SCREENING WITH PERSUASION

By

Dirk Bergemann, Tibor Heumann and Stephen Morris

September 2023

COWLES FOUNDATION DISCUSSION PAPER NO. 2338R2

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

Screening with Persuasion*

Dirk Bergemann† Tibor Heumann‡ Stephen Morris§

September 23, 2023

Abstract

We consider a general nonlinear pricing environment with private information. The seller can control both the signal that the buyers receive about their value and the selling mechanism. We characterize the optimal menu and information structure that jointly maximize the seller's profit. The optimal screening mechanism has finitely many items even with a continuum of values. We identify sufficient conditions under which the optimal mechanism has a single item. Thus the seller decreases the variety of items below the efficient level in order to reduce the information rents of the buyers.

JEL CLASSIFICATION: D44, D47, D83, D84.

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

*We acknowledge financial support from NSF grants SES-2001208 and SES-2049744, and ANID Fondecyt Iniciacion 11200365. We thank seminar audiences at Boston College, Brown University, Emory University, MIT, NBER Market Design, Pontificia Universidad Católica de Chile, University of Chile, the Warwick Economic Theory Workshop and the UCL memorial conference for Konrad Mierendorff for valuable suggestions, and, in particular, Mark Armstrong for his discussion at the latter conference. We benefited from discussions with Ian Ball, Roberto Corrao, Stefano DellaVigna, Piotr Dworczak, Fedor Sandomirskiy, Matt Gentzkow, Anton Kolotolin, Ellen Muir, Leon Musolff, Christopher Sandmann, Philipp Strack, and Alex Wolitzky. We thank Pedro Feijo de Moraes, Jack Hirsch, Hongcheng Li, Shawn Thacker, Michael Wang and Nick Wu for excellent research assistance.

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

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

§Massachusetts Institute of Technology, semorris@mit.edu

1 Introduction

1.1 Motivation

In a world with a large variety of products and hence feasible matches between buyers and products, information can fundamentally affect the match. A notable feature of the digital economy is that sellers, or platforms and intermediaries that sellers use to place their products, commonly have information about the value of the match between any specific product and any specific buyer. In particular, by choosing how much information is disclosed to the buyer about the value of the match between product and buyer, 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 thereby 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, and thus the seller screens while engaging in Bayesian persuasion.

We characterize the information structure and menu of choices that maximize the expected profit of the seller. The buyers have a continuum of possible values–their willingness-to-pay for quality. In the absence of any information design, the optimal menu typically offers a continuum of qualities to the buyers who then select given their private information (e.g., as in Mussa & Rosen (1978) and Maskin & Riley (1984)). By contrast, in the current setting, the seller controls the selling mechanism *and* the information structure. Importantly, the seller remains constrained in that he cannot observe either values or signal realizations of the buyers. The selling mechanism could be any (possibly stochastic) menu.

Our main result is a characterization of the optimal information and mechanism. 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. Each pool of buyers is allocated to a good of common quality (or perhaps a lottery over goods giving rise to that expected quality). Buyers are pooled into finitely many intervals and are offered a menu with finitely many items. This contrasts with the optimal menu in the absence of information design that contains a continuum of items.

We establish this result in three parts. The first, and most original, part establishes that the optimal mechanism is discrete (Theorem 1), pooling buyers into a discrete number of pools. This is shown by a variational argument. Suppose that there was an interval of expected values in the support of the information structure. We ask what happens to profit if we pool the allocation of a small interval of expected values in that larger interval. 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 value,* then this causes a first-order decrease in the information rents. Hence, screening an open set of expected 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 profit.

Second, we show that the mechanism is not only discrete but also finite (Theorem 2). Here, we argue that quality increments in an optimal discrete menu must be increasing, so that item qualities offered in the menu increase in a convex manner. This rules out accumulation points except at the bottom, which in turn establishes finiteness.

Third, we establish that information is optimally given by a monotone partition (Corollary 1). This follows from fixing the pooled distribution of qualities and observing that the problem with one majorization constraint is linear in the distribution of expected values. Thus, a maximum is attained at an extreme point of the set of feasible distributions (Kleiner et al. (2021)).

Having established that the optimal menu with persuasion is finite, we next ask how many elements the optimal menu will have. A second set of results establishes sufficient conditions for there to be a single item menu or to bound the number of items. Our earlier results allowed for general cost functions. In particular, the argument covers the case where the seller produces goods with different qualities with a convex cost function, as in the classic analysis of Mussa & Rosen (1978) and Maskin & Riley (1984). But it also covers the case where the distribution of qualities to be sold by the seller is exogenously given as in the recent work of Loertscher & Muir (2022). Sufficient conditions for sharper predictions about the size of the menu vary with the cost environment.

For a fixed supply of qualities, 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, the number of items in the menu is no more than the ratio of the upper bound to the lower bound of the support of the exogenous quality distribution (Corollary 3). For the case of variable supply, Theorem 4 shows that under modest tail property of the value distribution and the convexity of marginal cost, the optimal menu will consist of a single item. More generally, the number of items in the optimal menu is sensitive to the convexity of the cost function and we report results for the case where the cost function has a constant elasticity. We show that a single-item menu will always be optimal for a sufficiently high level of cost elasticity, that is, any given information structure generates less profit than pooling all values (i.e., providing no information) when the cost elasticity is high enough. Conversely, any given information structure will generate less profit than complete disclosure if the cost elasticity is low enough, i.e., approaching unit elasticity which corresponds to linear costs (Proposition 2).

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 sellers. Our analysis considers the case where the buyers only know the prior distribution and the seller has access to all feasible signals. A second feature is that buyers have the ability to find which items are available at what prices 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 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 8, we show that the optimal mechanism can be implemented as an indirect mechanism in the form of a recommender system. We show that it is sufficient for the optimal mechanism to convey all the information through recommended product qualities (Proposition 3). We then offer an alternative interpretation of the informational environment. We suppose that the seller does in fact observe the buyers' values. But for regulatory or business reasons, the seller is unable to offer prices for items that depend on the buyers' value. Thus, the seller cannot engage in perfect price discrimination (or third-degree price discrimination). Proposition 4 establishes that the resulting optimal outcome is in fact equivalent to the one where the seller is information rather than price constrained. We establish the equivalence by showing that the solution of the optimal direct and an optimal indirect mechanism, the recommender system, are identical.

Finally, we use the structure of the recommender system to establish that our main result, Theorem 1, remains valid in environments with general nonlinear preferences. Theorem 5 shows that the optimal mechanism remains discrete. In order to maintain some weak monotonicity conditions in the nonlinear setting we restrict attention to a large class of stochastic information structures that include among others the monotone partitional distributions. The appeal to a recommender system may be of independent interest to solve Bayesian persuasion problems in nonlinear environments.

1.2 Related Literature

We analyze a model of nonlinear pricing with endogenous or exogenous distributions of qualities, as in Mussa & Rosen (1978) and Loertscher & Muir (2022), respectively. They show that bundling different qualities, or randomizing the quality assignment via lotteries, can increase the revenue in the presence of irregular type distributions. Pooling is optimal in our setting for all regular and irregular distributions.

In our analysis, the seller can control the information of the buyers and the selling mechanism in which the pooling of qualities is feasible. 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 other work on optimal pricing combining 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 a single aggregate market. In doing so, the seller makes any item intended for one segment less attractive to the other segments in the same market. By contrast, Bergemann et al. (2015) and Haghpanah & Siegel (2022) allow many markets 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.

There are a number of related papers that investigate how information may influence the menu offered by multi-product sellers. Mensch & Ravid (2022) and Thereze (2022) consider a Mussa & Rosen (1978) style model and allow the buyer to acquire information about their willingness-to-pay. Each paper considers a different cost model of information and then derives the optimal menu that the seller offers in anticipation of the endogenous response of the buyers. The resulting menu offers a continuum of choices in which the buyer is implicitly compensated for the cost of information acquisition. More distant, Johnson & Myatt (2003) and Ellison (2005) also ask what determines the variety of a menu. They provide sufficient conditions under which the seller may offer fewer or more products as a function of cost structure and the distribution of tastes and advertising costs, respectively. The resulting conditions are quite distinct from ours as the seller cannot control the information in these models. Given the control of the information by the seller, one could allow the seller to offer the information directly for the sale to the agents, see Eső & Szentes (2003), (2007), and then possible extract an even larger surplus. We deliberately refrain from given the seller this additional instrument. In the digital economy, the leading application of this paper, the platforms are selling differentiated products and typically bundle the information with the sale of products (through recommendations) rather than offer the information as a separate service at a separate price.

Rayo (2013) considers a model of social status provision that shares some features with our model. The utility function of an agent before any transfer is a product of their type (or an increasing function of their type) and a social status which is equal to their *expected type* given some information structure. Rayo (2013) establishes the optimal information structure informed by techniques from Rayo & Segal (2010). The allocation in Rayo (2013) is an information structure rather than a quality allocation. Importantly, the information structure only affects the allocation but not the expectation of the agents regarding their own type.

There is some earlier work that asks when second-degree price discrimination may optimally resolve in a single-item menu. Anderson & Dana (2009) impose an *a priori* finite upper bound on the quality in the setting of Mussa & Rosen (1978). They state conditions under which all values receive the same quality, namely the quality at the upper bound. Sandmann (2022) shows that their sufficient condition requires that high value buyer's surplus is more concave than that of the low value buyer. Sandmann (2022) shows that a necessary condition for a single-item menu to be profit maximizing is that the single-item menu constitutes the socially optimal allocation. By contrast, in the current environment a continuum of qualities is socially optimal. Hence there would be no reason to restrict the menu and offer a bunching solution in the absence of information design.

2 Model

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

$$v \cdot q - p.$$

The buyers' values $v \in [\underline{v}, \bar{v}]$ are distributed according to a common prior distribution:

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

where $0 \leq \underline{v} < \bar{v} \leq \infty$, with strictly positive density (i.e., $f(v) > 0$ for all v). Initially, the buyers and the seller only know the common prior distribution F of values.

For the seller, producing a mass 1 of goods with a distribution of qualities $R \in \Delta\mathbb{R}_+$, has a cost

$$C(R) : \Delta\mathbb{R}_+ \to \mathbb{R}_+.$$

By defining the cost function in terms of distribution of qualities, we can encompass both the classic model of variable supply of qualities as well as the case of fixed supply of qualities. We only assume that C is monotone with respect to the *increasing convex order*. Formally, for any distributions $Q, R \in \Delta\mathbb{R}_+$, we say Q is greater than R in the increasing convex order, and denote this by $Q \succ_c R$, if for all increasing and convex functions $h : \mathbb{R}_+ \to \mathbb{R}_+$:

$$\mathbb{E}_R\left[h\left(x\right)\right] \leq \mathbb{E}_Q\left[h\left(x\right)\right], \tag{1}$$

see Definition 3.A.1 in Shaked & Shanthikumar (2007). Thus if $R \prec_c Q$, then R is both "lower" and "less variable" than Q in some stochastic sense. We assume that for any pair of distributions, $R \prec_c Q$ implies that $C(R) \leq C(Q)$.

There are two classes of cost functions that are of particular interest to us. Both classes are monotone in the increasing convex order. First, if there is a convex function $c : \mathbb{R} \to \mathbb{R}$ such that:

$$C(R) = \int c(q)dR(q), \tag{2}$$

we have that each quality good is produced independently. In this case, we recover the model of Mussa & Rosen (1978), with endogenous or *variable inventory*. By contrast, if there is an exogenous distribution of qualities Q and the seller can only pool and discard qualities from an exogenously given inventory, we recover the model of Loertscher & Muir (2022), the case of exogenous or *fixed inventory*. In this case the cost is:

$$C(R) = \begin{cases} 0, & \text{if } R \prec_c Q; \\ \infty, & \text{otherwise}; \end{cases} \tag{3}$$

for some exogenous Q. That is, the seller can restrict supply and bundle goods, which generates a (endogenous) distribution of qualities R that is "lower" and "less variable" than the original (exogenous) distribution of qualities Q. In this second case, R is the distribution of *expected* qualities, but since the utility is linear in q the expected quality is a sufficient statistic for the buyer's preferences. Hence, we omit the qualifier "expected" and write simply quality.

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 menu) that specifies the (expected) quality and payments for any reported (expected) value. We now describe these elements in turn.

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

$$S : [\underline{v}, \bar{v}] \rightarrow \Delta\mathbb{R}_{+},$$

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

$$w \triangleq \mathbb{E}[v \mid s]. \tag{4}$$

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.

A menu (or direct mechanism) with qualities $q(w)$ at prices $p\left(w\right)$ for every value w is given by:

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

The menu has to satisfy the usual incentive compatibility and participation constraints:

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

$$wq(w) - p(w) \geq 0, \quad \forall w \in \mathrm{supp}G. \tag{6}$$

The distribution of qualities R needed to supply the buyers' demand is given by:

$$R(r) = \int\limits_{\{w:q(w)\leq r\}} dG(w), \tag{7}$$

which determines the corresponding cost $C(R)$.

We refer to a mechanism as a pair (M, S) of menu M and information structure S. The seller's problem is to maximize expected profit subject to the above incentive compatibility and participation constraints, (5)-(6):

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

In the next section, we will see how the seller's problem can be stated entirely in terms of a choice of a pair of distributions (G, R) over expected values and qualities, respectively, subject to appropriate constraints.

3 The Optimality of Discrete Mechanisms

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 problem (8) as a joint optimization over two distributions, the distribution of the values and the distribution of qualities (i.e., the allocation). The distributions correspond to the information design and mechanism design problems that we are integrating. We then provide our main result, which shows that the presence of two distributions being simultaneously optimized leads to significant distinctive structure, i.e., an endogenous restriction to finite menus.

3.1 Optimization Over Two Distributions

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:

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

where G is the distribution of expected values and the second term inside the integral is the buyers' information rent. 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.

Given that $q(w)$ is non-decreasing, we can write the distribution of qualities in (7) in terms of the quantile t:

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

It is a convention to describe a mechanism as an allocation function that maps (expected) values w to qualities $q(w)$ and a distribution $G(w)$ of (expected) values. For our purposes, it is instead useful to describe the allocation as a function that maps the quantile t rather than the value w for several related reasons. By fixing the allocation as a function of quantiles $R^{-1}(t)$ and varying the distribution of expected values, the distribution of qualities will not change (as the distribution over quantiles is always uniform on $[0,1]$). Yet as we vary the information structure and keep the allocation described in terms of quantiles constant $R^{-1}(t)$, we are implicitly changing the allocation $q(w)$. Describing the allocation and the information structure in terms of quantiles has the added benefit that the revenue function will be bilinear in both distributions.

Thus, using the change of variables $t = G(w) \Leftrightarrow G^{-1}(t) = w$, and integrating by parts twice, we write (9) as follows:

$$\begin{aligned}
\mathbb{E}[p(w)] &= \int_0^1 \left(G^{-1}(t)R^{-1}(t) - \int_0^t R^{-1}(s)dG^{-1}(s) \right) dt \\
&= \int_0^1 \left(G^{-1}(t) - \frac{dG^{-1}(t)}{dt}(1-t) \right) R^{-1}(t)dt \\
&= \int_0^1 G^{-1}(t)(1-t)dR^{-1}(t) + R^{-1}(0)G^{-1}(0).
\end{aligned}$$

Of course, the above is only the revenue, to write the profit we need to include the cost. We now characterize the feasible distributions of expected values.

The buyers' information structure is summarized by the distribution of expected values 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.,

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

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 now report a characterization of the increasing convex order defined earlier in (1), which is closely related to the majorization constraint. For any pair of distributions $R, \hat{R} \in \Delta\mathbb{R}$, we have that $R \prec_c \hat{R}$ if and only if

$$\int_x^1 R^{-1}(t)dt \leq \int_x^1 \hat{R}^{-1}(t)dt, \, \forall x \in [0,1], \tag{10}$$

where we do not require that the inequality becomes an equality at $x = 0$ (see Theorem 4.A.3. in Shaked & Shanthikumar (2007)). The inequality (10) does not need to be satisfied with equality anywhere because R could be generated by discarding goods in the distribution $\hat{R}$ (and not only pooling). In this case, we say $\hat{R}^{-1}$ *weakly majorizes* R^{-1} and we write $R^{-1} \prec_c \hat{R}^{-1}$ (hence, the weak majorization order of the quantile function is equivalent to the increasing convex order of the associated distributions).

The seller's problem is then to maximize profit (revenue minus cost):

$$\max_{\substack{G^{-1} \prec F^{-1} \\ R^{-1} \in \Delta \mathbb{R}_+}} \int_0^1 G^{-1}(t)(1-t)\, dR^{-1}(t) + R^{-1}(0)\, G^{-1}(0) - C(R), \tag{11}$$

subject to R^{-1} being measurable with respect to $G(v)$. The additional measurability condition is to guarantee that we can implement the allocation rule using a direct mechanism $q(w) = R^{-1}(G(w))$. Hence, whenever G^{-1} is constant, R^{-1} must also be constant.

We additionally impose that G^{-1} is measurable with respect to R. In fact, if G^{-1} were increasing in some interval (t_1, t_2) and R^{-1} were constant, then we could consider the following information structure $\hat{G}$:

$$\hat{G}(t) = \begin{cases} \frac{\int_{t_1}^{t_2} G^{-1}(s)ds}{t_1 - t_2}, & \text{if } t \in (t_1, t_2); \\ G^{-1}(t), & \text{otherwise.} \end{cases}$$

The mechanism $(\hat{G}, R)$ would generate weakly larger profit than (G, R), and strictly larger profit if R is discontinuous at t_1. This second measurability condition implies that buyers are not provided information that is not relevant for making the decision of what item on the menu to buy. In fact, this second measurability condition can be interpreted as requiring that the mechanism can be either implemented as a direct mechanism or as an indirect mechanism in the form of a tariff of qualities and prices together with a recommendation system. We expand on this second interpretation in Section 7 after we provide our results.

We thus optimize among mechanisms that satisfy both of these measurability conditions. In other words, at any $t \in [0,1]$, $G^{-1}(t)$ is strictly increasing if and only if $R^{-1}(t)$ is strictly increasing. When a pair of distributions satisfy both of these measurability conditions, we say they have a *common quantile support*. In this case, the mechanism can be implemented as a direct mechanism or as an indirect mechanism, a tariff of qualities and prices together with a recommendation system. We denote by (G^*, R^*) a solution to this problem.

3.2 Optimality of Discrete Menu

Our main result shows that the mechanism offers a finite set of distinct qualities and therefore provides only partial information to buyers. Formally, we say that a given distribution H has discrete support if there is no open interval $(t_1, t_2) \subset [0, 1]$ such that H^{-1} is strictly increasing in (t_1, t_2). We say a mechanism (G, R) is discrete if it consists of distributions with discrete support: this concept captures the fact that information and the menu offered to consumers is discrete.

Theorem 1 (Optimality of Discrete Menu)
Every optimal mechanism is discrete.

We first provide an intuition and then provide the proof. Consider a mechanism in which G^{*-1} and R^{*-1} are strictly increasing in some interval $[t_1, t_2]$. Suppose the seller pools the allocation of all values corresponding to quantiles $[t_1, t_2]$, so that all values get the average quality on this interval. How much lower would the profit be? We begin the argument with the virtual values given by:

$$\phi(t) \triangleq G^{*-1} - (1-t)\frac{dG^{*-1}(t)}{dt}. \tag{12}$$

This is the standard formulation for the virtual values but written in terms of the quantile function instead of the distribution of values. The revenue generated is the expectation of the product of the virtual values and the qualities, so profit are given by:

$$\Pi^* \triangleq \int_0^1 R^{*-1}(t)\phi(t)dt - C(R^*).$$

We denote the (conditional) mean and variance of virtual values and qualities in the quantile $[t_1, t_2]$ by:

$$\mu_\phi \triangleq \frac{\int_{t_1}^{t_2} \phi(t)dt}{t_2 - t_1}; \qquad \mu_q \triangleq \frac{\int_{t_1}^{t_2} R^{*-1}(t)dt}{t_2 - t_1}; \tag{13}$$

$$\sigma_\phi^2 \triangleq \frac{\int_{t_1}^{t_2} (\phi(t) - \mu_\phi)^2 dt}{t_2 - t_1}; \qquad \sigma_q^2 \triangleq \frac{\int_{t_1}^{t_2} (R^{*-1}(t) - \mu_q)^2 dt}{t_2 - t_1}. \tag{14}$$

As we only compute conditional mean and variance in the interval $[t_1, t_2]$, we can safely omit an index referring to the interval $[t_1, t_2]$ in the expression of μ and σ.

The first step of the proof shows that the revenue losses due to pooling the qualities in the interval $[t_1, t_2]$ are bounded by:

$$\sigma_\phi \sigma_q (t_2 - t_1).$$

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). Note also that pooling (weakly) reduces the production cost.

If in addition to pooling the qualities we also 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 $\mu_q - R^{*-1}(t_1)$ 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:

$$\mu_v \triangleq \mathbb{E}[v \mid G^*(v) \in [t_1, t_2]].$$

Hence, pooling the values increases the payments for every value corresponding to quantiles higher than t_1 by an amount:

$$(\mu_v - G^{*-1}(t_1))(\mu_q - R^{*-1}(t_1)).$$

Here the first two terms being multiplied are small when the interval is small. However, payments are marginally increased for all values higher than t_1, which is a non-negligible mass of values (i.e., $(1 - t_1)$ is not small). In other words, pooling values increases the price of the quality improvement $(\mu_q - R^{*-1}(t_1))$ for all values higher than $G^{*-1}(t_1)$. Hence, pooling values generates a second-order benefit which always dominates the third-order distortions.

Proof of Theorem 1. We consider a candidate optimal mechanism and an interval (t_1, t_2) such that G^{*-1} and R^{*-1} are strictly increasing in this interval. We attain a contradiction by establishing that there is an improvement. It is useful to write the interval (t_1, t_2) in terms of its mid-point and width:

$$\hat{t} \triangleq \frac{v_1 + t_2}{2}; \quad \Delta \triangleq \frac{t_2 - v_1}{2}. \tag{15}$$

So, we have that

$$(t_1, t_2) = (\hat{t} - \Delta, \hat{t} + \Delta) \tag{16}$$

and we will eventually take the limit $\Delta \to 0$. We consider two cases: (i) ϕ is weakly increasing in $[t_1, t_2]$, and (ii) ϕ is not weakly increasing $[t_1, t_2]$. In the first case, we assume that $\hat{t}$ is such that

$\phi(t)$ and $G^{*-1}(t)$ are differentiable at $\hat{t}$ (any monotonic function is differentiable almost everywhere so we can find such $\hat{t}$). In the second case we assume that $\phi(t)$ is strictly decreasing at $\hat{t}$.

We consider qualities:

$$\tilde{R}^{-1}(t) = \begin{cases} \mu_q, & \text{if } t \in [t_1, t_2); \\ R^{*-1}(t), & \text{if } t \notin [t_1, t_2). \end{cases} \tag{17}$$

The difference in profit between the optimal policy Π^* and the variation $\tilde{\Pi}$ is given by:

$$\Pi^* - \tilde{\Pi} = \int_{t_1}^{t_2} \phi(t) R^{*-1}(t) dt - (t_2 - t_1)\mu_\phi \mu_q - \left(C(R^*) - C(\tilde{R}) \right). \tag{18}$$

Note that we only need to consider the qualities in the interval $[t_1, t_2]$ to compute the difference in revenue. We can write this expression more conveniently as follows:

$$\Pi^* - \tilde{\Pi} = \int_{t_1}^{t_2} (\phi(t) - \mu_\phi)(R^{*-1}(t) - \mu_q) dt - \left(C(R^*) - C(\tilde{R}) \right).$$

Since $\tilde{R} \prec_c R^*$, we have $C(\tilde{R}) \leq C(R^*)$. If $\phi(t)$ is strictly decreasing at $\hat{t}$, we have that:

$$\Pi^* - \tilde{\Pi} \leq \int_{t_1}^{t_2} (\phi(t) - \mu_\phi)(R^{*-1}(t) - \mu_q) dt < 0,$$

for a small enough Δ. Hence, in what follows we assume that $\phi(t)$ is increasing. Using the Cauchy-Schwarz inequality and the fact that $C(\tilde{R}) \leq C(R^*)$, we can bound the first integral (and thus the whole expression) as follows:

$$\Pi^* - \tilde{\Pi} \leq \sigma_q \sigma_\phi (t_2 - t_1).$$

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

$$\sigma_q \leq \sqrt{(\mu_q - R^{*-1}(t_1))(R^{*-1}(t_2) - \mu_q)},$$

and similarly for ϕ. So we have that:

$$\Pi^* - \tilde{\Pi} \leq \sqrt{(\mu_q - R^{*-1}(t_1))(R^{*-1}(t_2) - \mu_q)} \sqrt{(\mu_\phi - \phi_1)(\phi_2 - \mu_\phi)} (t_2 - t_1),$$

and we conclude that, using the earlier definition in (15)-(16):

$$\lim_{\Delta \to 0} \frac{\Pi^* - \tilde{\Pi}}{\Delta^3} \leq \frac{dR^{*-1}(\hat{t})}{dt} \frac{d\phi(\hat{t})}{dt}. \tag{19}$$

The efficiency losses are therefore of order Δ^3.

We now consider a variation of the mechanism (G^*, R^*) denoted by $\left(\hat{G}, \hat{R}\right)$ where the allocation $\hat{R}^{-1}$ remains the same as in $\tilde{R}^{-1}$ (see (17) above):

$$\hat{R}^{-1}(t) = \begin{cases} \mu_q, & \text{if } t \in [t_1, t_2); \\ R^{*-1}(t), & \text{if } t \notin [t_1, t_2); \end{cases} \tag{20}$$

but we change the information structure regarding v so that all values in (t_1, t_2) are pooled:

$$\hat{G}^{-1}(t) = \begin{cases} \mu_v & \text{if } t \in [t_1, t_2); \\ G^{*-1}(t), & \text{if } t \notin [t_1, t_2). \end{cases} \tag{21}$$

Observe that the total surplus generated by $(\hat{G}, \hat{R})$ and by $(G^*, \tilde{R})$ is the same. Then, the difference in the generated profit is equal to the difference in the expected buyers' surplus:

$$\hat{\Pi} - \tilde{\Pi} = (\mu_q - q_1^-)(\mu_v - G^{*-1}(t_1))(1 - t_1), \tag{22}$$

where

$$q_1^- \triangleq \lim_{t \uparrow t_1} R^{*-1}(t),$$

so the limit is taken from below. Since, $q_1^- \leq R^{*-1}(t_1)$, we have that:

$$\hat{\Pi} - \tilde{\Pi} \geq (\mu_q - R^{*-1}(t_1))(\mu_v - G^{*-1}(t_1))(1 - t_1).$$

We conclude that:

$$\lim_{\Delta \to 0} \frac{\hat{\Pi} - \tilde{\Pi}}{\Delta^2} \geq \frac{dR^{*-1}(\hat{t})}{dv}(1 - t_1). \tag{23}$$

Here we used that $(\mu_v - G^{*-1}(t_1))/\Delta \to 1$, as $\Delta \to 0$. The efficiency losses are of order Δ^2. We conclude that for Δ small enough, the new policy generates a profit improvement. ■

Our variation argument for Theorem 1 relates to an insight of Wilson (1989). He studied the *surplus-maximizing* mechanism and asked what losses are associated with restricting the number of items to at most N. He showed that the surplus losses are of order $1/N^2$. We complement this argument with the fact that the gains that come from reducing informational rents always have a larger order of magnitude. We thus conclude that there is always some amount of pooling in the *profit-maximizing* mechanism.

We use a perturbation argument to establish our result, i.e., we show a revenue gain from pooling a small interval if there were a region of full separation. In contrast, Bergemann & Pesendorfer

(2007) analyze finite mechanisms with an exogenous bound on the number of items, and then show that as the bound grows, the number of items used in the optimal mechanism eventually stays constant. Hence, as well as working in different settings, their argument does not provide an explicit variation that constitutes an improvement of a continuous mechanism. The present argument identifies the general logic of jointly optimizing information and allocation. For example, we show that the logic of our argument can be extended to nonlinear environments in Section 8.

By Theorem 1, we can describe the optimal mechanism by a countable collection of quantiles $\{t_k\}_{k\in K}$, where at each quantile t_k there is a discontinuous jump in both qualities and expected values and both distributions are constant everywhere else. As shorthand notation, we denote by w_k the expected value and Δr_k the quality increment at quantile t_k:

$$w_k \triangleq G^{-1}(t_k); \qquad \Delta r_k \triangleq r(t_k) - \lim_{t\uparrow t_k} R^{-1}(t).$$

Following (11) the profit are given by:

$$\Pi = \sum_{k=I} w_k \Delta r_k (1 - t_k) - C(R). \tag{24}$$

This formula restates the profit formula in (11) in terms of a countable collection of quantiles. We shall use this representation in the next section.

4 The Optimality of a Finite Mechanism

So far, we have established that the optimal mechanism pools many values and correspondingly many qualities. The optimal mechanism generates a discrete distribution of values and qualities. But how many elements will there be in the partition, or, correspondingly, how many items will there be on the menu? We now establish that typically there are in fact only finitely many items.

To prove the optimal mechanism is finite, we will impose additional assumptions (on the cost function or the distribution of values) that guarantee that the qualities offered by an optimal mechanism are bounded above and then prove that these qualities have no accumulation points. The arguments in Section 3.2 allow us to prove that there are no accumulation points, except possibly at the top of the distribution. The arguments do not apply when the accumulation point is at the top of the distribution: in the proof of Theorem 1, we considered an interval (t_1, t_2) that is small enough relative to the mass of values above the interval (that is, small relative to $(1 - t_2)$). When the accumulation point is at the top of the distribution, the arguments no longer go through because we must consider an interval of the form $(t_1, 1)$.

We begin this section by showing that the qualities offered by an optimal mechanism must exhibit increasing differences. This result is of independent interest, but it also rules out the possibility of accumulation points at the top of the distribution.

4.1 The Convexity of Optimal Qualities

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

The convexity of menus arises with some frequency in nonlinear pricing. In mobile phone pricing, the data packages offered are frequently convex. Similarly, mobile phones themselves are offered with a variety of memory chips with convex structure.[1]

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

Proof. We now consider three consecutive discontinuous jumps $t_{l-1},\ t_l,\ t_{l+1}$. We fix a mechanism (G, R) and assume that there exists l such that $\Delta r_l > \Delta r_{l+1}$, and prove that there exists another mechanism that generates higher profit. To make the notation more compact, we define:

$$g_{l-1} \triangleq t_l - t_{l-1} \text{ and } g_l \triangleq t_{l+1} - t_l.$$

Note that there is an atom with probability g_l and that the expected value is w_l.

The mechanism is defined as follows. First, for all $k \notin \{l-1, l\}$, the quantiles and the level of the quality and value jump at the same quantile:

$$\widetilde{w}_k = w_k; \quad \widetilde{r}_k = r_k; \quad \widetilde{t}_k = t_k.$$

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

We modify the information structure as follows:

$$\begin{aligned} \tilde{t}_{l-1} &= t_{l-1}, & \tilde{t}_l &= t_l + \varepsilon; \\ \tilde{w}_{l-1} &= \frac{g_{l-1} w_{l-1} + \varepsilon w_l}{g_{l-1} + \varepsilon}, & \tilde{w}_l &= w_l; \\ \tilde{r}_{l-1} &= \frac{g_{l-1} r_{l-1} + \varepsilon r_l}{g_{l-1} + \varepsilon}, & \tilde{r}_l &= r_l. \end{aligned}$$

Note that:

$$\begin{aligned} \widetilde{g}_{l-1}\widetilde{w}_{l-1} + \widetilde{g}_l \widetilde{w}_l &= g_{l-1} w_{l-1} + g_l w_l; \\ \widetilde{g}_{l-1}\widetilde{r}_{l-1} + \widetilde{g}_l \widetilde{r}_l &= g_{l-1} r_{l-1} + g_l r_l. \end{aligned}$$

We additionally have that the difference between $\tilde{G}^{-1}$ and G^{-1} changes its sign at most once, and then from positive to negative. Similarly, for the difference $\tilde{R}^{-1} - R^{-1}$, it changes its its sign at most once, and then from positive to negative. We can then appeal to Shaked & Shanthikumar (2007), Theorem 3.A.44 to conclude that $\tilde{G}^{-1} \prec G^{-1}$ and $\tilde{R}^{-1} \prec R^{-1}$. So $\tilde{G}$ is feasible and $C(\tilde{R}) \leq C(R)$.

Taking the derivative and evaluating at 0, we get:

$$\begin{aligned} \frac{\partial}{\partial \varepsilon}\widetilde{\Pi}|_{\varepsilon=0} &= \frac{\partial}{\partial \varepsilon}\big(\widetilde{w}_{l-1}\Delta\widetilde{r}_{l-1}(1-\widetilde{t}_{l-1}) + \widetilde{w}_l \Delta \widetilde{r}_l (1-\widetilde{t}_l) - C(\tilde{R})\big)|_{\varepsilon=0} \\ &= -\frac{(1-t_l)(w_{l+1}-w_l)(\Delta r_{l+1} - \Delta r_l))}{(t_{l+1}-t_l)} - \frac{\partial}{\partial \varepsilon} C(\tilde{R})|_{\varepsilon=0}. \end{aligned}$$

The first equality simply identifies the terms in (24) that change with ε and the second equality explicitly computes the derivative and evaluates it at $\varepsilon = 0$. The cost is decreasing in ε because $\tilde{R}^{-1} \prec R^{-1}$, so we get:

$$\frac{\partial}{\partial \varepsilon}\widetilde{\Pi}|_{\varepsilon=0} \geq -\frac{(1-t_{l-1})(w_l - w_{l-1})(\Delta r_l - \Delta r_{l-1}))}{(t_l - t_{l-1})}.$$

Hence, if $(\Delta r_l - \Delta r_{l-1}) < 0$, we have that

$$\frac{\partial}{\partial \varepsilon}\widetilde{\Pi}|_{\varepsilon=0} > 0,$$

which implies that the mechanism (G, R) is not optimal. We thus conclude that in any optimal mechanism $\left(\Delta r_l^* - \Delta r_{l-1}^*\right) \geq 0$. This concludes the proof. ■

The intuition for the result is that between any two consecutive quality increments levels Δq_k, Δq_{k+1}, informational rents depend on the lower quality increment level Δq_k while the surplus gains from separation depend on the higher quality increment level Δq_{k+1}. To gain a formal intuition for this result, suppose we consider a model with two levels of values and qualities, that is, $v \in \{\underline{v}, \overline{v}\}$, $q \in \{\underline{q}, \bar{q}\}$, both uniformly distributed. We ask whether offering a one-item menu via pooling or offering a two-item menu via separating generates higher profit. We also assume excluding the low type is not optimal.

If the items are offered separately, the high value buys the high quality, and the low value buys the low quality. In this environment, the first quality increment is the lower quality level and the second quality increment is the difference between the quality levels:

$$\Delta q_1 = \underline{q}; \quad \Delta q_2 = \bar{q} - \underline{q}.$$

The total surplus generated when offering a menu is:

$$S_M = \frac{1}{2}(\underline{v} \cdot \underline{q} + \bar{v} \cdot \bar{q}).$$

If the goods and values are pooled, the total surplus generated is:

$$S_P = \frac{(\bar{v} + \underline{v})}{2} \frac{(\bar{q} + \underline{q})}{2}.$$

The difference is then given by:

$$S_M - S_P = \frac{1}{4}\Delta q_2(\bar{v} - \underline{v}).$$

In contrast, when offering a menu, the informational rents are the rents the high value buyer gains by mimicking the low value buyer:

$$U_M = \frac{1}{2}(\bar{v} - \underline{v})\Delta q_1.$$

Separating is optimal only when the reduction in informational rents – proportional to Δq_1– is smaller than the surplus gains – proportional to Δq_2. In this example, we get that separating is optimal only if $2\Delta q_1 \leq \Delta q_2$, so the second quality increment must be twice as large as the first quality increment. When doing the variational analysis in the proof, we consider the surplus gains and informational rents reduction when pooling a small fraction of some interval $k + 1$ with the immediate predecessor interval k. Thus we obtain a weaker condition, namely, increasing differences. However, the intuition is the same: the surplus losses from pooling are proportional to Δq_{k+1} while the informational rents reduction is proportional to Δq_k.

The convexity in the qualities provides the key insight to establish that there will be finitely many rather than countable many items. In particular, the convexity of the menu will exclude the possibility of an accumulation point at the top range of the menu.

4.2 The Optimality of a Finite Item Menu

We say that a mechanism (G, R) is finite if both distributions have finite support. The optimal mechanism is finite if there is an upper bound on the qualities provided. Towards this end, we introduce weak conditions that guarantee that there exists $\bar{q} < \infty$ such that for any distribution $R \in \Delta\mathbb{R}_+$,

$$C(R) < \infty \Rightarrow R(\bar{q}) = 1. \tag{25}$$

If there exists $\bar{q} < \infty$ such that (25) holds, then we say that the *cost function C is bounded.* Hence, the seller never supplies a quality higher than $\bar{q}$. If we consider a model with a fixed inventory (see (3)), then this is a restriction on the exogenous distribution of qualities. If we consider a model with separable cost $c(q)$ (see (2)) and the distribution of values is bounded by $\bar{v}$, then one can always appropriately truncate the cost function. More precisely, in any optimal mechanism the seller never supplies any quality higher than the efficient quality with the highest value:

$$\bar{q} = c^{'-1}(\bar{v}).$$

Hence, by considering an alternative cost function:

$$\hat{c}(q) \triangleq \begin{cases} c(q), & \text{if } q \leq \bar{q}; \\ \infty, & \text{otherwise;} \end{cases}$$

we can satisfy (25) without changing the nature of the optimal mechanism.

Theorem 2 (Finite Item Menu)
If the cost function is bounded, then the optimal mechanism (G^*, R^*) is finite.

Proof. By the boundedness of the cost function, the qualities offered by an optimal mechanism are in a compact space $[0, \bar{q}]$. The finiteness property of the optimal mechanism is then equivalent to the property that the optimal mechanism has no accumulation points. Proposition 1 implies that there cannot be any accumulation points, except possibly at some quantile $\hat{t}$ satisfying $R^{-1*}(\hat{t}) = 0$. Hence, it is a decreasing accumulation point (that is, the limit of expected values converges to $\hat{t}$ from the right). We denote the expected value in the limit by $\hat{w}$.

We now consider any *infinite* optimal mechanism and show that it cannot be optimal. We denote by $\{t_k\}_{k \leq K}$ the discontinuity quantiles, with $t_k < t_{k+1}$. The index runs from $-\infty$ to K. Analogous to (12), we define:

$$\phi_k \triangleq w_k - (w_{k+1} - w_k)\frac{1 - t_{k+1}}{g_k}, \tag{26}$$

and extend (13)-(14) 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.

We consider the impact on revenue of pooling to consecutive intervals and define $\hat{\Pi}$ and $\tilde{\Pi}$ in the same way as in Section 3.2. Following the same steps as in Section 3.2, we have that:

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

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:

$$\frac{\Pi^* - \tilde{\Pi}}{\hat{\Pi} - \tilde{\Pi}} \leq \frac{g_k(\phi_{k+1} - \phi_k)(g_k + g_{k+1})}{g_{k+1}(w_{k+1} - w_k)(1 - t_{k-1})}. \tag{27}$$

We prove that we can find intervals such that the right-hand-side of this inequality is less than 1.

We now define:

$$h_k \triangleq \frac{g_k}{w_{k+1} - w_k},$$

and note that we can write ϕ_k as follows:

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

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

$$\hat{h} \triangleq \lim_{k \to -\infty} h_k.$$

Since ϕ_k is positive we must have that $\hat{h} > 0$.

We now prove that $\hat{h} < \infty$. For this we define the following function:

$$\Psi(t_k) \triangleq (w_k - \hat{w})(1 - t_k).$$

Recall that $\hat{w}$ denotes the limit $w_k \to \hat{w}$ as $k \to -\infty$. We can write the function alternatively as follows:

$$\Psi(t_k) = -\sum_{\ell \leq k} g_k(\phi_k - \hat{w}).$$

If $\hat{h} = \infty$, then in the limit $k \to -\infty$, $\phi_k \to \hat{w}$. But, for every k away from the limit, we have that $\Psi(t_k) > 0$ (since $w_k > \hat{w}$). We thus must have that there exists k such that

$$\phi_k < \hat{w} = \lim_{k \to -\infty} \phi_k.$$

This prove that the virtual values are not increasing, which is a contradiction. We thus have that $\hat{h} < \infty$.

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

$$\frac{\Pi^* - \tilde{\Pi}}{\hat{\Pi} - \tilde{\Pi}} \leq \frac{h_k(\phi_{k+1} - \phi_k)(w_{k+2} - w_k) \max\{h_k, h_{k+1}\}}{(w_{k+2} - w_{k+1}) h_{k+1}(1 - t_{k-1})}.$$

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

$$\frac{\Pi^* - \tilde{\Pi}}{\hat{\Pi} - \tilde{\Pi}} \leq \frac{h_k(\phi_{k+1} - \phi_k) 2 \max\{h_k, h_{k+1}\}}{h_{k+1}(1 - t_{k-1})}.$$

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

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

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

While we have established that the optimal mechanism is finite, we have not so far identified any structure of which values and qualities are pooled with which other values and qualities. In the next section we leverage the fact that the objective function (11) is linear in G^{-1} to provide a complete characterization of the pooling structure.

5 The Optimality of Monotone Partitions

We use the structure of the extremal points of the set of distributions that are majorized by F (i.e., the set $\{G^{-1} : G^{-1} \prec F^{-1}\}$) to show that the distribution of expected values in a optimal mechanism always pools intervals of values. In fact, revenue is linear in the allocation R^{-1}, so when the seller has a fixed inventory and maximizes revenue (i.e., when the cost is (3)) the distribution of qualities will have a similar structure.

The argument we report in this section is one that has been widely used in the existing literature to solve problems that correspond to maximizing a linear function subject to a single majorization constraint. By contrast, our arguments to establish the optimality of discrete and then finite mechanisms critically relied on the fact that we can employ two instruments–information and allocation–which interact. After showing that the optimal mechanism is monotone partitional, we shall review the related literature that considers maximization problems over a one dimensional distribution.

5.1 Monotone Partition

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

We begin by explaining the extreme points of the set of quantile distributions that are majorized by F^{-1}, that is, the set $\{G^{-1} : G^{-1} \prec F^{-1}\}$. A distribution of values G is said to be *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

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

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

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

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

$$G^{-1}(t) \triangleq \begin{cases} F^{-1}(t), & \text{if } t \in [x_i, x_{i+1}) \text{ for some } i \in J; \\ w_i, & \text{if } t \in [x_i, x_{i+1}) \text{ for some } i \notin J. \end{cases} \tag{28}$$

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 (28).

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

$$\Pi = \max_{\{G^{-1}:G^{-1}\prec F^{-1}\}} \int_0^1 G^{-1}(t)(1-t)dR^{*-1}(t) + R^{*-1}(0)G^{*-1}(0). \tag{29}$$

The optimization problem (29) is an upper semi-continuous linear functional of G^{-1}. Upper semi-continuity can be verified by noting that the quantile function G^{-1} is (by definition) upper semi-continuous. Hence, if $\hat{G}^{-1} \to G^{-1}$ (taking the limit using the L^1 norm), we have that $\limsup \hat{G}^{-1}(t) \leq G^{-1}(t)$ for all $t \in [0,1]$. Hence, $\limsup \int_0^1 \hat{G}^{-1}(t)(1-t)dR^{*-1}(t) \leq \int_0^1 G^{-1}(t)(1-t)dR^{*-1}(t)$.

Following Bauer's maximum principle (Bauer (1958)), the maximization problem attains its maximum at an extreme point of $\{G^{-1} : G^{-1} \prec F^{-1}\}$. By Theorem 1, we excluded the possibility of intervals of complete disclosure. The distribution of expected values G is said to be *monotone pooling* if all intervals are pooling.

Corollary 1 (Monotone Pooling of Values)
There exists an optimal mechanism (G^{*-1}, R^{*-1}) in which G^{*-1} is monotone pooling.

The corollary shows that the optimal information structure can be constructed in a straightforward manner: the value space is partitioned into intervals, and buyers are only told to which interval their value belongs.

Suppose now the seller has a fixed inventory of goods (as in (3)). The profit maximization problem is then reduced to revenue maximization subject to two majorization constraints:

$$\max_{\substack{G^{-1}\prec F^{-1} \\ R^{-1}\prec_c Q^{-1}}} \int_0^1 G^{-1}(t)(1-t)dR^{-1}(t) + R^{-1}(0)G^{-1}(0), \tag{30}$$

We thus obtain an optimization problem subject to two majorization constraints that is bilinear in R and G. Since the problem is linear in R^{-1} we obtain the same characterization as with G^{-1} except we need to account for the fact that the majorization constraint is a weak majorization constraint.

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 (as the seller can discard some items). 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:

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

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 a 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.

Corollary 2 (Monotone Pooling of Qualities)
If the supply of qualities is exogenous (see 3), there exists a optimal mechanism (G^{*-1}, R^{*-1}) in which R^{*-1} is monotone pooling.

The argument is similar to the one offered for Corollary 1, except that the extreme points are now the set of feasible allocations including the possibility of discarding some goods. Additionally, we recall that in any optimal mechanism G^{*-1} and R^{*-1} are on a common quantile support. Thus when the cost is given by (3), the mechanism is determined by a finite number of cutoff quantiles $\{x_0, ..., x_K\}$, with $x_0 = 0$ and $x_K = 1$ such that, for all $x \in [x_k, x_{k+1})$:

$$G^{*-1}(x) = \frac{\int_{x_k}^{x_{k+1}} F^{-1}(t)dt}{x_{k+1} - x_k} \text{ and } R^{*-1}(x) = \frac{\int_{x_k}^{x_{k+1}} Q^{-1}(t)dt}{x_{k+1} - x_k}.$$

We illustrate the results below in Figure 1 and 2 for the case of fixed inventory. 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 1 below. More directly, when we plot the quantile distributions G^{*-1} and R^{*-1} as in Figure 2 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 qualities but the jumps occur at the same quantiles which reflects that values and qualities are assorted in monotone manner.

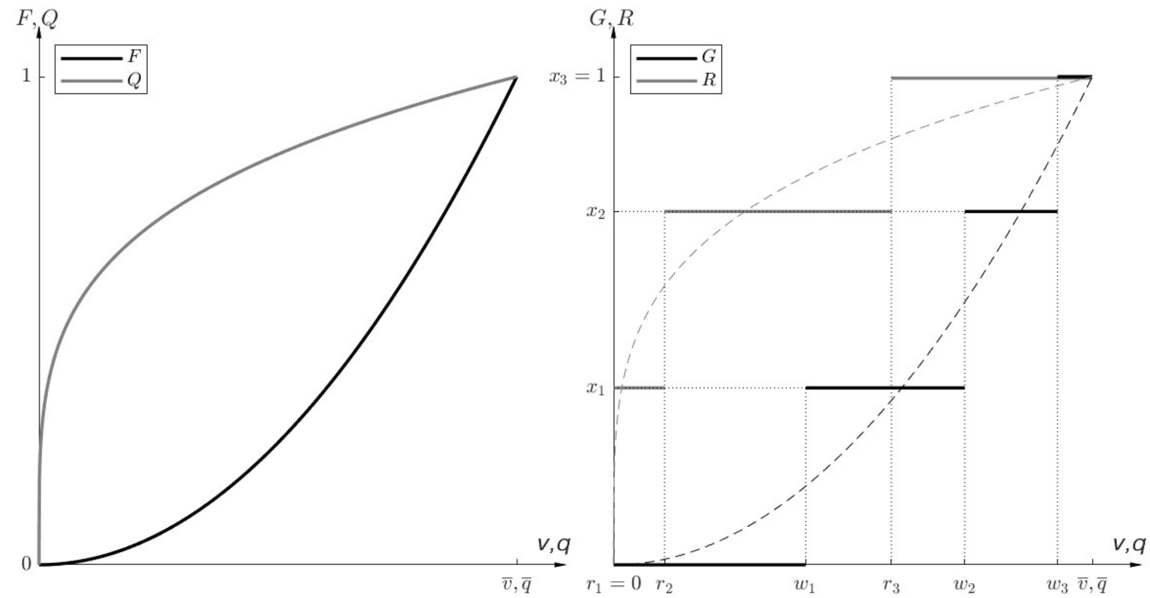

Figure 1: The give distributions of values $F(v) = v^2$ and $Q(q) = q^{1/4}$ are depicted on the left. The associated optimal monotone pooling distributions G and R are depicted on the right.

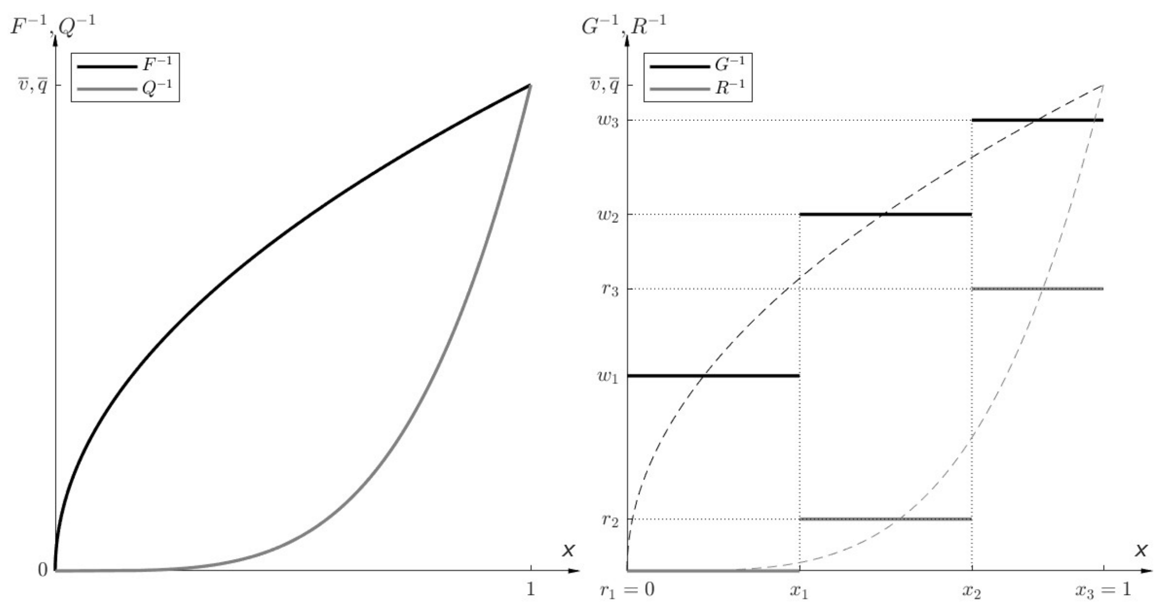

Figure 2: The given quantile distribution $F^{-1}(t) = t^{1/2}$ and $Q^{-1}(t) = t^4$. The corresponding optimal quantile distributions G^{-1} and R^{-1} have common support and display jumps at the same quantiles.

5.2 One vs. Two Majorization Constraints

The objective function in (30) is obtained without making use that G and R are optimization variables. Hence, we can consider alternative problems in which we maximize over one distribution and fix the other distribution exogenously. This allows obtaining different maximization problems with their own distinct economic interpretation. We now show how this allows placing several papers in a common framework. These papers have in common that the original economic problem reduces to a maximization over one distribution and the objective function is the same as in (30).

For example, if we take (30) 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, or in other words bidders in 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).[2] If instead we impose that $R = Q = q^{\frac{1}{N-1}}$ and optimize only over G, we recover the setting of Bergemann et al. (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 maximizes buyers' surplus in a two-sided matching market (see Proposition 4 in Kleiner et al. (2021)). All the above optimization problems (with one majorization constraint) can be solved using versions of ironing techniques first developed by Myerson (1981) and later generalized by Kleiner et al. (2021), combining both separating and pooling regions in the optimum solutions.

The argument for monotone partitional information in our paper follows this earlier literature. Interestingly, this argument is entirely separable from earlier result, Theorem 1 regarding the dis-

[2] 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 object given the report of all 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 $Q \succ_w q^{\frac{1}{N-1}}$, as developed in detail in Kleiner et al. (2021).

creteness and finiteness of the optimal mechanisms. For this result the fact that two distributions were joint instruments in the optimal design was key.

6 On the Number of Items in the Optimal Menu

We established the necessity of finite menus. To provide further results on the number of items in the optimal mechanism we need to focus our attention on more specific cost functions. We first analyze the case in which the seller has a fixed inventory (see (3)). We show that a single-item menu is optimal if the density of qualities is increasing; and we show that the number of items offered is bounded above by the ratio of upper bound of quality to the lower bound of quality. We then examine the case in which the cost is separable (see (2)). We again provide conditions under which a single-item menu is optimal: the marginal cost must be convex and the distribution of values must satisfy a "modest-tails" condition. Also, we establish a relationship between the cost elasticity (when the elasticity is constant) and the number of items in the optimal menu. In particular, any given information structure is dominated by complete pooling if the elasticity is sufficiently high; and is dominated by complete disclosure if the elasticity is sufficiently low. The proofs of the results in this section are relegated to the Appendix.

6.1 Optimality of Single-Item Menu with Fixed Inventory

First, consider the fixed inventory case (see (3)). The first result offers a sufficient condition for the optimal menu to contain a single-item menu.

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

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. More specifically, the surplus loss is:

$$\Delta S = g_H(1 - g_H)(w_H - w_L)(r_H - r_L).$$

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

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

Furthermore, the losses in surplus are smaller than the gains from eliminating buyers surplus when the low quality r_L is not too small relative to the high quality r_H. Note that both expressions are proportional to $(w_H - w_L)$ so the distribution of values does not play a role in the comparison, which provides some intuition for why the distribution of values does not play a role in Theorem 3.

When analyzing how to segment a continuous distribution of qualities, we can use the above insights to understand the structure of the optimal policy. 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 at 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 also bound the number of items offered in any mechanism. The bound relies on the upper and lower bound of the support of values. For the following result we assume that $\underline{q} > 0$.

Corollary 3 (Finite Upper Bound on the Number of Items)
The number of items K offered by an optimal mechanism is bounded above by

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

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

6.2 Optimality of Single-Item Menu with Variable Inventory

We now assume that the cost function is separable as in (2) and establish when single-item mechanisms are optimal. We say a distribution has *modest tails* if

$$f'(v) < 0 \Rightarrow f''(v) \leq 0. \tag{32}$$

This condition states that f must be concave when it is decreasing. For example, any distribution with (weakly) increasing density satisfies (32). It is also satisfied if the density is linearly decreasing. In contrast, the condition cannot be satisfied by any distribution with unbounded support.

Theorem 4 (Optimality of Single-item Mechanisms with Variable Inventory) *If the distribution satisfies modest tails and $c'''(q) \geq 0$, then the optimal mechanism is a single-item menu.*

Theorem 4 shows that in a range of settings the optimal mechanism is a single-item menu. We remark that this does not immediately mean that all values will be pooled. It is possible that the conditions of the theorem are satisfied and the optimal mechanism consists of a single-item but some buyers are excluded (and buy 0 quality at 0 price).

As before, to gain intuition, we consider a mechanism that generates two possible expected valuations (w_L, w_H) and consider the possibility of pooling both values. The trade-off is the same is we previously saw: the benefit of pooling is that the information rents will be eliminated and the cost is that the social surplus will be reduced. The reduction in informational rents is given by (31), where now r_L is endogenously determined. However, when $w_H \approx w_L$, we have that $r_L \approx c^{-1}(w_L)$, so we can compute the following first-order effect:

$$U' \mid_{w_H = w_L} = g_H c^{-1}(w_L).$$

When the seller offers a menu, the total surplus is strictly smaller than the total surplus generated by the efficient mechanism. This bound is sufficient to check that:

$$\frac{\partial \Delta S}{\partial w_H} \mid_{w_H = w_L} = 0.$$

In other words, at a first-order approximation the gains in total surplus 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. Hence, when the difference in value is small, it is better to pool both values.

The argument illustrates why it is necessary to have enough dispersion in values to make it optimal to separate values. For example, if the distribution has unbounded support then f must be decreasing and convex (at least in part of the domain), and so the modest tails condition (32) will not be satisfied. In such case, a single-item mechanism will not be optimal.

6.3 Constant Elasticity Cost Function

In this section we continue to assume that the cost function is separable as in (2).. The main novel insight that comes from this is understanding the role of the cost elasticity in determining the number of items being offered. To understand the role of the cost function $c(q)$ in the determination of the optimal menu, we now focus on a one-parameter family of cost functions with

$$c(q) = q^{\eta}/\eta$$

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 profit by offering infinitely large quantities to the buyers.

Consider some fixed information structure S that generates finitely many values $w_1, ..., w_K$ with associated probabilities $g_1, ..., g_K$. We define the corresponding virtual values as in (26). 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 profit generated by an information structure S is then given by:

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

The profit corresponds 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 generate a profit equal to:

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

We can now compare the profit generated by different information structures.

Proposition 2 (Cost Elasticity and Information Policy)
Consider some finite information structure S with $K > 1$ signals:

1. *There exists $\overline{\eta}_S$ such that information structure S generates less profit than complete pooling if and only if $\eta \geq \overline{\eta}_S$.*
2. *There exists $\underline{\eta}_S$ such that information structure S generates less profit than complete disclosure if $\eta \leq \underline{\eta}_S$.*

Proposition 2 considers any given information structure S and evaluates how S performs against either complete pooling or complete disclosure as the cost elasticity changes. For every S there are finite upper and lower bounds , $\overline{\eta}_S$ and $\underline{\eta}_S$, respectively, such that either complete pooling or complete disclosure dominate the given information structure S. Thus, eventually either the suppression of information rent or the social efficiency argument dominate any interior trade-off between these two objectives.

Alternatively, we might consider any given cost function and associated elasticity η and ask if the efficiency gains from screening are eventually dominated by the profit gains that come with the suppression of the information rent. In the working paper version, Bergemann, Heumann & Morris (2022), we establish as Theorem 5 that if the support of the distribution of values is sufficiently narrow, then the gains from a more efficient allocation are dominated by the reduction in the information rent. Namely, pooling is optimal for every distribution F with support in $[\underline{v}, \overline{v}]$ if and only if

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

The result is established by a sequence of improvement arguments that terminate with the complete analysis (and explicit solution) of the binary type model. We refer the reader to the working paper for the details of the proof.

7 Implementation

Our analysis so far assumed a direct mechanism where the buyers observed a signal identified with an expected value which they reported to the mechanism. And we maintained the assumption that the seller did not observe individual buyers' realized signals. In this section, we first describe an implementation of the direct mechanism via a recommender system. We then describe how the assumption that the seller is uninformed can be replaced with the assumption that the seller is unable to offer personalized prices. Both exercises clarify the relevance of the model's finding for digital markets.

7.1 Recommender System

We now show that the direct mechanism can also be realized by a specific indirect mechanism. Namely, we define a recommender system by a tariff, namely a menu of qualities and prices, and a recommendation policy. The recommender system conveys all relevant information through a recommended quality. We thus offer a natural implementation of the optimal mechanism through which the seller can convey the information to the buyers.

We now define a recommender system (P, Z) as a tariff P and a recommendation policy Z. The seller offers a tariff $P : \mathbb{R}_+ \to \mathbb{R}_+$ where $P(q)$ is the price of quality q. The tariff is assumed to include the option not to participate, thus buy a quality $q = 0$ at price $P(0) = 0$. Given a set of qualities and a tariff function P the seller offers a recommendation:

$$Z : [\underline{v}, \bar{v}] \to \Delta\mathbb{R}_+,$$

which (stochastically) suggests a quality level $Z(v) = q$ to a buyer with type v. Thus, we consider a specific information structure Z in which the signal is expressed in terms of a product quality recommendation. The information of each buyer now consists of the common prior and the realized product recommendation. The recommender system (P, Z) is assumed to satisfy the following *obedience constraint*:

$$\mathbb{E}[vq \mid Z(v) = q] - P(q) \geq \mathbb{E}[vq' \mid Z(v) = q] - P(q'), \text{ for all } q, q' \in \mathbb{R}_+. \tag{34}$$

That is, given the information contained in the quality recommendation $Z(v) = q$, it is optimal for the buyer to purchase the recommended quality q at price $P(q)$ rather than any other quality $q' \neq q$ at price $P(q')$.

If we consider the optimal choice of the seller among all recommender systems that satisfy the above obedience constraints, then the seller maximizes expected profit:

$$\max_{\substack{P:\mathbb{R}_+\to\mathbb{R}_+ \\ Z:[\underline{v},\bar{v}]\to\Delta\mathbb{R}_+}} \mathbb{E}[P(Z\left(v\right)) - c(Z\left(v\right))], \tag{35}$$

subject to the obedience constraints (34). By evoking the revelation principle for information design and the taxation principle of nonlinear pricing we obtain the following equivalence.

Proposition 3 (Equivalence of Direct Mechanism and Recommender System)
For every direct mechanism (M, S) which satisfies individual rationality and incentive constraints, there is an outcome equivalent recommender system (P, Z) which satisfies obedience constraints, and conversely.

Proof. By the revelation principle for information design, see Proposition 1 in Bergemann & Morris (2019), it is sufficient to provide information in terms of recommended actions only. Therefore, it is sufficient to provide signals in terms of product quality recommendations, thus the incentive compatible choice of $q\left(S\left(v\right)\right)$ can be recommended or $q\left(S\left(v\right)\right) = Z\left(v\right)$. By the taxation principle for nonlinear pricing, see Proposition 1, Guesnerie & Laffont (1984), every direct mechanism is outcome equivalent to an indirect mechanism in terms of a tariff, thus $P\left(Z\left(v\right)\right) = p\left(S\left(v\right)\right)$. Thus a nonlinear tariff $P\left(\cdot\right)$ and a recommender system $Z\left(\cdot\right)$ are sufficient to implement the product choices of any abstract information structure S and menu M. Conversely, the revelation principle for information design tells that any indirect mechanism, in particular (P, Z) can be implemented by some direct mechanism (M, S). ■

We stated the equivalence between a direct mechanism and a recommender system for the multiplicative preference, $v \cdot q$. But the argument only uses incentive compatibility and the taxation principle for nonlinear pricing, so the equivalence results holds for all nonlinear utility environments. We will make use of this more general insight later in Section 8.

The equivalence result suggests that it is sufficient to convey the information expressed in terms of the product choice. Consistent with this, eBay for example personalizes the search results for each buyer through a machine learning algorithm and determines a personalized default order of search results in a process referred to as "Best Match," see eBay (2022).

7.2 Personalized Pricing

Our maintained interpretation of the informational environment is that the seller determines (stochastically) the information that the buyers receive about their values, but that the seller *does not observe* the realized information (signal) of the buyers.

We now offer an alternative interpretation of the informational environment (which leaves the analysis unchanged). Suppose that the seller in fact *does observe* the buyers' values. But for regulatory or business reasons, the seller has to offer a public menu of qualities at corresponding prices. Thus he is unable to offer prices for items that depend on the individual buyer's value. Therefore the seller cannot engage in price discrimination, in the sense of preventing some buyers from buying at a price available to other buyers. However, the seller can convey information about the buyers' values. This may be implemented by a recommendation of a specific item on the menu, without preventing the buyer from purchasing a different item. The optimal mechanism is then a menu of qualities and prices and a recommendation policy of products that conveys information about the values.

We refer to the first environment as an uninformed seller environment and to the second (and alternative) environment as no personalized pricing environment. In the first environment, the seller is constrained by the private information of the buyers which the seller has to elicit in a direct (or indirect) mechanism. In the second environment, the seller is constrained in the choice of pricing policy while having complete information about the buyers' values.

We now establish that the optimal solutions of these two informational environments coincide. In fact, the same recommendation policy offers an implementation of the optimal mechanisms in both problems.

Proposition 4 (Uninformed Seller and No Personalized Pricing)
Every optimal uninformed seller mechanism is outcome equivalent to a no personalized pricing mechanism, and conversely.

Proof. In the no personalized pricing environment, the seller knows the value v of each buyer and thus could make an offer as function of the value and a specific quality to each buyer, thus offer $(p(v, q(v)), q(v))$ to a buyer with value v. In the no personalized pricing setting, the seller is constrained to offer the same price to all buyers and thus cannot condition the price on v. Thus the price for any offered quality q has to be the same $p(q)$ across all buyers. But the seller can convey the information through a recommendation $q(v)$. Now the profit maximizing menu is simply the solution to the earlier problem of the recommender system (35), thus establishing an equivalence of the optimal solution in the uninformed seller and no personalized pricing environment. ■

In this alternative interpretation, the seller uses their informational advantage only to recommend different products to different consumer segments but offers the same price for every product across all segments. Indeed, 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).

8 Optimality of Discrete Menu with Nonlinear Utility

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 multiplicative separable function of willingness-to-pay and quality, $v \cdot q$. It is immediate that the results of Theorem 1 and 2 will continue to hold for any multiplicative separable value function:

$$u_v(v)\, u_q(q) - p,$$

where the component functions are increasing in v and q respectively. This would merely involve a rescaling of the values or qualities.

A more substantial question is whether the results of Theorem 1 and 2 extend to a general supermodular and nonlinear environment. In this section we provide a generally positive answer with qualifications. Throughout this section, we assume that the seller has a separable cost function as in (2) and we make mild assumptions on the buyer's utility that we explain next.

We now consider a setting where the buyer's utility is nonlinear in quality q and type v:

$$u(q, v) - p.$$

where u is a strictly increasing in both arguments v and q, and strictly supermodular:

$$\frac{\partial^2 u(q, v)}{\partial q \partial v} > \varepsilon, \tag{36}$$

for some $\varepsilon > 0$. We additionally assume that the functions have bounded concavity:

$$\left|\frac{\partial^2 u(v,q)}{\partial q^2}\right|, \left|\frac{\partial^3 u(v,q)}{\partial q^2 \partial v}\right|, \left|\frac{\partial^2 c(q)}{\partial q^2}\right| \leq D, \tag{37}$$

for some finite constant $D > 0$. With the bounded concavity of the objective function we can approximate and bound locally the behavior of the revenue function.

In this nonlinear setting the variable v does not directly present the value or (marginal) willingness-to-pay, and so we refer to v in this section as type v. Correspondingly, the expectation of the value given the signal, $\mathbb{E}[v|s]$, is not a sufficient statistic anymore for the expected utility given the signal, $\mathbb{E}[u(v,\cdot)|s]$. Thus we cannot describe the information structure S in terms of the distribution of posterior expectations $\mathbb{E}[v|s]$.

We therefore take a different route to establish the optimality of discrete information structures. We write the seller's problem as finding a tariff function and a recommender system that satisfies the obedience constraints that we introduce earlier in Section 7.1. Formally, both problems are the same as established in Proposition 3 but analytically the latter one is more tractable. We thus denote by R the distribution of qualities offered by the recommender system:

$$R(q) = \mathbb{P}\{Z(v) \leq q\}.$$

Theorem 5 establishes that discrete menus remain optimal.

We examine the optimal mechanism (i.e., tariff and recommender system) that maximizes profit when the seller is restricted to mechanisms that allocate higher qualities to higher values in a stochastic sense. More precisely, we assume that the seller is restricted to recommender systems that generate positive interdependence between value and qualities. More precisely, we require that v is *positively regression dependent* on $Z(v)$ as introduced by Lehmann (1966), (5.3). Formally, we require that for any $q \leq q'$, the distribution of v conditional on the recommendation $Z(v) = q'$ first-order stochastically dominates the distribution of v conditional on the recommendation $Z(v) = q$:

$$\mathbb{P}\{v \leq \hat{v} \mid Z(v) = q'\} \leq \mathbb{P}\{v \leq \hat{v} \mid Z(v) = q\}, \tag{38}$$

for all $\hat{v}$. A sufficient condition for (38) is that the density is log-supermodular (see Lehmann (1966) (8.1)) , which in turn, is equivalent to requiring that $(v, Z(v))$ are affiliated (see Milgrom & Weber (1982)). As we demonstrate in Lemma 1, this condition guarantees that the expected utility of the buyer maintains a supermodularity property. Note that the information structures that consist of pooling intervals or intervals of full disclosure, which generate the extremal distributions of expected values (28), do naturally satisfy (38).

Theorem 5 (Optimality of Discrete Menu)
In every optimal mechanism (P^*, Z^*) *that satisfies* (38)*,* R^* *is nowhere absolutely continuous.*

This theorem is the direct counterpart of Theorem 1. It shows that the optimal menu is discrete (when restricting attention to monotone recommender systems). However, it does not provide a general characterization of the optimal information structure – i.e., we do not know whether the optimal information structure consists of pooling intervals as in the bilinear case. The condition (38) of positive regression dependence also provides a mild restriction of the feasible information structures.

Before we provide the main proof of the theorem, we provide an auxiliary result that helps explain the role of positive regression dependence. In the presence of (38), the expected utility generated by the recommender system satisfies the strict supermodularity property. We denote by $U : \mathbb{R}_+ \times [0,1] \to \mathbb{R}_+$ the expected utility of a buyer who is recommended the quality in quantile t of the distribution (that is, $R(Z(v)) = t$) and who buys a quality q:

$$U(q,t) = \mathbb{E}[u(q,v) \mid R(Z(v)) = t]. \tag{39}$$

We write the expectation in terms of the quantile t, that is, $U(q,t)$. Expressing the utility in terms of quantile t rather than quality q is convenient as it distinguishes the recommendation t from a choice of quality q. We define the expected value of the type who is recommended to buy quality at quantile $t = R(q)$:

$$w(t) \triangleq \mathbb{E}[v \mid R(s) = t].$$

This notation mimics the earlier notation of expected values, see (4). We have that (38) implies that $w(t)$ is increasing in t.

Lemma 1 (Supermodularity Property)
The expected utility $U(q,t)$ *is increasing in* t *and strictly supermodular in* (q,t)*, that is,*

$$(U(q',t') - U(q,t')) - (U(q',t) - U(q,t)) > \varepsilon(q'-q)(w(t') - w(t)), \tag{40}$$

for all $t < t'$ *and* $q < q'$ *(where* ε *is the lower bound in* (36)*).*

Proof. To prove the result, we first define:

$$n(q,t) \triangleq U(q,t) - \varepsilon w(t) q,$$

and show that it is supermodular. To verify this, we note that for any $q' > q$, we have that:

$$n(q',t) - n(q,t) = \mathbb{E}\left[\int_q^{q'} \left(\frac{\partial u(x,v)}{\partial q} - \varepsilon v\right) dx \mid R(Z(v)) = t\right].$$

The integrand is increasing in v–it suffices to take derivative with respect to v and use (36)). Thus $n(q',t') - n(q,t)$ is increasing in t which follows as the utility function is increasing in v and the distribution of v conditional on t is ordered according to first-order stochastic dominance. We thus have that:

$$(n(q',t') - n(q,t')) - (n(q',t) - n(q,t)) \geq 0.$$

Replacing the definition of $n(q,t)$ we get the result. ■

The positive dependence between q and v (via $Z(v)$) is required to guarantee that the supermodularity (40) is also satisfied in expectation through the recommendation $Z(v)$. If one were to impose more stringent conditions on the utility function then it would be possible to relax the assumptions on the distribution of (q,v) as to continue satisfying (40). When the utility is linear in q and v, then (40) is always satisfied and we do not need to impose any a priori restriction on the information structure (hence, recovering our main result).

Since the expected utility satisfies the supermodularity property, we can use the Envelope condition to compute the revenue. More precisely, for any mechanism and recommendation system, the revenue generated is given by:

$$\Pi = \int_0^1 \Phi(R^{-1}(t),t)dt, \tag{41}$$

where Φ is the virtual utility:

$$\Phi(R^{-1}(t),t) \triangleq U(R^{-1}(t),t) - c(R^{-1}(t)) - \frac{\partial U(R^{-1}(t),t)}{\partial t}(1-t),$$

see Guesnerie & Laffont (1984). We can now conclude the proof of Theorem 5 following analogous arguments as in Theorem 1, where the details are relegated to the Appendix.

Thus, we obtain the optimality of discrete menus even in a general environment of nonlinear preferences. We obtain this result with some qualification in terms of a mild restriction on the feasible information structures. We also do not provide a complete characterization of the optimal information structure. We are merely showing that discreteness is always part of the optimal mechanism.

In the Appendix, we examine what happens when any monotonicity condition as in (38) is removed entirely. We show by means of two examples that a critical argument used to prove

Theorem 1 (namely that when the menus are continuous the seller can increase profit by pooling information), may not go through anymore.

9 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 and the preferences of the buyers. Motivated by this, we considered a canonical nonlinear pricing problem that gave the seller control over the disclosure of information regarding the value of the buyers for the products offered.

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

We analyzed a canonical model of second-degree price discrimination as in Mussa & Rosen (1978) or Maskin & Riley (1984). These models largely consider (pure) vertical differentiation among the buyers and in consequence in the choice and price of products. While vertical differentiation captures an important economic aspect, other specifications, in particular horizontal differentiation, might be of interest as well. Towards this end, we briefly discuss why horizontal differentiation is likely to lead to very different implications regarding the optimal information policy. Thus, consider a model of pure horizontal differentiation where there are many varieties of the product, and for each type of buyer there is some variety that attains the maximum value and all other varieties generate a lower surplus. Thus for example a utility function $u(v,q) = u - (v-q)^2$ would represent such a model of pure horizontal differentiation where the quadratic loss function expresses the fact that for every 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.

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

10 Appendix

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

Proof of Theorem 3. Recall that in a finite-item menu, the profit can be written as in (24) and the optimal information structure and allocation are determined by the cutoff quantiles $\{x_1, ..., x_K\}$. We consider the optimality conditions of the highest two intervals of an optimal mechanism. For this, we define the profit from the highest two items:

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

which are the last two terms of the summations in (24). If the last two intervals are pooled, then the profit 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:

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

To make the expressions more compact, we denote the boundary of the partitions in quality space as follows:

$$q_{K-2} \triangleq Q^{-1}(x_{K-2}); \quad q_{K-1} \triangleq Q^{-1}(x_{K-1})$$

and recall that $\bar{q} = Q^{-1}(x_K) = Q^{-1}(1)$. If the distribution Q has increasing density, then we have that:

$$r_{K-2} \leq q_{K-2}; \quad r_{K-1} \geq \frac{q_{K-2} + q_{K-1}}{2}; \quad r_K \leq \frac{\frac{\bar{q}+q_{K-1}}{2}\left(\frac{(\bar{q}-q_{K-1})g_{K-1}}{q_{K-1}-q_{K-2}}\right) + \bar{q}\left(g_K - \frac{(\bar{q}-q_{K-1})g_{K-1}}{q_{K-1}-q_{K-2}}\right)}{g_K}.$$

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:

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

Thus, pooling the last two intervals increases revenue. 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 $g_K(q_{K-1} - q_{K-2}) + g_{K-1}(\bar{q} - q_{K-1}) = 0$. ■

We establish Theorem 4 through a sequence of optimal menus for successively richer environments. We begin with the optimal single-item menu and proceed to a quadratic payoff environment with linear density. Suppose that the seller were constrained to only offer a single item. Which item would he then offer and at what price?

Lemma 2 (Single Item Menu)
The optimal single item menu satisfies the following first-order conditions:

$$\mu^* = c'(q^*) \quad \textit{and} \quad c(q^*) = v^* q^*, \tag{42}$$

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

$$(q^*, v^*) \in \arg\max_{q,v} \mathbb{P}[v' \geq v](\mathbb{E}[v' \mid v' \geq v]q - c(q)). \tag{43}$$

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

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

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

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

$$\mu^* = c'(q^*) \quad \text{and} \quad c(q^*) = v^* q^*, \tag{44}$$

The first condition states that the quality is efficiently supplied given that the (expected) value of the buyer who buys the good is μ^*. The second condition states that the threshold v^* is also efficiently chosen: given that q^* units are going to be supplied, it is efficient to sell to a buyer with value v if and only if the utility he obtains from this quality is larger than the cost of producing it. We note that the second equality might eventually be satisfied by some $v^* < \underline{v}$, which means there is no exclusion. ■

The reason there are no distortions in the quality supplied and in the threshold is that in a single-item mechanism there is zero buyer surplus. So these quantities are not distorted to reduce consumer surplus. In general, when the optimal mechanism is a multi-item mechanism, both the thresholds and the qualities provided are distorted to reduce the consumer surplus. The rest of the proof of Theorem 4 proceeds as follows. We first show that when the cost is quadratic and the density is linearly decreasing, the optimal mechanism is a single-item mechanism. We then show that the optimal mechanism is a single-item mechanism when the marginal cost is convex and the density is linearly decreasing. Finally, we show that the distributions (32) are mean-preserving contractions of an (appropriately constructed) linearly-decreasing density, which we use to prove that the optimal mechanism is a single-item mechanism.

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

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

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

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

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

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

$$c(q) \triangleq \alpha q + \frac{\beta q^2}{2} + \gamma. \tag{45}$$

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

Recall that in a finite-item menu, the profit can be written as in (24). We consider the optimality conditions of the highest two intervals of an optimal mechanism. For this, we define the profit from the highest two items:

$$\begin{aligned}\Pi_{K-1,K}(v_{K-1},\Delta q_{K-1},\Delta q_K) \;\triangleq\; & \big((g_{K-1}+g_K)\Delta q_{K-1}w_{K-1} - g_{K-1}c(q_{K-2}+\Delta q_{K-1}) \\ & +g_K\left(\Delta q_K w_K - c(q_{K-2}+\Delta q_{K-1}+\Delta q_K)\right)\big),\end{aligned} \tag{46}$$

which are the last two terms of the summations in (24). If the optimal mechanism is a multi-item mechanism, the solution to the following problem:

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

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

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

$$\Delta q_{K-1} = \max\left\{\frac{w_{K-1} - \alpha - \beta q_{K-2}}{\beta} - \frac{(w_K - w_{K-1})g_K}{\beta g_{K-1}}, 0\right\} \text{ and } \Delta q_K = \frac{w_K - \alpha - \beta q_{K-1}}{\beta}. \quad (47)$$

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

$$\frac{w_K - \alpha - \beta q_{K-2}}{w_{K-1} - \alpha - \beta q_{K-2}} < \frac{g_{K-1} + g_K}{g_K}. \quad (48)$$

To write expressions that are compact, we define:

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

Note that in an optimal mechanism we must have that $\kappa \geq 0$, as otherwise the mechanism would be offering a quality q_{K-2} whose marginal cost is higher than the value for all values $v \in [v_{K-3}, v_{K-2}]$, which is clearly suboptimal. Using these definitions, we have that (48) is satisfied if and only if:

$$\frac{3 - 2z + \kappa}{1 - \frac{2z^2}{1+z} + \kappa} < \frac{1}{z^2}. \quad (49)$$

And, for every z satisfying (49), (46) can be written as follows:

$$\Pi_z \triangleq \frac{(g_{K-1} + g_K)(\bar{v} - v_{K-2})^2}{18\beta\,(1 - z^2)}\left((3 - 2z + \kappa)z^2\left(1 - 2z + \frac{4z^2}{1+z} - \kappa\right) + \left(1 + \kappa - \frac{2z^2}{1+z}\right)^2\right). \quad (50)$$

Hence, Π_z is equal to $\Pi_{K-1,K}$ when (48) is satisfied. If the optimal mechanism is a multi-item mechanism, there must exist $z \in [0, 1]$ satisfying (49) that maximizes (50).

If z^* maximizes (50) and satisfies (49) with strict inequality, then z^* must satisfy the first- and second-order conditions. However, there is no $z^* \in [0, 1]$ that satisfies the first- and second-order conditions:

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

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

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

$$(v^*(\alpha,\beta), \Delta q_K^*(\alpha,\beta)) \triangleq \underset{0\leq\Delta q, v_{K-2}\leq v\leq\bar{v}}{\arg\max}\ \Pi_{K-1,K}(v,0,\Delta q), \tag{51}$$

where we define the optimal quality for the last interval:

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

Thus, we consider a restricted optimization problem where the seller takes as given the first $K-2$ intervals and allocations. The restricted problem (51) is then to find an interval $(v_{K-1},\bar{v}] = (v^*(\alpha,\beta),\bar{v}]$ and an allocation $q_K^*(\alpha,\beta)$ so as to maximize the profit from all types in the given interval $(v_{K-2},\bar{v}]$. This restricted maximization problem allows the interval $(v_{K-1},\bar{v}]$ to be a strict inclusion of $(v_{K-2},\bar{v}]$: that is, $(v_{K-1},\bar{v}] \subsetneq (v_{K-2},\bar{v}]$. In this case, all the types in the interval $(v_{K-2}, v_{K-1}]$ will receive the allocation q_{K-2}. Now, from Proposition 5, we know that when the cost is linear-quadratic:

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

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

We now analyze the entire class of convex cost functions with $c'''(q) \geq 0$. We assume that the optimal mechanism consists of multiple items and reach a contradiction. We denote by $\widehat{c}(q)$ a linear-quadratic cost function (as in (45)). We note that $c(q)$ and $\widehat{c}(q)$ intersect three times at most. Furthermore, if $c(q)$ and $\widehat{c}(q)$ are equal at qualities $q_1,\ q_2,\ q_3$, then the difference $\widehat{c}(q) - c(q)$ satisfies:

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

We use this for the following result.

Lemma 3 (Dominating Cost Function)
For every convex cost function with $c'''(q) \geq 0$ and for every (q_{K-2}, q_{K-1}, q_K) with $q_{K-2} \leq q_{K-1} \leq q_K$, there exists (α,β,γ) satisfying $c(q_{K-2}) = \widehat{c}(q_{K-2})$ and one of the following three conditions:

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

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

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

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

$$\widehat{c}(q_{K-2}) = c(q_{K-2});\ \widehat{c}(q_{K-1}) = c(q_{K-1});\ \widehat{c}(q_K) = c(q_K). \tag{52}$$

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

$$\begin{aligned}
\alpha =& \frac{c(q_K)\left(q_{K-2}^2 - q_{K-1}^2\right) + c(q_{K-1})(q_K^2 - q_{K-2}^2) + c(q_{K-2})\left(q_{K-1}^2 - q_K^2\right)}{(q_K - q_{K-1})(q_K - q_{K-2})(q_{K-1} - q_{K-2})} \\
\beta =& \frac{2(c(q_K)(q_{K-1} - q_{K-2}) + c(q_{K-1})(q_{K-2} - q_K) + c(q_{K-2})(q_K - q_{K-1}))}{(q_K - q_{K-1})(q_K - q_{K-2})(q_{K-1} - q_{K-2})} \\
\gamma =& \frac{c(q_K)q_{K-1}q_{K-2}(q_{K-1} - q_{K-2}) + c(q_{K-1})q_K q_{K-2}(q_{K-2} - q_K) + c(q_{K-2})q_K q_{K-1}(q_K - q_{K-1})}{(q_K - q_{K-1})(q_K - q_{K-2})(q_{K-1} - q_{K-2})}
\end{aligned}$$

These are the coefficients one obtains from the interpolation of a second-degree polynomial.

Since $\widehat{c}$ is a linear-quadratic cost function and since $c''' \geq 0$, we have that for all $q \geq q_{K-2}$:

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

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

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

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

Suppose that:

$$q^*(\alpha,\beta) < q_{K-1}, \tag{53}$$

where (α,β) satisfy (52). We then need to find different parameters α, β. We consider parameters α, β as a function of q implicitly defined as follows:

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

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

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

where the first inequality follows from (53) and the second inequality follows from the fact that q^* by definition is larger than q_{K-2} (see (51)) . Thus, following the intermediate value theorem, there exists a $\hat{q} \in [q_{K-2}, q_K]$ such that:

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

Furthermore, note that $q_K > \max\{\hat{q}, q_{K-1}\}$, so we have that $\widehat{c}(q_K) < c(q_K)$. Thus, we are in Case 2 of Lemma 3.

Finally, if

$$q^*(\alpha, \beta) > q_K,$$

we can find α, β, γ such that Case 3 is satisfied in an analogous way to the case when (53) was satisfied. In particular, we consider parameters α, β as functions of q implicitly defined as follows:

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

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

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

This concludes the proof. ■

With Lemma 3 we can extend the optimality result to convex cost functions.

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

Proof. We now suppose that the optimal mechanism satisfies $\Delta q_{K-1}, \Delta q_K > 0$ and reach a contradiction. In the same manner as (46), we define:

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

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

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

We now consider the three cases in Lemma 3.

If we take (α, β) so that the first case in Lemma 3 holds, then we have:

$$\Pi_{K-1,K}(v_{K-1}, \Delta q_{K-1}, \Delta q_K) = \hat{\Pi}_{K-1,K}(v_{K-1}, \Delta q_{K-1}, \Delta q_K); \tag{54}$$

$$\Pi_{K-1,K}(v^*_{K-1}(\alpha,\beta), 0, \Delta q^*_K(\alpha,\beta)) > \hat{\Pi}_{K-1,K}(v^*_{K-1}(\alpha,\beta), 0, \Delta q^*_K(\alpha,\beta)). \tag{55}$$

We thus have that:

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

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

If we consider (α, β) that satisfy the cases 2 or 3 of Lemma 3, then the argument is analogous but (55) will hold with equality and (54) will hold with strict inequality. ■

We now analyze distributions with modest tails. We begin with an important property of the optimal single-item mechanism when the distribution has a linearly-decreasing density. For these distributions, the first-order conditions (44) are necessary and sufficient conditions for optimality when $c'''(\cdot) \geq 0$.

Proposition 7 (Sufficient Conditions for Optimality)
If $c'''(q) \geq 0$, the distribution is $L(v; \underline{v}, \bar{v})$, and $(\hat{q}, \hat{v})$ satisfy the first-order condition (44)*, then $(\hat{q}, \hat{v})$ solves* (43)*, i.e. $(\hat{q}, \hat{v}) = (q^*, v^*)$.*

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

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

Hence, if $(\hat{q}, \hat{v})$ satisfy the first-order condition (44) we have that:

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

We have that:

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

We now note that:

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

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

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

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

We verify that the first-order condition is sufficient for optimality. For this, we check that the solution is always interior, and since there is only one point that satisfies the first-order condition, this must be the optimum. We first note that $\bar{q} \in \{0, \infty\}$ is clearly never optimal. It is also easy to see that $v = \bar{v}$ cannot be an optimum as then the objective function of (43) is 0. We finally note that $v^* = 0$ is never optimal, which can be checked by noting that the first-order condition with respect to the cutoff gives $c(q^*) \leq v^* q^*$. Hence, the solution is always interior and it must be the only point that satisfies the first-order conditions. ■

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

Towards this end, we consider a distribution $L(v; \underline{z}, \overline{z})$ with a linearly-decreasing density where the lower and upper bounds of the distribution L, namely $\underline{z}, \overline{z}$, are chosen to satisfy the following properties relative to the distribution F and the optimal single-item threshold v^* under F given by (43):

$$L(v^*; \underline{z}, \overline{z}) = F(v^*) \text{ and } \mathbb{E}_L[v \mid v \geq v^*] = \mathbb{E}_F[v \mid v \geq v^*], \tag{56}$$

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

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

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

$$\hat{F}(v) = \begin{cases} L(\hat{v}; \underline{z}, \overline{z}), & \text{if } v \in [0, \hat{v}]; \\ L(v; \underline{z}, \overline{z}), & \text{if } v \in [\hat{v}, \overline{z}]; \end{cases} \tag{57}$$

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

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

In the proof of Lemma 4 we will show that indeed such a $\hat{v}$ exists. Thus, $\hat{F}(v)$ is constructed by taking the mass of the lower tail of $L(v; \underline{z}, \overline{z})$ and moving it to 0. In other words, $\hat{F}$ is equal to $L(v; \underline{v}, \bar{v})$ for $v \geq \hat{v}$, and $\hat{F}$ has an atom of size $L(\hat{v}; \underline{z}, \overline{z})$ at $v = 0$.

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

Lemma 4 (Distribution Comparison)
Distribution $\hat{F}$ is a mean-preserving spread of F and $\hat{F}$ is first-order stochastically dominated by $L(v; \underline{z}, \overline{z})$.

Proof. We first compare $L(v, \bar{z}, \underline{z})$ with F. Since f satisfies (32) and $L'(v, \bar{z}, \underline{z})$ is linearly decreasing, we must have that f and $L'(v, \bar{z}, \underline{z})$ intersect at most twice. However, by construction L is constructed to satisfy (56), so they intersect exactly twice at two values $v_1, v_2 \geq v^*$. Hence, (32) implies that for all $v' \in [0, \infty)$:

$$\int_{F(v')}^1 F^{-1}(v) dv \leq \int_{F(v')}^1 L^{-1}(v; \overline{z}, \underline{v}) dv. \tag{58}$$

If the inequality is satisfied with equality for $v' = 0$, we have that $\hat{v} = \underline{z}$ and, otherwise, $\hat{v} > \underline{z}$ (where $\hat{v}$ is used to construct $\hat{F}$ in (57)).

Since $\hat{F}$ is constructed by taking the mass of the lower tail of $L(v; \underline{z}, \overline{z})$ and moving it to 0, it is transparent that $\hat{F}$ is first-order stochastically dominated by $L(v; \underline{z}, \overline{z})$. We have that (58) implies that for all $v' \geq \hat{v}$:

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

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

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

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

We can now conclude the proof by establishing Theorem 4.

Final Step of the Proof of Theorem 4. We first verify that the optimal single-item mechanism when the distribution is $L(v; \underline{z}, \overline{z})$ is the same as when the distribution is F. By construction of $L(v; \underline{z}, \overline{z})$, the first-order condition that is satisfied for F is also satisfied for $L(v; \underline{z}, \overline{z})$. Following Proposition 7, for the linearly decreasing density the first-order condition is sufficient for optimality, and thus (v^*, q^*) given by (43) do in fact form the optimal mechanism for L.

We have that $L(v; \underline{z}, \overline{z})$ generates at least as much profit as $\hat{F}$, and $\hat{F}$ generates at least as much profit as F. Since the optimal mechanism for distribution $L(v; \underline{z}, \overline{z})$ is a single-item mechanism, and this mechanisms generates the same profit (by construction) under distribution F, this must also be the optimal mechanism under distribution F. ■

Proof of Proposition 2. 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 expectation 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 $\overline{\eta}_S$ such that

$$\Pi_S \geq \Pi_P \iff \eta \geq \overline{\eta}_S.$$

This proves the first statement.

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

$$\hat{\Pi} = \int_{\underline{v}}^{\bar{v}} \pi(\max\{\phi(v), 0\}) f(v) dv,$$

where ϕ is defined in (12). We bound the ratio between the profit generated by S and complete disclosure, respectively, as follows:

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

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

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

We thus have that:

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

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

Before we proceed with the proof of Theorem 5, we establish an auxiliary result that bounds the payoff variation. Suppose R^* is the distribution of qualities induced by the optimal mechanism and suppose R^{*-1} is strictly increasing in some interval $[t_1, t_2]$. We have that $w(t)$ and $R^*(t)$ are increasing in this interval, so without loss of generality we fix some $\hat{t}$ at which they are both differentiable and take $\Delta = (t_2 - t_1)/2$. As we take $\Delta \to 0$, we keep $\hat{t}$ fixed and adjust $t_1 = \hat{t} - \Delta$ and $t_2 = \hat{t} + \Delta$. In interval $[t_1, t_2]$ the first-order condition must be satisfied:

$$\frac{\partial \Phi(t, R^{-1}(t))}{\partial q} = 0 \text{ and } \frac{\partial^2 \Phi(t, R^{-1}(t))}{\partial q^2} \leq 0.$$

We provide a preliminary lemma that shows that the virtual values Φ have bounded concavity.

Lemma 5 (Bounded Concavity)
There exists $\bar{d} > 0$ such that:

$$\left| \frac{\partial^2 \Phi(t, R^{-1}(t))}{\partial q^2} \right| \leq \bar{d}.$$

for all $t \in [t_1, t_2]$.

Proof. We now use (37) to show that the seller's objective function has bounded concavity. We have that $w(t)$ is increasing and we can take interval $[t_1, t_2]$ such that $w(t)$ has bounded derivative:

$$\frac{\partial w(t)}{\partial t} \leq \delta,$$

for some $\delta > 0$. The derivative is given by:

$$\frac{\partial^2 \Phi(t, R^{-1}(t))}{\partial q^2} = \frac{\partial^2 U(t, R^{-1}(t))}{\partial q^2} - \frac{\partial^2 c(R^{-1}(t))}{\partial q^2} - \frac{\partial^3 U(t, R^{-1}(t))}{\partial t \partial q^2}(1-t).$$

Bounding the derivatives, we get:

$$\left| \frac{\partial^2 U(t, R^{-1}(t))}{\partial q^2} \right| = \left| \mathbb{E}\left[\frac{\partial^2 u(v,q)}{\partial q^2} \mid R(s) = t \right] \right| \leq D,$$

and

$$\left| \frac{\partial}{\partial q^2} \frac{\partial U(t, R^{-1}(t))}{\partial t} \right| = \left| \frac{\partial}{\partial t} \mathbb{E}\left[\frac{\partial^2 u(v,q)}{\partial q^2} \mid R(s) = t \right] \right| \leq D \frac{\partial w(t)}{\partial t} \leq D\delta.$$

Hence,

$$\left| \frac{\partial^2 \Phi(t, R^{-1}(t))}{\partial q^2} \right| \leq 2D + D\delta.$$

We thus prove the result. ■

Proof of Theorem 5. We now consider an alternative allocation which we show to improve the revenue of the seller:

$$\tilde{R}^{-1}(t) = \begin{cases} \mu_q, & \text{if } t \in [t_1, t_2]; \\ R^{*-1}(t), & \text{otherwise.} \end{cases}$$

where

$$\mu_q = \frac{\int_{t_1}^{t_2} R^{*-1}(s) ds}{t_2 - t_1}.$$

We keep the information structure unchanged. Hence, the preference $U(q,t)$ remains the same, but instead we change the allocation:

$$\tilde{R}^{-1}(v) = \begin{cases} \mu_q, & \text{if } t \in (t_1, t_2); \\ R^{*-1}(t), & \text{if } t \notin (t_1, t_2); \end{cases}$$

which is the analog of (17). The profit difference between the optimal policy and the variation is given by:

$$\Pi^* - \tilde{\Pi} = \int_{t_1}^{t_2} \Phi(R^{*-1}(t), t) dt - \int_{t_1}^{t_2} \Phi\left(\frac{\int_{t_1}^{t_2} R^{*-1}(s) ds}{t_2 - t_1}, t \right) dt \leq \bar{d} \int_{t_1}^{t_2} (R^{-1}(t) - \mu_q)^2 dt.$$

Note that we are not changing the information structure, so the change in profit corresponds simply to the change in virtual surplus computed with the preferences $U(q,t)$. We thus have that:

$$\frac{\Pi^* - \tilde{\Pi}}{\Delta^3} \to \bar{d} \left(\frac{\partial R^{*-1}(t)}{\partial t} \right)^2,$$

which is the analog of (19).

We now consider policy (20). We additionally change the information structure so that all values get the same recommendation

$$\hat{s}(v) = \begin{cases} Z(v), & \text{if } R^*(Z(v)) \notin [t_1, t_2]; \\ \mu_q, & \text{if } R^*(Z(v)) \in [t_1, t_2]. \end{cases}$$

As before, the total surplus generated does not change as we are only pooling the information of buyers who were already getting the allocation μ_q. We then have that:

$$\begin{aligned} \hat{\Pi} - \tilde{\Pi} =& (1-t_1)\left[\mathbb{E}[U(t,\mu_q) - U(t, R^{*-1}(t_1)) \mid t \in [t_1,t_2]] - \big(U(t_1,\mu_q) - U(t_1, R^{*-1}(t_1))\big)\right] \\ &\geq (1-t_1)\varepsilon\mathbb{E}[(w(t)-w(t_1))(R^{*-1}(t) - R^{*-1}(t_1)) \mid t \in [t_1,t_2]], \end{aligned}$$

where the inequality follows from (40). As in Section 3.2, we can then conclude that:

$$\lim_{\Delta \to 0} \frac{\hat{\Pi} - \tilde{\Pi}}{\Delta^2} \geq \varepsilon \frac{\partial w(\hat{t})}{\partial t} \frac{dR^{*-1}(\hat{t})}{\partial t}(1-t_1).$$

We thus have that the reduction in informational rents are of order Δ^2. We thus conclude that for Δ small enough, the new policy generates higher profit. ■

Counter-Examples to Theorem 5 With weaker conditions than (36) and (38), the arguments used in Theorem 5 may fail to go through. Consider first the following example.

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

$$u(q,v) = \min\{q, v\} \tag{59}$$

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

This first example shows that the seller may sometimes extract the full surplus if the type does not affect the marginal value (in some quality range). In this first example, (36) is not satisfied, so we provide a second example in which (36) is satisfied, (38) is not satisfied and the buyer is left with zero surplus even though there is non-trivial screening.

Suppose now the buyer has the following preferences:

$$u(q,v)=\begin{cases} q+\varepsilon vq & \text{if } q\leq v \\ v+\varepsilon qv & \text{if } q\geq v \end{cases} \tag{60}$$

and the marginal cost of production of quality is normalized to zero. We assume that v is uniformly distributed in $[0,1]$. Note that if $\varepsilon=0$, we recover (59), but for $\varepsilon>0$ we have a strictly positive cross-derivative:

$$\frac{\partial^2 u(v,q)}{\partial v \partial q}>\varepsilon \tag{61}$$

for every (v,q) (the cross derivative is infinitely large at $q=v$, but this is not crucial for the arguments).

We consider the following recommendation system:

$$Z(v)=\frac{1}{2}-|v-1/2|; \tag{62}$$

and associated tariff:

$$p(q)=\left(1+\frac{\varepsilon}{2}\right)q. \tag{63}$$

The recommendation system pools high and low values in order to exactly average $1/2$, so $s\in[0,1/2]$. That is, for any s, the value is $v=s$ or $v=1-s$ with equal probability (and $s\in[0,1/2]$ in every realization). On the other hand, the tariff is linear in q so it corresponds to a constant price per-unit of quality. In the proof of Proposition 8 we show that the recommender system satisfies the obedience constraint.

Importantly, the recommender system recommends a qualities over a continuous menu. We now show that there is no additional pooling that can improve the seller's profit.

Proposition 8 (No Profitable Bundling)

There does not exist an recommendation system and tariff $(\hat{s},\hat{p})$ that generates more profit than (s,p) (defined in (62) and (63)) and satisfies either one of these conditions:

1. *$\hat{s}$ is Blackwell less informative than $\hat{s}$, or*

2. *$\hat{s}$ generates less total surplus than s.*

Proof. For any information structure $s:[0,1]\to\Delta\mathbb{R}$, we define:

$$U(q,s)\triangleq\mathbb{E}[u(q,v)\mid s].$$

This is the natural analog of (39). Note that:

$$\mathbb{E}[v \mid s] = \frac{1}{2}, \text{ for all } s \in [0, 1/2].$$

We then have that:

$$U(q,s) = \begin{cases} q + \varepsilon \frac{1}{2} q, & \text{if } q \leq s; \\ \frac{1}{2}(s + \varepsilon s q) + \frac{1}{2}(q + q(1-s)), & \text{if } q \in [s, 1-s]; \\ \frac{1}{2} + \varepsilon q \frac{1}{2}, & \text{if } q \geq 1-s; \end{cases}$$

where this is obtained by simply taking expectation of $u(v, q)$. We then have that:

$$\frac{\partial^2 U(q,s)}{\partial q \partial s} = 0, \text{ for all } q \notin [s, 1-s].$$

Note that the recommendation (62) satisfies the obedience constraint. We also note that the buyers are left with 0 surplus. Thus (63) implements allocation (62). In consequence the buyer can be left with 0 surplus even though (61) is satisfied.

If s is Blackwell more informative than $\hat{s}$, then (p, s) generates higher total surplus than $(\hat{p}, \hat{s})$. Since (p, s) leaves the buyer with zero surplus, it follows that profit equals the total surplus generated. Hence, we get the result. ■

Proposition 8 shows that we can construct a mechanism (that is, a recommendation and a tariff) that implements a continuous menu and does not allow the seller to improve by generating a small amount of pooling. While we have not provided a characterization of the optimal mechanism, this example suffices to show that the arguments in Section 8 may not go through if we relax the assumption that values and qualities are positively regression dependent (see (38)).

References

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

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

Bergemann, D., 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. & Morris, S. (2019). Information design: A unified perspective, *Journal of Economic Literature* **57**: 44–95.

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/* .

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

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

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

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

Haghpanah, N. & Siegel, R. (2022). The limits of multi-product discrimination, *American Economic Review: Insights* **4**: 443–458.

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

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

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

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

Lehmann, E. (1966). Some concepts of dependence, *Annals of Mathematical Statistics* **37**: 1137–1153.

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

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

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

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

Milgrom, P. & Weber, R. (1982). A theory of auctions and competitive bidding, *Econometrica* **50**: 1089–1122.

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.

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

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.

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

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

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