and a single item) to the buyers. Importantly, the sufficient condition for the distribution of qualities holds for all possible distribution of values. The sufficient condition is tight to the extent that for *every* distribution Q of qualities that has linear decreasing, but nearly constant density, there exist distribution of values for which a multi-item mechanism is optimal. So even "slightly" decreasing densities can lead to optimal multi-item mechanisms.

A first intuition for the result can be gained by considering a two-item mechanism as follows. Suppose we fix a two-item mechanism, offering low and high qualities to low and high expected value buyers, thus (r_L, r_H) to (w_L, w_H) respectively. We then consider the possibility of pooling both items and consequently both values. The benefit of pooling is that the information rents will be eliminated and the cost is that the social surplus will be reduced. However, both of these terms are both proportional to the difference in expected values generated by the two-item mechanism. More specifically, the surplus loss will be:

$$\Delta S = g_H(1 - g_H)(w_H - w_L)(r_H - r_L). \tag{30}$$

If the the two items r_L and r_H are pooled, then the buyers will loose all of their surplus. The reduction in buyers' surplus will be:

$$\Delta U = -g_H(w_H - w_L)r_L. \tag{31}$$

In above two term, the change in surplus is proportional to the difference in expected values $(w_H - w_L)$. Furthermore, the losses in surplus will be smaller than the gains from eliminating buyers surplus when the low quality r_L is not too small large relative to the high quality r_H.

When analyzing how to segment a continuous distribution of qualities bound of the value and quality distribution. If the density of qualities is increasing, then no matter how small the segment is which we try to separate from the lower part of support, the difference in qualities, $r_H - r_L$, will be small relative to the quality of the lower end of the distribution. In contrast, if the density is decreasing, then by separating a small interval around the top of the value distribution, one can generate a large difference between the low and high quality items.

We can bound the number of items offered in any mechanism. The bound relies in the upper and lower bound of the support of values. In this Subsection, we assume that $\underline{q} > 0$.

Proposition 2 (Finite Upper Bound on the Number of Items)
The number of items K offered by an optimal mechanism is bounded above by

$$K \leq \frac{\bar{q}}{\underline{q}}.$$

Proof. The lowest quality item will be at least $r_1 \geq \underline{q}$. Using Proposition 1, the k-th must have quality of at least $k\underline{q}$. However, the last item K-th cannot have a quality higher than $\bar{q}$, so we have that $K\underline{q} \leq \bar{q}$, which proves the result. ■

6 Endogenous Qualities

In this section we analyze a model in which the seller can produce vertically differentiated goods with an increasing and convex cost for quality. Thus, we consider the seminal model of Mussa & Rosen (1978) where the supply of qualities is endogenous and determined by the seller to maximizes the revenue. We first formally introduce the model. We then explain how our results from the previous section with an exogenous supply of qualities extend to this model with an endogenous supply of qualities. We then provide some novel results that describe how the optimal mechanism changes with the cost function of quality. In this section, we restrict attention to cost functions with a constant elasticity. We show that if the distribution of values has narrow support relative to the cost elasticity then the optimal mechanism is a single-item menu. Finally, we consider the comparative statics of how the cost elasticity impacts the nature of the optimal mechanism and information policy.[2]

6.1 Cost of Quality

We maintain the same environment as described in Section 2 except for the fact that the seller can choose to produce any quality level q at an increasing and convex cost cost $c(q)$. The seller chooses a menu (or direct mechanism) with qualities $q\left(w\right)$ at prices $p\left(w\right)$:

$$M \triangleq \left\{(q(w), p(w))\right\}_{w\in[\underline{v},\overline{v}]},$$

subject to the earlier incentive compatibility and participation constraints, (5) and (6) respectively. The seller's problem is to maximize expected profits, revenues minus cost:

$$\Pi \triangleq \max_{G,M} \mathbb{E}_G[p(w) - c(q(w))].$$

Thus, the only change is that we have added the production cost $c\left(q\right)$. The qualities are now chosen endogenously by the seller rather than given exogenously by the distribution of qualities Q.

Following the same steps as in Section 3, we can write the seller's problem analogous to (10) as

[2] In the Appendix we provide a complete characterization of the optimal mechanism when the distribution of values has binary support. The binary value setting is both a leading example for our general analysis as well as source for some auxiliary yet foundational results for the results in this Section.

follows:

$$\max_{G,R} \int_0^1 G^{-1}(t)(1-t)dR^{-1}(t) - \int_0^1 c(R^{-1}(t))dt.,$$
$$\text{subject to } F \prec G \text{ and } R^{-1} \text{ is non-decreasing,}$$

Relative to earlier formulation in (10), we have added the production cost $c(q)$ and removed the constraint requiring that the quality distribution R must be majorized by Q. Instead we simply require that R is non-decreasing in the quantile to satisfy incentive compatibility.

We can verify that Theorem 1 remains valid in this new environment as well. The objective function continues to be linear in G. And, in fact, the distribution of expected values G only changes the revenue, not the expected cost (once we have fixed the allocation rule to be R). In the proof of Theorem 1 we showed that the pooling of values always improves the seller's revenue. Of course, now the pooling of values will also change the cost of the provided qualities. However, if the cost of supplying a given quality is convex, then then pooling will reduce the total cost, and so a fortiori the pooling of values remains optimal.

Theorem 4 (Structure of the Optimal Mechanism)
In every optimal mechanism the optimal distribution G^* is a finite monotone pooling distribution.

6.2 Constant Elasticity Cost Function

To understand the role of the cost function $c(q)$ in determination of the optimal menu, we now focus on a one parameter family of cost functions with

$$c(q) = \frac{1}{\eta} q^{\eta}$$

for $\eta > 1$. The parameter η represents the (constant) cost elasticity. Note that if costs were linear in quality, i.e., $\eta = 1$, then it would be possible for the seller to make infinite profits by offering infinitely large quantities to the buyers.[3]

We can now state that, if the support of the distribution is narrow enough, the optimal mechanism will necessarily be pooling.

[3]In the working paper version (Bergemann, Heumann & Morris (2022)), we conduct the analysis for general increasing and convex cost function rather than the constant elasticity cost functions. In particular, we show that a single-item menu is optimal if the marginal cost is convex and the distribution of values has either increasing or concave density (similar to Theorem 5 here). We also provide a complete solution of the binary value environment.

Theorem 5 (Optimality of Single-Item Menu for Distributions with Narrow Support) *Pooling is optimal for every distribution F with support in $[\underline{v}, \overline{v}]$ if and only if*

$$\frac{\overline{v}}{\underline{v}} < \eta. \tag{32}$$

Proof. The proof of this result partially relies on the analysis of the binary value model we provide in Section 9. As a direct corollary of Proposition 5 we conclude that, if the distribution of values has binary support $v \in \{\underline{v}, \overline{v}\}$, and the ratio between the high and low value is lower than η :

$$\frac{\overline{v}}{\underline{v}} < \eta,$$

then the optimal mechanism pools both values. Furthermore, if (32) is not satisfied, then there is a binary value distribution with support in $\{\underline{v}, \overline{v}\}$ for which pooling is not the optimal mechanism.

We now consider a distribution F (possibly with continuous support) that satisfies (32) and prove the optimal mechanism is pooling. 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, as previously explained, 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. ■

To gain some intuition, we consider the case in which $\overline{v}/\underline{v} \approx 1$ and we denote by $\overline{f} = \Pr(v = \overline{v})$. We ask whether pooling or offering a two-item menu generates higher profits. Note that when $\overline{v}/\underline{v} \approx 1$, excluding the low value is not optimal. Of course, when $\overline{v}/\underline{v} = 1$, 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:

$$\left.\frac{\partial \Delta S}{\overline{v}}\right|_{\overline{v}=\underline{v}} = 0.$$

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 buyers is $U_P = 0$. By contrast, the information rents when offering a menu are:

$$U_M = \overline{f}\underline{q}(\overline{v} - \underline{v}).$$

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 $\overline{v}$. 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.

Anderson & Dana (2009) consider a model of second-degree price discrimination as in Mussa & Rosen (1978) but where there is a priori a finite upper bound on the quality provided. They provide conditions under which all values receive the same quality, namely the quality at the upper bound. Sandmann (2022) shows that their sufficient condition requires that high valuation buyer's surplus is more concave than that of the low valuation buyers. Sandmann (2022) shows that a necessary condition for a single-item menu to be revenue optimal is that the single-item menu constitutes the socially optimal allocation. By contrast, in the current environment a continuum of qualities is socially optimal. Hence there would be no reason to restrict the menu and offer a bunching solution in the absence of persuasion.

Consider some fixed information structure S that generates finitely many values $w_1, ..., w_K$ each with probability $g_1, ..., g_K$ as in (24). We define the corresponding virtual values as:

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

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).

As the cost function is a power function $c(q) = q^\eta/\eta$, we can express the profit as a function of the virtual value ϕ as follows:

$$\pi(\phi) \triangleq \begin{cases} \frac{\eta-1}{\eta}\phi^{\frac{\eta}{\eta-1}} & \text{if } \phi \geq 0; \\ 0 & \text{otherwise.} \end{cases} \tag{33}$$

The profits generated by an information structure S are then given by:

$$\Pi_S = \sum_{k=1}^{K} g_k \pi(\phi_k).$$

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 $\phi \geq 0$ is given by:

$$\frac{\pi''(\phi)\phi}{\pi'(\phi)} = \frac{1}{\eta - 1},$$

so that the hypothetical risk-loving agent is more risk loving as η is closer to 1. For comparison, if we were to pool all values v into a single interval with expectation $\mu_v = \mathbb{E}[v]$, then the resulting information structure would generates profits equal to:

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

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

Proposition 3 (Profit Comparisons as a Function of Cost Elasticity)
Consider some finite information structure S with $K > 1$ values:

1. *There exists η_S such that information structure S generates less profit than complete 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 $\eta \leq \eta_S$.*

Proof. To prove this result, we first compare the profits generated by some finite information structure S and the complete 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:

$$\Pi_S \geq \Pi_P \iff \eta \geq \eta_S.$$

This proves the first statement.

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

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

where ϕ is defined in (13). We bound the ratio between the profits generated by S and complete disclosure as follows:

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

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

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

We thus have that:

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

The limit is obtained from observing that when $\eta \to 1$, the exponent in (33) converges to infinity, so the integrand diverges to infinity. ■

7 Discussion and Interpretation

Recommender System as an Indirect Mechanism Our leading interpretation is that the seller can influence the information that the buyers have about their value but does not observe the realization of the information structure or the value. Although we did not pursue it formally here, an alternative interpretation of our model is that the seller does in fact observe the buyers' value but is unable, for regulatory or business model reasons, to offer prices for item that depend on the buyers' 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 buyers that recommends one item on the menu.

Formally, 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 $S : \mathbb{R}_+ \to \Delta Q$. 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 buyers is steered to a specific alternative at a price that is common to all consumers. This implementation arises in our model if we impose the interim obedience constraint that recommendations are optimal for the buyers conditional on the recommendation received.

Consistent with this interpretation, eBay personalizes the search results for each buyers 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 & 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).

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

Beyond the Multiplicative Environment We stated our main results in the environment of nonlinear pricing first proposed by Mussa & Rosen (1978). There, the buyers' value is given by a multiplicatively separable function of willingness-to-pay and quality, θq. The results of Theorem 1 and 2 will continue to hold for and multiplicative separable value function

$$u\left((v, q, p\right) = h_v\left(v\right) h_q\left(q\right) - p$$

where the component functions are monotone increasing in v and q respectively. Moreover, if the value would be generated by a general supermodular function $h\left(v, q\right)$:

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

where h is a strictly increasing function and supermodular in both arguments v and q, then a version of Theorem 1 would still remain to hold. Namely, complete disclosure would never be optimal. However, the pooling of values may no longer be monotone. In the working paper version, Bergemann, Heumann & Morris (2022) we provide an instance where the optimal information structure is a non-monotone partition. Finally, in the absence of supermodularity, the finite menu result of Theorem 2 may disappear and there we provide an instance where complete disclosure is optimal.

8 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 buyers for the products offered.

We showed that in the presence of information and mechanism design, the seller offers a menu with only a small 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 buyers 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 we obtained our results in the model of nonlinear pricing pioneered by Mussa & 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 buyers. 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 buyers, the product offered is therefore *opaque*, as its exact properties are not known to the buyers who is only guaranteed certain distributional properties of the product. This practice is sometimes referred to as *opaque pricing,* see Jiang (2007) and Shapiro & Shi (2008) for applications to services and transportation and Bergemann, Heumann, Morris, Sorokin & 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 buyers opaque to find the optimal trade-off between efficient matching of quality and taste against the revenues from surplus extraction.

9 Appendix

Throughout this section, we study the model introduced in Section 6.1 and additionally assume that the values v of the buyers 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 buyers and subsequently screens the different values through different qualities the *screening solution*; or, (ii) the seller does not disclose any information and subsequently pools all values 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 4 (Complete Screening or Complete Pooling)
The optimal mechanism either exhibits complete screening *or* complete pooling.

Proof. We can write the expected profits 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 & 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 (36)-(38), but replacing f_H, v_H, v_L with g_H, w_H, w_L. To make the notation more compact, we define:

$$\alpha \triangleq \frac{\eta}{\eta - 1}. \tag{34}$$

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 values 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 values 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

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

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

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

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 derivative, 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.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 buyers' 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 buyers. 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 buyers 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:

$$q_L = v_L^{\frac{1}{\eta-1}}\left(1 - \frac{(\frac{v_H}{v_L} - 1)f_H}{f_L}\right)^{\frac{1}{\eta-1}}, \text{ and } \ q_H = v_H^{\frac{1}{\eta-1}}. \tag{35}$$

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:

$$\Pi_P = v_L^{\frac{\eta}{\eta-1}} \frac{\eta-1}{\eta} (f_L + f_H \frac{v_H}{v_L})^{\frac{\eta}{\eta-1}}; \tag{36}$$

$$\Pi_H = v_L^{\frac{\eta}{\eta-1}} \frac{\eta-1}{\eta} f_H (\frac{v_H}{v_L})^{\frac{\eta}{\eta-1}} \ ; \tag{37}$$

$$\Pi_M = v_L^{\frac{\eta}{\eta-1}} \frac{\eta-1}{\eta} \left(\left(\frac{(1-f_H \frac{v_H}{v_L})^{\frac{\eta}{\eta-1}}}{(1-f_H)^{\frac{1}{\eta-1}}} \right) + f_H (\frac{v_H}{v_L})^{\frac{\eta}{\eta-1}} \right). \tag{38}$$

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 5 (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

$$v_L < v_H f_H. \tag{39}$$

Lemma 5 provides a well-known trade-off in screening problems. Serving the low value increases efficiency but it also increases the buyers' information rents. Hence, the quality offered to the low value is distorted downwards. The distortion is increasing in the probability that the buyers has a high value, and if this probability is too high, then the low value buyers 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 (39) 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 buyers' 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:

$$r(f_H, \eta) \triangleq \{r \in \mathbb{R}_+ \mid \Pi_P \geq \max\{\Pi_H, \Pi_M\} \iff v_H/v_L \leq r\}. \tag{40}$$

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.

Lemma 6

For all (f_H, v_H) such that $g_H v_H \geq 1$, it is optimal to pool values if and only if

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

Proof. Recall that the profits of the three strategies are given by (36)-(38). 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 $f_H v_H \geq 1$,

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

which completes the proof. ■

Proposition 5 (Optimality of Pooling)

The correspondence $r(f_H, \eta)$ as defined in (40) is not empty, it is single valued, and is increasing in η and decreasing in f_H, with $r(1, \eta) = \eta$.

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

$$\frac{1 - \tilde{f}_H}{\tilde{f}_H^{\frac{\eta-1}{\eta}} - \tilde{f}_H} \triangleq \frac{1}{\tilde{f}_H}. \tag{41}$$

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:

$$\tilde{\Pi} \mid_{v_H=1} = 0; \tag{42}$$

$$\frac{\partial \tilde{\Pi}}{\partial v_H} \mid_{v_H=1} = -f_H \alpha < 0; \tag{43}$$

$$\frac{\partial^2 \tilde{\Pi}}{\partial v_H^2} = (\alpha - 1)\alpha f_H \left(f_H (1 - f_H)^{1-\alpha} (1 - f_H v_H)^{\alpha-2} - f_H (f_H (v_H - 1) + 1)^{\alpha-2} + v_H^{\alpha-2} \right) > 0, \tag{44}$$

where α is defined in (34). The last inequality can be verified as follows. If $\alpha < 2$, then:

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

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 (42)-(44) 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:

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

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

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

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 (35). 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

$$q_L v_L - c(q_L) - (q_P v_L - c(q_P)) < 0. \tag{45}$$

The left-hand-side of the inequality is the difference between the intercept of Π_M and Π_P. We now prove that (45) 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. ■

Propositions 4 and 5 jointly give us a complete description of the optimal mechanism. Figure 5 illustrates the qualitative properties of the optimal mechanism for different values of (f_H, v_H).

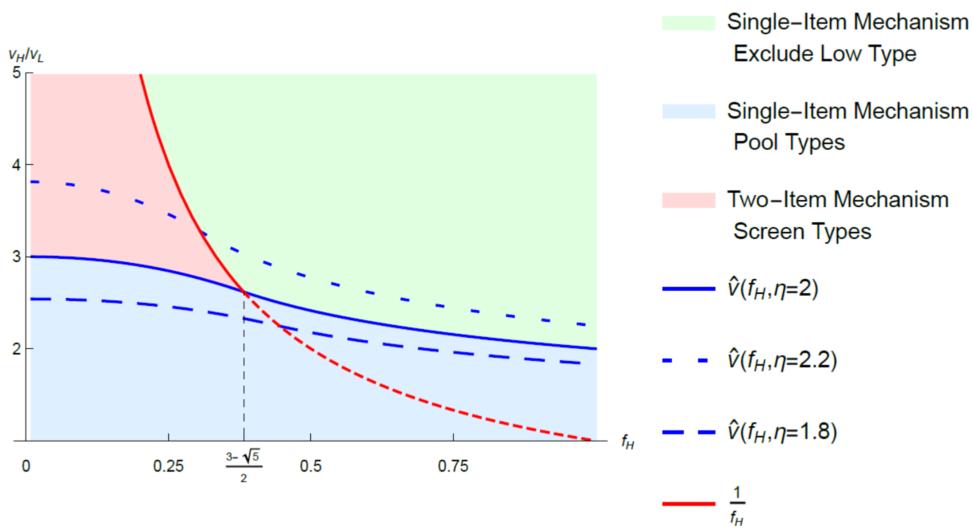

Figure 5: Optimal Mechanism

Figure 5 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:

$$r(f_H, 2) = \begin{cases} 1 + \frac{1}{\sqrt{f_H}}, & \text{if } f_H > \frac{3-\sqrt{5}}{2}; \\ \frac{f_H^2 - 3f_H + 3}{f_H^2 - f_H + 1}, & \text{if } f_H \leq \frac{3-\sqrt{5}}{2}. \end{cases} \tag{46}$$

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 1 (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.

References

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

Bergemann, D., Brooks, B. & Morris, S. (2015). The limits of price discrimination, *American Economic Review* **105**: 921–957.

Bergemann, D., Heumann, T. & Morris, S. (2022). Screening with persuasion, *Technical Report 2338*, Cowles Foundation for Research in Economics, Yale University.

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

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

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

Cavallo, A. (2017). Are online and offline prices similar? evidence from large multi-channel retailers, *American Economic Review* **107**(1): 283–303.

Cavallo, A. (2019). More amazon effects: Online competition and pricing behaviors, *in* F. R. B. of Kansas City (ed.), *Jackson Hole Economic Symposium Proceedings*, pp. 291–355.

DellaVigna, S. & Gentzkow, M. (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. & Siegel, R. (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. & Gentzkow, M. (2011). Bayesian persuasion, *American Economic Review* **101**: 2590–2615.

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

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

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

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

Mussa, M. & Rosen, S. (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. & Szentes, B. (2017). Buyer-optimal learning and monopoly pricing, *American Economic Review* **107**: 2072–2080.

Sandmann, C. (2022). When are single-contract menus profit maximizing?, *Technical report*.

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

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

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**.

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